CHAPTER 2

ELECTRONIC COMPONENTS AND SEMICONDUCTOR DEVICES

2.1 INTRODUCTION

In packaging electronic systems, it is often necessary to design cabinets, drawers, cages or gates to house the components and to permit easy and rapid access for maintenance. Printed circuit boards (PCB's) are designed to support hundreds of small components with thousands of interconnections. A cooling system is incorporated into the cabinets to dissipate the heat generated by the electronic devices. Interconnections and heat removal with small thermal penalties are the two most important features in the design of highly reliable products. There is a wide variety of hardware both mechanical and electrical which must be included in the design of even a relatively simple electronic system. A partial listing of some of the most common components used in design is given below:

Discrete transistors	Power supplies	Logic type IC's
Transformers	Memory type IC's	Plasma panels
Diodes	Liquid crystal displays	Resistors
Cathode ray tubes	Capacitors	Vacuum tubes
Inductors	Disk drives	Potentiometers
Tape drives	Relays	Lamps
Switches	Fans	Circuit breakers
Cold plates	Connectors	Cable harnesses

The emphasis in this textbook is placed on the packaging of microelectronic circuits which are used extensively in the design of high performance digital systems, laboratory instrumentation and automated manufacturing systems. To package microelectronic circuits, it is necessary to understand in at least a qualitative manner the functional behavior of microelectronic devices. For this reason, the basic principles of semiconductor theory are reviewed and the operation of semiconductor diodes and transistors is described. The coverage is then extended to introduce logic gates and the Metal-Oxide-Semiconductor-Field-Effect-Transistor (MOSFET) that are used in most electronic systems today. The last part of the chapter treats the scale of integration with projections into the future which will markedly affect packaging strategies in the later years of this decade. Of particular importance are the significant changes in packaging design that will occur because of the introduction of new electronic devices. These changes in packaging design will be driven by higher I/O count and higher heat dissipation that will be typical of the newer high performance components. These changes will require designers to create PCBs with higher density pads and circuit lines and to develop more effective heat dissipation methods in order to fully utilize the logic devices produced with new chip fabricating technologies.

2.2 CONDUCTORS, INSULATORS AND SEMICONDUCTORS

Conductivity in materials depends on the structure of the atoms of elements which are combined to give an alloy and the resulting atomic bonding between these atoms. Consider the metal aluminum with the atomic structure illustrated in Fig. 2.1. The aluminum atom has a full inner shell (the K shell) and a full L shell with 2 and 8 electrons filling these shells. However, the outer M shell contains only 3 electrons and they are loosely bound to the nucleus. These nearly free M shell electrons act as negative charge carriers in conducting current in a wire made of aluminum. The resistance R of a conductor in the shape of a wire or a rectangular conductor (line) formed on the surface of a PCB is given by:

$$R = \frac{\rho L}{A} \tag{2.1}$$

where L is the conductor length (cm), A is its cross section area (cm) and ρ is its resistivity (ohm-cm).

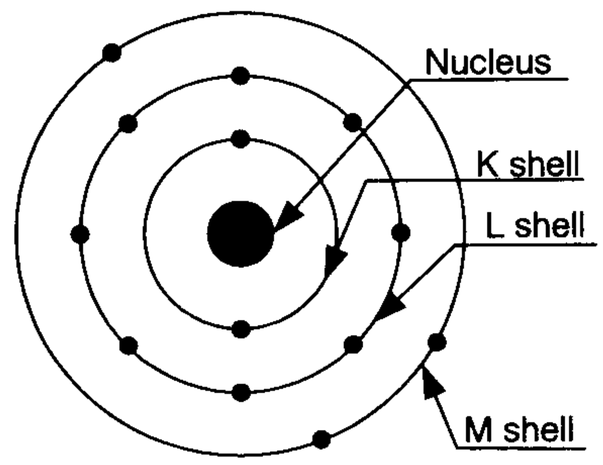


Fig. 2.1 Atomic structure of aluminum.

The resistivity ρ depends on the atomic structure of the element. If there are a large number of loosely bound electrons in the outer shell, the resistivity is low as indicated in Table 2.1. It is evident from the results shown in this table that the resistivity of different types of materials can vary over 18 orders of magnitude. For metal conductors ρ is of order 10^{-6} Ω-cm but for insulators ρ is of order 10^{12} Ω-cm or higher.

Table 2.1
Resistivity of select materials

Material	Classification	Resistivity (Ω-cm)
Silver	Conductor	1.63×10^{-6}
Copper	Conductor	1.72×10^{-6}
Aluminum	Conductor	2.83×10^{-6}
Nickel	Conductor	6.9×10^{-6}
Platinum	Conductor	9.8×10^{-6}
Silicon	Semiconductor	1.56×10^{5}
Aluminum oxide	Insulator	1×10^{15}
Silicon dioxide	Insulator	1×10^{14}
Epoxy	Insulator	1×10^{15}
Polyethylene	Insulator	1×10^{18}

The atomic structure of insulating materials clearly shows the reason for the large values of resistivity ρ. Consider the atomic structure of silicon dioxide SiO_2 shown in Fig. 2.2. Silicon has a full K shell, a full L shell and four electrons in the outer M shell. Oxygen has a full K shell and six of eight electrons necessary to fill the L shell. The silicon atom combines with two oxygen atoms to provide the correct number of electrons to fill the outer shells of the two oxygen atoms when the four silicon electrons are shared. The shared electrons are covalently bonded, and the free silicon electrons are not

available to act as charge carriers. Materials like SiO_2, Al_2O_3, MgO, BN, Si_3N_4 and BeO, which are classified as ceramics, are all covalently bonded with complete outer shells. They act as insulators with very high resistivity.

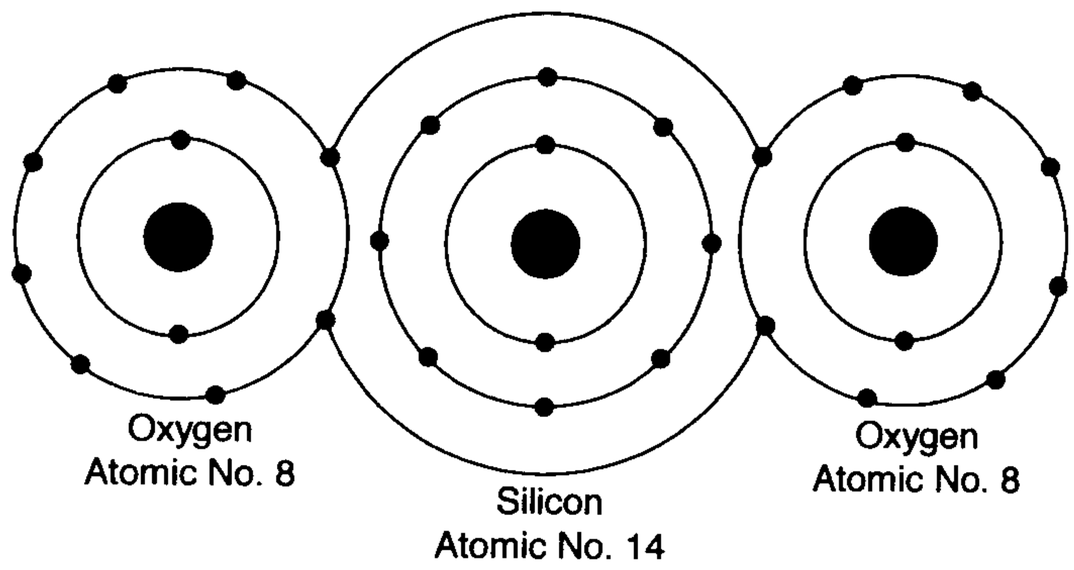


Fig. 2.2 Atomic structure of SiO_2.

Elements such as carbon, silicon and germanium each have four electrons in their outer shells. The atoms in these elements form covalent bonds with four neighboring atoms to form stable structures similar to that shown for silicon in Fig. 2.3. The electrons are tightly bound due to the covalent structure and free electrons are not available to act as charge carriers. However, the atomic structure is not perfect and some of the electrons have sufficient energy to jump from the valence state to the conduction state. For this reason, materials such as silicon, germanium and gallium arsenide are classified as semiconductors and exhibit resistivity much higher than the metal conductors but significantly lower than the ceramic insulators. The resistivity of silicon is 1.56×10^5 Ω-cm and is due to the presence of one electron out of 2×10^{13} which has sufficient energy (1.1eV) to jump from the valence band to the conduction band. The ability of these electrons in the conduction band to carry charge is termed intrinsic conduction.

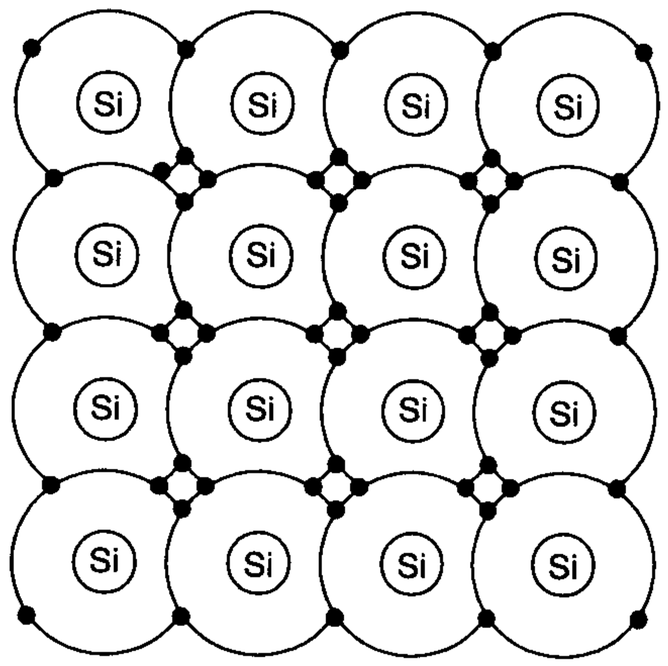


Fig. 2.3 Covalent bonding of electrons in silicon.

2.3 EXTRINSIC SEMICONDUCTORS

The intrinsic semiconducting capability of silicon was described in the previous section. The semiconducting properties of silicon, geranium or gallium-arsenide can be modified by changing the atomic structure of single crystals of these materials. To show the modification of the crystal lattice, consider the structure of silicon with its covalent bonding as shown in Fig. 2.3. If one or more of the silicon atoms in the lattice is replaced with an impurity atom, the conductivity of the modified structure is changed. Impurity atoms called dopants are of two basic types:

1. Elements from the V column of the periodic table with five valence electrons in their outer shell such as phosphorus, arsenic and antimony
2. Elements from the III column of the periodic table with three valence electrons including boron, aluminum and gallium in their outer shells.

For example, if phosphorus with its five valence electrons is introduced into the silicon lattice, the fifth electron is not used to complete covalent bonding. This extra electron, an extrinsic charge, is relatively free to carry current. Hence, the resistivity of the doped silicon is lower than that of pure

silicon. Silicon with dopant elements from the V column of the periodic table are classified as type N semiconductors, because they contain extra electrons which are negative charge carriers.

If silicon atoms in the crystal lattice are replaced with a dopant atom with three valence electrons in its outer shell such as boron, then covalent bonding occurs between silicon and boron but the outer shell is not filled—a single vacancy or hole exists. This hole is an accepter of electrons, and it acts as a positive charge carrier. The semiconductors with dopant elements selected from column III in the periodic table are classified as type P because of the positive charge, which is carried by the holes, is capable of moving through the atomic lattice.

Conduction occurs in a semiconductor when either the electrons or the holes move through the lattice due to the application of an electric field. The resistivity of the semiconductor is given by:

$$\rho = \frac{1}{eN\eta} \tag{2.2}$$

where N is the number of charge carriers, e is the charge on the carrier and η is the carrier's mobility.

The velocity v of the charge carrier is dependent upon:

$$v = \eta E \tag{2.3}$$

where E is the electric field.

It is interesting to note that the velocity of the electrons in N type silicon is about three times greater than the velocity of the holes in P type silicon under the same applied field. The reduced mobility of the holes is due to the mechanics of the motion of the holes. The manner in which the holes move through the lattice is presented in Fig. 2.4. The movement of a hole is due to the jumping of an electron from an adjacent atom to fill that hole, while at the same time creating a new hole displaced by one lattice spacing. This cumbersome motion of the holes accounts for their lower velocity through the lattice structure.

