FUTURE VISION BIE

One Stop for All Study Materials

& Lab Programs

Future Vision

By K B Hemanth Raj

Scan the QR Code to Visit the Web Page

Or

Visit : https://hemanthrajhemu.github.io

Gain Access to All Study Materials according to VTU,
CSE – Computer Science Engineering,
ISE – Information Science Engineering,
ECE - Electronics and Communication Engineering
& MORE...

Join Telegram to get Instant Updates: https://bit.ly/VTU_TELEGRAM

Contact: MAIL: futurevisionbie@gmail.com

INSTAGRAM: www.instagram.com/hemanthraj_hemu/

INSTAGRAM: www.instagram.com/futurevisionbie/

WHATSAPP SHARE: https://bit.ly/FVBIESHARE

Eastern Economy Edition

Third Edition

Basic VLSI Design

Douglas A. Pucknell
Kamran Eshraghian

PHI

Contents

A Review of Microelectronics and An Introduction to MOS Technology

Chapter 1

If you would have the kindness to begin at the beginning, I should be vastly obliged; all these stories that begin in the middle simply fog my wit.

— Count Anthony Hamilton

OBJECTIVES

This chapter 'sets the scene' by reviewing the evolution of integrated circuits (ICs), and comparing the general characteristics of currently available technologies, including BiCMOS and GaAs as well as nMOS and CMOS.

Basic MOS transistor action is briefly reviewed and an overview of fabrication processes is given to help appreciate the nature of the technologies.

1.1 INTRODUCTION TO INTEGRATED CIRCUIT TECHNOLOGY

There is no doubt that our daily lives are significantly affected by electronic engineering technology. This is true on the domestic scene, in our professional disciplines, in the workplace, and in leisure activities. Indeed, even at school, tomorrow's adults are exposed to and are coming to terms with quite sophisticated electronic devices and systems. There is no doubt that revolutionary changes have taken place in a relatively short time and it is also certain that even more dramatic advances will be made in the next decade.

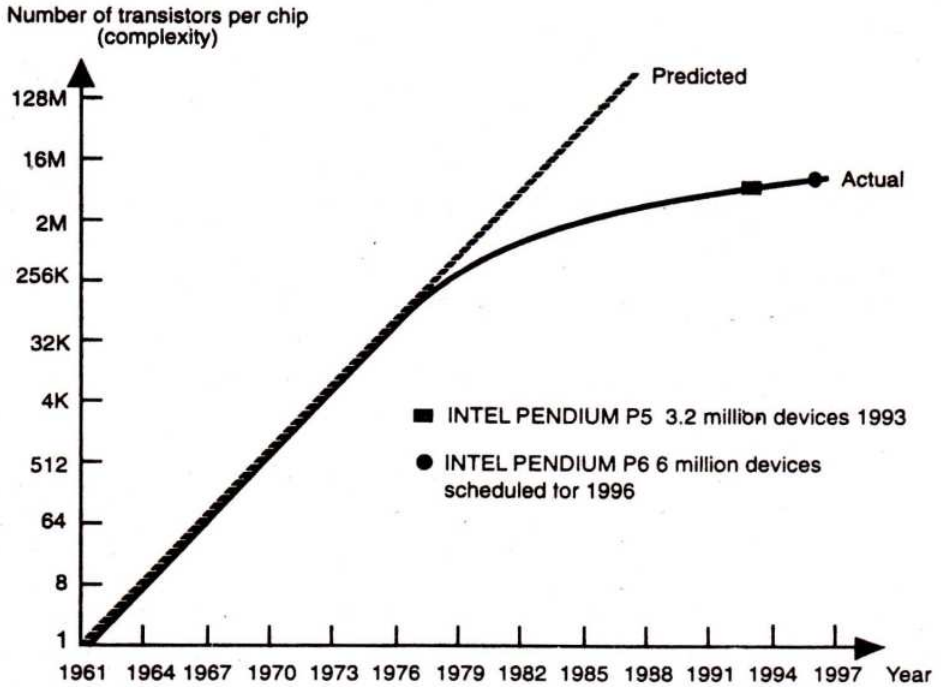
Electronics as we know it today is characterized by reliability, low power dissipation, extremely low weight and volume, and low cost, coupled with an ability to cope easily with a high degree of sophistication and complexity. Electronics, and in particular the integrated circuit, has made possible the design of powerful and flexible processors which provide highly intelligent and adaptable devices for the user. Integrated circuit memories have provided the essential elements to complement these processors and, together with a wide range of logic and analog integrated circuitry, they have provided the system designer with components of considerable capability and extensive application. Furthermore, the revolutionary advances

in technology have not yet by any means run their full course and the potential for future developments is exciting to say the least.

Up until the 1950s electronic active device technology was dominated by the vacuum tube and, although a measure of miniaturization and circuit integration did take place, the technology did not lend itself to miniaturization as we have come to accept it today. Thus the vast majority of present-day electronics is the result of the invention of the transistor in 1947.

The invention of the transistor by William B. Shockley, Walter H. Brattain and John Bardeen of Bell Telephone Laboratories was followed by the development of the Integrated Circuit (IC). The very first IC emerged at the beginning of 1960 and since that time there have already been four generations of ICs: SSI (small scale integration), MSI (medium scale integration), LSI (large scale integration), and VLSI (very large scale integration). Now we are beginning to see the emergence of the fifth generation, ULSI (ultra large scale integration) which is characterized by complexities in excess of 3 million devices on a single IC chip. Further miniaturization is still to come and more revolutionary advances in the application of this technology must inevitably occur.

Over the past several years, Silicon CMOS technology has become the dominant fabrication process for relatively high performance and cost effective VLSI circuits. The revolutionary nature of this development is indicated by the way in which the number of transistors integrated in circuits on a single chip has grown as indicated in Figure 1.1. Such progress is highlighted by recent products such as RISC chips in which it is possible to process some



Note: K signifies a multiplier of 1024 and M a multiplier of 1,048,576.

FIGURE 1.1 Moore's first law: Transistors integrated on a single chip (commercial products).

35 million instructions per second. In order to improve on this throughput rate it will be necessary to improve the technology, both in terms of scaling and processing, and through the incorporation of other enhancements such as BiCMOS. The implications of this approach is that existing silicon technology could effectively facilitate the tripling of rate. Beyond this, i.e., above 100 million instructions per second, one must look to other technologies. In particular, the emerging Gallium Arsenide(GaAs) based technology will be most significant in this area of ultra high speed logic/fast digital processors. GaAs also has further potential as a result of its photo-electronic properties, both as a receiver and as a transmitter of light. GaAs in combination with silicon will provide the designer with some very exciting possibilities.

It is most informative in assessing the role of the currently available technologies to review their speed and power performance domains. This has been set out as Figure 1.2 and the potential presented by each may be readily assessed.

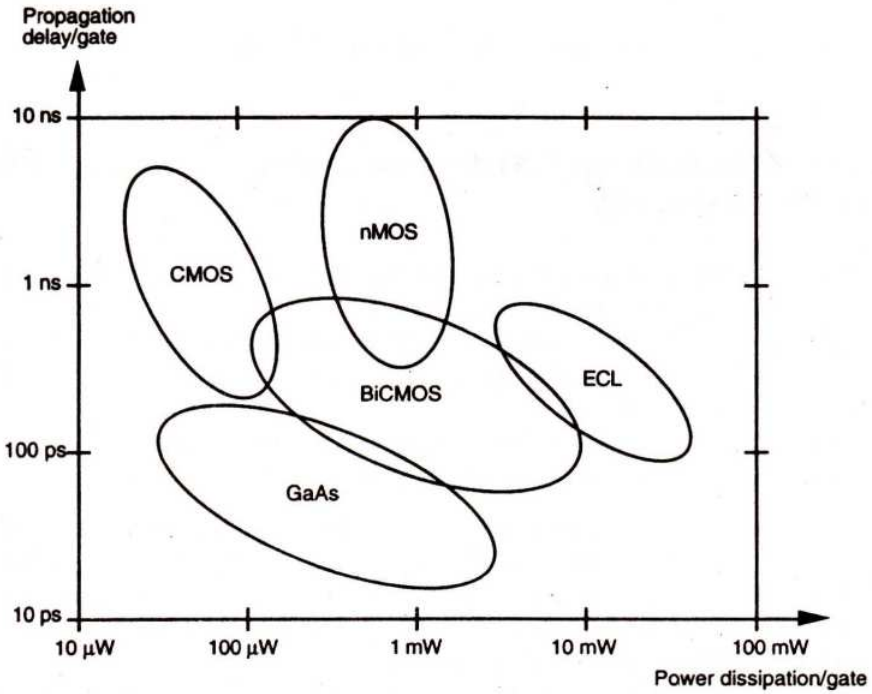


FIGURE 1.2 Speed/power performance of available technologies.

