CHAPTER

1

PRACTICAL CONSIDERATIONS

The field of VLSI design is a resource-intense engineering discipline. Project and product definitions are economically motivated, and competition on a worldwide basis is very keen. The market potential for innovative designs is very large, but the market window is often short due to both competition and changing consumer demands. Financial gain potential for both individuals and companies in this field is phenomenal, but the risks can also be very large. Some of the most advanced equipment and CAD resources available in any discipline are focused toward VLSI design and production; this focus makes the field very dynamic but also necessitates a continuing training and learning effort on the part of the design engineers to remain current and productive in this field.

It is our goal in this book to introduce basic electronic principles needed by the integrated circuit designer and to discuss engineering tradeoffs and practical considerations that are necessary for the student to make the transition from the classroom to industry as an integrated circuit designer. Although it is impossible to discuss all the practical aspects considered by experienced designers, it is our hope that through the discussions and comments presented in this book, the student will develop a sense of what types of practical questions must be addressed throughout the design process.

This chapter gives a brief historical overview, followed by a discussion of some of the terminology and jargon specific to the VLSI design field. (We have chosen to adopt the jargon used in the field because this is the language used by VLSI designers to communicate.) Size and complexity perspectives of VLSI circuits are discussed, and the basic types of processes used in IC and VLSI design are qualitatively summarized. The design process itself and the tools available to the designer are covered. Finally, a brief discussion of economics is presented to give the reader a basic appreciation of design costs and fabrication costs of

integrated circuits. Included is a simple discussion of the relationship between yield and chip area, which often is the key factor in determining whether a design will be economically viable.

1.0 INTRODUCTION

Historically, several events trace the evolution of what is currently termed VLSI technology. In the early 1930s, theoretical developments by Lilienfeld[1] and Heil[2] discussed the predecessor to what is now commonly called the field effect transistor (FET). Technological challenges delayed the practical utilization of this device for nearly three decades. In 1947 and 1948, three researchers at Bell Laboratories—Brattin, Bardeen, and Schockley—introduced the bipolar junction transistor (BJT). This development marked the practical beginning of the microelectronics industry. For the next 15 years, large numbers of different BJTs were produced and applied in a wide range of instrumentation systems. The BJTs replaced vacuum tubes in many applications and provided the impetus for a host of new electronic systems.

In the summer of 1958 Jack Kilby, an engineer at Texas Instruments, invented the first integrated circuit. Early the following year Robert Noyce of Fairchild independently reported on a procedure that more closely resembles integrated circuits of today. The specific details of Kilby's circuit are inconsequential, but the impact of his approach has been phenomenal.[3,4] The work of Kilby and Noyce marked the beginning of what has become the field of VLSI design.

Germanium was widely used as a semiconductor in some of the early discrete devices. Silicon has been the dominant semiconductor material used for integrated circuit fabrication for the past two decades, and most experts agree that it will remain dominant for the next decade. Since over 25% of the earth's crust is made of silicon, a real silicon shortage is highly unlikely! Other materials, such as gallium arsenide, are gaining acceptance in niche markets, which may be quite profitable.

Improvements in technology—ranging from improvements in materials and photolithography to advancements in processing—have been propelled by the significant financial gains offered to groups that excel in this area. Many integrated circuits of today contain a very large number of transistors, over 1 million in some designs. Conventional methods for circuit design that involved iteration at the breadboard level proved impractical for designing integrated circuits. This is due to poor designer productivity and the high cost associated with fabricating ICs. Methods of efficiently handling large quantities of design data were needed. Models of transistors that accurately predict experimental performance were required. Methods were needed for increasing designer productivity and reducing the design cycle time as the size and complexity of circuits increased.

The tools available now to the IC designer are very powerful and dynamic. Most require the use of large computers or, more recently, powerful graphics-intense workstations. The continued investment in both hardware and software needed for current integrated circuit design is high but is also crucial to remaining competitive. As powerful and dynamic as these tools may be, the fierce competition in the marketplace has resulted in the evolution of user-friendly software

with which the engineer can establish proficiency with a modest investment of time and effort. In the following chapters the tools needed to design VLSI circuits are investigated.

In spite of the sophistication and cost of both the hardware and software necessary to remain competitive in the field of VLSI design, most major contributions to this field are based upon relatively simple and basic innovations by the engineer. These innovations occur in circuit design, processing, and modeling as well as in the evolution of the CAD tools themselves. They are often made by young engineers. Because of the economic impact of innovations in the VLSI design field, advancement potential and rewards for talented and ambitious engineers are essentially unlimited. This potential exists for both young and old in institutions ranging from small start-up companies to the giants of the industry.

Integrated circuit fabrication requires the use of mechanical and optical equipment and materials capable of precisely maintaining close tolerances and small geometries. As with any high-technology field, a large amount of technical jargon has evolved, which must be mastered by anyone wishing to be conversant with those working in the area. The balance of this chapter is devoted to practical considerations and an introduction of terminology associated with the field of integrated circuit design.

An *integrated circuit (IC)* is a combination of interconnected circuit elements inseparably associated on or within a continuous substrate.

The *substrate* is the supporting material upon or within which an IC is fabricated or to which an IC is attached.

A *monolithic IC* is an IC whose elements are formed in place upon or within a semiconductor substrate with at least one of the elements formed within the substrate.

A *hybrid IC* consists of a combination of two or more IC types or an IC with some discrete elements.

A *wafer* (or *slice*) is the basic physical unit used in processing. It generally contains a large number of identical ICs. Typically, the wafer is circular; production wafers have a diameter of 4, 5, or 6 in.

The *chip* is one of the repeated ICs on a wafer. A typical production wafer may contain as few as 20 or 30 ICs or as many as several hundred or even several thousand, depending upon the complexity and size of the circuit being fabricated. The terms *die* and *bar* are used interchangeably for *chip* in some companies.

A *test plug*, or process control bar (PCB), or process control monitor (PCM), is a special chip that is repeated only a few times on each wafer. It is used to monitor the process parameters of the technology. After processing, the validity of the process is verified by measuring, at the wafer probe level, the characteristics of devices and/or circuits on the test plug. If the measurements of key parameters at the test plug level are not acceptable, the wafer is discarded.

A *test cell*, or *test lead*, is a special chip repeated only a few times on each wafer. It differs from the test plug in that the circuit designer includes this cell specifically to monitor the performance of elementary subcircuits or subcomponents.

Considerable effort has been expended toward using the entire wafer as a single IC, but this approach is challenged by defects in processing and the associated decline in yield and by the inherent delay in signals that must transverse the wafer. Those efforts are in the field termed wafer scale integration (WSI)[5-10].

1.1 SIZE AND COMPLEXITY OF INTEGRATED CIRCUITS

Integrated circuits are typically classified in terms of the number of devices or potential devices used in the design of the circuit and in terms of the feature size of the process. The device count is generally restricted to the number of active devices (either FETs or BJTs). As will be seen later, most integrated circuits contain large numbers of BJTs or FETs but contain few, if any, passive components. The classification of integrated circuits by device count is summarized in Table 1.1-1.

Classifications based upon feature size are also common. This classification is in terms of a typical minimum feature size (such as minimum gate length or minimum polysilicon width or minimum metal width) or in terms of the pitch (minimum of the sum of the minimum width of a feature and minimum spacing between similar features). The pitch is often nearly twice the minimum feature size. In the early to middle 1970s, the minimum feature size was typically 7 μ to 10 μ. In the late 1970s and early 1980s feature sizes to 5 μ were popular. In the mid-1980s, the minimum feature size had shrunk to under 2 μ, with some groups producing 1 μ and $1\frac{1}{4}$ μ circuits. The early 1990s should see practical submicron processes in production with feature sizes between 0.75 μ and 0.25 μ.

The impact of shrinking the feature size on silicon warrants discussion. For reference purposes, a sketch of a FET appears in Fig. 1.1-1. The FET is composed of a conductive gate region, which is separated from the surface of the substrate by a very thin insulating layer. Diffusions on either side of the gate form what are termed the drain and source regions. The minimum feature size of this

TABLE 1.1-1
Classification of integrated circuits by device count

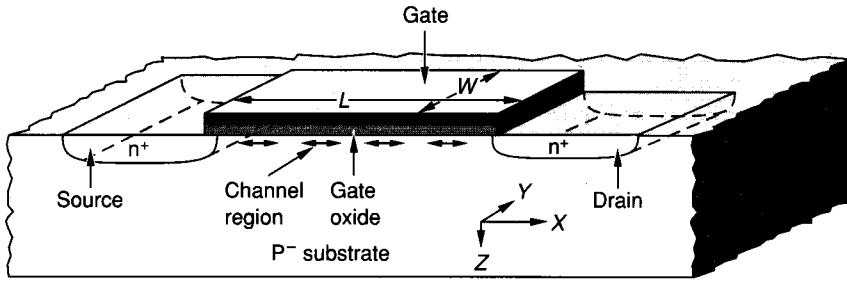| Nomenclature	Active device count	Typical functions
SSI	1–100	Gates, op amps, many linear applications
MSI	100–1000	Registers, filters, etc.
LSI	1000–100,000	Microprocessors, A/D, etc.
VLSI	10^5–10^6	Memories, computers, signal processors

FIGURE 1.1-1
Simplified 3-dimensional view of a FET.

process is roughly the minimum allowable value for L and W. For example, in a 5 μ process the minimum permissible value of L and W would be 5 μ. The area required for the gate of the transistor in such a process would be 25 μ^2. Even though the lateral dimensions of the FET (x and y directions in Fig. 1.1-1) are small, the vertical dimensions are typically much smaller. For example, the thin insulating layer under the gate in a typical 5 μ process is about 1000 Å thick. The relative perspective of the lateral and vertical (z direction) dimensions is grossly underemphasized in Fig. 1.1-1. Because of the large differences in lateral and vertical dimensions of FETs, the lateral dimensions are generally expressed in microns (or occasionally mils) and the vertical dimensions in angstroms. It is very important that the designer have an appreciation for both lateral and vertical feature sizes in any process. Conversions from meters to angstroms as well as a comparison with English units are given in Table 1.1-2. In this table, and throughout this book, the term *micron* and the abbreviation μ, which corresponds to the industry-accepted jargon for the micrometer, will be used.