Electric field + ⇒ −
Electron direction ⇐
Hole direction ⇒
Current direction ⇒

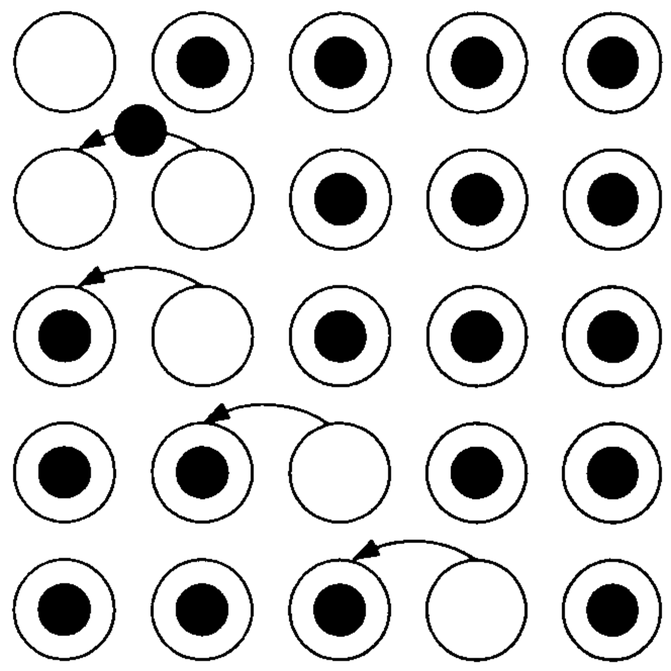


Fig. 2.4 Movement of a hole to the right by a sequence of electron movements to the left.

The resistivity of both N and P type silicon can be controlled over five orders of magnitude 10^2 to 10^{-3} (Ω-cm) by adjusting the amount of dopant (its concentration) added to the silicon as indicated in Fig. 2.5. The ability to control the resistivity is of critical importance in fabricating diodes, transistors or resistors in silicon. Methods have been developed for producing single crystal silicon with the dopant impurities of different concentrations introduced at local sites on the surface of a wafer cut from a single crystal.

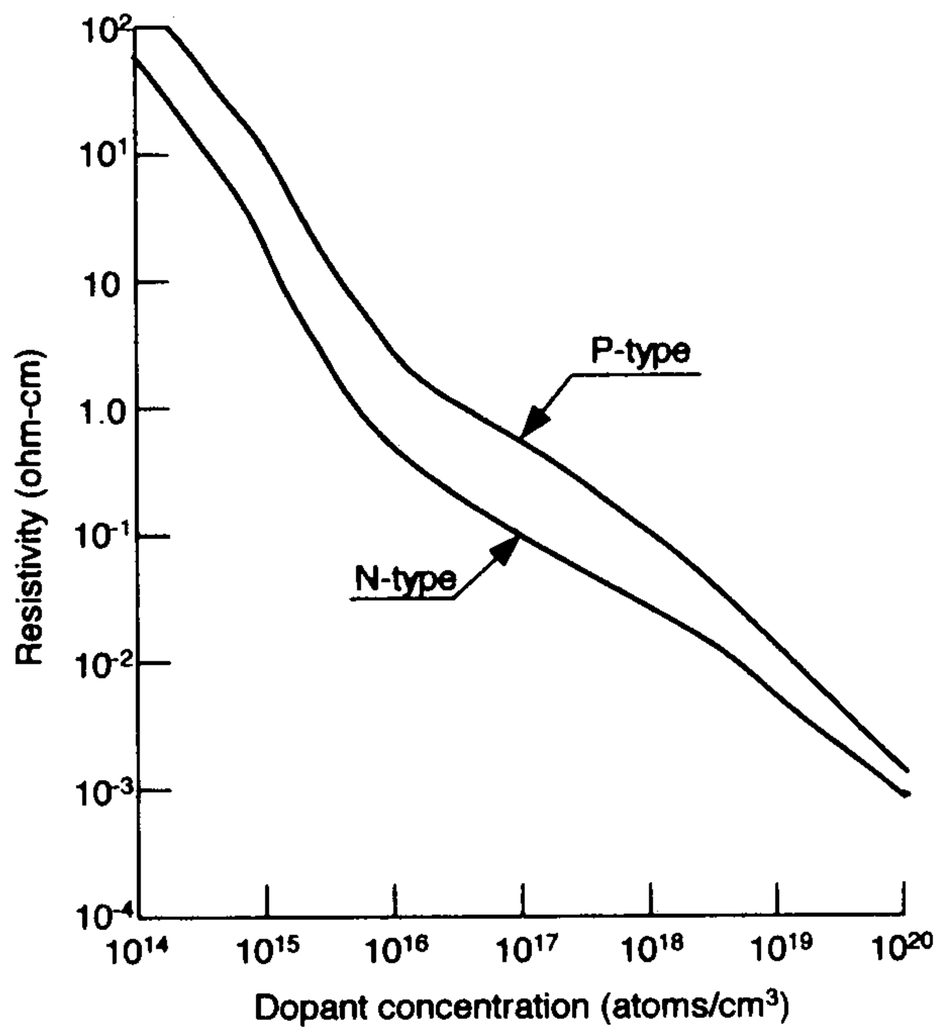


Fig. 2.5 Resistivity of P- and N-type silicon as a function of dopant concentration.

2.4 THE P-N JUNCTION

An interface between P and N type silicon can be formed by taking N type material and implanting boron into one of its surfaces to create a local region of P type material. This procedure produces a P-N junction, which is illustrated schematically in Fig. 2.6a. The P type material contains holes which act as accepters for electrons and the N type material contains electrons which serve as donors to fill these holes.

At the interface of the P and N type semiconductors, the holes and electrons combine and eliminate each other forming a thin region free of charge carriers. The layer where the holes have accepted the donor electrons is termed the depletion region indicating that the available charge carriers have been depleted. The remaining holes in the P material create a negative charge on the electrode opposite the junction and the remaining electrons in the N material create a positive charge on its electrode. In this state the junction is electrically neutral.

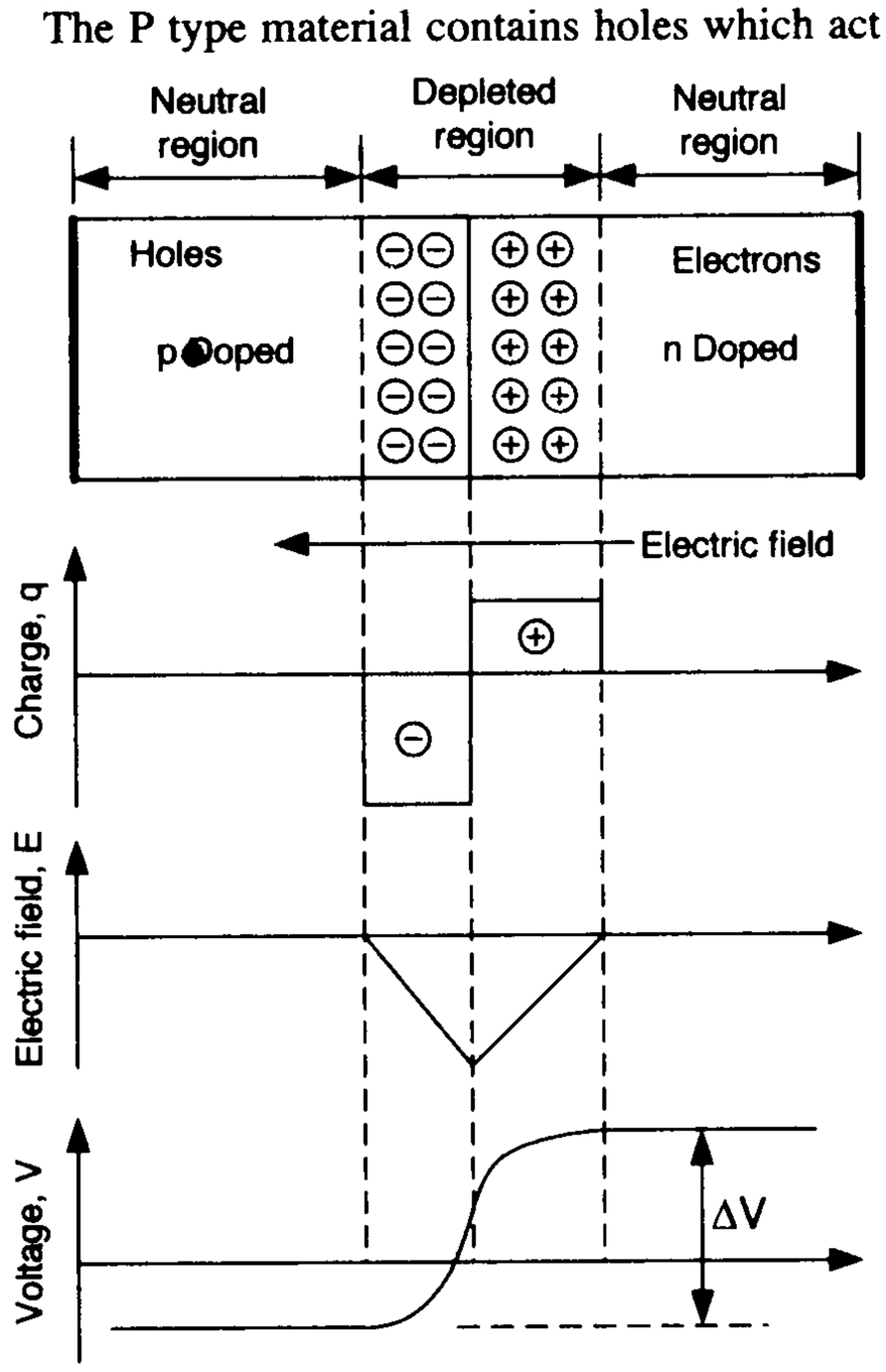


Fig. 2.6 Electrical characteristics of a P-N junction.

The presence of positive charge carriers (holes) on one side of the depletion layer and negative charge carriers on the other side of the layer develops an internal electric field across the junction. Essentially the depletion layer behaves like pure silicon and is capable of only intrinsic conduction. In this state, the depletion layer limits the flow of electrons across the junction.

The thickness of the depletion layer depends on the concentration of the impurity dopants. High concentrations lead to thin layers and lower concentrations yield thick layers. This influence of concentration on the thickness of the depletion layer is illustrated in Fig. 2.7. The ability to control the thickness of the depletion layer is useful in the development of diodes and transistors.

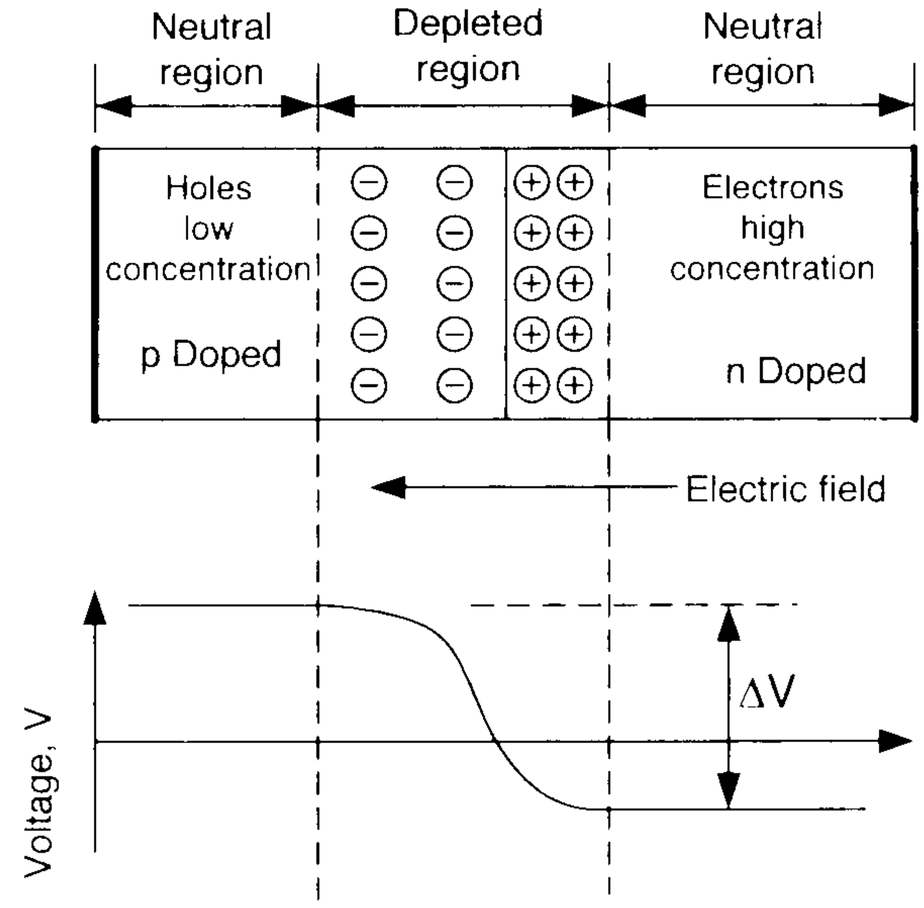


Fig. 2.7 Influence of the concentration of dopant on the thickness of the P-N junction.