This text deals mostly with silicon-based VLSI, including BiCMOS, but also introduces GaAs-based technology. ECL-based technology is not covered here, but much of the material given is relevant to the general area of the design of digital integrated circuits.

1.2 THE INTEGRATED CIRCUIT (IC) ERA

Such has been the potential of the silicon integrated circuit that there has been an extremely rapid growth in the number of transistors (as a measure of complexity) being integrated into circuits on a single silicon chip. In less than three decades, this number has risen from tens to millions as can be seen in Figure 1.1. The figure sets out what has become known as "Moore's first law" after predictions made by Gordon Moore (of Intel) in the 1960s. It may be seen that his predictions have largely come true except for an increasing divergence between "predicted" and "actual" over the last few years due to problems associated with the complexities involved in designing and testing such very large circuits.

Such has been the impact of this revolutionary growth that IC technology now affects almost every aspect of our lives. More is still to come since we have not yet reached the limits of miniaturization and there is no doubt that tens of millions of transistors will be readily integrated onto a single chip in the future. This evolutionary process is reflected in Table 1.1.

Truly the 1970s, the 1980s and now the 1990s may well be described as the integrated circuit era.

1.3 METAL-OXIDE-SEMICONDUCTOR (MOS) AND RELATED VLSI TECHNOLOGY

Within the bounds of MOS technology, the possible circuit realizations may be based on pMOS, nMOS, CMOS and now BiCMOS devices.

However, this text will deal with nMOS, then with CMOS (which includes pMOS transistors) and BICMOS, and finally with GaAs technology, all of which may be classed as leading Integrated circuit technologies.

Although CMOS is the dominant technology, some of the examples used to illustrate the design processes will be presented in nMOS form. The reasons for this are as follows:

- For nMOS technology, the design methodology and the design rules are easily learned, thus providing a simple but excellent introduction to structured design for VLSI.
- nMOS technology and design processes provide an excellent background for other technologies. In particular, some familiarity with nMOS allows a relatively easy transition to CMOS technology and design.
- For GaAs technology some arrangements in relation to logic design are similar to those employed in nMOS technology. Therefore, understanding the basics of nMOS design will assist in the layout of GaAs circuits.

Not only is VLSI technology providing the user with a new and more complex range of 'off the shelf' circuits, but VLSI design processes are such that system designers can readily design their own special circuits of considerable complexity. This provides a new degree of freedom for designers and it is probable that some very significant advances will result. Couple this with the fact that integration density is increasing rapidly, as advances in technology shrink the feature size for circuits integrated in silicon. Typical manufacturers'

TABLE 1.1 Microelectronics evolution

Year	*1947*	*1950*	*1961*	*1966*	*1971*	*1980*	*1990*	*2000*
Technology	*Invention of the transistor*	*Discrete components*	*SSI*	*MSI*	*LSI*	*VLSI*	*ULSI**	*GSI†*
Approximate numbers of transistors per chip in commercial products	1	1	10	100–1000	1000–20,000	20,000–1,000,000	1,000,000 10,000,000	>10,000,000
Typical products	—	Junction Transistor and diode	Planar devices Logic gates Flip-flops	Counters Multiplexers Adders	8 bit micro-processors ROM RAM	16 and 32 bit micro-processors Sophisticated peripherals GHM Dram	Special processors, Virtual reality machines, smart sensors	

* Ultra large-scale integration
† Giant-scale integration

Note: The boundary lines between technologies in the table are not artificially created. Crossing each boundary requires new design methodology, simulation approaches, and new methods for determining and routing communications and for handling complexity.

commercial IC products have shown this trend quite clearly as shown in Figure 1.3 and, simultaneously, the effectiveness of the circuits produced has increased with scaling down. A common measure of effectiveness is the speed power product of the basic logic gate circuit of the technology (for nMOS, the *Nor* gate, with *Nand* and *Nor* gates for CMOS). Speed power product is measured in picojoules (pJ) and is the product of the gate switching delay in nanoseconds and the gate power dissipation in milliwatts. Typical figures are implied in Figure 1.2.

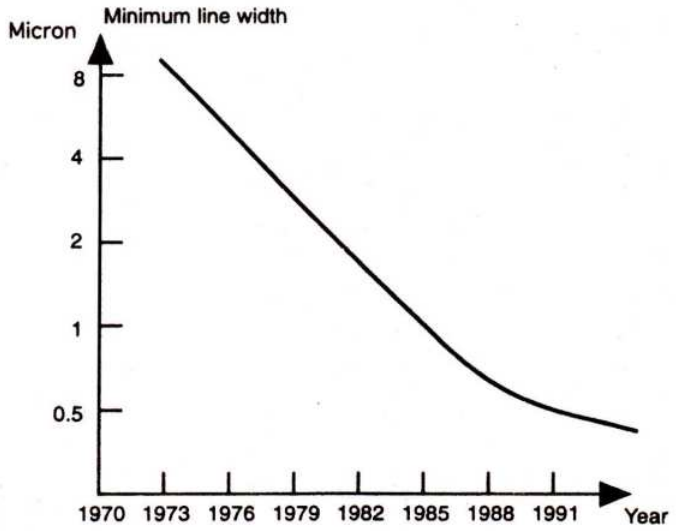


FIGURE 1.3 Approximate minimum line width of commercial products versus year.

1.4 BASIC MOS TRANSISTORS

Having now established some background, let us turn our attention to basic MOS processes and devices. In particular, let us examine the basic nMOS enhancement and depletion mode transistors as shown in Figures 1.4(a) and (b).

nMOS devices are formed in a p-type substrate of moderate doping level. The source and drain regions are formed by diffusing n-type impurities through suitable masks into these areas to give the desired n-impurity concentration and give rise to depletion regions which extend mainly in the more lightly doped p-region as shown. Thus, source and drain are isolated from one another by two diodes. Connections to the source and drain are made by a deposited metal layer. In order to make a useful device, there must be the capability for establishing and controlling a current between source and drain, and this is commonly achieved in one of two ways, giving rise to the enhancement mode and depletion mode transistors.

Consider the enhancement mode device first, shown in Figure 1.4(a). A polysilicon gate is deposited on a layer of insulation over the region between source and drain. Figure 1.4(a) shows a basic enhancement mode device in which the channel is not established and the device is in a non-conducting condition, $V_D = V_S = V_{gs} = 0$. If this gate is connected to a

suitable positive voltage with respect to the source, then the electric field established between the gate and the substrate gives rise to a charge inversion region in the substrate under the gate insulation and a conducting path or channel is formed between source and drain.

The channel may also be established so that it is present under the condition $V_{gs} = 0$ by implanting suitable impurities in the region between source and drain during manufacture and prior to depositing the insulation and the gate. This arrangement is shown in Figure 1.4(b). Under these circumstances, source and drain are connected by a conducting channel, but the channel may now be closed by applying a suitable negative voltage to the gate.