  We are now in a position to develop a realistic perspective for the number of devices (transistors) that can be fabricated on a given piece of silicon. Assume initially that the area required for a single FET is essentially equal to the area

TABLE 1.1-2
Conversion of parameters used for device characterization in semiconductor industry

		Conversion				
Unit	Symbol	Angstroms	Microns	Mils	Meters	Inches
Angstrom	Å	—	10^{-4} μ	3.94×10^{-6} mil	10^{-10} m	3.94×10^{-9} in
Micron	μ	10^4Å	—	0.0394 mil	10^{-6} m	3.94×10^{-5} in
Mil	mil	2.54×10^5 Å	25.4 μ	—	2.54×10^{-5} m	0.001 in
Meter	m	10^{10} Å	10^6 μ	3.9×10^4 mil	—	39 in
Inch	in	2.54×10^8 Å	2.54×10^4 μ	10^3 mil	2.54×10^{-2} m	—

required for the gate (i.e., $W \cdot L$ in Fig. 1.1-1). With this assumption, a 4 inch wafer used in a 5 μ process can accommodate

$$N_{5\mu} = \frac{\pi(2 \text{ in})^2}{25\mu^2} \cdot \left(\frac{2.54 \times 10^4 \mu}{\text{in}}\right)^2 = 3.24 \times 10^8$$

transistors. Actually, due to spacing restrictions and interconnection requirements, the number of transistors that can be placed on this wafer will be from one to two orders of magnitude less. Regardless, it should be apparent that a very large number of transistors can be fabricated on such a wafer. The impact of shrinking the feature size can now be appreciated. If we could build transistors that were $0.5 \mu \times 0.5 \mu$, the number of transistors that could be accommodated by the same 4 inch wafer in the 0.5 μ process becomes

$$N_{0.5\mu} = \frac{\pi(2 \text{ in})^2}{0.25\mu^2} \cdot \left(\frac{2.54 \times 10^4 \mu}{\text{in}}\right)^2 = 3.24 \times 10^{10}$$

subject to the same reduction for spacing and interconnections as in the 5 μ process. Nonetheless, the 100-fold increase in device count is very significant.

To obtain an appreciation for the significance of a 100-fold increase in device count, assume one piece of silicon was used to design a small computer system. The same piece of silicon could be used to build 100 identical computer systems if a 100-fold increase in device count were obtained. Correspondingly, with a fixed chip area, the high-density circuit could perform the work of 100 of the small computer systems. In addition, from an economics viewpoint, the cost of fabricating wafers has increased only modestly as the device geometries have decreased.

Beyond the increase in device count, two other major benefits are derived by shrinking device sizes. First, as the device sizes decrease, the speed of circuits increases approximately linearly with feature size reduction. In the previous "small computer" example, it can be observed that in addition to obtaining a significant increase in the number of "equivalent computers" with decreasing device sizes, each of the smaller computers will work much faster! The other major benefit relates to yield, size, and complexity. It will be seen later that the yield depends primarily upon the silicon area (more precisely, active silicon area) of a chip and is relatively independent of the number or size of transistors in this area. Correspondingly, decreasing feature size makes possible some useful designs, which were either physically too large or which had low yields in a large feature size process.

There are some limitations associated with shrinking the feature size. These include a deterioration in matching characteristics, increased cost of equipment required for processing the wafers, additional capability requirements for software design aids, and an increased impact of interconnection delays. Concerns about increased power dissipation density and processing complications associated with heat cycling limitations during fabrication also exist. It is generally agreed,

however, that the benefits of shrinking the minimum feature size far outweigh the limitations, and a major worldwide research effort is ongoing to further shrink device sizes.

The number of devices that could potentially be placed on a wafer (calculated above) is strongly dependent upon the wafer size. Of more importance is the number of devices that can be placed on a chip, which represents a small portion of the area of a wafer as indicated in Fig. 1.1-2. From a practical viewpoint, the chip size seldom is (reference 1989) much in excess of 1 cm^2. The chip area of Texas Instruments' (TI) 1M DRAM is 0.54 cm^2. That of the Motorola 68020 microprocessor is 0.85 cm^2. Even in 1 cm^2, a large number of devices can be utilized. For example, in a 5 μ process, the 1 cm^2 chip can accommodate the gates of about 4 million 5 $\mu \times$ 5 μ transistors. As mentioned previously, the realistic number of practical devices is from one to two orders of magnitude smaller. For example, the 68020 microprocessor has about 200,000 transistor sites and was designed in a 1.8 μ process. The TI 1M DRAM, designed in a 1 μ process, has 1,048,576 transistors and an equal number of capacitors in the basic memory array, along with about 52,000 transistors in the control circuit. The TI 16M DRAM, which should be in volume production in 1991, will have a die area of nearly 1 cm^2, will have 16,770,000 transistors and an equal number of capacitors in the basic array along with over 150,000 transistors in the control circuit, and will be fabricated in a 0.6 μ process.

An analogy between the features on an integrated circuit and the features on the map of a large city is often drawn to obtain a realistic appreciation of the complexity of existing integrated circuits. This is motivated, in part, by the observation that under a high-power microscope, a dense integrated circuit shows a resemblance to a street map of a city with the interconnections corresponding to the city streets. Assuming that the pitch of a process maps to one city block, that a city block is 200 meters on a side, and that the pitch equals twice the minimum feature size, it follows that the magnification factor is $10^8{:}x$ where x
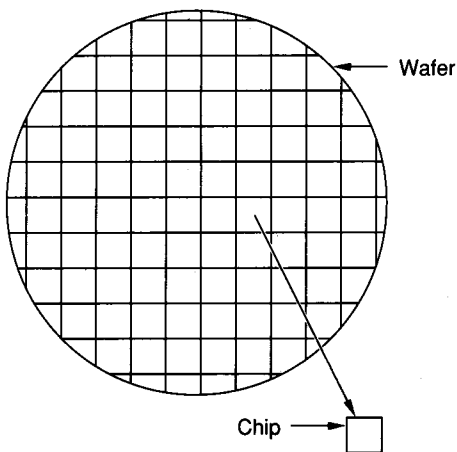

Wafer

Chip

FIGURE 1.1-2
Sketch of a wafer showing repeated "chips." (See Plate 1 in the color insert of this book for a color photograph of a commercial wafer.)

is the minimum feature size of the process in microns. For the 10 μ processes of the early 1970s with a typical die 2.5 mm on a side, this magnification would map the die to a city about 14 miles on a side, which corresponds to a city about the size of Tulsa, Oklahoma. For a 5 μ process with a die 5 mm on a side, which was popular in the late 1970s, the map increases to that of a city nearly 60 miles on a side. This corresponds to a city the size of the greater Chicago metropolitan area. For the 1 μ processes of the late 1980s, with a 1 cm^2 die size, the mapping is to a city 600 miles on a side, which would correspond to a city that is 30% larger than the entire state of Texas. Finally, for the projected 0.25 μ processes with a die 2 cm on a side, this same mapping would be to a city nearly 5,000 miles on a side. This would correspond to a city with an area equal to nearly half the earth's land surface.

Many integrated circuits require a silicon area that is considerably less than the maximum practical chip size. Several advantages are offered by using smaller-sized die. First, since the cost of processing a wafer is essentially independent of the size of the die, a smaller chip size will result in fabrication of more chips per wafer and, thus, a reduction in the effective cost per chip. Second, the yield (percentage of chips that are good) decreases rapidly with increasing chip size; details of this are discussed in Sec. 1.5. In addition, since rectangular chips are fabricated on round wafers, the amount of wafer wasted around the periphery is reduced with smaller chips.

Example 1.1-1. Assume an operational amplifier (op amp) requires an area 100 mil × 100 mil and a microprocessor requires an area 1 cm × 1 cm. (*a*) How many of each type of chip can be fabricated on a 5 inch wafer? (*b*) If the yield for the op amp is 98% and that for the microprocessor is 30%, compare the average number of good chips per wafer of each device that can be anticipated. (*c*) If the fabrication cost per wafer is $400, what is the effective cost per good chip for each device?

Solution. (*a*) Neglecting the area loss on the periphery of the wafer, we calculate the number of op amps and microprocessors as

$$n_{\text{opamp}} = \frac{\pi(2.5 \text{ in})^2}{(0.1 \text{ in})^2} = 1963$$

$$n_{\mu\text{proc}} = \frac{\pi(2.5 \text{ in})^2}{(1 \text{ cm})^2} = 126$$

(*b*)

$$n_{\text{opamp,effective}} = (0.98)(1963) = 1923$$

$$n_{\mu\text{proc,effective}} = (0.3)(126) = 37$$

(*c*)

$$C_{\text{opamp}} = \frac{\$400}{1923} = 20.8¢$$

$$C_{\mu\text{proc}} = \$10.53$$

One naturally poses the question: How small will device sizes ultimately become? Although we will not quantitatively answer this question, we will address some of the major factors which place limits on decreasing device dimensions.