2.5 SEMICONDUCTOR DIODES AND RESISTORS

2.5.1 Semiconductor Diodes

The simplest semiconductor device—the diode—is based on the properties of a P-N junction. A diode is a device which permits the flow of current in one direction and blocks current flow in the other. To show the essential features of a semiconductor diode, consider the P-N junction, described in Section 2.4, with a positive voltage V^+ on the P electrode and a negative voltage V^- on the N electrode. This arrangement, which is shown in Fig. 2.8, represents the diode under a forward bias voltage. The positive voltage on the P electrode repels the positively charged holes and drives them toward the junction. The negative voltage on the N electrode repels the negatively charged electrons driving them toward the junction. Additional recombination of holes and electrons take place, the depletion layer becomes thinner and the internal electric field which serves as a barrier to conduction is reduced. Further increases in the forward bias ($V^+ - V^-$) overcomes the potential barrier permitting the flow of electrons from the N material across the junction to the P electrode and the flow of holes from the P material across the junction to the N electrode. The forward current increases non-linearly with increasing forward bias. It should be noted that current does not flow until the potential barrier is overcome by the forward bias voltage.

Next, consider the same semiconductor diode but with the polarity reversed so that the V^+ is applied to the N electrode and the V^- is applied to the P electrode as indicated in Fig. 2.9. In this case, the electrons in the N material are attracted to its electrode and the holes in the P material are attracted to its electrode. The depletion layer is widened and the potential barrier due to the internal field increases. Current flow due either to the motion of electrons or holes across the junction cannot occur because of this increase in the potential barrier due to the reverse bias. Some leakage current flows but

this is due to i
measured in na
reversed current

Fig. 2.8
semiconductor

Fig. 2.10 F
as a functi

this is due to intrinsic conduction resulting from lattice imperfections. Typical leakage currents are measured in nanoamperes; whereas, the forward currents are measured in milliamperes. Forward and reversed currents are shown as a function of voltage bias for a typical semiconductor diode in Fig. 2.10.

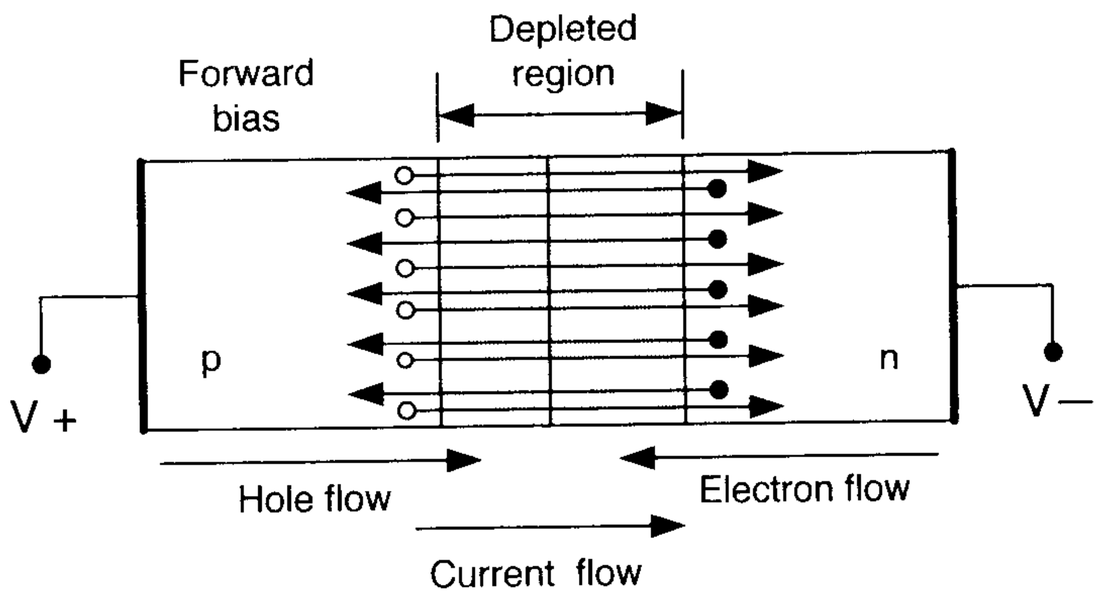


Fig. 2.8 Characteristics of a semiconductor diode under forward bias voltage.

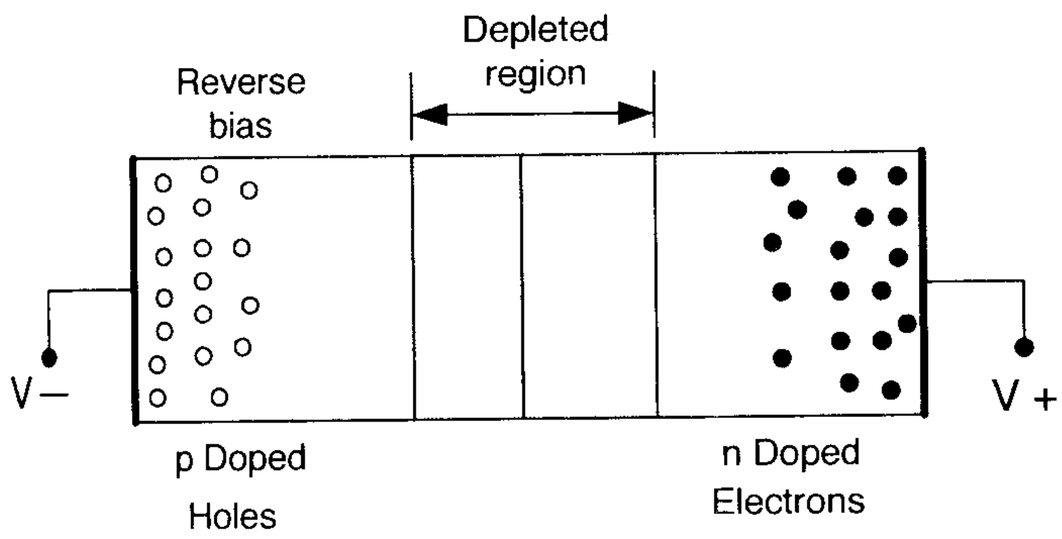


Fig. 2.9 Characteristics of a semiconductor diode under reverse bias.

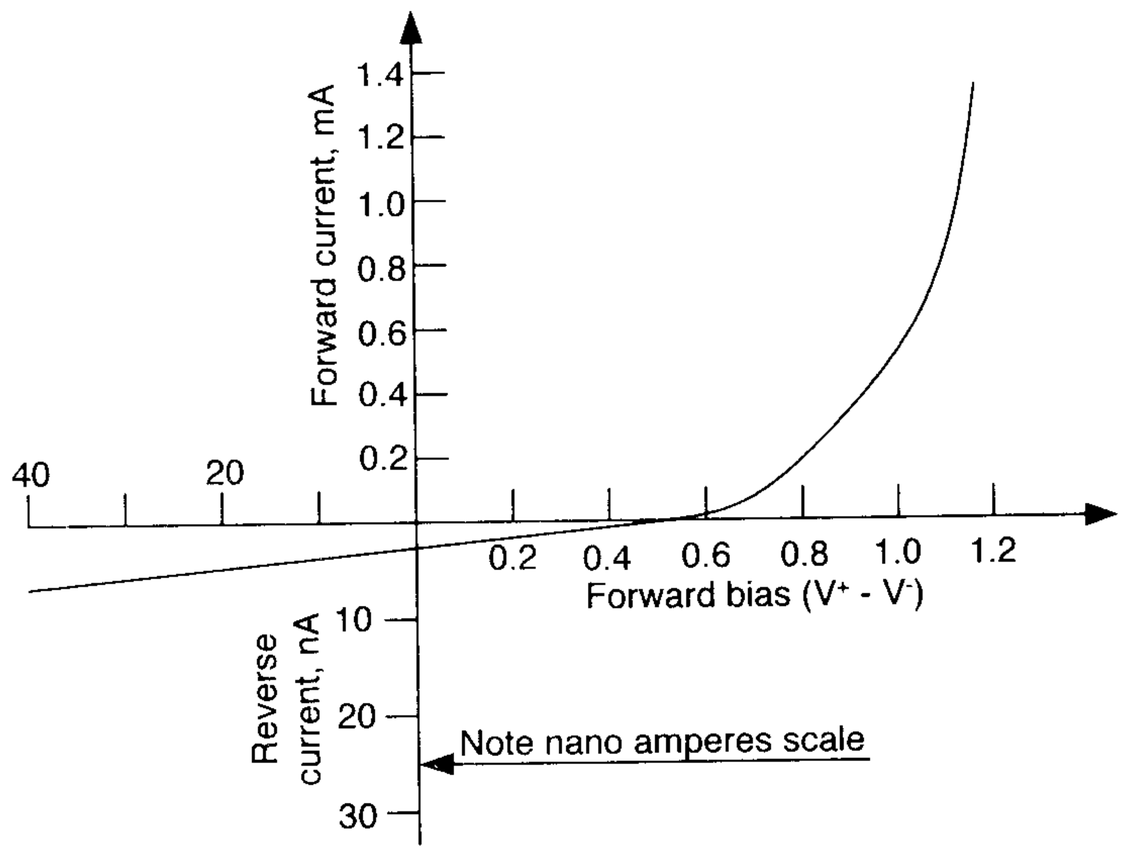


Fig. 2.10 Forward and leakage current as a function of voltage bias on a P-N junction diode.

2.5.2 Semiconductor Resistors

Resistors are usually discrete devices added to the circuit to control the characterisitc impedance of signal linmes or to limit currents. These discrete resistors are fabricarted from wire, carbon, or thin metalic films. Many resistors are formed on insulating cylindrical cores with the resistive material placed in a spiral on the surface of the cylinder. Wire leads extend from the axis of the cylinder at each end. Power resistors are housed in larger packages in order to better dissipate the heat they generate more efficiently. Resistors used in digital systems are usually much smaller and are housed in leadless surface mounted packages.

Resistors can also be inplanted into integrated circuits as part of the chip making process during wafer fabrication. In this instance, the semiconductor material employed to form the a resistor. Recall from Fig. 2.5 that the resistivity of the silicon can be adjusted by changing the concentration of the dopant material. By reserving small areas on the surface of the chip, and adjusting the dopant concentration in each of these areas integral resistors can be fabricated directly on the chip. However, the fabrication process is difficult and they consume a signifiant amount of chip area. For these reasons, chip designers usually use other means to control current flow on the chip and discrete resistors are mounted on the printed circuit board to control impedance and off-chip current flows.

In some applications semiconductor resistors serve as sensors because their resistance is markedly affected by both strain and temperature. In some research projects these sensors are implanted on a chip to study the strains developed in the silicon as a function of either globl or local temperature changes. The semiconductor sensors have an advantage of high sensitivity.

2.6 TRANSISTORS

Transistors are solid state switches which can be open to effectively block current flow or closed to permit current flow as illustrated in Fig. 2.11. Solid state switches differ in two important aspects from mechanical switches. Firstly, the transistor has no moving parts and is activated by a control signal. As a consequence of this feature, the switch can be activated very quickly with typical switching times for modern chip designs of the order of 0.1 ns. Secondly, the transistor acts as a current amplifier since the current passing through the switch $I_c + I_b$ is 10 to 100 times larger than the control current I_b required to activate the transistor. These two characteristics of the transistor permit it to be used extensively in digital logic circuits where logic design requires an extremely large number of switches to execute even simple digital functions. Transistors are also used extensively in analog circuits where the amplification of currents is extremely important.