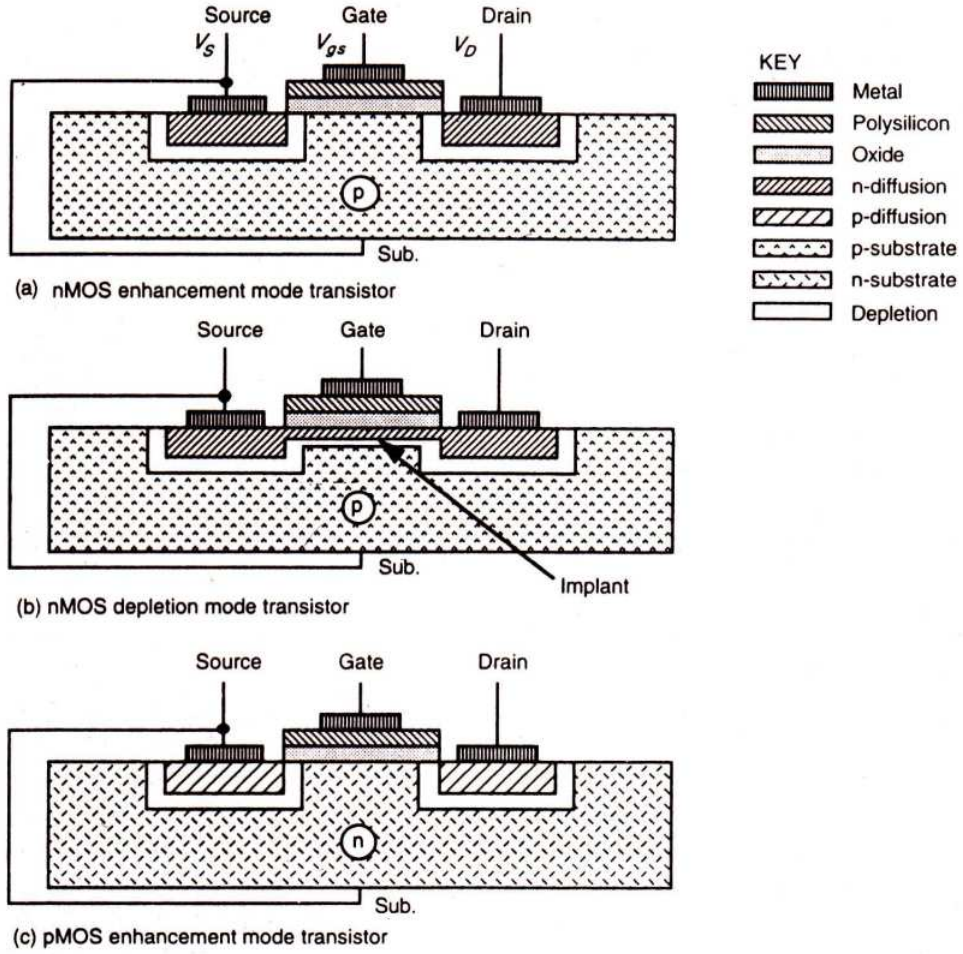


FIGURE 1.4 MOS transistors (V_D = 0 V. Source gate and substrate to 0 V).

In both cases, variations of the gate voltage allow control of any current flow between source and drain.

Figure 1.4(c) shows the basic pMOS transistor structure for an enhancement mode device. In this case the substrate is of n-type material and the source and drain diffusions are

consequently p-type. In the figure, the conditions shown are those for an unbiased device; however, the application of a *negative* voltage of suitable magnitude ($> |V_t|$) between gate and source will give rise to the formation of a channel (p-type) between the source and drain and current may then flow if the drain is made negative with respect to the source. In this case the current is carried by holes as opposed to electrons (as is the case for nMOS devices). In consequence, pMOS transistors are inherently slower than nMOS, since hole mobility μ_p is less, by a factor of approximately 2.5, than electron mobility μ_n. However, bearing these differences in mind, the discussions of nMOS transistors which follow relate equally well to pMOS transistors.

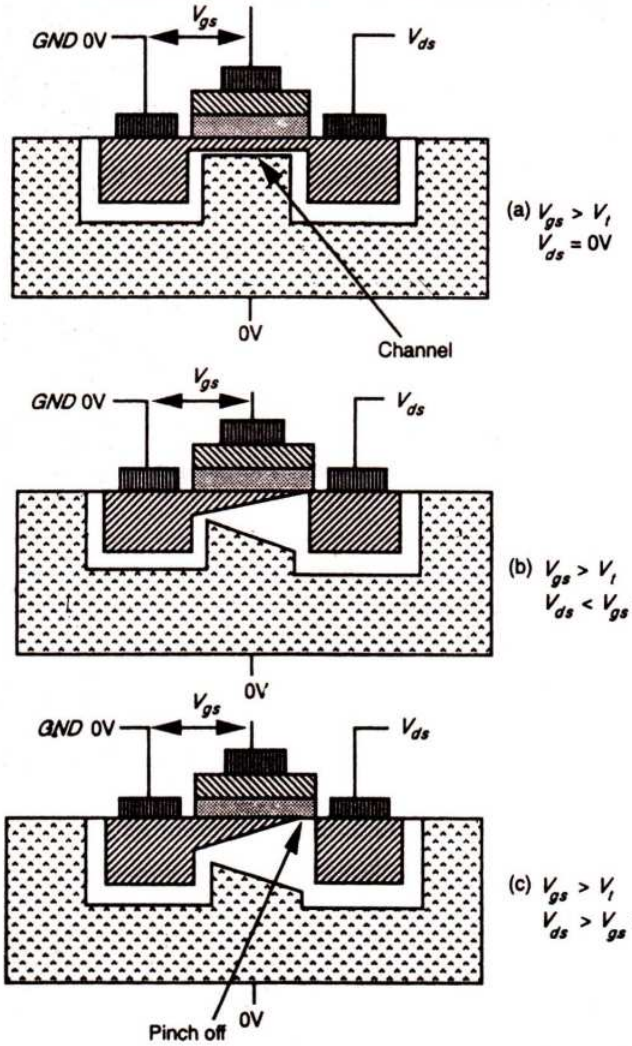
1.5 ENHANCEMENT MODE TRANSISTOR ACTION

To gain some understanding of this mechanism, let us further consider the enhancement mode device, as in Figure 1.5, under three sets of conditions. It must first be recognized that in order to establish the channel in the first place a minimum voltage level of *threshold voltage* V_t must be established between gate and source (and of course between gate and substrate as a result). Figure 1.5(a) then indicates the conditions prevailing with the channel established but no current flowing between source and drain ($V_{ds} = 0$). Now consider the conditions prevailing when current flows in the channel by applying a voltage V_{ds} between drain and source. There must, of course, be a corresponding IR drop = V_{ds} along the channel. This results in the voltage between gate and channel varying with distance along the channel with the voltage being a maximum of V_{gs} at the source end. Since the effective gate voltage is $V_g = V_{gs} - V_t$, (no current flows when $V_{gs} < V_t$) there will be voltage available to invert the channel at the drain end so long as $V_{gs} - V_t \geq V_{ds}$. The limiting condition comes when $V_{ds} = V_{gs} - V_t$. For all voltages $V_{ds} < V_{gs} - V_t$, the device is in the non-saturated region of operation which is the condition shown in Figure 1.5(b).

Consider now what happens when V_{ds} is increased to a level greater than $V_{gs} - V_t$. In this case, an IR drop = $V_{gs} - V_t$ takes place over less than the whole length of the channel so that over part of the channel, near the drain, there is insufficient electric field available to give rise to an inversion layer to create the channel. The channel is, therefore, 'pinched off' as indicated in Figure 1.5(c). Diffusion current completes the path from source to drain in this case, causing the channel to exhibit a high resistance and behave as a constant current source. This region, known as *saturation,* is characterized by almost constant current for increase of V_{ds} above $V_{ds} = V_{gs} - V_t$. In all cases, the channel will cease to exist and no current will flow when $V_{gs} < V_t$. Typically, for enhancement mode devices, $V_t = 1$ volt for $V_{DD} = 5$ V or, in general terms, $V_t = 0.2\ V_{DD}$.

1.6 DEPLETION MODE TRANSISTOR ACTION

For depletion mode devices the channel is established, due to the implant, even when $V_{gs} = 0$, and to cause the channel to cease to exist a negative voltage V_{td} must be applied between gate and source.



Note: V_{ds} is the drain-to-source voltage. Substrate assumed connected to 0 V.

FIGURE 1.5 Enhancement mode transistor for particular values of V_{ds} with ($V_{gs} > V_t$).

V_{td} is typically $< -0.8\ V_{DD}$, depending on the implant and substrate bias, but, threshold voltage differences apart, the action is similar to that of the enhancement mode transistor.

Commonly used symbols for nMOS and pMOS transistors are set out in Figure 1.6.