Up to now, limitations were imposed primarily by limitations in resolution of processing equipment. We are approaching the point, however, where the physics of the semiconductors themselves are starting to cause problems. Gate oxides (the insulating layer under the gate) below 100 Å thick are being investigated in conjunction with submicron research efforts. The density of silicon atoms in single crystal silicon is 5×10^{22} atoms/cm^3. This corresponds to an "average" atom spacing of 2.71 Å. The nearest-neighbor distance is 1.18 Å and the lattice constant is 5.43 Å. It should be apparent that the sub–100 Å gate oxide layers have dimensions approaching the dimensions of the atomic structure of the semiconductor crystalline structure itself. Silicon dioxide, which is typically used as the insulating layer, is even coarser than silicon. The silicon dioxide density is about 2.3×10^{22} molecules/cm^3, with an "average" molecular spacing of 3.52 Å. It should be apparent that irregularities in surfaces of the order of magnitude of a few molecules become significant in sub–100 Å oxide layers. Quantum mechanical tunneling occurs if oxide thicknesses become thinner than about 50 Å, thus placing a practical lower bound on oxide thicknesses.[11]

High electric field strengths, which may cause device failures, also are of concern. Voltages up to 5 V are regularly applied across the 1000 Å silicon dioxide insulating layers. This corresponds to electric fields of the order of magnitude of

$$E_{1000 \, Å} = \frac{5 \text{ V}}{1000 \text{ Å}} = \frac{500 \text{ kV}}{\text{cm}}$$

This electric field is *very large* but still less than the breakdown field of silicon dioxide, which is in the 5–10 MV/cm range depending on how the oxide was grown. If, however, the same 5 V were applied to the 100 Å oxide layer, the electric field strength would be in the neighborhood of the breakdown field for the oxide. Furthermore, the irregularities in the very thin oxide layers can cause further significant local increases in field strength. The only option is to decrease the voltage applied to the oxide layer, but this is unattractive for two reasons. First, as the voltage across this oxide decreases, noise effects become more significant, thus increasing the chance of occasional errors in circuits using these devices. Second, existing systems have well-defined voltage levels, which a large number of manufacturers adhere to. Since parts made by various manufacturers are often interconnected to form complex systems, interfacing parts with nonstandard signal levels causes a significant increase in design complexity.

It is generally desirable to scale the vertical as well as the lateral dimensions when decreasing device sizes. The method under which this scaling occurs affects both the reliability and performance specifications of the process. Various scaling strategies are possible as the lateral dimensions of the MOSFETs decrease. In the *constant field scaling strategy*, the vertical dimensions typically decrease at the same rate as the lateral dimensions. To maintain a fixed electric field, the operating voltage also decreases at the same rate. The *constant voltage scaling strategy* is attractive because electrical compatibility with existing circuits is maintained. In a constant voltage scaling strategy that has been proposed, the vertical dimensions decrease quadratically relative to the lateral dimensions.[12]

Obviously tradeoffs among performance, yield, compatibility with existing technology, reliability, process complexity, device performance, and impact of parasitics must be made when selecting a scaling strategy.

1.2   THE MICROELECTRONICS FIELD

The microelectronics field is quite broad. Many different types of processes and approaches have found niches in the microelectronics market place. Figure 1.2-1 depicts the major processes that have received a reasonable degree of acceptance.

The first division in process types occurs between active and inert substrates. The high-volume integrated circuits typically utilize the active substrates. Some of the more demanding requirements as well as specialized and/or low-volume circuits use the inert substrates. The inert substrates are also used in most hybrid ICs. These latter circuits are often noted for requiring relatively modest investments in processing equipment, but the consumer cost of the hybrid ICs themselves is quite high.

Two types of processes that utilize inert substrates are particularly important. These are the thin and thick film processes. These processes are capable of producing good resistors with attractive temperature characteristics. This is difficult to achieve with the standard active substrate processes.

The active substrate is generally silicon or doped silicon although considerable research effort has been expended over the last decade in using gallium arsenide (GaAs). Two primary separate types of silicon processes have evolved. The bipolar process uses the BJT as the basic active device whereas the MOS processes use the metal oxide semiconductor field effect transistor, or MOSFET (sometimes termed IGFET for insulated gate FET), as the basic active device.

The bipolar process was the most popular through the 1960s and early 1970s. The bipolar process offers potential for operation at very high frequencies and offers some performance advantages such as large transconductances, which are of benefit in many linear applications. The power dissipation in bipolar integrated circuits is, however, often quite high, and the device density is not as high as that attainable with the MOS processes. The popular TTL logic family, ECL, and I^2L all fall under the bipolar label. Many linear ICs also are fabricated in the bipolar processes. Although the relative amount of research effort in the bipolar area is small compared to that focusing on the MOS processes, the production volume of bipolar ICs is still very large, and some new developments still use the bipolar process. There is, however, considerable development work ongoing in the design of smart bipolar power devices, which contain the control circuitry along with the power devices.

The MOS process is often divided into three categories: NMOS, PMOS, and CMOS. The basic devices in MOS processes are the p-channel and n-channel MOSFETs discussed in Chapter 2. The term PMOS refers to a MOS process that uses only p-channel FETs. The PMOS process was used in some of the earlier MOS designs, but is rarely used today primarily because the electrical characteristics of p-channel MOSFETs are not as attractive as those of n-channel MOSFETs. This is because the mobility (discussed in Chapter 3) of p-type material is considerably poorer than that of the n-type material. The term
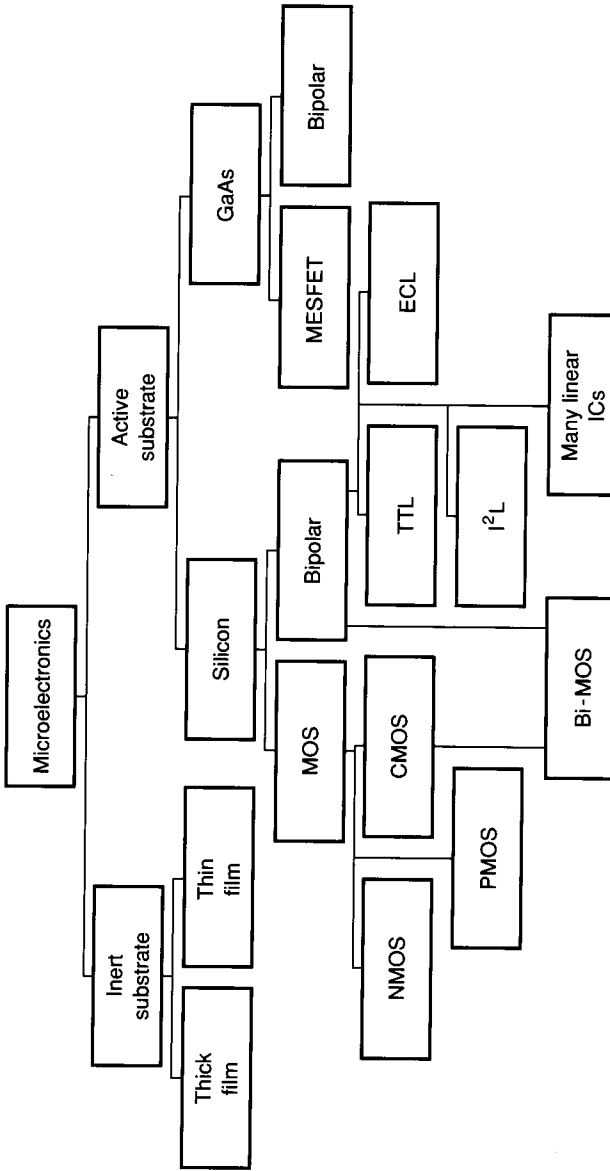
FIGURE 1.2-1
Types of major processes used in IC fabrication.

11

NMOS refers to a MOS process using only n-channel FETs. Excellent density and reasonable performance characterize the NMOS process. The term CMOS (complementary MOS) refers to a MOS process that simultaneously provides both n-channel and p-channel devices. The availability of both types of FETs offers the designer considerable additional flexibility over that attainable with either an NMOS or PMOS process. In digital applications, the availability of complementary devices offers potential for *very* low static power consumption. In analog applications, the circuit complexity can often be reduced in the CMOS process relative to what is attainable with either the NMOS or PMOS processes. The increased flexibility of CMOS is partially offset by increased fabrication costs and an increase in the silicon area required to implement basic digital functions. The tradeoffs generally favor the CMOS process over NMOS in most new designs. MOS processes are used for most VLSI scale circuits. Applications include memories, interfacing, microprocessors, basic logic functions, and a host of linear and mixed linear and digital applications.

Recently there has been a major effort toward combining both bipolar and MOS devices in a single process. This more complex and expensive process, termed Bi-MOS, is becoming cost-effective in a growing class of applications. Some processes also include thin film components with MOS and/or bipolar devices, but the expense associated with adding the thin film layer is justifiable only in specialized applications.

1.3 IC DESIGN PROCESS

It is generally the goal of the IC designer to design an integrated circuit that meets a given set of specifications while expending minimal labor and physical resources in a short time frame. Furthermore, the production yield should be high, the process simple, and the die area small. The conventional approach to circuit design often involves much iteration at the breadboard level. Because of the complexity of many VLSI designs and the cost of resources associated with IC design and fabrication, the conventional approach is totally unacceptable. A simple example is useful for obtaining an appreciation for the magnitude of the task facing the VLSI designer.

> **Example 1.3-1.** Assume that the productivity of a "conventional" discrete component circuit designer is measured in terms of the average number of transistors per day that the designer produces for a circuit and that the productivity is independent of the complexity of the circuit. If it is assumed that a two person-month effort is required to design a 20-transistor circuit following a conventional approach, how long will it take the same designer to design a circuit that has 500,000 transistors using the same design approach? How large will the circuit schematic be if it requires an average of 2 cm^2 of space for each transistor in the circuit?
>
> *Solution.* The productivity rate of the designer is 20 transistors/2 person-months = 10 transistors/person-month. Thus, the 500,000-transistor VLSI circuit would require 50,000 person-months, or about 4200 person-years. Note that 4200 person-years is equivalent to about 105 productive person-lifetimes! The schematic would occupy $(500,000)(2)$ cm$^2 = 10^6$ cm^2 of area. This is the area of a square 10 m on a side!