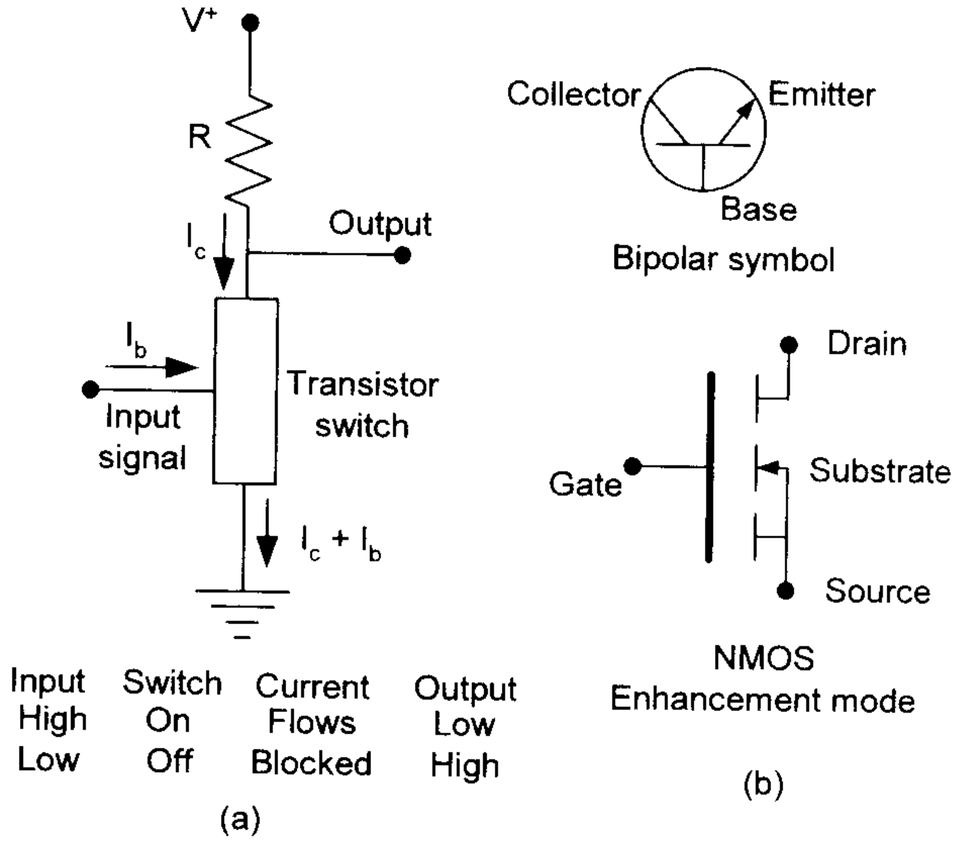


Fig. 2.11 (a) A transistor acting as a solid state switch. (b) Symbols for bipolar and MOSFET transistors.

There are two
semiconductor
described in the

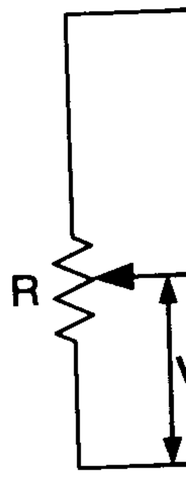
2.6.1 Bipolar

Consider the N
interfaces. Me
emitter, base a
material used
which share a

The ba
base are wide
and V^- is app
conduction be
as an open sw
on the base-en

If a p
junction incre
injected into t
the positively
on the transis
collector to t
transistor a
potentiometer
to emitter vo
V_{BE} is less th
the transistor
the forward
equals V_s.
greater than
switch will
flows and V_c

Fig. 2.12 A



There are two basic types of transistors that are in use today—the bipolar type and the metal oxide semiconductor field effect transistor (MOSFET). Key features of both of these basic transistors will be described in the next section.

2.6.1 Bipolar Transistors

Consider the N-P-N bipolar transistor, defined in Fig. 2.12a, which has three elements with solid state interfaces. Metal electrodes are placed on each element to provide for electrical connections to the emitter, base and collector. Note, that the P type silicon in the base is relatively thin compared to the N material used in the collector and emitter. This arrangement corresponds to two back-to-back diodes which share a common base.

The base is lightly doped and the depletion regions at the P-N junctions on both sides of the base are wide and effectively cover the entire thickness of the base when V^+ is applied to the collector and V^- is applied to the emitter. The depleted P material in the base acts as a potential barrier to conduction between the emitter and the collector as shown in Fig. 2.12b. In this state, the transistor acts as an open switch, because the base-collector diode is operating in reverse bias. Also the forward bias on the base-emitter diode is less than the threshold voltage required to conduct the forward current.

If a positive voltage is applied to the base electrode, the forward bias on the emitter-base junction increases and the potential barrier at this interface is overcome. Electrons from the emitter are injected into the base. These electrons find the base depleted, diffuse through it and then are attracted to the positively charged collector electrode. Thus, it is clear that a small positive voltage on the base turns on the transistor switch, electrons flow from the emitter to the collector and current flows from the collector to the emitter. The simple circuit, presented in Fig. 2.13, shows the action of a N-P-N transistor as a switch. The potentiometer at the left permits the base to emitter voltage V_{BE} to be varied. If V_{BE} is less than the threshold voltage of the transistor, the switch will be open the forward current blocked and V_{CE} equals V_s. However, when V_{BE} is greater than the threshold voltage, the switch will close, the forward current flows and V_{CE} approaches zero.

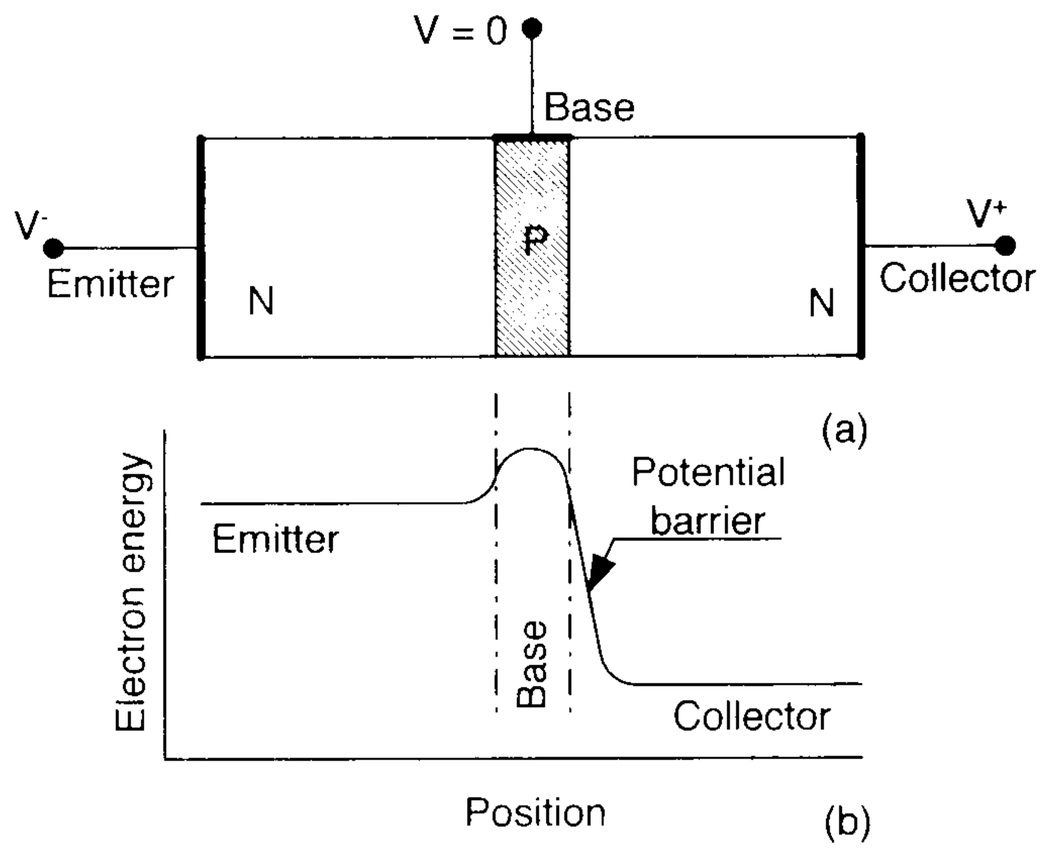


Fig. 2.12 A bipolar transistor acting as an open switch.

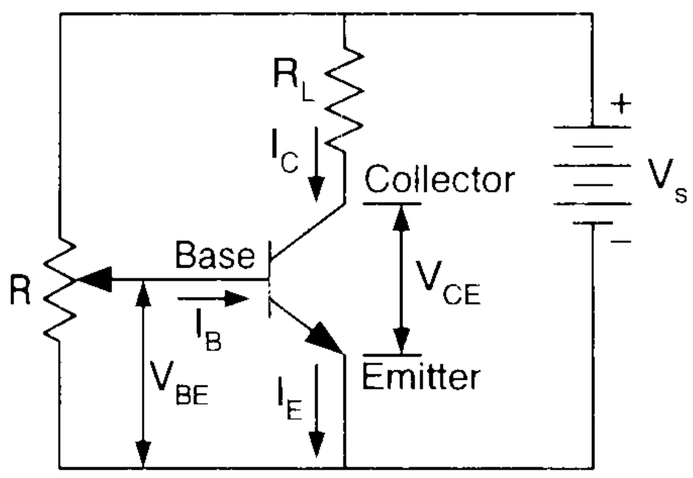


Fig. 2.13 Circuit showing the adjustment of the base voltage until the transistor acts as a closed switch.

An analysis of the circuit shown in Fig. 2.13 using Kirchhoff's law gives the current I_E as:

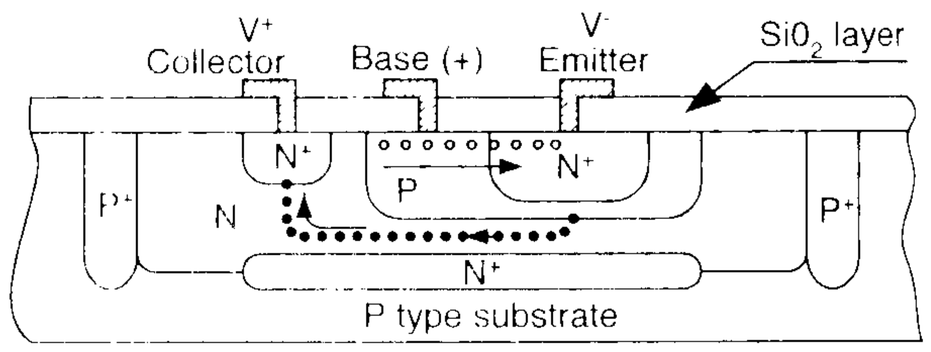
$$I_E = I_B + I_C \qquad (2.4)$$

The current gain G for the transistor is defined as:

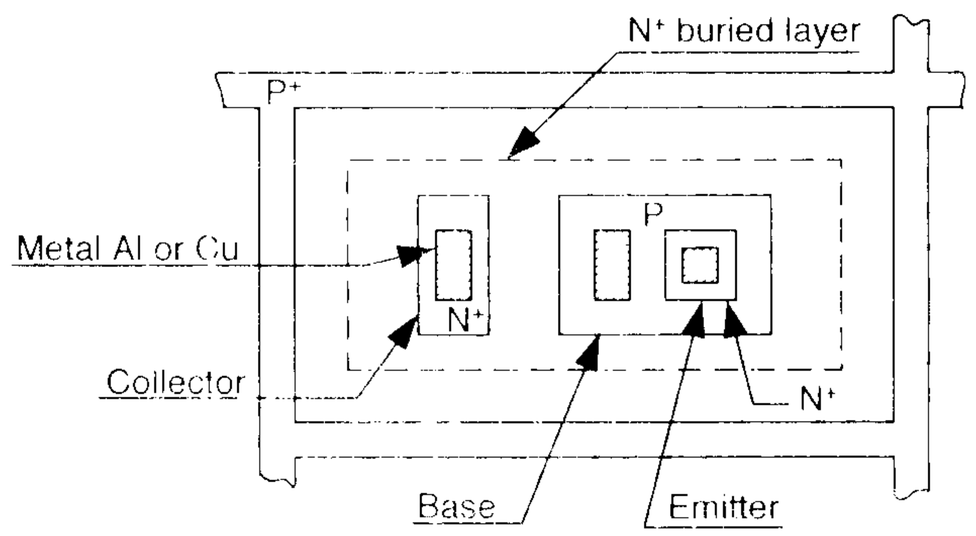
$$G = I_E / I_B \qquad (2.5)$$

The current gain is dependent primarily on the thickness and area of the base and the concentration of the doping on both sides of the P-N junctions. Current gains of 10 to 100 are commonly achieved in bipolar transistors.

The illustration of the N-P-N bipolar transistor, presented in Fig. 2.12, is a schematic used to show the concept of back to back diodes with forward and reverse bias. The actual construction details of a transistor differ significantly from this simple diagram. Production techniques for integrated circuits are based on placing 10^4 to 10^8 transistors as well as other components such as resistors, capacitors and diodes on a wafer cut from a very large single crystal of silicon. The wafers used in production facilities today are usually from 4 to 12 in. (100 to 300 mm) in diameter and 21 to 31 mils (525 to 775 μm) thick. The transistors are arranged in a planar array on one surface of the wafer. The N-P-N structure, described previously, is formed through the surface of the wafer using ion implantation techniques to vary the local concentration of the N and P type impurities. A cross section of a single N-P-N bipolar transistor more representative of current technology is represented in Fig. 2.14a.