1.7 nMOS FABRICATION

A brief introduction to the general aspects of the polysilicon gate self-aligning nMOS fabrication process will now be given. As well as being relevant in their own right, the fabrication processes used for nMOS are relevant to CMOS and BiCMOS which may be viewed as

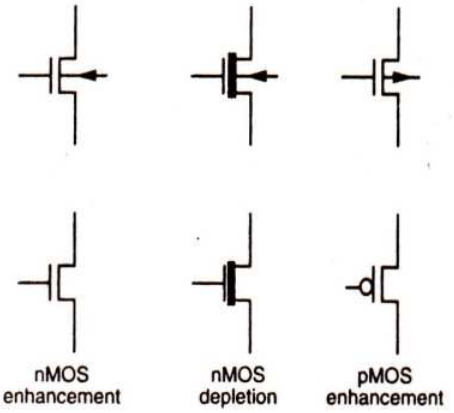


FIGURE 1.6 Transistor circuit symbols.

involving additional fabrication steps. Also, it is clear that an appreciation of the fabrication processes will give an insight into the way in which design information must be presented and into the reasons for certain performance characteristics and limitations. An nMOS process is illustrated in Figure 1.7 and may be outlined as follows:

1. Processing is carried out on a thin wafer cut from a single crystal of silicon of high purity into which the required p-impurities are introduced as the crystal is grown. Such wafers are typically 75 to 150 mm in diameter and 0.4 mm thick and are doped with, say, boron to impurity concentrations of $10^{15}/cm^3$ to $10^{16}/cm^3$, giving resistivity in the approximate range 25 ohm cm to 2 ohm cm.
2. A layer of silicon dioxide (SiO_2), typically 1 μm thick, is grown all over the surface of the wafer to protect the surface, act as a barrier to dopants during processing, and provide a generally insulating substrate on to which other layers may be deposited and patterned.
3. The surface is now covered with a photoresist which is deposited onto the wafer and spun to achieve an even distribution of the required thickness.
4. The photoresist layer is then exposed to ultraviolet light through a mask which defines those regions into which diffusion is to take place together with transistor channels. Assume, for example, that those areas exposed to ultraviolet radiation are polymerized (hardened), but that the areas required for diffusion are shielded by the mask and remain unaffected.
5. These areas are subsequently readily etched away together with the underlying silicon dioxide so that the wafer surface is exposed in the window defined by the mask.
6. The remaining photoresist is removed and a thin layer of SiO_2 (0.1 μm typical) is grown over the entire chip surface and then polysilicon is deposited on top of this to form the gate structure. The polysilicon layer consists of heavily doped polysilicon deposited by chemical vapor deposition (CVD). In the fabrication of fine pattern devices, precise control of thickness, impurity concentration, and resistivity is necessary.
7. Further photoresist coating and masking allows the polysilicon to be patterned (as shown in Step 6) and then the thin oxide is removed to expose areas into which

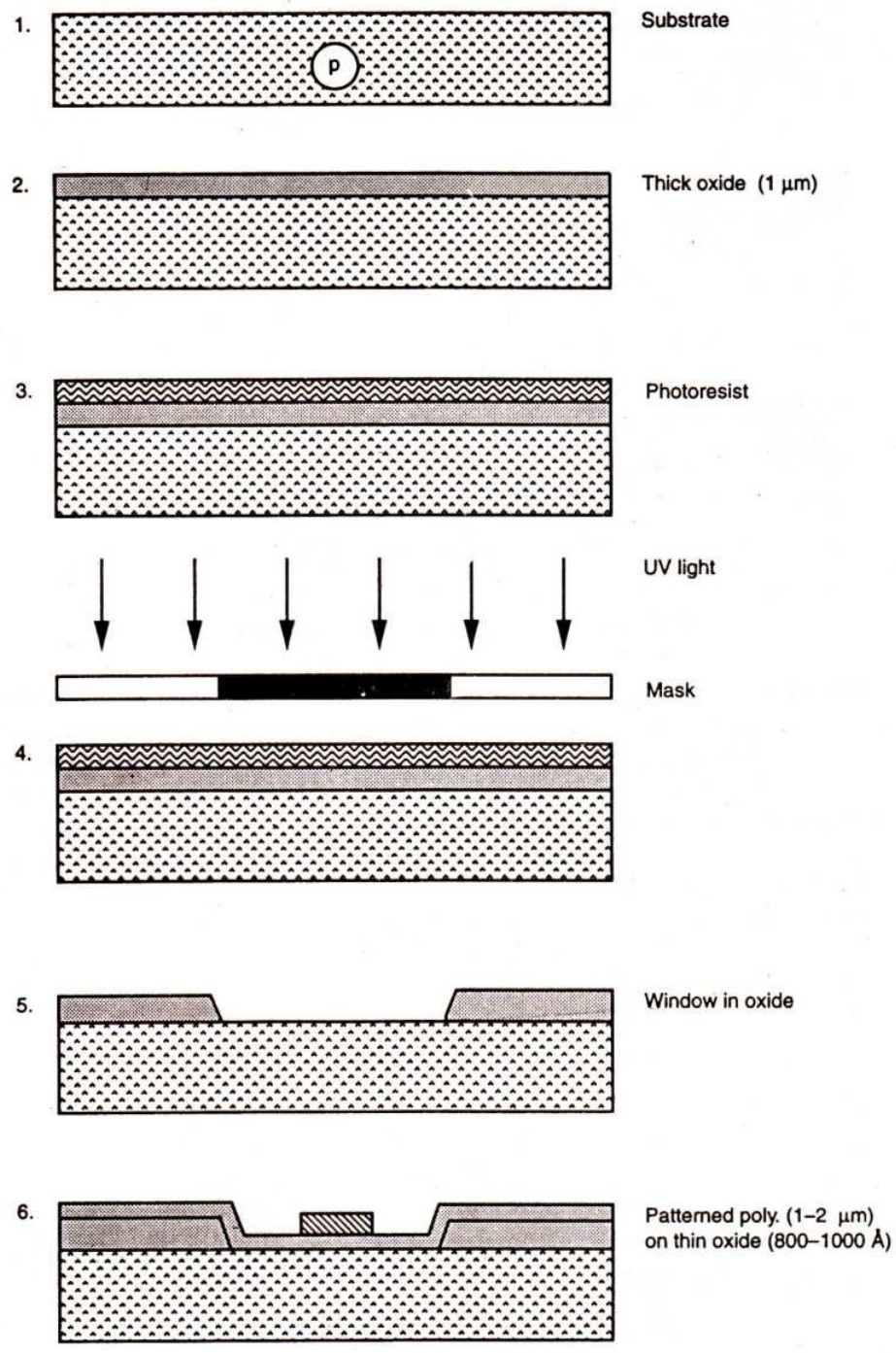


FIGURE 1.7 Continued

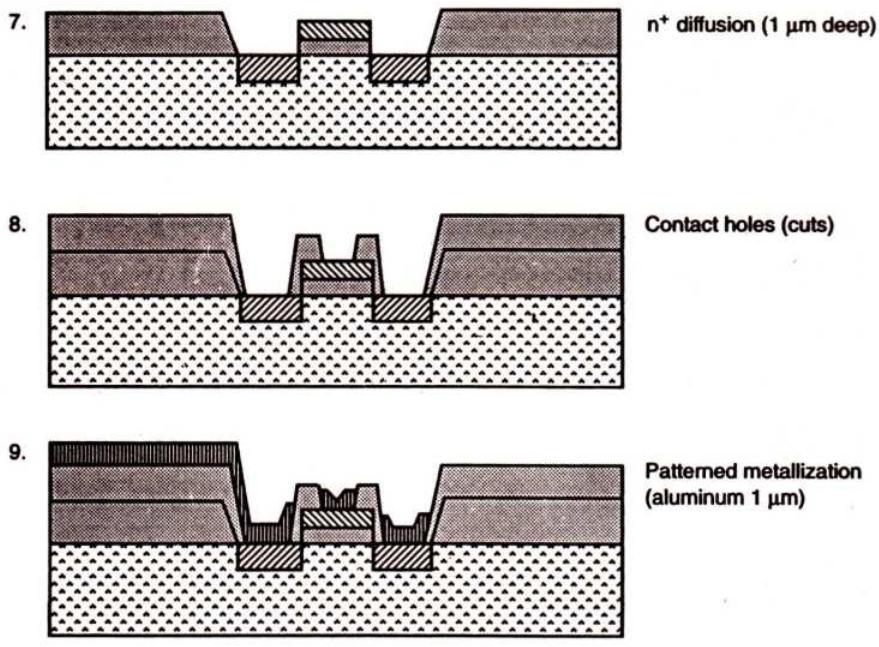


FIGURE 1.7 nMOS fabrication process.

n-type impurities are to be diffused to form the source and drain as shown. Diffusion is achieved by heating the wafer to a high temperature and passing a gas containing the desired n-type impurity (for example, phosphorus) over the surface as indicated in Figure 1.8. Note that the polysilicon with underlying thin oxide act as masks during diffusion—the process is self-aligning.

8. Thick oxide (SiO_2) is grown over all again and is then masked with photoresist and etched to expose selected areas of the polysilicon gate and the drain and source areas where connections (i.e. contact cuts) are to be made.