From Example 1.3-1, two things should be apparent. First, designer efficiency *must* be improved in the field of VLSI design. Second, much more efficient methods of handling large amounts of design data associated with schematic drafting, layout, and simulation are required.

At the outset, one might suspect that designer productivity will actually decrease with circuit complexity. Although this is typically the case for unstructured designs, most existing VLSI circuits are regularly structured and utilize a small number of basic circuits a large number of times. Powerful design aids, mostly in the form of powerful computer programs, are crucial for the successful design of VLSI circuits.

Two approaches to IC design are philosophically identifiable. In the first, called a *bottom-up approach*, the designer starts at the transistor or gate level and designs subcircuits of increasing complexity, which are then interconnected to realize the required functionality. In the second, termed the *top-down approach*, the designer repeatedly decomposes the system-level specifications into groups and subgroups of simpler tasks. The lowest-level tasks are ultimately implemented in silicon, either with standard circuits that have been previously designed and tested (often termed *standard cells*) or with low-level circuits designed to meet the required specifications. In the extreme case, the top-down approach results in a *silicon compiler*, discussed later, in which all blocks are automatically designed with a computer.

The top-down approach is used for some digital designs and often results in a significant increase in designer productivity. Considerable effort has been expended at following the top-down approach for analog design, but analog design requirements are sufficiently specialized that the top-down approach is currently practical only in certain classes of analog designs. It is often the case that both analog and digital system designs use varying combinations of top-down and bottom-up design concepts.

A block diagram of the conventional IC design process is shown in Fig. 1.3-1. The starting point is a set of design specifications. On complicated designs, a major effort is required to obtain a complete set of circuit and system specifications.

Preliminary designs are based upon simple models of devices or subcircuits. These are typically at the behavioral or logic level for digital circuits and at the component or device level for analog circuits. A preliminary computer simulation using more accurate models is used to verify the performance of the preliminary design. Good device and subcircuit models are crucial. A model is "good" if it accurately predicts experimental performance after fabrication and is sufficiently simple to avoid the requirement of excessive computer time during the simulation. Considerable time is often invested in the initial computer simulation and preliminary design loop.

Once the preliminary design is deemed acceptable, the actual layout takes place. The layout phase is often entered on subcircuits prior to the completion of the preliminary design phase. A good overall floorplan is obtained early in the design after a good estimate of the overall architecture and size of the subcircuits is obtained. The floorplan contains all major busing and cell (subcircuit) placement information as well as I/O pad designations.
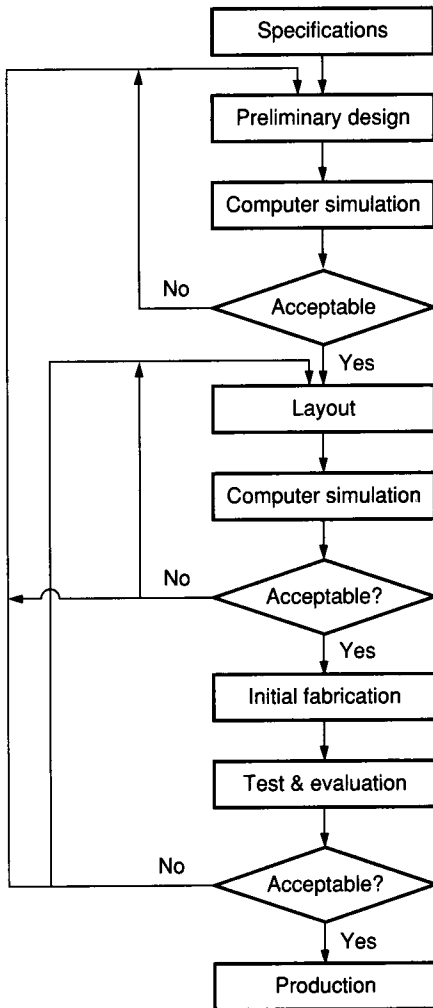
```
                    ┌─────────────────┐
                    │  Specifications │
                    └─────────────────┘
                             │
                    ┌─────────────────┐
                    │ Preliminary design │
                    └─────────────────┘
                             │
                    ┌─────────────────┐
                    │ Computer simulation │
                    └─────────────────┘
                             │
         No          ◇ Acceptable ◇
                             │ Yes
                    ┌─────────────────┐
                    │     Layout      │
                    └─────────────────┘
                             │
                    ┌─────────────────┐
                    │ Computer simulation │
                    └─────────────────┘
                             │
         No          ◇ Acceptable? ◇
                             │ Yes
                    ┌─────────────────┐
                    │ Initial fabrication │
                    └─────────────────┘
                             │
                    ┌─────────────────┐
                    │ Test & evaluation │
                    └─────────────────┘
                             │
         No          ◇ Acceptable? ◇
                             │ Yes
                    ┌─────────────────┐
                    │   Production    │
                    └─────────────────┘
```

FIGURE 1.3-1
Block diagram of conventional IC design process.

Additional computer simulation is undertaken on subcircuits after layout and ultimately on the entire circuit. These simulations are crucial since parasitic effects associated with the layout play a significant role in both analog and digital circuits. In analog circuits the parasitics tend to degrade performance specifications whereas in digital circuits the parasitics generally cause additional unwanted delays or potentially disastrous race conditions. Software also exists at this level in the design for verifying that the layout violates no major design rules and that the circuit in the layout corresponds to the initial circuit schematic. The results of the computer simulation often necessitate changes in the layout or changes in the design itself if the effects of the parasitics cannot be resolved with changes in the layout.

Following an acceptable computer simulation of the entire circuit, the circuit is committed to fabrication. In evolving processes or complicated designs,

subcircuits and/or test structures are often fabricated early in the design process to provide modeling information and/or verify functionality of subcircuits. The commitment of the design to fabrication is an expensive proposition. Dollar costs for a single fabrication run of $20,000 to $40,000 are common, and delays of 1 to 4 months from submission of the design for fabrication to physical silicon for testing are typical. A *single* error in the circuit design, simulation, or layout generally makes the circuit either partially or totally nonfunctional. Based upon the experimental evaluation, either the circuit is released to production or the appropriate step of the design process is re-entered. Although first silicon is often not acceptable for production, the timeliness of the market window and the costs associated with fabrication make repeated iteration at the silicon level unacceptable. Improvements in design tools and methodologies are, however, producing a rapidly growing trend toward fully functional first silicon.

Significant resources have been invested on a worldwide basis at increasing the effectiveness of the computer-based design aids. A common goal is to automate repetitive and tedious tasks and minimize the amount of human interaction required in the design process. The rationale for this goal is twofold. First, it helps improve designer productivity, and second, it helps reduce errors that are often directly attributable to the designer. Additional information about the CAD tools available to the IC designer is provided throughout this text. Some of the tools and alternative approaches available to the designer deserve mention at this point.

Gate arrays and "seas of gates" are becoming quite popular for certain applications such as replacing large number of SSI chips with a single IC. Such arrays are integrated circuits containing a large number (several hundred) of digital gates or transistor cells, which can be used to implement complicated logic functions. These circuits are initially processed up to but not including the interconnection layers to form a versatile and generic building block. Computer programs that determine the interconnections necessary to implement customer-specific logic functions are widely available and are used to determine the required interconnections. One or more interconnection layers are then added to the generic array circuit, resulting in an economical, production-ready integrated circuit that can be fabricated in a few days. These mask programmable arrays are most practical in low- to medium-volume applications where custom design and production costs cannot be justified or in applications that require very quick turn-around. Silicon area is not, however, efficiently utilized in these arrays, making them less attractive for high-volume applications, in which silicon costs become a major factor in overall fabrication costs.

Some groups have been successful at automating the design process in restricted applications to the extent that the input to a computer program is the system specification and the output is a layout that should yield production-ready silicon. These programs are called silicon compilers, and they partially remove the conventional IC designer from the design cycle. The resulting simplification on the typical block diagram of Fig. 1.3-1 should be apparent. These programs support the premise that the circuits will be correct by construction, thus min-

imizing the need for verification by simulation. Although it might appear that silicon compilers will eventually make the IC designer obsolete, this is highly unlikely. These types of CAD tools will, rather, provide the experienced designer with additional capability for meeting more complicated and challenging design goals and provide basic silicon access to other design engineers.

1.4   ECONOMICS

Although the field of VLSI design is challenging from both engineering and scientific viewpoints, neither government nor industry is willing to support its evolution on these merits alone. Both research and developmental efforts are generally focused toward areas where investments can be recouped in the marketplace within a relatively short time. It is thus crucial that the designer be familiar with the economics of IC production.

Two questions naturally arise prior to developmental efforts on any VLSI circuit. First, does a sufficiently large market exist to justify development? Second, can a product be developed and produced so that a reasonable profit will be realized over the expected life of the product? We will not attempt to consider the first question in this text, but the second question is easier to address. On a product that is anticipated to have a relatively small total sales volume, the developmental costs will typically dominate. On a product with a large projected sales volume over the life of the product, the actual production costs dominate.

The developmental costs can be estimated once the amount of engineering effort required to bring a product into production is known. It should be pointed out that even for seemingly simple designs, the developmental costs may become quite large. For example, a design that requires a 12-month effort by an experienced designer may accrue a total project development cost of $350,000 to $450,000 or more even though the base salary of the designer may be only $40,000/year. The burden (multiplier on the base salary of the designer) is quite high because of various expenses: standard fringe benefits the employer generally provides, technician support, applications engineering, computer time charges, equipment amortization, documentation preparation, test procedure development, mask generation, pilot production, and so on. The burden factor varies considerably from project to project and from company to company. Detailed estimates of the specific costs will generally be made because of differing project requirements. Information about design costs within a company is highly proprietary. Nevertheless, it should be apparent that IC design is a very expensive proposition.