(a) Section view



(b) Top view

Fig. 2.14 Structure of a N-P-N transistor fabricated on a wafer of P doped silicon.

The structure in Fig. 2.14 is developed beginning with a wafer cut from a P doped silicon crystal. Islands of N^+ (high concentration of N impurities) and more localized islands of N and P are formed by a series of production steps which utilize advanced lithographic processes to define feature sizes which are measured in nanometers. A layer of SiO_2 which serves as an insulator to prevent surface currents is formed by oxidizing the top surface of the wafer. Channels to the N^+ and P islands are etched through the SiO_2 layer and filled with a vapor deposited aluminum or copper which provide the electrodes for the collector, base and emitter. Deep islands of P^+ are placed around the structure to provide transistor to transistor isolation. The electron flow from the N^+ emitter region, through the thin P base to the N island and N^+ collector, is shown in Fig. 2.14a.

A top view of the structure is illustrated in Fig. 2.14b where the planar dimensions of the islands of doped silicon are indicated. As manufacturing processes are improved the dimensions of the features are reduced and the overall planar area required for a transistor, capacitor, resistor or a gate is decreased.

With feature si
silicon necessa
very small fea
chip.

Featur
chip. Feature
RAM memory
half pitch of 6
technology in
this generatior

2.6.2 Metal

The MOSFE
bipolar transi
electrons or h
fabricated by
islands are im
between the t
form the gate.
N island (the
N island (the
switching act
attracts the e
channel beca
which blocks
acts as an op
are repelled.
the opposite
transistor act
NMOS trans
mode, becau
flow of electi

Fig. 2.15

NM
depletion m
fabrication d
and the drai
the N chann

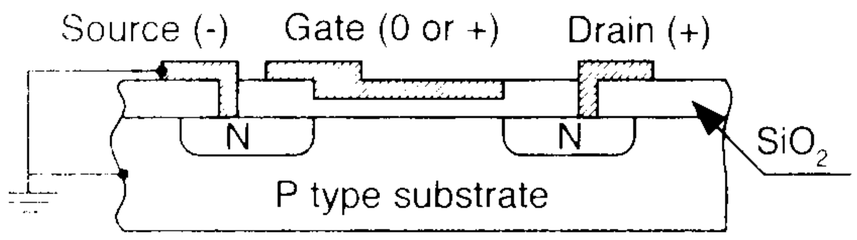
[1] The data on
their predicti

With feature sizes on gate lengths of 40 nm and the pitch on metal lines of about 170 nm, the area of silicon necessary for a bipolar transister with its isolating boundary is of the order of 25×10^{-14} m^2. The very small feature sizes[1] permit the placement of about 400 million transistors on a 10 × 10 mm size chip.

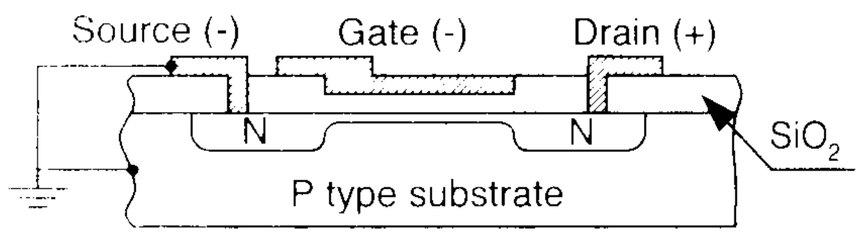
Feature size is a common measure of the technology used in developing and manufacturing a chip. Feature size is defined in technical dictionaries as half the distance between cells in a dynamic RAM memory chip. For example, in 2006 the technology used to produce a dynamic RAM yielded a half pitch of 65 nm. This fact implies that the smallest feature size on a chip manufactured with this technology in 2006 is smaller than the dynamic RAM half pitch of 65 nm. Hence, the gate lengths on this generation of chips will be smaller than 65 nm.

2.6.2 Metal Oxide Semiconductor Field Effect Transistor

The MOSFET (metal-oxide-semiconductor-field-effect-transistors) devices were developed after the bipolar transistor. As the name implies, these transistors utilize field effects to control the flow of electrons or holes to perform the switching action. A typical structure of a NMOS transistor, which is fabricated by implanting N type islands in a P type substrate, is presented in Fig. 2.15a. After the two islands are implanted in the substrate, the entire surface is covered with an oxide layer. The P substrate between the two islands forms the channel region. A metal electrode is placed over the SiO_2 layer to form the gate, which is essentially a small capacitor coupled to the channel region. The electrode to one N island (the source) is connected to the P substrate and then it is grounded. The electrode to the other N island (the drain) is connected to a positive supply voltage. The voltage on the gate controls the switching action of the transistor. When the gate voltage is zero, the positive voltage on the drain attracts the electrons in the N material at the source, but these electrons cannot flow through the P channel because it contains many holes that combine with the electrons to produce a depletion layer, which blocks the further flow of electrons. With zero or negative gate voltage, the NMOS transistor acts as an open switch. However, if a positive voltage is applied to the gate, the holes in the P channel are repelled, the electrons from the source travel across the P channel to the drain and current flows in the opposite sense. With a positive gate voltage the transistor acts as a closed switch. This type of NMOS transistor action is called the enhancement mode, because positive gate voltage enhances the flow of electrons from the source to the drain.



(a) Enhancement mode



(b) Depletion mode

Fig. 2.15 Section view of a NMOS transistor showing key features

NMOS transistors are fabricated differently if they operate in the depletion mode. The depletion mode NMOS transistor is illustrated in Fig. 2.15b. The primary difference between the fabrication details is the presence of a thin layer of N material under the gate which connects the source and the drain. In this configuration, a zero gate voltage permits the normal passage of electrons through the N channel from the source to the drain. However, the application of a negative voltage to the gate

[1] The data on feature sizes were taken from the International Technology Roadmap for Semiconductors and reflect their predictions for Application Specific Integrated Circuit (ASIC) chips produced in 2006.

repels the electrons from the thin N channel converting it to an insulating channel blocking the flow of electrons from the source to the drain and effectively opening the transistor switch.

PMOS transistors are also designed to operate in a manner similar to the NMOS transistors as illustrated in Fig. 2.16. In this case, two islands of P material are placed in an N substrate. With PMOS the polarities of the source, gate and the drain are reversed and the charge carriers are holes rather than electrons. The PMOS transistors are not commonly employed, because their switching time is 2 to 3 times longer than the NMOS transistors. The longer switching time is due to the lower mobility (velocity) of the holes in switching PMOS in comparison to the velocity of the electrons that are utilized in NMOS switching.

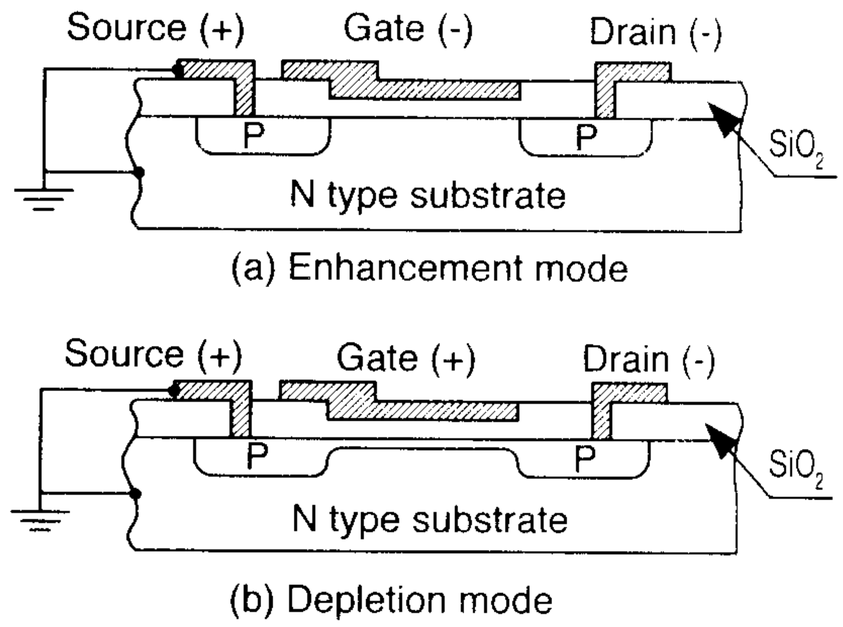


Fig. 2.16 Section view of a PMOS transistor showing key features.

A third type of MOS transistor is the complementary MOS device known as CMOS. This device incorporates both a NMOS and a PMOS transistor connected in series as shown in Fig. 2.17. The two gates are connected together and are activated by a single control voltage. The two drains are also connected together to provide an output signal, which can be either high or low. The two sources are wired separately with the source on the NMOS transistor grounded and the source on the PMOS transistor connected to the V^+ supply. This series connection of the two transistors results in very small power dissipation in the steady state. The two transistors require control signals of opposite polarity to conduct: consequently, one or the other of the transistor switches is always open. Because of this fact, the V^+ supply and ground are not connected together except for the period during switching from one state to the other. When the control signal is negative, the PMOS transistor conducts and the output goes high; however, the current flow is very small because of the of the high impedance and the low voltage difference between the source and the output. When the control signal is positive, the NMOS transistor conducts and the output goes low. Again the current flow is minimized due to the high impedance and low potential difference between the gate and the source. Large current flow and power loses occur only during the switching of the CMOS transistor[2]. During switching, a conducting path exists between the V^+ supply and ground while one transistor is turned on and the other is turned off. The power dissipated in a CMOS transistor depends on the frequency of operation of the transistor with increasing power dissipation as the frequency increases.

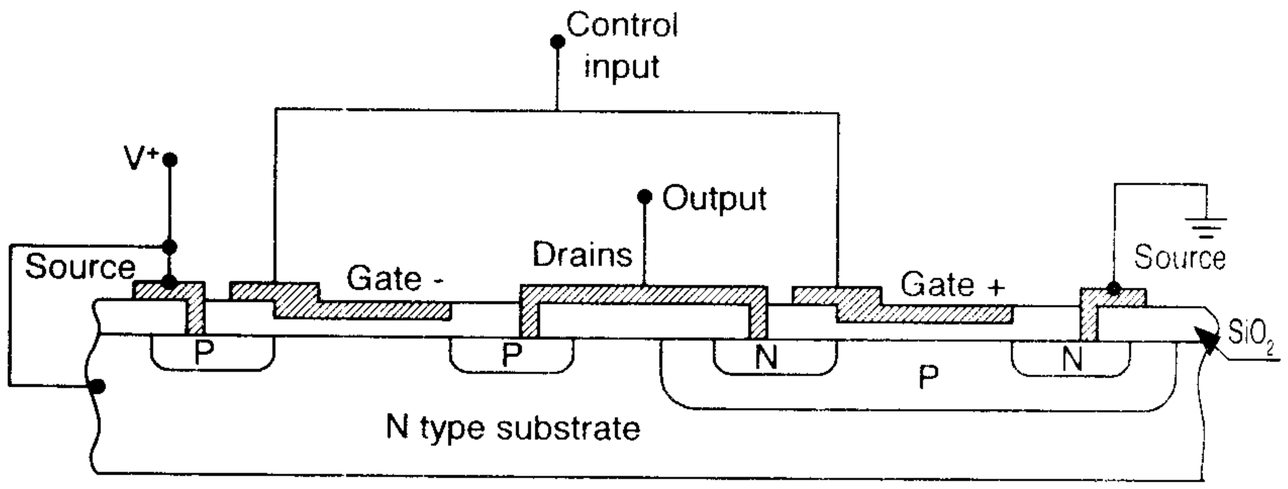


Fig. 2.17 A CMOS transistor employs a series connected pair of NMOS and PMOS transistors.

[2] Power loses that occur due to leakage current have been ignored in this discussion.

2.7 COMPARISON OF TRANSISTOR TYPES

Bipolar transistors are low impedance devices which require relatively high currents to operate and dissipate large amounts of heat. Bipolar transistors switch rapidly, and for equal feature size and power density, they operate at a higher frequency than the MOS devices. Bipolar transistors must be isolated to prevent interaction with adjacent devices on the chip. This isolation is achieved by using isolation barriers as demonstrated in Fig. 2.14; however, placement of these barriers utilizes a sizable area on the chip. This placement reduces the number of bipolar transistors that can be placed on a chip. Consequently, the density (number of components/area) is lower with bipolar than with MOSFET technology. Because of the very high heat dissipation and the large chip area required for a bipolar transistor, MOSFET technology has become dominant. In particular, the CMOS transistor with its low power dissipation is the most widely employed technology in a wide range of products.