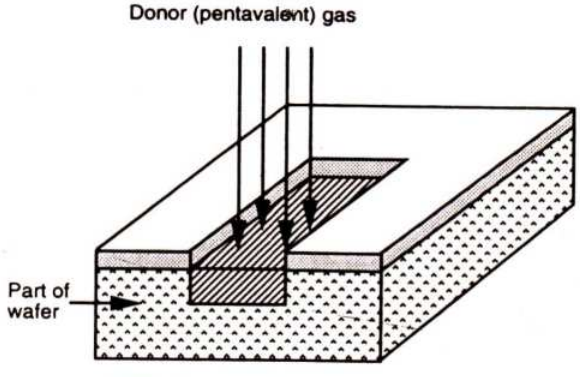


FIGURE 1.8 Diffusion process.

9. The whole chip then has metal (aluminum) deposited over its surface to a thickness typically of 1 μm. This metal layer is then masked and etched to form the required interconnection pattern.

It will be seen that the process revolves around the formation or deposition and patterning of three layers, separated by silicon dioxide insulation. The layers are diffusion within the substrate, polysilicon on oxide on the substrate, and metal insulated again by oxide.

To form depletion mode devices it is only necessary to introduce a masked ion implantation step between steps 5 and 6 or 6 and 7 in Figure 1.7. Again, the thick oxide acts as a mask and this process stage is also self-aligning.

Consideration of the processing steps will reveal that relatively few masks are needed and the self-aligning aspects of the masking processes greatly ease the problems of mask registration. In practice, some extra process steps are necessary, including the overglassing of the whole wafer, except where contacts to the outside world are required. However, the process is basically straightforward to envisage and circuit design eventually comes down to the business of delineating the masks for each stage of the process. The essence of the process may be reiterated as follows.

1.7.1 Summary of An nMOS Process

- Processing takes place on a p-doped silicon crystal wafer on which is grown a 'thick' layer of SiO_2.
- *Mask* 1—Pattern SiO_2 to expose the silicon surface in areas where paths in the diffusion layer or gate areas of transistors are required. Deposit thin oxide over all. For this reason, this mask is often known as the *'thinox' mask* but some texts refer to it as the *diffusion mask.*
- *Mask* 2—Pattern the ion implantation within the thinox region where depletion mode devices are to be produced—*self-aligning.*
- *Mask* 3—Deposit polysilicon over all (1.5 μm thick typically), then pattern using Mask 3. Using the same mask, remove thin oxide layer where it is not covered by polysilicon.
- Diffuse n^+ regions into areas where thin oxide has been removed. Transistor drains and sources are thus self-aligning with respect to the gate structures.
- *Mask* 4—Grow thick oxide over all and then etch for contact cuts.
- *Mask* 5—Deposit metal and pattern with Mask 5.
- *Mask* 6—Would be required for the overglassing process step.

1.8 CMOS FABRICATION

There are a number of approaches to CMOS fabrication, including the p-well, the n-well, the twin-tub, and the silicon-on-insulator processes. In order to introduce the reader to CMOS design we will be concerned mainly with well-based circuits. The p-well process is widely used in practice and the n-well process is also popular, particularly as it is an easy retrofit to existing nMOS lines, so we will also discuss it briefly.

For the lambda-based rules set out later, we will assume a p-well process.

1.8.1 The p-well Process

A brief overview of the fabrication steps may be obtained with reference to Figure 1.9, noting that the basic processing steps are of the same nature as those used for nMOS.

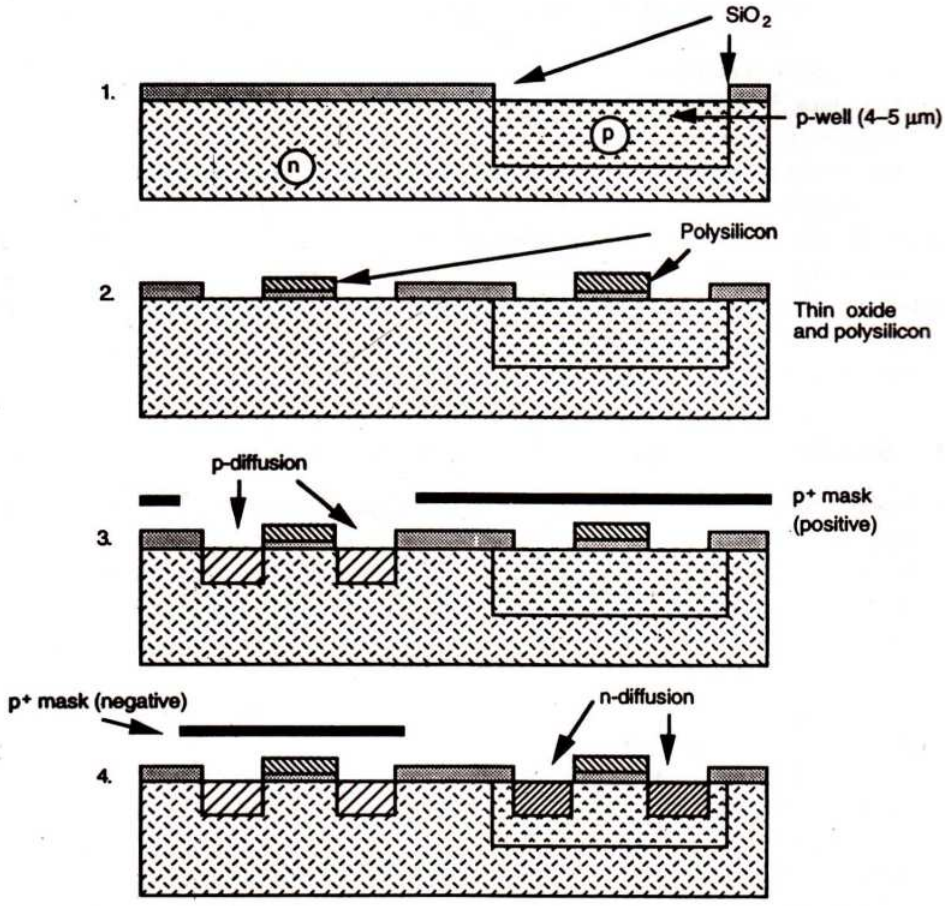


FIGURE 1.9 CMOS p-well process steps.

In primitive terms, the structure consists of an n-type substrate in which p-devices may be formed by suitable masking and diffusion and, in order to accommodate n-type devices, a deep p-well is diffused into the n-type substrate as shown.

This diffusion must be carried out with special care since the p-well doping concentration and depth will affect the threshold voltages as well as the breakdown voltages of the n-transistors. To achieve low threshold voltages (0.6 to 1.0 V) we need either deep-well diffusion or high-well resistivity. However, deep wells require larger spacing between the n- and p-type transistors and wires due to lateral diffusion and therefore a larger chip area.

The p-wells act as substrates for the n-devices within the parent n-substrate, and, provided that voltage polarity restrictions are observed, the two areas are electrically isolated. However,

since there are now in effect two substrates, two substrate connections (V_{DD} and V_{SS}) are required, as shown in Figure 1.10.

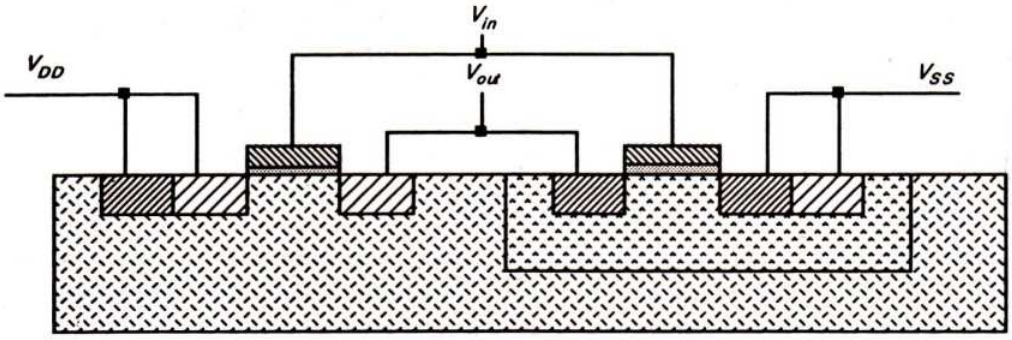


FIGURE 1.10 CMOS p-well inverter showing V_{DD} and V_{SS} substrate connections.