The production costs are somewhat easier to estimate. These can be grossly decomposed into the costs listed in Table 1.4-1; typical values are listed in Table 1.4-2. The wafer fabrication costs are dependent on the size of the wafer, the number of mask steps, and the type of processing done at each step. These costs are determined by labor costs, materials and maintenance costs, and amortization costs of the processing equipment. Correspondingly, the wafer probe and final test results are dependent on both type and complexity of the circuit itself. Because good testers are quite expensive ($500,000 to $1.5 million for good digital testers

TABLE 1.4-1
Major costs associated with wafer processing and fabrication

		Cost	
		Per wafer	Per die
Wafer fabrication			
Blank wafer	x_1	✓	
Wafer processing	x_2	✓	
Wafer probe	x_3	✓	
Wafer sawing	x_4	✓	
Die attach, bonding,			
and packaging	x_5	✓	
Packaging	x_6		✓
Final test	x_7		✓

TABLE 1.4-2
Typical processing and packaging costs for 12-mask, $3\ \mu$ CMOS process (1988) based upon volume production

	Processing costs	
	4" process	5" process
Wafer fabrication		
Blank wafer	$x_1 = \$10$	$x_1 = \$15$
Wafer processing	$x_2 = \$140$	$x_2 = \$150$
Wafer probe (per wafer)	$x_3 = \$25$	$x_3 = \$40$
Wafer sawing (per wafer)	$x_4 = \$3$	$x_4 = \$3$
Die attach and		
bonding (per wafer)	$x_5 = \$3$	$x_5 = \$5$
Packaging	x_6 (see below)	x_6 (see below)
Final test (per package)	$x_7 = 30¢/cm^2$	$x_7 = 30¢/cm^2$

	Package costs†	
Plastic DIP	8 pin	$0.032
Plastic DIP	16 pin	0.048
Plastic DIP	24 pin	0.091
Plastic DIP	64 pin	0.70
Ceramic side brazed	16 pin	1.05
Ceramic side brazed	24 pin	1.50
Ceramic side brazed	64 pin	4.95
Ceramic CERDIP	16 pin	0.096
Ceramic CERDIP	24 pin	0.26
Ceramic CERDIP	40 pin	0.64
Ceramic pin grid array	68 pin	6.40
Ceramic pin grid array	84 pin	7.50
Ceramic pin grid array	132 pin	10.15
Ceramic pin grid array	224 pin	18.00

†Packaging cost estimates courtesy Dr. W.E. Loeb of W.E. Loeb and Associates.[13]

in 1989), these costs become significant if much time on the tester is required. The die attach and bonding costs shown in Table 1.4-1 are listed as a per-wafer cost. Some groups prefer to figure these as per-die costs instead.

The packaging costs depend on both the type and size of the package. As is shown in Table 1.4-2, low pin count plastic packages cost a few cents, whereas ceramic packages with a large number of pins could cost several dollars. Pin styles also differ. The standard dual inline packages (DIP) have recently been replaced, in some applications, with pin grid arrays and surface mount packages. The plastic surface mount small outline integrated circuit (SOIC) packages are a little less expensive than the DIP structures shown in Table 1.4-2, as are the plastic lead chip carriers (PLCC). For example, a 68-pin PLCC would cost about $0.30 compared to $0.70 for the 64-pin plastic DIP.

Since the processing equipment is quite expensive (a single line for processing standard CMOS 5 inch wafers in a 2 μ process may cost $30 million) and the labor costs to keep a line operational are quite high, the amortization costs are also dependent on the accounting procedures used by the individual companies. For example, during periods when the demand for semiconductors is high and the production facilities are running near capacity, the fabrication costs per wafer are reduced if the costs are based upon instantaneous amortization costs. Correspondingly, when the demand for semiconductors is soft, this method of accounting results in significant increases in the per-wafer fabrication costs, which may be somewhat misleading. These economy-dependent variances could be as much as 2:1 or more.

It may appear from the information presented in Table 1.4-2 that elimination of the wafer probe step could reduce system costs. In general, this is far from true, as indicated by Example 1.4-1.

> **Example 1.4-1.** Compare the costs of producing an integrated circuit in a conventional 3 μ, 12-mask CMOS process with the options for processing, testing, and packaging as shown in Table 1.4-3. Assume the die size is 0.5 cm × 0.5 cm, the yield at wafer probe is 50%, the packaging yield (sawing, die attach, bonding, and packaging) is 90%, and that Table 1.4-2 gives realistic values for both fabrication and packaging costs. Approximate the number of potential dies by the wafer/die-area ratio, and assume that dies are discarded as soon as they are determined to be defective.

TABLE 1.4-3
Processing, testing, and packaging options for Example 1.4-1

Option	Wafer size	Wafer probe	Type of package
1	4 in	Yes	16-pin plastic
2	5 in	Yes	16-pin plastic
3	4 in	Yes	16-pin sidebrazed ceramic
4	4 in	No	16-pin plastic
5	4 in	No	16-pin sidebrazed ceramic

Solution. The number of potential dies for the 4 inch and 5 inch wafers are approximately

$$N_4 = \frac{\pi(2 \text{ in})^2}{(0.5 \text{ cm} \times 0.5 \text{ cm})} = 324$$

$$N_5 = \frac{\pi(2.5 \text{ in})^2}{(0.5 \text{ cm} \times 0.5 \text{ cm})} = 506$$

The actual number of potential dies is somewhat smaller since all potential dies around the perimeter of the wafer are incomplete and hence useless. These effects become more significant as the die size increases and as the wafer size decreases.

For options 1, 2, and 3, the production cost per good package becomes

$$C = \left[\left(\frac{x_1 + x_2 + x_3 + x_4 + x_5}{N} \right) \left(\frac{1}{\theta_{\text{probe}}} \right) + x_6 + x_7 \frac{A_W}{N} \right] \left[\frac{1}{\theta_{\text{package}}} \right] \quad (1.4\text{-}1)$$

where θ_{probe} and θ_{package} represent probe and package yields respectively, x_1–x_7 are as in Table 1.4-1, A_W is the wafer area, and N is the number of potential dies.

For options 4 and 5, the production cost per good package becomes

$$C = \left(\frac{x_1 + x_2 + x_4 + x_5}{N} + x_6 + x_7 \frac{A_W}{N} \right) \left(\frac{1}{\theta_{\text{probe}}} \right) \left(\frac{1}{\theta_{\text{package}}} \right) \quad (1.4\text{-}2)$$

Substituting from Table 1.4-2, we obtain the following costs for options 1–5 respectively, $C_1 = \$1.38$, $C_2 = \$1.07$, $C_3 = \$2.49$, $C_4 = \$1.34$, and $C_5 = \$3.57$.

This example clearly demonstrates several points. First, very significant differences in packaging costs exist. The ceramic packages are much more expensive than plastic. Ceramic packages do, however, provide better isolation from the ambient external environment and are necessary in some high-reliability applications. The importance of the wafer probe step for low-yield parts packaged in expensive packages should be apparent. This allows defective dies to be identified and culled prior to packaging, thus avoiding the expensive proposition of packaging defective dies. The wafer fabrication and packaging costs considered in the example are reasonably independent of circuit complexity. The fabrication yield, however, is strongly a function of die size; it decreases rapidly with increasing die size, as will be demonstrated later in this chapter. In this example, the value of the die itself constitutes a relatively small portion of the overall IC cost. This is typical for small dies. For large dies, the die value generally dominates because of the significant decrease in yield. The probe and final test costs were also relatively small in this example, but testing costs become very significant for larger complicated circuits.

1.5 YIELD

Generally, some of the dies on any wafer do not meet the performance specifications. The percentage of dies that do meet performance specifications is termed the *wafer yield*. In general, the yield decreases rapidly with increasing die area. Consequently, die area plays a major role in the economics of IC design, to

the extent that estimates about the die area and yield are almost always considered when making decisions about whether IC development for specific needs will be economically justifiable. A basic understanding about the economic impact of yield is very important, and since some of the terminology needed to discuss yield is first introduced in Chapters 2 and 3, the reader may need to make an occasional forward reference to those chapters while reading this section.

Accurate IC yield prediction is difficult to obtain due to the variety of factors that impact yield. Yield is dependent on the specific type of circuit, the design methodology followed by the designer, the layout itself, and the physical fabrication process. The factors that affect yield include *dust* particles (by definition, any unwanted foreign objects in solid, liquid, or gaseous state), crystal defects, mask defects, alignment errors, breakage and human handling errors, and parameter drifts. The dust particles on wafers, which introduce local defects, are foreign particles that adversely affect either the photolithographic process or the subsequent processing steps. If large enough, these dust particles will cause problems such as failure of a transistor, breaks in an interconnection, or shorting of two adjacent devices, levels, or interconnections. These types of defects are generally assumed to be randomly distributed over the surface of the wafer and from wafer to wafer. Dust particles on a mask or reticle, if large enough, will also cause failures that are repeated every time that portion of the reticle or mask is used. These defects typically occur at the same geometrical position on all wafers in a lot.

Crystal defects are present in the wafer prior to fabrication and cause local circuit defects such as failure of transistors. Mask defects may be local or global, depending on the cause. Alignment errors will typically be limited to a single wafer. It is, however, often the case that all dies on the wafer will be defective if alignment errors occur.