The MOSFET devices exhibit high impedance and consequently they operate with lower currents. The isolation of the MOSFET transistors is inherent and no isolation barriers are required. This feature saves chip area and component density is higher than that which can be achieved with bipolar technology. Processing of NMOS or PMOS is generally easier than with bipolar and the design scales readily to smaller feature sizes. The channels are shallow and parallel to the surface and their length is dependent upon feature size. This lateral deployment adds capacitance and limits the switching speed. The very shallow diffusion layers yield small cross-sectional areas for current flow near the surface at relatively high current density.

PMOS transistors are rarely used because their switching speed is low relative to NMOS transistors. The mobility of holes as charge carriers is much less than the mobility of electrons and because PMOS depends largely on the movement of holes it is inherently slow. NMOS operates with electrons as charge carriers and is much faster than PMOS. The primary advantages of NMOS over bipolar is density, ease in scaling to smaller feature sizes and less complexity in manufacturing.

CMOS transistors consist of two series connected transistors, one PMOS and the other NMOS. In this arrangement, the forward current from the V^+ supply to ground occurs only during the switching operation and not during steady state in either the high or the low mode. This feature greatly reduces their power dissipation. Also relatively high switching speed can be achieved at low power levels. Feature size is critical in determining MOSFET switching speeds, and as feature size has progressively decreased in the past four decades, CMOS has become the most commonly used technology in the design and manufacturing of new chips for new products.

2.8 LOGIC GATES

In most digital instrumentation or computers, logic gates are employed in large numbers to perform complex operations at extremely high frequencies. The circuits involved are large often containing 10^5 to 10^8 logic gates. However, while these circuits are large, they are simple because they contain only three different types of basic gates. The basic gates include the AND, OR and NOT. Other more complex logic elements are often used, but they consist of combinations of these three gates.

The AND gate may be represented by the circuit shown in Fig. 2.18 where two switches A and B are placed in the line from the source to the load. The voltage V_s is applied to the load Z_L only if switch A and switch B are both closed. The possibilities for the AND gate are listed in a truth table shown below in Table 2.2. Note that (0) is used to represent a false statement and (1) to represent a true statement. With regard to the voltage applied to the load, the number (1) indicates true—V_s is applied to the load.

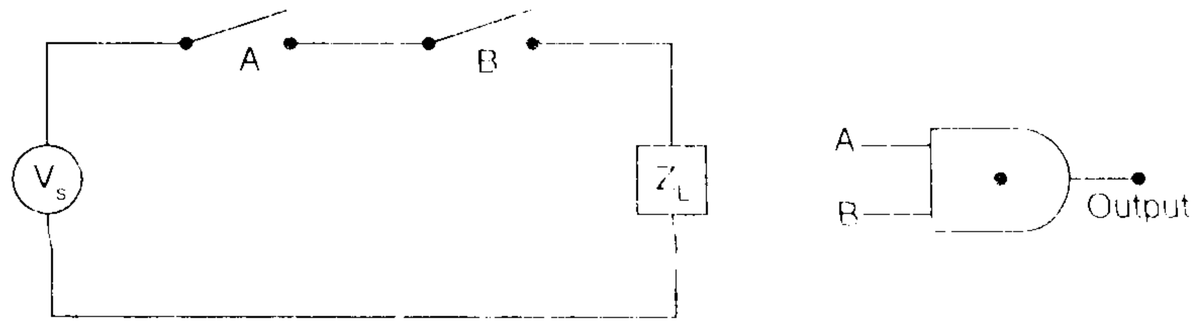


Fig. 2.18 Circuit representing the AND gate and its symbol.

Table 2.2
Truth table for the AND gate A – B = T

Switches or inputs		Output
A	B	T
0	0	0
0	1	0
1	0	0
1	1	1

The OR gate is represented by the circuit given in Fig. 2.19, where two switches A and B are placed in parallel in the line between the voltage source and the load. When A is closed **or** when B is closed, the voltage is applied to the load Z_L and T = 1.
The truth table for an OR gate with two switches is presented in Table 2.3.

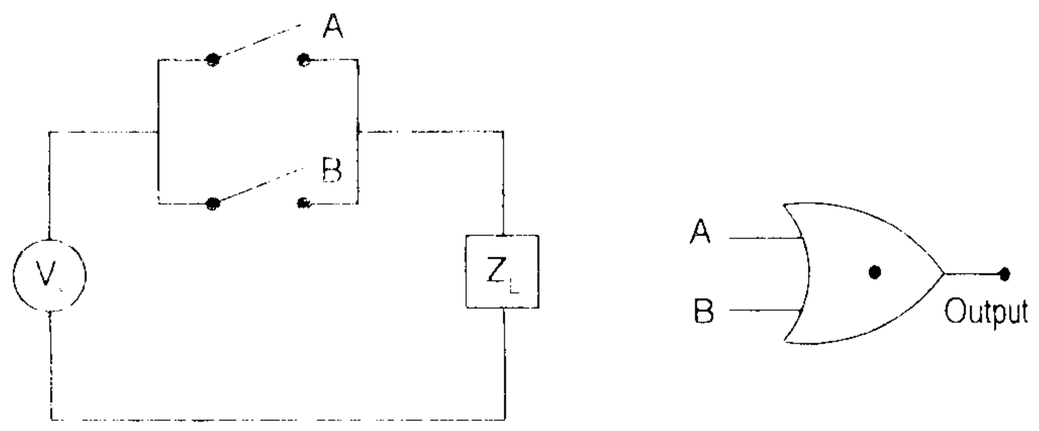


Fig. 2.19 Circuit representing the OR gate and its symbol

Table 2.3
Truth table for the OR gate A + B = T

Switches or inputs		Output
A	B	T
0	0	0
0	1	1
1	0	1
1	1	1

The NOT gate which is illustrated in Fig. 2.20 is an inverter. In this case, the mechanical switch has been replaced by a transistor which is turned on (closed) by a positive input voltage. If the input signal to the transistor is (0), the transistor act as an open switch, no current flows and the output voltage is V_s or (1). When the input signal goes to (1), the transistor conducts acting like a closed switch and the output is grounded giving the low state or (0). It is clear from this circuit that the input is high (A) the output is low ($\overline{A}$) and changing the input to low ($\overline{A}$) results in an output which is high (A).

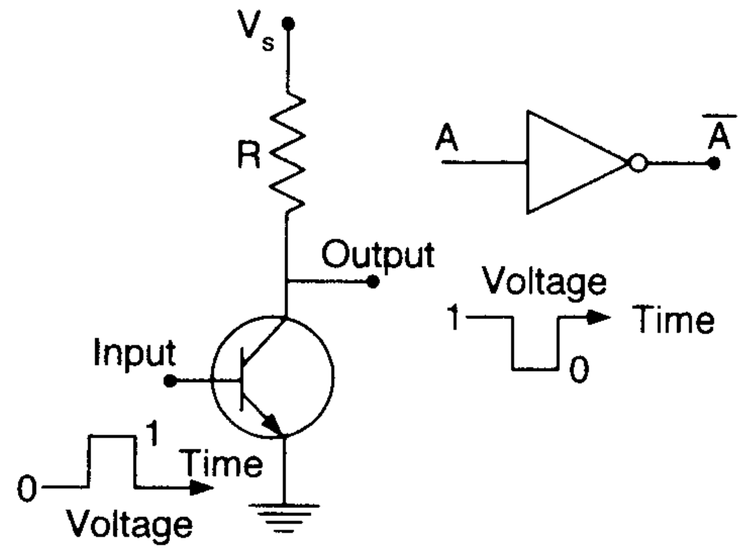


Fig. 2.20 Circuit representing the NOT gate and its symbol.

These basic gates are arranged in circuits to perform digital functions. For example, the arrangement of AND, OR and NOT gates shown in Fig. 2.21 represents a binary adder. A digital system is composed of many of these digital functions and may contain a million or more of the simple basic gates. The number of chips use to build the logic circuits depends on the scale of integration used to fabricate the chip. With ULSI (ultra large scale integration) and larger chip sizes, it is possible to place of the order of 10^8 gates on a single chip thus permitting the development of extremely large digital systems with relatively few chips.

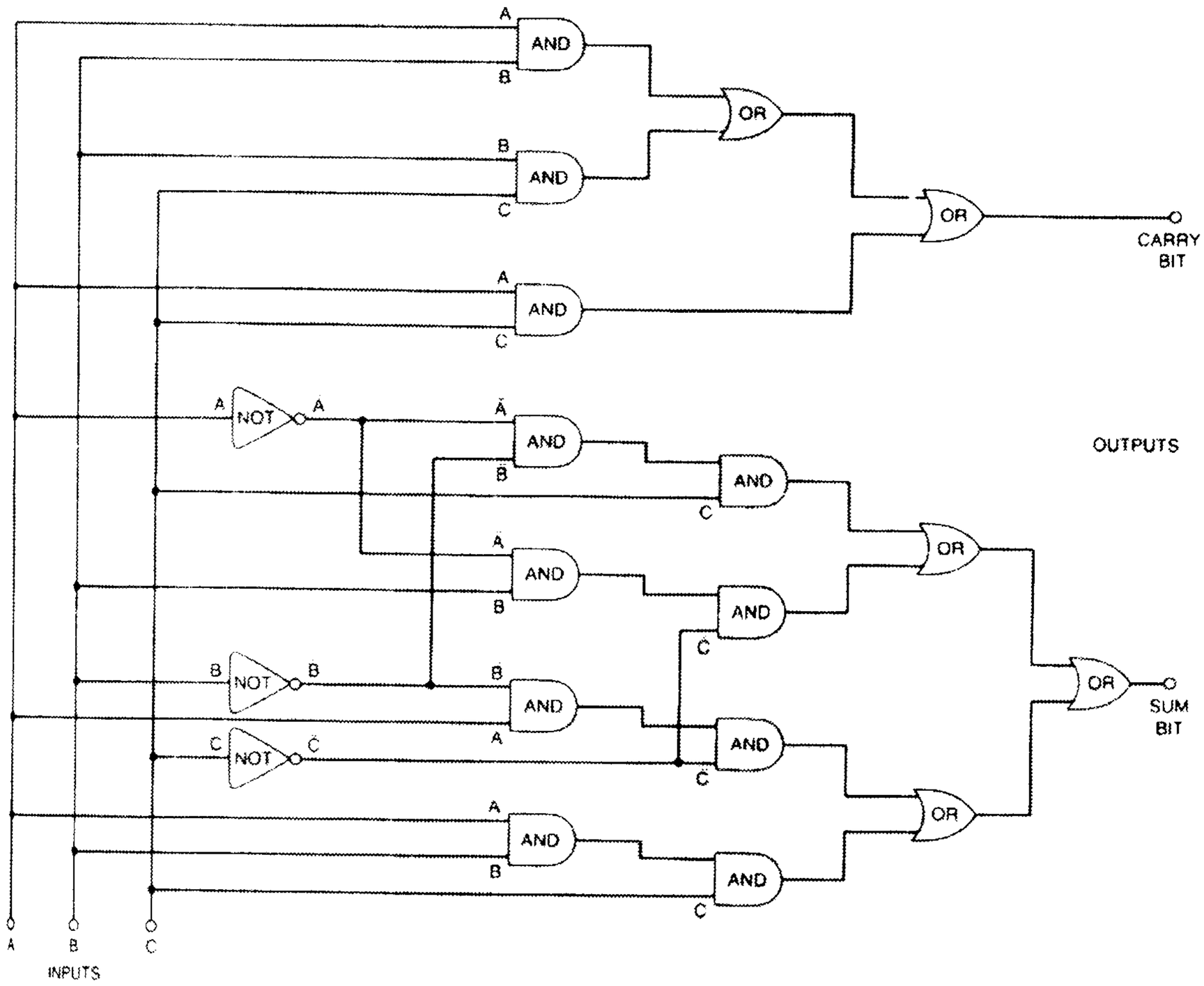


Fig. 2.21 The binary adder, a functional element of a digital system, is constructed from many simple AND, OR and NOT gates.

2.9 GATE TECHNOLOGIES

Transistors are used to replace mechanical switches shown in the logic gates represented in Fig. 2.19 and 2.20. Other electrical components such as resistors, capacitors and diodes are implanted on the surface of a silicon wafer to produce an integrated circuit (IC). Since the advent of the transistor in the early 1950's, there has been an evolution in the design of logic gates to improve their performance. The objective of the design changes have been to:

1. Increase the switching speed of the gate.
2. Improve density by reducing the number of components required for the gate
3. Reduce the chip area needed for the gate components.
4. Reduce the noise generated in switching.
5. Improve the drive capability.
6. Reduce power requirements.