In all other respects—masking, patterning, and diffusion—the process is similar to nMOS fabrication. In summary, typical processing steps are:

- *Mask* 1 — defines the areas in which the deep p-well diffusions are to take place.
- *Mask* 2 — defines the thinox regions, namely those areas where the thick oxide is to be stripped and thin oxide grown to accommodate p- and n-transistors and wires.
- *Mask* 3 — used to pattern the polysilicon layer which is deposited after the thin oxide.
- *Mask* 4 — A p-plus mask is now used (to be in effect "Anded" with Mask 2) to define all areas where p-diffusion is to take place.
- *Mask* 5 — This is usually performed using the negative form of the p-plus mask and defines those areas where n-type diffusion is to take place.
- *Mask* 6 — Contact cuts are now defined.
- *Mask* 7 — The metal layer pattern is defined by this mask.
- *Mask* 8 — An overall passivation (overglass) layer is now applied and Mask 8 is needed to define the openings for access to bonding pads.

1.8.2 The n-well Process

As indicated earlier, although the p-well process is widely used, n-well fabrication has also gained wide acceptance, initially as a retrofit to nMOS lines.

N-well CMOS circuits are also superior to p-well because of the lower substrate bias effects on transistor threshold voltage and inherently lower parasitic capacitances associated with source and drain regions.

Typical n-well fabrication steps are illustrated in Figure 1.11. The first mask defines the n-well regions. This is followed by a low dose phosphorus implant driven in by a high temperature diffusion step to form the n-wells. The well depth is optimized to ensure against p-substrate to p^+ diffusion breakdown without compromising the n-well to n^+ mask separation. The next steps are to define the devices and diffusion paths, grow field oxide, deposit and

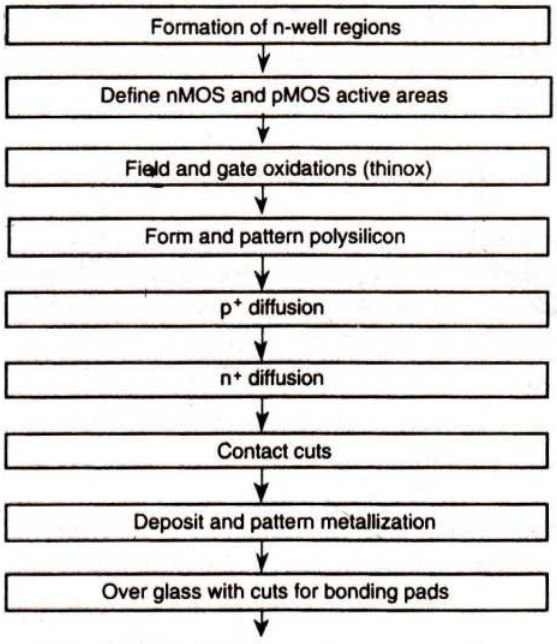


FIGURE 1.11 Main steps in a typical n-well process.

pattern the polysilicon, carry out the diffusions, make contact cuts, and finally metalize as before.

It will be seen that an n^+ mask and its complement may be used to define the n- and p-diffusion regions respectively. These same masks also include the V_{DD} and V_{SS} contacts (respectively). It should be noted that, alternatively, we could have used a p^+ mask and its complement, since the n^+ and p^+ masks are generally complementary.

By way of illustration, Figure 1.12 shows an inverter circuit fabricated by the n-well process, and this may be directly compared with Figure 1.10.

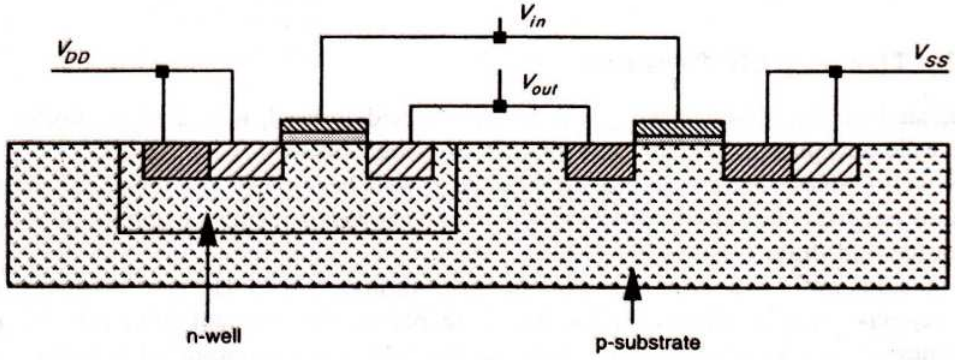


FIGURE 1.12 Cross-sectional view of n-well CMOS inverter.

Owing to differences in charge carrier mobilities, the n-well process creates non-optimum p-channel characteristics. However, in many CMOS designs (such as domino-logic and dynamic-logic structures), this is relatively unimportant since they contain a preponderance of n-channel devices. Thus the n-channel transistors are mainly those used to form logic elements, providing speed and high density of elements.

Latch-up problems can be considerably reduced by using a low-resistivity epitaxial p-type substrate as the starting material, which can subsequently act as a very low resistance ground-plane to collect substrate currents.

However, a factor of the n-well process is that the performance of the already poorly performing p-transistor is even further degraded. Modern process lines have come to grips with these problems, and good device performance may be achieved for both p-well and n-well fabrication.

The design rules which are presented for 1.2 μm and 2 μm technologies in this text are for Orbit™ n-well processes.

1.8.2.1 The Berkeley n-well process

There are a number of p-well and n-well fabrication processes and, in order to look more closely at typical fabrication steps, we will use the Berkeley n-well process as an example. This process is illustrated in Figure 1.13.

1.8.3 The Twin-Tub Process

A logical extension of the p-well and n-well approaches is the twin-tub fabrication process.

Here we start with a substrate of high resistivity n-type material and then create both n-well and p-well regions. Through this process it is possible to preserve the performance of n-transistors without compromising the p-transistors. Doping control is more readily achieved and some relaxation in manufacturing tolerances results. This is particularly important as far as latch-up is concerned.

In general, the twin-tub process allows separate optimization of the n- and p-transistors. The arrangement of an inverter is illustrated in Figure 1.14, which may, in turn, be compared with Figures 1.10 and 1.12.

1.9 THERMAL ASPECTS OF PROCESSING

The processes involved in making nMOS and CMOS devices have differing high temperature sequences as indicated in Figure 1.15.

The CMOS p-well process, for example, has a high temperature p-well diffusion process (1100 to 1250°C), the nMOS process having no such requirement. Because of the simplicity, ease of fabrication, and high density per unit area of nMOS circuits, many of the earlier IC designs, still in current use, have been fabricated using nMOS technology and it is likely that nMOS and CMOS system designs will continue to co-exist for some time to come.