Parameter drifts will cause transistors to have characteristics different from those desired. These parameter drifts are introduced during processing and are attributable to the inherent practical and physical limitations associated with processing steps such as photolithography, deposition, and diffusion. These changes may affect all wafers in a lot or may affect individual transistors on a wafer. The average values of key parameters at the lot level are closely monitored. Lots are rejected if these values lie outside a predefined acceptability region. On a single wafer, the parameters are considered to vary statistically from transistor to transistor. Statistical circuit parameter variations large enough to cause circuit failures are more common in analog circuits than in digital circuits. Failures due to parameter variations are termed *soft* faults, and those failures that cause complete transistor failure or interconnection errors are termed *hard* faults. Whereas a hard fault generally results in nonfunctionality of part or all of the circuit, circuits with soft faults may well be functional but fail to meet specifications. A single hard fault or soft fault will often cause rejection of a die. Most digital circuits are designed so that they are unaffected by all but the most serious parameter variations. Thus, failures in digital circuits generally result from hard faults. Both soft faults and hard faults are of concern in analog circuits.

Faults are bad for two reasons. First, since they cause failure of the die, they drive up the effective costs of a good die. Second, they are often difficult and

expensive to detect. In fact, questions about whether some complicated circuits are even testable arise. It is now an accepted responsibility of the designers to address the testing problem at the design stage. ICs often contain additional circuitry, which is used in the testing of the circuit itself. The testing question is definitely not trivial for complicated designs, and it represents an active area of research. Related to this problem is the field of fault-tolerant technology, also an active area of research, in which circuits are designed so that they remain functional even if some faults do exist.

The probability of having one or more hard faults on any die associated with particulate material or crystal defects generally increases with both die area and circuit complexity. The reason for the increase with die area should be apparent; it is more likely that a die will contact some fixed defects on a wafer if the die area is increased. This is illustrated in simplified form in Fig. 1.5-1, where varying die sizes are considered with a fixed defect distribution. It should be apparent from this example that with 52 potential dies, 45 will be free of defects (Fig. 1.5-1b). With 12 potential dies (Fig. 1.5-1c), 7 will be free of defects, and with 4 potential dies none will be free of defects (Fig. 1.5-1d). In this example, yield decreases from 86% to 0% as the die size increases.

Although not accounted for in the previous example, the density of defects which will cause die failure also increases with die complexity. If a die is
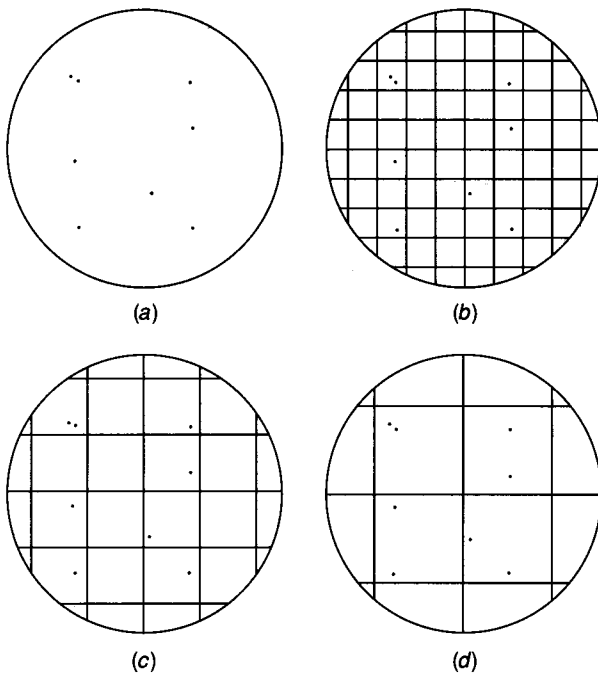

FIGURE 1.5-1
Effects of defects on yield: (a) Defect locations, (b) Small die size, (c) Larger die size, (d) Die size with zero yield.

not heavily utilized, some defects may occur in areas where no components or interconnections exist, so die failure will not result. Likewise, some types of defects are more likely to cause a fault if they occur where a transistor is located than where interconnections are made. It can be shown, however, that for a fixed circuit schematic overall yield decreases with increasing die size and increases if the devices are densely packed on the die. The reader is thus cautioned not to draw the conclusion that the yield can be improved by decreasing the density if more silicon area for a given circuit is used. In general, the decrease in failure density on a die with a loosely packed circuit will be more than offset by the total number of defects anticipated in the die with larger area.

Several models are used to predict yield associated with crystal- and particulate-related defects. A simple model based upon using the Poisson distribution to predict the probability of K defects in a die area, A, with an average defect density of D defects per unit area results in the expression for the probability that a given die of area A has zero defects:

$$P = e^{-AD} \tag{1.5-1}$$

The defect density, D, is typically in the range of 1 to 2/cm^2.

Since many defects cause a failure only if the defect occurs in an active area, it may be preferable to use the active area and the corresponding average active area defect density in Eq. 1.5-1. Alternatively, even better results should be obtainable if the active and interconnection defects are considered separately. If A_A and A_I denote the active and interconnect areas, respectively, and D_A and D_I denote the average defect densities in those two regions, then the probability that a die is good can be obtained by modifying Eq. 1.5-1

$$P = e^{-(A_A D_A + A_I D_I)} \tag{1.5-2}$$

Example 1.5-1. If the average defect density is 1/cm^2, what is the average cost per good die for a 4" wafer if the die size is 1 cm^2? 0.2 cm^2? What would be the price per good die if the average defect density increased to 2/cm^2? Neglect die losses associated with the wafer edges. Assume the cost for the 4" wafer and processing is $200 and that Eq. 1.5-1 models the probability that a device is good.

Solution. The average cost per good die is 200/NP where P is the probability the die is good and N is the number of potential dies per wafer. The number of potential 1 cm^2 dies in a 4" wafer is (neglecting edge losses)

$$N_{1cm} = \frac{\pi \left[(2 \text{ in})(2.5 \text{ cm/in}) \right]^2}{1 \text{ cm}^2} = 81$$

Likewise

$$N_{0.2cm} = 405$$

For a 1/cm^2 defect density, it follows from Eq. 1.5-1 that the probability that the 1 cm^2 die is good is 36%, whereas the probability that the 0.2 cm^2 die is good is 82%. The corresponding cost per good 1 cm^2 die is thus $6.86. The cost per good 0.2 cm^2 die is 60¢. If the defect density were to increase to 2/cm^2, the average

cost per good 1 cm^2 die increases to \$18.24 and the 0.2 cm^2 die increases to 74¢. From this example it should be apparent that increasing either the defect density or die area significantly increases the average die cost once the product of A and D exceeds unity.

Example 1.5-2. If the average defect density is 1.5/cm^2, what is the yield if the die size is increased to that of an entire 4" wafer? On the average, how many such wafers must be fabricated to give one good die? If the cost of each wafer is \$200, what would be the average cost per good die? Use the simple model of Eq. 1.5-1 to predict yield.

Solution. From Eq. 1.5-1, the probability that the die is good is

$$P = e^{-\pi(2 \text{ in})^2(2.54 \text{ cm/in})^2(1.5)/\text{cm}^2}$$

$$P = 1.53 \times 10^{-53}$$

It thus requires, on the average, $1/P = 6.5 \times 10^{52}$ wafers to yield one good die. At a cost of \$200/wafer, the average cost per good die would be $\$1.3 \times 10^{55}$.

Two important observations can be made from the simplified calculations of the previous example. First, the effective cost of good large die is very high. Second, the concept of using the entire wafer as a single IC is totally impractical if the circuit is not tolerant of any faults. The concept of using the entire wafer as a single IC (termed wafer scale integration or WSI), is sufficiently attractive, however, that considerable effort has been devoted to research in this area. These circuits must, of course, provide for a mechanism of repair—either through inherent redundancy or some form of mechanical reconfiguration (e.g., laser trimming)—to be viable, since some defects will occur during processing.

For large devices the yield predicted by Eq. 1.5-1 is somewhat pessimistic. Two other models used are the Seeds model[14] and Murphy model[15], which are characterized respectively by the following expressions for the probability that a given die is good:

$$P = e^{-\sqrt{AD}} \tag{1.5-3}$$

$$P = \left(\frac{1 - e^{-AD}}{AD}\right)^2 \tag{1.5-4}$$

Additional information about yield prediction can be found in references 16–19.

Soft faults due to parameter drifts also affect yield in a statistical sense. The soft faults play a major role in yield of analog circuits. The effects of parameter drifts can best be appreciated by considering Fig. 1.5-2, in which the statistical distributions of the parameter X are shown. In Fig. 1.5-2a the statistical distribution of the average of the parameter over repeated processing runs is shown. X_{MAX} and X_{MIN} define a process window within which the average value of the parameter must reside. Both design and processing groups must agree on an acceptable process window. The cost of fabricating an IC increases with a decreasing process window and approaches infinity as the width of the process window converges to zero. It is often the case that X_{MAX} and X_{MIN} differ from
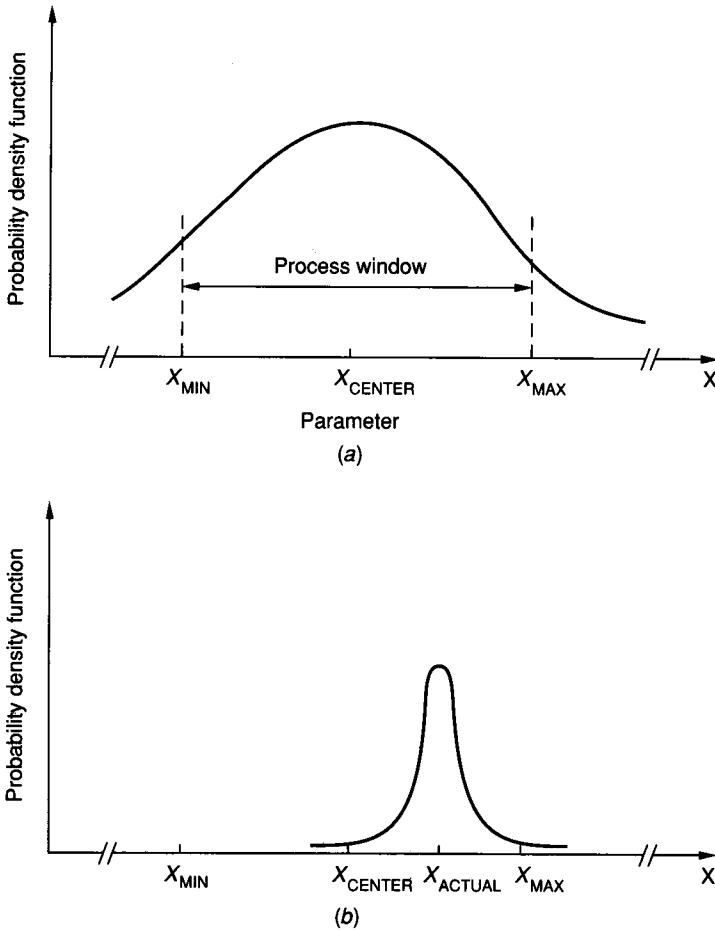
FIGURE 1.5-2
Statistical parameter spreads in a variable X: (a) Process window, (b) Wafer level variations.