2.9.1 MOSFET Gates

MOSFET circuits were introduced for switches in gates in the late sixties, with a FET gate incorporating a resistive load as shown in Fig. 2.22a. The resistor of this circuit was replaced with a second MOSFET transistor as shown in Fig. 2.22b to lower the resistance of the gate and to improve performance. Next, the enhancement type load transistor was replaced with a depletion type transistor as shown in Fig. 2.22c to accomplish switching at lower gate voltages. Finally, two complementary transistors one PMOS and the other NMOS were connected in series as illustrated in Fig. 2.22d to minimize power dissipation. This configuration is the most popular of the MOSFET devices in use today and it is known as CMOS.

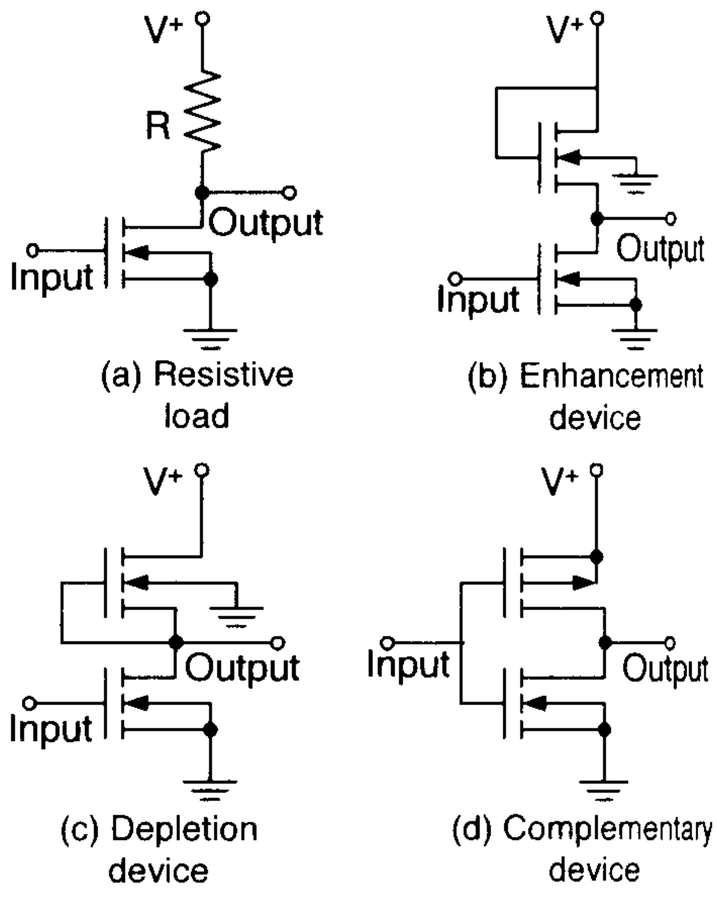


Fig. 2.22 Progressive development of MOSFET logic devices.

When the two complementary transistors are connected in series they operate from a single input and very little power is dissipated as the two transistors require input signals of opposite polarity for conduction to occur. For this reason, they are never both on at the same time and the supply voltage V^+ is not connected to ground through the transistors. The power dissipated depends on the frequency of switching, the feature size and the leakage currents due to quantum tunneling [1]. Power density for very-high-performance logic chips approaches 1000 W/cm^2, but drop dramatically for more moderate-to-low performance chips to 1.0 to 10 W/cm^2.

The fabrication details of a CMOS gate, which are presented in Fig. 2.23, show the use of polysilicon (highly doped amorphous silicon) as the gate material and the use of relatively shallow N and P channels. Because isolation channels between transistors are not required, the density (gates/unit

area) that can be achieved in MOSFET technology is higher than that which can be obtained with bipolar technology.

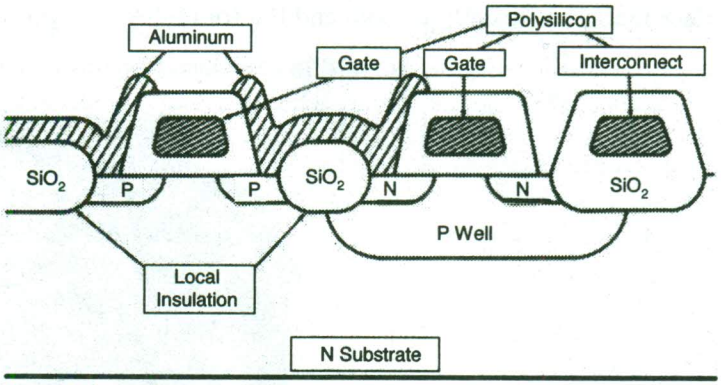


Fig. 2.23 Fabrication details revealed in a section view of a CMOS transistor.

2.10 CHIP AND WAFER FABRICATION

Most chips are fabricated from doped silicon to form products like a central processing unit (CPU) or an Application Specific Integrated Circuit (ASIC). An example of the Intel P-6 CPU with 5.5×10^6 transistors is presented in Fig. 2.24. The chips are small rectangular pieces of silicon that are cut from a wafer of silicon. The wafers are double-side polished and have a notch to identify the orientation of the grain structure. The thickness of the wafer varies with its diameter with the thickness increasing with the diameter. The 300 mm thick wafers are supplied with thickness ranging from 0.75 to 0.80 mm.

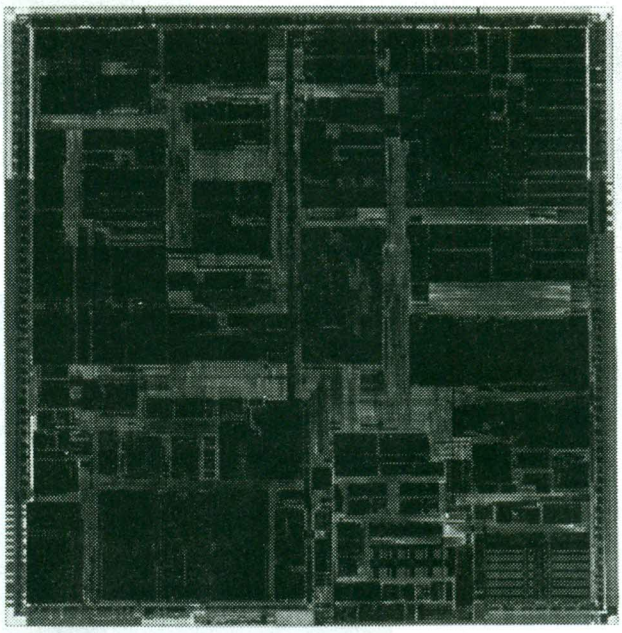


Fig. 2.24 Example of a recent CPU from Intel that contains 5.5 million transistors.

2.10.1 Wafer Fabrication

Single crystal ingots of silicon are usually produced by the Czochralski (CZ) method. In the CZ process, polycrystalline silicon, purified to less than 1 part per billion of trace impurities, is molten together with dopants in a quartz crucible. The dopants, such as boron and phosphorus, are used for adjusting the resistivity of the silicon. A small single crystal silicon rod (seed), about 5 mm in diameter by 200 mm long, is placed in the molten silicon with an inert gas atmosphere at a temperature of about 1421° C. The seed is slowly rotated and pulled up from the melt forming a single crystalline ingot with the same orientation as the seed. An illustration of growing a crystal by the CZ process is presented in Fig. 2.25.

Ingot diameters vary considerably with 200 to 300 mm commonly employed by most semiconductor manufactures[3]. The ingots are ground to the finished diameter and then sliced into thin

[3] It is anticipated that wafer diameters will increase to 450 mm by 2012.

wafers with a diamond saw. The slices are sorted and those with the same thickness are then lapped and etched to remove the surface damage caused by the saw. Finally, the wafer is polished to produce a surface that is sufficiently smooth and flat for high yield optical photolithography.

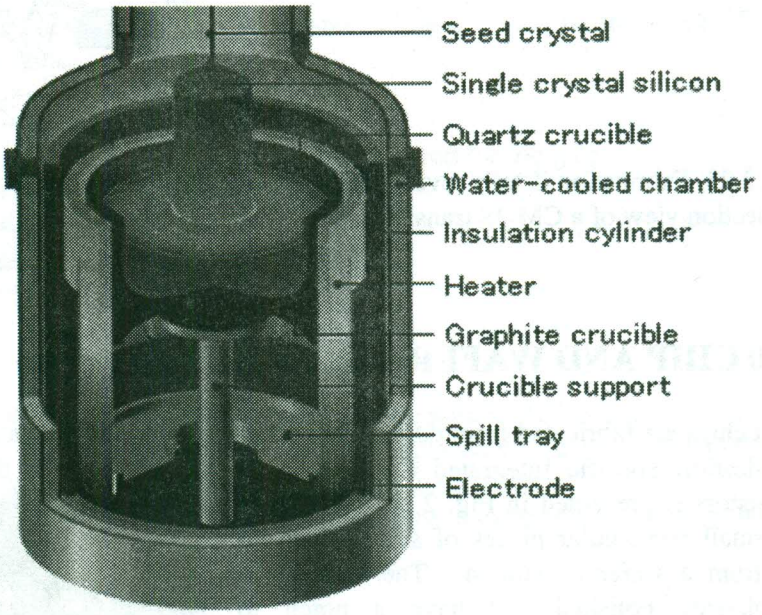


Fig. 2.25 Growing single crystal silicon ingots in a CZ crucible and furnace.

A photograph of a several wafers cut from a finished ingot of silicon is presented in Fig. 2.26.

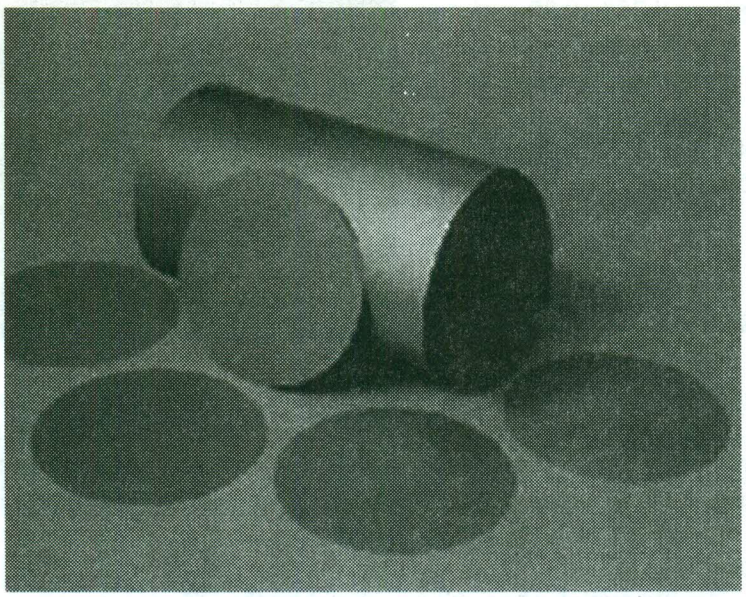


Fig. 2.26 Photograph of wafers cut from an ingot of single crystal silicon.

Mechanical properties of interest for silicon are presented in Table 2.4.

Table 2.4
Mechanical and thermal properties of Silicon.