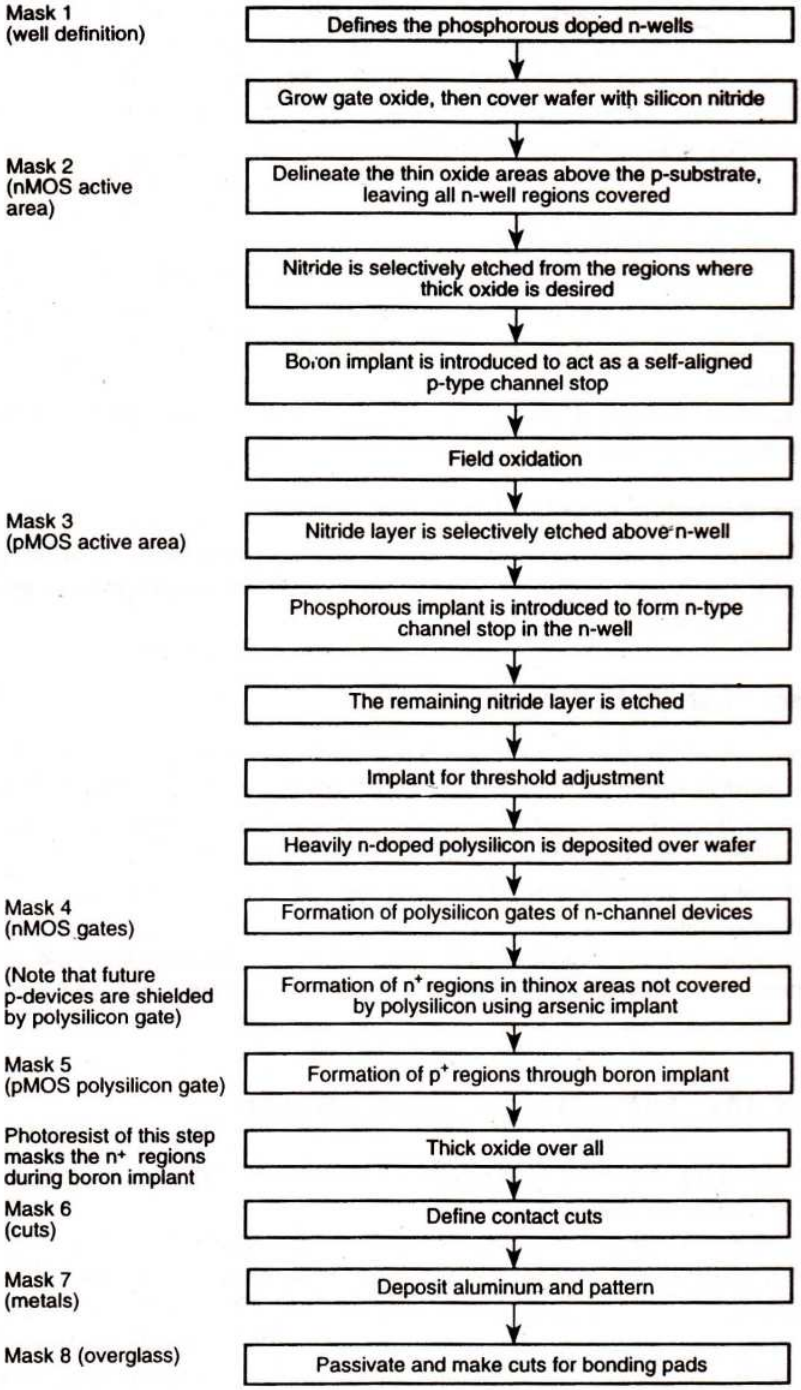


FIGURE 1.13 Flow diagram of Berkeley n-well fabrication.

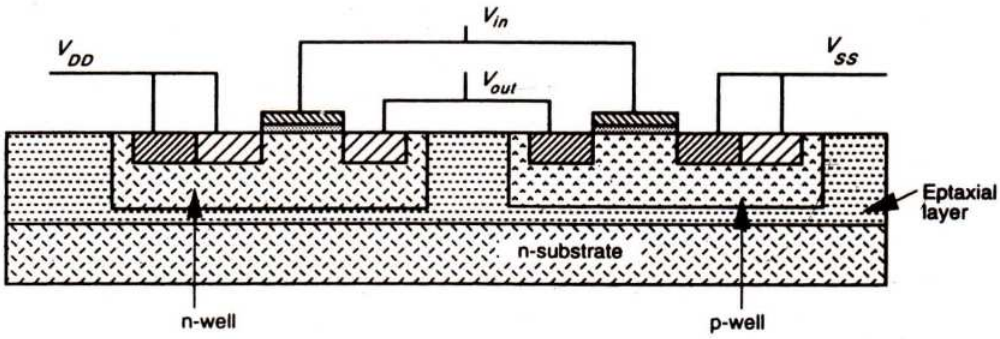


FIGURE 1.14 Twin-tub structure.

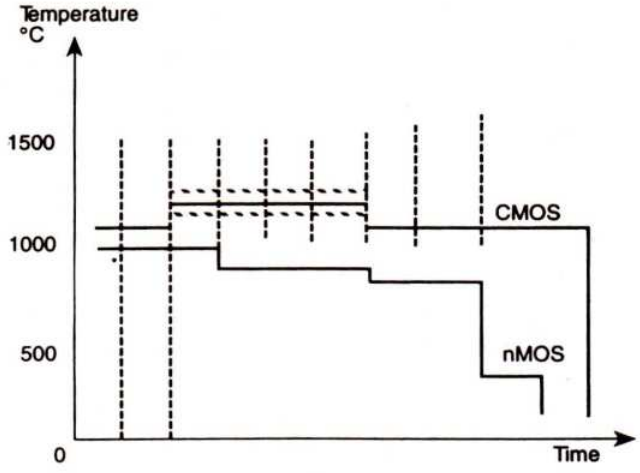


FIGURE 1.15 Thermal sequence difference between nMOS and CMOS processes.

1.10 BiCMOS TECHNOLOGY

A known deficiency of MOS technology lies in the limited load driving capabilities of MOS transistors. This is due to the limited current sourcing and current sinking abilities associated with both p- and n-transistors and although it is possible, for example, to design so called super-buffers using MOS transistors alone, such arrangements do not always compare well with the capabilities of bipolar transistors. Bipolar transistors also provide higher gain and have generally better noise and high frequency characteristics than MOS transistors and it may be seen (Figure 1.2) that BiCMOS gates could be an effective way of speeding up VLSI circuits. However, the application of BiCMOS in sub-systems such as ALU, ROM, a register-file, or, for that matter, a barrel shifter, is not always an effective way of improving speed. This is because most gates in such structures do not have to drive large capacitive loads so that the BiCMOS arrangements give no speed advantage. To take advantage of BiCMOS, the whole functional entity, not just the logic gates, must be considered. A comparison between the characteristics of CMOS and bipolar circuits is set out in Table 1.2 and the

TABLE 1.2 Comparison between CMOS and bipolar technologies

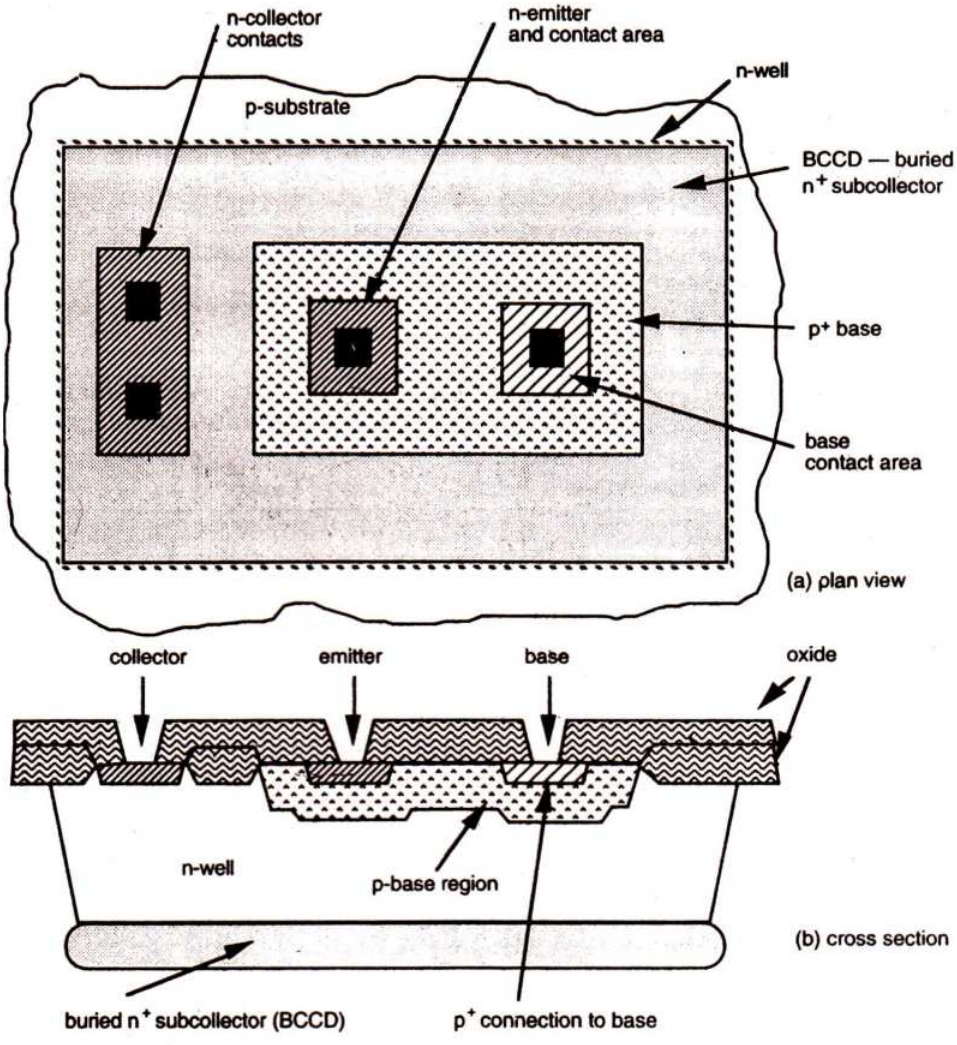
CMOS technology	*Bipolar technology*
• Low static power dissipation	• High power dissipation
• High input impedance (low drive current)	• Low input impedance (high drive current)
• Scalable threshold voltage	
• High noise margin	• Low voltage swing logic
• High packing density	• Low packing density
• High delay sensitivity to load (fan-out limitations)	• Low delay sensitivity to load
• Low output drive current	• High output drive current
• Low g_m ($g_m \alpha V_{in}$)	• High g_m ($g_m \alpha e^{V_{in}}$)
	• High f_t at low currents
• Bidirectional capability (drain and source are interchangeable)	• Essentially unidirectional
• A near ideal switching device	

differences are self-evident. BiCMOS technology goes some way towards combining the virtues of both technologies.