X_{CENTER} by 10% to 50% for most of the major process parameters. More details about parameter spreads can be found in Chapter 2. Good designs will meet specifications over a wide process window. It often requires considerable effort on the part of the designer to produce designs that meet specifications over a wide process window. This is particularly true of analog designs.

Even at the wafer level, statistical variations in the parameter occur across the wafer. These variations are somewhat dependent on proximity of devices and, because of the way parameters are defined, the symmetry in layout. This variation, not including either device proximity or layout, is shown in Fig. 1.5-2b, where X_{ACTUAL} represents the average value of the parameter X on a given wafer. X_{ACTUAL} is thus considered a random variable relative to processing but a fixed parameter once processing is complete. The statistical variance at the wafer level is generally much smaller than the variance at the processing level.

For example, the capacitance density of oxide capacitors may have a processing variation (variance) of 10% of the mean whereas the wafer-level variation may be less than 0.1% of the mean. Two interesting observations can be made from the curves in Fig. 1.5-2. First, after fabrication the parameters are distributed around X_{ACTUAL} rather than the process center, X_{CENTER}. In fact, after processing the probability that a parameter is close to X_{CENTER} may be quite small. Second, it should be apparent that designs should meet specifications over *both* process window and wafer-level variations.

Most good digital designs are functionally tolerant to large process parameter variations of all key parameters and are quite insensitive to wafer-level variations. This insensitivity is inherent in most of the basic digital circuit structures. These variations, however, play a major role in the speed specifications of digital circuits, with faster circuits commanding a premium price in the marketplace.

The analog designer, on the other hand, must pay considerable attention to both process window and wafer-level variations. The process window variations are actually so large that many design approaches used in discrete component design are totally impractical in monolithic designs. For reasons that will become apparent in later chapters, the performance of many analog designs is strongly dependent on close matching of devices and device characteristics and relatively tolerant of nominal parameter variations within the process window. Consequently, the curve of Fig. 1.5-2*b* plays a key role in many analog designs.

Although the exact shape of the parameter distributions shown in Fig. 1.5-2 is of interest, that information is not widely distributed. Rather, a single parameter that in some sense characterizes the parameter distribution is more commonly available. For example, a comment about the threshold voltage, V_T, such as $V_T = 0.5 \text{ V} \pm 20\%$ means that the value of V_T after processing will "usually" be within $\pm 20\%$ of 0.5 V. Actually, depending on the process acceptance criteria, the average value of V_T may be within this window for all accepted wafers. Correspondingly, a comment such as "V_T matching is to 5 mV" means that the wafer-level parameter variations such as those shown in Fig. 1.5-2*b* typically differ by less than 5 mV from the measured average value. Whereas the bounds on the process window may be fixed due to rejection of wafers that do not meet the acceptance criterion, the wafer-level variations are essentially not bounded. Unless stated differently, it will be assumed throughout this text that a wafer-level or process window distribution as specified above corresponds to a window within which *99.73% of the devices* will fall. This corresponds to a ± 3 standard deviation window if the random variable is normally distributed—an assumption that is often made.

Example 1.5-3. Assume the die of a circuit is 6 mm on a side with a defect density of $0.7/\text{cm}^2$. Assume also that the die includes an analog section, which includes 100 transistors in which the characteristics must all be matched to within 0.5% of the die average value to meet specifications. Assume that the matching characteristics of the transistor are dominated by a single parameter that is normally distributed with a ± 3 standard deviation window, characterized by a variation from the die average of $\pm 0.5\%$.

(a) Using the Seeds model for defect density, determine the approximate yield if potential soft faults in the analog section are neglected.

(b) Estimate the yield if both hard and soft faults are considered.

(c) Repeat (b) if matching to 0.25% is required in the analog section.

Solution. (a) From (Eq. 1.5-3), the hard yield is the probability that a die is good, which is given by

$$P_H = e^{-\sqrt{AD}} = e^{-\sqrt{(6\text{mm})^2(7\times10^{-3})/\text{mm}^2}} = 60.5\%$$

(b) The probability that a die has no soft faults is equal to the probability that *all* 100 transistors meet the matching specifications. The probability that each transistor meets the matching requirements is $P = .9973$. Thus, the probability that all 100 transistors meet the specifications is

$$P_S = (.9973)^{100} = .763$$

The overall yield is thus the product of the hard and soft fault yield, which is

$$P = P_H P_S = 46\%$$

(c) For a matching to 0.25%, it follows from a standard investigation of the normal distribution that the probability that a given transistor meets the specifications is equal to the probability that the normal random variable is within ± 1.5 standard deviations of the mean. From a probability density table for the normal distribution, it follows that the probability that any one transistor meets the specifications is 0.8664. Thus, the probability that all 100 transistors meet specifications is

$$P_S = (0.8664)^{100} = 5.9 \times 10^{-7}$$

Thus, from part (a), the overall yield is

$$P = P_S P_H = 3.5 \times 10^{-6} = 3.5 \times 10^{-5}\%$$

Note that this modest increase in matching requirement from 0.5% to 0.25% results in a yield that is so small as to make such a design totally impractical.

The previous example demonstrates a very important point. If very much matching is required in analog circuit design, then the matching requirements must not be too severe if the yield is to be high enough to make the circuit economically feasible.

Since considerable variation from the process center occurs, as indicated in Fig. 1.5-2a, one naturally asks the question: What value should be used for the parameters at the design stage? Stated alternately, the problem is to find the design target, or the design center, for each parameter. The process center is widely used for the design center. It can, however, be shown that in some cases, the yield can be improved if something other than the process center is used for the design center. The field of design centering and yield optimization addresses these questions. Unless stated otherwise, it will be assumed throughout this book that the desired design center corresponds to the process center.

The reader should be cautioned to avoid the temptation of using experimental parameter centers obtained from measurements made on one or two wafer lots

in lieu of the design center that has evolved with the process itself. It should be apparent from Fig. 1.5-2 that using experimentally measured parameters rather than the process center as design centers will statistically bias designs in a way that will often be accompanied by a decrease in yield.

Industry is keenly aware of the relationship among the specified process window (which dictates the design specification range), the actual wafer-level variations, and the corresponding yield. A *capability index*, which relates these variables for any process parameter, is defined by the expression[20]

$$C_p = \frac{\text{Design specification width}}{\text{Process width}} \qquad (1.5\text{-}5)$$

where the process width is generally defined to be ± 3 standard deviations (σ) about the mean. Assuming a normally distributed process around the mean in which the process width is characterized by the $\pm 3\sigma$ window, it follows that if the process window is centered in the design specification window, then the probability that a parameter of a device lies inside the design specification window is

$$P = \frac{1}{\sqrt{2\pi}} \int_{-3C_p}^{3C_p} e^{-(x^2/2)} dx \qquad (1.5\text{-}6)$$

This error function, erf (x), integral is readily obtainable from tables of the normal probability distribution, which appear in most elementary statistics texts. If C_p is small, the probability that a device parameter is actually within the process window is small; the yield will be poor and will deteriorate even more if the mean of a parameter for a wafer is not centered within the design specification window. Correspondingly, if C_p is large, the yield will be high and the process will be reasonably tolerant to small shifts in the wafer-level parameter center from the design specification center. Setting the design specification width so that $C_p = 2$ will result in reasonable yield in many applications, although this may place unrealistic performance demands on the designer for some design projects.

Example 1.5-4 clearly demonstrates the yield implications associated with properly establishing the capability index.

> **Example 1.5-4.** If 1000 devices on a chip must have a specific parameter within the specified design process window, determine the soft yield if the process has been characterized by a capability index of (a) $C_p = 0.5$, (b) $C_p = 1.0$, (c) $C_p = 1.5$, and (d) $C_p = 2.0$.

> *Solution.* With $C_p = 0.5$, it follows from Eq. 1.5-6 and a normal random variable table that $P = .8664$. Thus, the probability that all 1000 devices have a parameter within the design specification window is $P_{1000} = (.8664)^{1000} \approx 0$. If $C_p = 1.0$, $P = .9973$ and $P_{1000} = .067$, which represents about a 6.7% yield. If $C_p = 1.5$, $P = .999993$ and $P_{1000} \approx .993$. Finally, if $C_p = 2.0$, $P = .999999998$ and $P_{1000} \approx 1.0$, indicating essentially a 100% yield.

1.6   TRENDS IN VLSI DESIGN

Several trends in the production of VLSI circuits are readily identifiable. Some of these have already been discussed in this chapter. The most visible is the continual shrinking of the minimum geometrical feature size. Although this trend will continue, the rate at which the minimum feature size decreases is slowing. This slowing is partially attributable to inherent physical limitations in the photolithographic process and the rapidly increasing costs associated with very fine resolution processing equipment. Problems associated with power dissipation density, the effects of additional diffusion in small devices due to subsequent heat cycle steps during processing, and concerns about a practical alternative to the widely used standard local oxidation (LOCOS) processing step are all becoming significant in submicron processes. The size of the silicon atom and the silicon dioxide molecule will also be of increasing concern in submicron processes. These concerns about size place very real lower bounds on the conventional approach to integrated circuit design.

A trend in increasing speed in digital circuits is readily identifiable. Some research efforts with gallium arsenide (GaAs) suggest that this material may ultimately supplant silicon at very high frequencies. GaAs is attractive because of higher electron mobility and, in some applications, because of reduced sensitivity to radiation. Continual improvement in the performance of silicon circuits raises doubts, however, about when and if a transition to GaAs will occur. The increase in speed and increase in circuit complexity are direct results of the reduction in feature size.

A third trend is the increasing complexity of circuit function and device count on a die. This trend is crucial for the development of new markets for integrated circuits.

A fourth trend is toward increased designer productivity and an ever-growing dependence on the computer in the design process. Design methods that were standard a decade ago would be totally unworkable in many current design projects.

A fifth trend is the continual shift of where design, production, and markets are geographically located. A decade ago most major and innovative design efforts and a significant portion of the production were done in the United States and Western Europe, with more mature technologies being transferred to the Far East. Production was generally most profitable in countries where labor costs were low. An increasing shift in the design efforts to the more developed Far Eastern countries has occurred in recent years, with the mature production shifting more into the less-developed Far Eastern countries, where labor costs remain low. The marketplace is also shifting, with a large number of less-developed countries now experiencing the data processing and telecommunications revolution that swept the West a decade earlier. A trend of the design and production activities geographically following the marketplace is also observable.

A sixth trend is a growing coupling of a specific process and its processing equipment. As feature sizes shrink and processes become more complex, the process is becoming increasingly dependent on the performance of specific pieces of equipment.

Other trends include the use of more powerful CAD tools, which extends VLSI design to the realm of the systems designer, and increasing the complexity of processes by using silicon on insulator (SOI) or combining both MOS and bipolar technologies into a single process.

REFERENCES

1. J. E. Lilienfeld, U.S. Patent 1,745,175; 1930.
2. O. Heil, British Patent 439,457; 1935.
3. T. R. Reid, *The Chip*, Simon & Schuster, New York, 1985.
4. J. D. Ryder and D. G. Fink, "Engineers and Electrons", IEEE Press, New York, 1984.
5. I. Catt, "Wafer-Scale Integration," *Wireless World (GB)*, vol. 87, no. 1546, pp. 57–59, July 1981.
6. W. R. Moore, "Introducing Wafer-Scale Integration," *Silicon Design*, vol. 6, p. 9, January 1986.
7. R. Aubusson and I. Catt, "Wafer-Scale Integration—A Fault Tolerant Procedure," *IEEE J. Solid State Circuits*, vol. SC-13, no. 3, pp. 339–344, June 1978.
8. D. L. Peltzer, "Wafer-Scale Integration: The Limits of VLSI?" *VLSI Design*, vol. IV, no. 5, pp. 43–47, September 1983.
9. R. R. Johnson, "The Significance of Wafer-Scale Integration in Computer Design," presented at the *IEEE Int. Conf. on Computer Design: VLSI in Computers*, Port Chester, NY, pp. 101–105, October 1984.
10. IEEE, "Whatever Happened to Wafer-Scale Integration?," *IEEE Spectrum*, vol. 19, no. 6, p. 18, June 1982.
11. R. H. Dennard, E. H. Gaensslen, H. -N. Yu, V. L. Rideant, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFETS with Very Small Physical Dimensions," *IEEE J. Solid State Circuits*, vol. SC-9, pp. 256–268, October 1974.
12. Eric Demoulin, "Fabrication Technology of MOS ICs for Telecommunications," *Design of MOS VLSI Circuits for Telecommunications*, ed. Y. Tsividis and P. Antognetti, Prentice-Hall, Englewood Cliffs, N.J. 1985, Chapter 1.
13. W. E. Loeb, "IC Packages, Costs, Trends and Forecasts," W. E. Loeb and Associates, Soquel, Calif., June 1987.
14. R. B. Seeds, "Yield and Cost Analysis of Bipolar LSI," IEEE Int. Electron Devices Meeting, Washington, D.C., p. 12, 1967.
15. B. T. Murphy, "Cost-Size Optima of Monolithic Integrated Circuits," *Proc. IEEE*, vol. 52, pp. 1537–1545, December 1964.
16. D. G. Ong, *Modern MOS Technology, Processes, Devices and Design*, McGraw-Hill, New York, 1984.
17. W. E. Ham, "Yield-Area Analysis: Part I—Diagnostic Tool for Fundamental Integrated Circuit Process Problems," *RCA Review*, vol. 39, pp. 231–249, June 1978.
18. T. Okabe, M. Nagata and S. Shimada, "Analysis on Yield of Integrated Circuits and a New Expression for Yield," *Electrical Engineering in Japan*, vol. 92, pp. 135–141, 1972.
19. C. H. Stapper, A. N. McLaren, and M. Dreckmann, "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product," *IBM J. Res. Develop.*, vol. 24, pp. 398–409, 1980.
20. V. E. Kane, "Process Capability Indices," *Journal of Quality Technology*, vol. 18, pp. 44–51, January 1986.

PROBLEMS

Section 1.1

1.1. How many SiO_2 molecules will be required to form the insulating layer under the gate of a MOSFET that has a 3 μ × 3 μ gate area if the silicon dioxide thickness is 800 Å? Repeat if the gate is 0.5 μ × 0.5 μ with a 100 Å SiO_2 layer.

1.2. What is the average number of vertically stacked SiO_2 molecules in an 80 Å gate oxide layer?

Section 1.4

1.3. If the average resistance density (resistance/unit area) is 1.2 Ω/μ^2 and the capacitance density is 0.7 fF/μ^2(1 fF $= 10^{-3}$ pf $= 10^{-15}$ F) and if a 5 inch wafer costs \$250 to produce, what is the cost (based upon area) of a 250 kΩ resistor? What is the cost of a 100 pF capacitor? How many 2 $\mu \times$ 2 μ transistors can be placed in the area required for the 250 kΩ resistor? For the 100 pF capacitor? What is the cost per transistor if spacing and interconnection area are neglected? What is the cost per transistor if spacing and interconnection increase the area per transistor by a factor of 10?

Section 1.5

1.4. If a 4 inch wafer costs \$200 and the defect density is $D = 2/cm^2$, what is the maximum permissible die area that can be used if the effective die cost (cost per good die) must not exceed 45¢? Assume the Seeds model characterizes the yield of the process and neglect area losses associated with placing square dies on a round wafer.

1.5. If an 8-bit flash A/D converter has 2^8 comparators and all must have an offset voltage less than 40 mV (-40 mV $< V_{offset} <$ 40 mV) for the device to operate properly, and if the offset voltage for the comparators in this process is 20 mV, what percentage of the dies will fail due to soft faults associated with the offset voltage? If the same approach is used to build a 10-bit converter, the offset voltage requirement is more stringent (10 mV). If the same process is used, what percentage of the dies will now fail due to soft faults associated with the process?

1.6. Derive an expression for and plot the effective die cost (cost per good die) versus die area if the defect density is 2/cm² for both the simple model of Eq. 1.5-1 and Seeds model. Also obtain an expression for and plot the cost per unit area and the cost per good unit area versus die area for both models. Assume the wafer size is 4 inch and the wafer cost after processing is \$200.

1.7. What is the average cost per good 24-pin plastic DIP package of a die that is 0.1 inch on a side if it is fabricated on a 5 inch wafer and the average defect density is 1.5/cm²? Assume that wafer probing is used to eliminate defective dies and that the packaging yield is 85%. Use the typical processing costs listed in Table 1.4-2 for your calculations.

1.8. Repeat Example 1.4-1 if the die size is decreased to 0.2 cm $\times$ 0.2 cm and the defect density is 1.5/cm². Assume the yield model of Eq. 1.5-1 and a packaging yield of 95%.

1.9. Assume an integrated circuit has a die size of 0.8 cm $\times$ 0.8 cm, defect density of 2/cm², a wafer size of 4 inch, a package requirement of 64 pins, and a packaging yield of 90%.
  (a) Calculate the average cost per good package if wafer probing is used and parts are packaged in plastic DIP packages.
  (b) Compare the results of part (a) with the average cost per good package if wafer probing is omitted and parts are packaged in ceramic side brazed packages.
  Use the processing and packaging cost estimates of Table 1.4-2.

1.10. A 5 inch wafer is yielding 85% at wafer probe with a die size of 0.2 cm $\times$ 0.2 cm. Determine the defect density assuming the process is characterized by the Seeds model.

1.11. Assume an area overhead per die of 250,000 μ^2 to allow for interconnection of a die to the outside world. Assume that a minimum size transistor is 3 μ^2 and an additional 60 μ^2 is needed for spacing and internal interconnection for each transistor. Assume also that the die is fabricated on a 4 inch wafer and that production costs are characterized by the values given in Table 1.4-2.

(a) Calculate and plot the average cost of fabrication per transistor on this die as the number of transistors on the die ranges between 1 and 100,000.

(b) Calculate and plot the average cost per good transistor on good dies if the fatal defect density in the external interconnection area is 0.1/cm^2 and that in the area devoted to transistors (transistor, spacing, and interconnection) is 1.5/cm^2. Use the model of Eq. 1.5-2.

1.12. Determine the minimum acceptable capability index for characterizing a process if a soft yield of 95% must be maintained on a die that contains 2000 devices.