When considering CMOS technology, it becomes apparent that theoretically there should be little difficulty in extending the fabrication processes to include bipolar as well as MOS transistors. Indeed, a problem of p-well and n-well CMOS processing is that parasitic bipolar transistors are inadvertantly formed as part of the outcome of fabrication. The production of npn bipolar transistors with good performance characteristics can be achieved, for example, by extending the standard n-well CMOS processing to include further masks to add two additional layers—the n^+ subcollector and p^+ base layers. The npn transistor is formed in an n-well and the additional p^+ base region is located in the well to form the p-base region of the transistor. The second additional layer—the buried n^+ subcollector (BCCD) is added to reduce the n-well (collector) resistance and thus improve the quality of the bipolar transistor. The simplified general arrangement of such a bipolar npn transistor may be appreciated with regard to Figure 1.16. Bipolar transistor characteristics will follow in Chapter 2 and the relevant design rules are dealt with in Chapter 3. A quick appraisal of Figures 3.13(f) will serve to further illustrate the actual geometry of a BiCMOS bipolar transistor in n-well technology. Since extra design and processing steps are involved, there is an inevitable increase in cost and this is reflected in Figure 1.17; which also includes ECL and GaAs gates for cost comparison.

1.10.1 BiCMOS Fabrication in an n-well Process

The basic process steps used are those already outlined for CMOS but with additional process steps and additional masks defining: (i) the p^+ base region; (ii) n^+ collector area; and (iii) the buried subcollector (BCCD).



Note: For clarity, the layers have not been drawn transparent but BCCD underlies the entire area and the p^+ base underlies all within its boundary.

FIGURE 1.16 Arrangement of BiCMOS npn transistor (Orbit 2 μm CMOS).

Table 1.3 sets out the process steps for a single poly, single Metal CMOS n-well process, showing the additional process steps for the bipolar devices.

1.10.2 Some Aspects of Bipolar and CMOS Devices

Clearly there are relative advantages and disadvantages when comparing bipolar technology with CMOS technology. A readily assimilated comparison of some key features was set out as Table 1.2.

TABLE 1.3 n-well BiCMOS fabrication process steps

Single poly. single metal CMOS	*Additional steps for bipolar devices*
• Form n-well	
	• Form buried n^+ layer (BCCD)
• Delineate active areas	
• Channel stop	
	• Form deep n^+ collector
• Threshold V_t adjustment	
• Delineate poly./gate areas	
• Form n^+ active areas	
• Form p^+ active areas	
	• Form p^+ base for bipolars
• Define contacts	
• Delineate the metal areas	

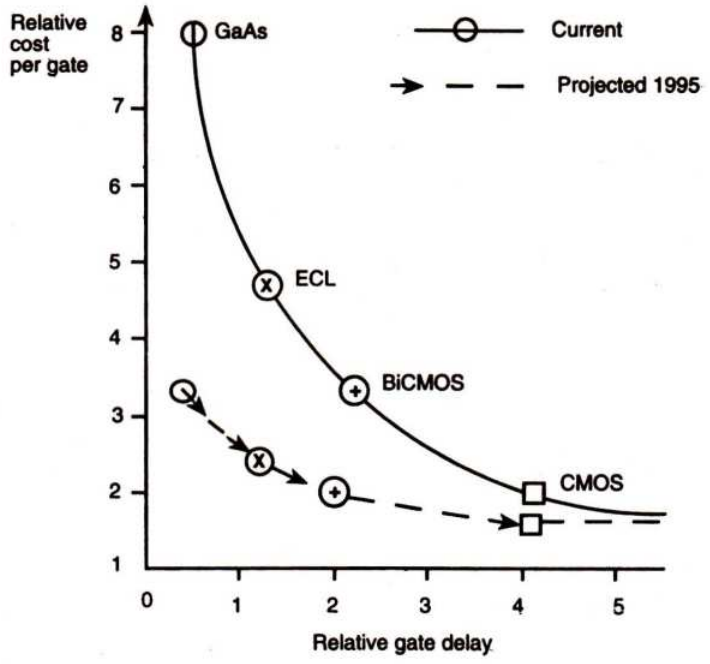


FIGURE 1.17 Cost versus delay for logic gate.

It will be seen that there are several advantages if the properties of CMOS and bipolar technologies could be combined. This is achieved to a significant extent in the BiCMOS technology. As in all things, there is a penalty which, in this case, arises from the additional process steps, some loss of packing density and thus higher cost.

A cost comparison of all current high speed technologies may be assessed from Figure 1.17.

A further advantage which arises from BiCMOS technology is that analog amplifier design is facilitated and improved. High impedance CMOS transistors may be used for the input circuitry while the remaining stages and ouput drivers are realised using Bipolar transistors.

To take maximum advantage of available Silicon technologies one might envisage the following mix of technologies in a silicon system.

CMOS	for logic
BiCMOS	for I/O and driver circuits
ECL	for Critical high speed parts of the system

However, in this text we will not be dealing with the ECL technology.

1.11 PRODUCTION OF E-BEAM MASKS

All the processes discussed have made use of masks at various stages of fabrication. In many processes, the masks are produced by standard optical techniques and much has been written on the photolithographic processes involved. However, as geometric dimensions shrink and also to allow for the processing of a number of different chip designs on a single wafer, other techniques are evolving. One popular process used for this purpose uses an E-beam machine. A rough outline of this type of mask making follows:

1. The starting material consists of chrome-plated glass plates which are coated with an E-beam sensitive resist.
2. The E-beam machine is loaded with the mask description data (MEBES).
3. Plates are loaded into the E-beam machine, where they are exposed with the patterns specified by the customer's mask data.
4. After exposure to the E-beam, the plates are introduced into a developer to bring out the patterns left by the E-beam in the resist coating.
5. The cycle is followed by a bake cycle and a plasma de-summing, which removes the resist residue.
6. The chrome is then etched and the plate is stripped of the remaining E-beam resist.

The advantages of E-beam masks are:

- tighter layer to layer registration;
- smaller feature sizes.

There are two approaches to the design of E-beam machines:

- raster scanning;
- vector scanning.

In the first case, the electron beam scans all possible locations (in a similar fashion to a television display) and a bit map is used to turn the E-beam on and off depending on whether the particular location being scanned is to be exposed or not.

For vector scanning, the beam is directed only to those locations which are to be exposed. Although this is inherently faster, the data handling involved is more complex.

1.12 OBSERVATIONS

This chapter has set the scene by introducing the basically simple MOS transistor structures and the relatively straightforward fabrication processes used in the manufacture of nMOS, CMOS and BiCMOS circuits. We have also attempted to emphasize the revolutionary spread of semiconductor technology which has, in the short space of 30 years, advanced to a point where we now see complex systems completely integrated onto a single chip.

Although this text concentrates on digital circuits and systems, similar techniques can be applied to the design and fabrication of analog devices. Indeed, the trends are toward systems of VLSI (and beyond) complexity which will in future include, on single chips, significant analog interfaces and other appropriate circuitry. This higher level of integration will lead to fewer packages and interconections and to more complex systems than today. There will be a marked beneficial effect on cost and reliability of the systems that will be available to all professions and disciplines and in most aspects of everyday life.

Our discussions of fabrication have in some instances simplfied the processes used in order to reveal or emphasize the essential features. Indeed, the fabrication of similar devices by different fabricators may vary considerably in detail. This is also the case with the design rules (Chapter 3) which are specified by the fabricator. Design rules will be introduced via the concept of "lambda-based" rules, which are a result of the work of Mead and Conway and, although not producing the tightest layouts, these rules are acceptable to many fabricators. A study of lambda-based rules also provide a good way of absorbing the essential concepts underlying any set of design rules. However, the text also gives an up-to-date set of real world "micron-based" rules for 2 μm and for 1.2 μm n-well CMOS technologies which may be used when the designer reaches an acceptable level of competence. The 2 μm rule set is for a BiCMOS process and thus also provides for bipolar npn transistors. It must be noted here that "2 μm technology", for example, means that the minimum line width (and, consequently, the typical feature size of the geometry) of the chip layout will be 2 μm.

In order to understand the basic features MOS and BiCMOS IC technologies, we must now look into the basic electrical properties.