Property	Value
Density	2.33 g/cm^3
Tensile Strength	7,000 MPa
Modulus of Elasticity	190 GPa
Poisson's Ratio	0.17
Knoop Hardness	850 kg/mm^2
Coefficient of Thermal Expansion	2.59×10^{-6}/°C
Thermal Conductivity at 300 °K	1.56 W/(cm-°C)
Thermal Diffusivity	0.9 cm^2 /s
Specific Heat	0.70 J/(g-°C)

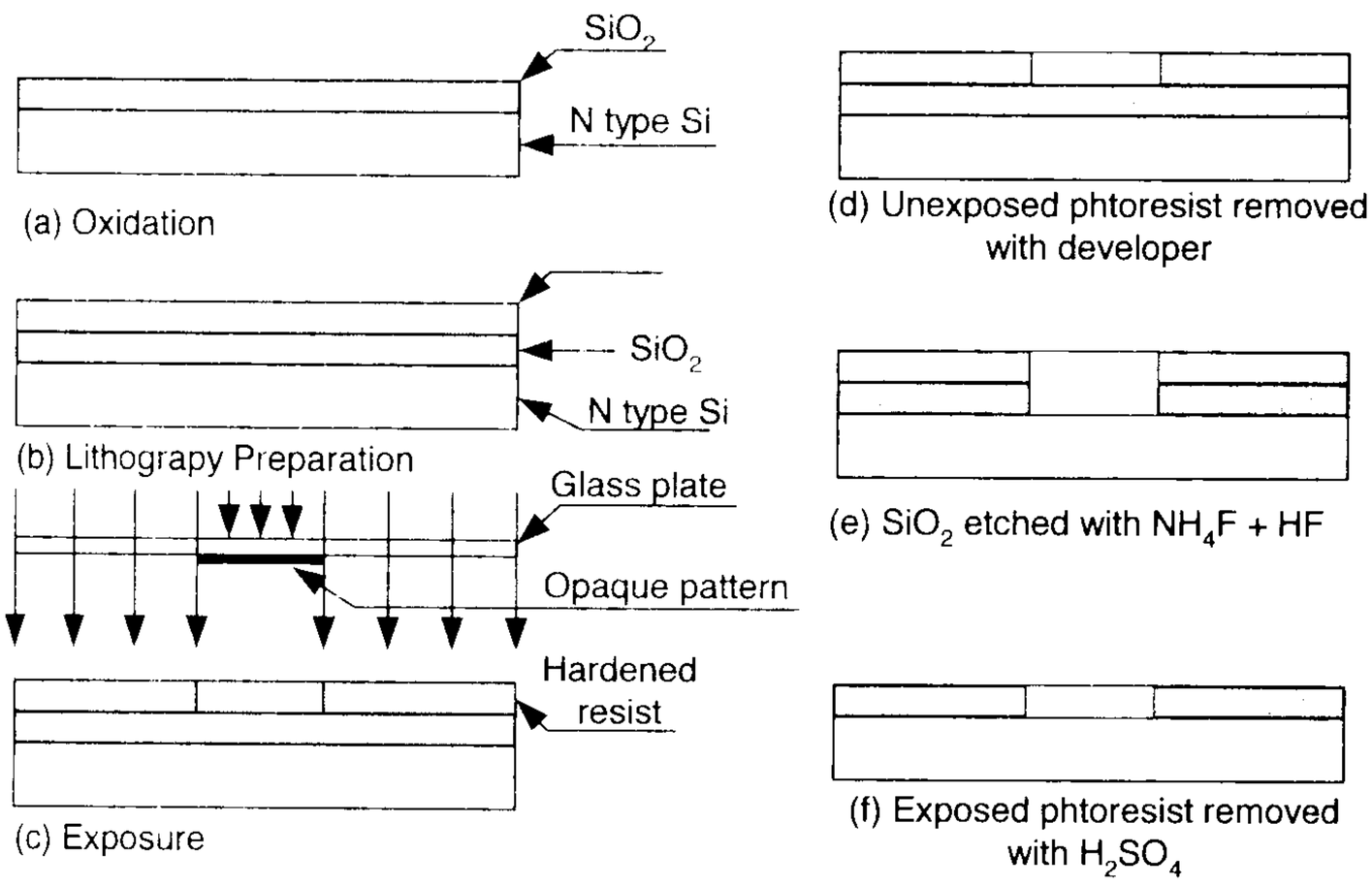


Fig. 2.27 Illustration of several steps in the lithographic process for producing the pattern of the features on a chip.

2.10.2 Masks and Lithography

Chips are produced on wafers by using a lithographic process that was originally developed to produce printing plates. This process, illustrated in Fig. 2.27, has been refined and is used today to produce chips as well as printed circuit boards. The steps in the process include:

1. Form a thin layer of oxide over the entire surface of the silicon wafer.
2. Apply a layer of photoresist over this surface.
3. Position a mask with the chip's surface features over this photosensitive layer.
4. Project UV light through the mask to image the surface features onto the photoresist.
5. Develop the photoresist opening small regions where chip features are to be applied.
6. Etch away the layer of oxide with hydrofluoric acid.
7. Remove the photoresist with sulfuric acid.

After the feature regions of the chip are defined, the P and N regions on the chip are form with diffusion doping or with ion implantation methods. At this stage in the production process, i necessary to apply the wiring to connect the various P and N islands to form gates and to wire the g together to produce the logic or memory elements. Photolithography is also used in making the requi interconnections. The surface of the wafer is covered with photoresist and a mask is used that has t image of the wiring lines and the bonding pads. The photoresist is exposed with UV light to form image and of the wiring lines and bonding pads. The wafer is then placed in a vacuum chamber a aluminum is sputtered over the entire surface of the wafer. When the photoresist is removed a sing layer of wiring connecting many of the features remains. To complete the wiring a layer of polyimide applied to insulate the first wiring layer from the second. Second, third and additional wiring layers a added as required to complete the interconnections of the gates and on-chip capacitors and to conn the I/O with the bonding pads.

2.10.3 Bonding Pad Preparation

Bonding pads are use for I/O connections and as such they are larger and thicker than the normal wiri used for interconnections of on-chip components. The type of bonding pad preparation depends on th technology used in connecting the chip to its first level package (the chip carrier). Three methods ar commonly employed—wire bonding, TAB bonding and flip-chip bonding. We will describe the pa preparation for each of these interconnection technologies in the following subsections.

Wire Bonding

The most widely used wire materials for connecting chip pads to chip carrier pads are gold an aluminum; however, silver and copper are also employed. Gold was the original material used whe wire bonding technology was developing because it is a noble metal that is easy to form and to weld Gold wire can be bonded in the shape of a either a wedge or a ball. Aluminum is still widely used because of its lower welding temperature, good electrical performance and lower cost than gold. More recently copper wire has been introduced for ball bonding, in spite of its higher welding temperature because of its lower costs and improved looping characteristics in automatic ball bonding machines.

Ball bonds have the advantage that they can be placed on very close centers consistent with newer chips with high I/O. On the other hand, aluminum wire can only be wedge bonded and the use of aluminum is limited when the bonding pads are on close centers. An example of ball bonds formed from gold wire on two rows of chip bonding pads is presented in Fig. 2.28.

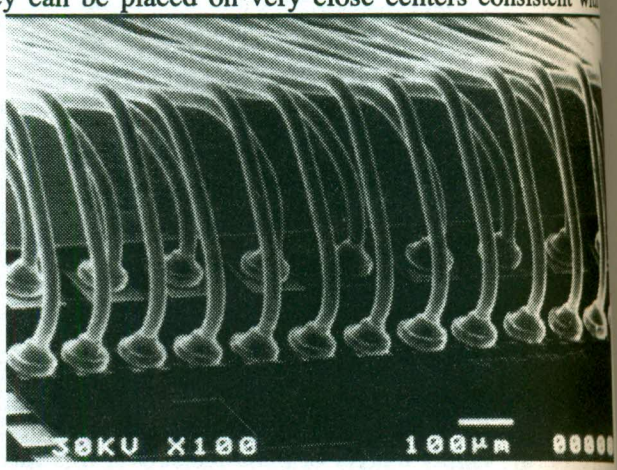


Fig. 2.28 Gold wires welded to chip bonding pads to make connections with the chip carrier.

The pad preparation depends on the wire material and the bonding process. The thermo-sonic process is used with gold wires and currently comprises about 90% of all wire bonding. It is performed at temperatures ranging from 100°C to 240°C depending on the wire and pad materials. The weld is formed when the ultrasonic energy combines with the capillary action of thermo-compression bonding.

After the feature regions of the chip are defined, the P and N regions on the chip are formed with diffusion doping or with ion implantation methods. At this stage in the production process, it is necessary to apply the wiring to connect the various P and N islands to form gates and to wire the gates together to produce the logic or memory elements. Photolithography is also used in making the required interconnections. The surface of the wafer is covered with photoresist and a mask is used that has the image of the wiring lines and the bonding pads. The photoresist is exposed with UV light to form an image and of the wiring lines and bonding pads. The wafer is then placed in a vacuum chamber and aluminum is sputtered over the entire surface of the wafer. When the photoresist is removed a single layer of wiring connecting many of the features remains. To complete the wiring a layer of polyimide is applied to insulate the first wiring layer from the second. Second, third and additional wiring layers are added as required to complete the interconnections of the gates and on-chip capacitors and to connect the I/O with the bonding pads.

2.10.3 Bonding Pad Preparation

Bonding pads are use for I/O connections and as such they are larger and thicker than the normal wiring used for interconnections of on-chip components. The type of bonding pad preparation depends on the technology used in connecting the chip to its first level package (the chip carrier). Three methods are commonly employed—wire bonding, TAB bonding and flip-chip bonding. We will describe the pad preparation for each of these interconnection technologies in the following subsections.

Wire Bonding

The most widely used wire materials for connecting chip pads to chip carrier pads are gold and aluminum; however, silver and copper are also employed. Gold was the original material used when wire bonding technology was developing because it is a noble metal that is easy to form and to weld. Gold wire can be bonded in the shape of a either a wedge or a ball. Aluminum is still widely used because of its lower welding temperature, good electrical performance and lower cost than gold. More recently copper wire has been introduced for ball bonding, in spite of its higher welding temperature, because of its lower costs and improved looping characteristics in automatic ball bonding machines.

Ball bonds have the advantage that they can be placed on very close centers consistent with newer chips with high I/O. On the other hand, aluminum wire can only be wedge bonded and the use of aluminum is limited when the bonding pads are on close centers. An example of ball bonds formed from gold wire on two rows of chip bonding pads is presented in Fig. 2.28.

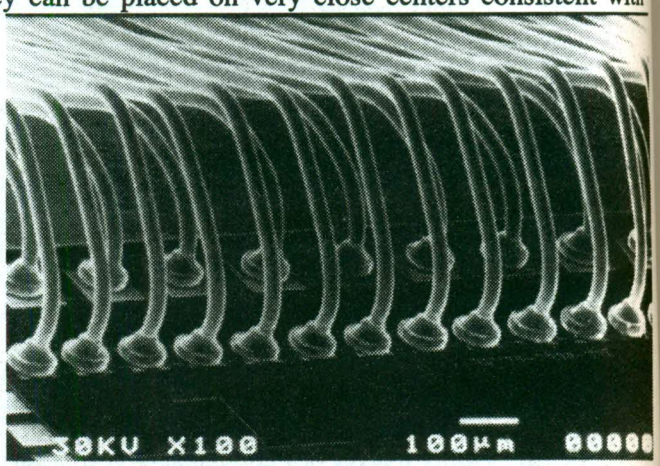


Fig. 2.28 Gold wires welded to chip bonding pads to make connections with the chip carrier.

The pad preparation depends on the wire material and the bonding process. The thermo-sonic process is used with gold wires and currently comprises about 90% of all wire bonding. It is performed at temperatures ranging from 100°C to 240°C depending on the wire and pad materials. The weld is formed when the ultrasonic energy combines with the capillary action of thermo-compression bonding.

An illustration showing the capillary tube and the clamp used with this process is presented in Fig. 2.29. In this instance, the bonding pads are square and they have been built up to increase their thickness by sputtering additional aluminum onto the pad surfaces. In some cases the aluminum pads are plated with gold to improve the thermo-compression bonding process.

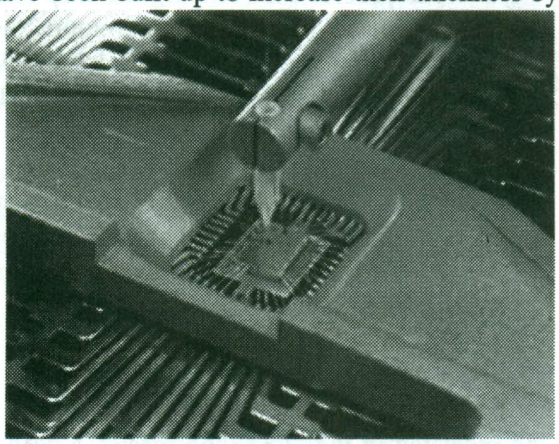


Fig. 2.29 A wire bonding machine.

Tape Automated Bonding (TAB)

Tape-automated bonding (TAB) is another method for connecting the bonding pads to the leadframe of a chip carrier. The interconnecting lines are patterned on a multilayer polyimide tape, as indicated in Fig. 2.30. The tape is positioned above the chip with ends of the copper lines in line with the bonding pads on the chip.

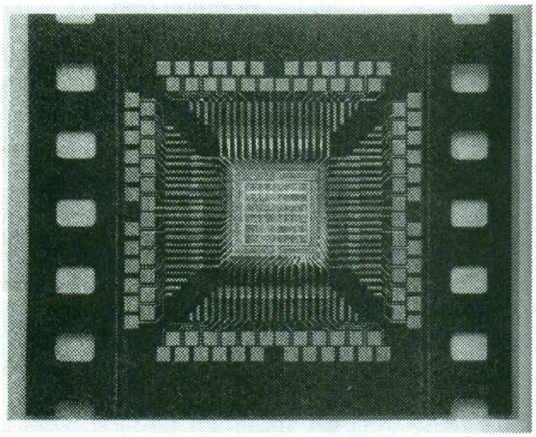


Fig. 2.30 Tape automated bonding (TAB) used to make connections between the chip bonding pads and the lead frame of a chip carrier.

The connections are made in a compression welding process, where a "thermode" is pressed down on the inner-lead-ends forcing them into contact with the bumped pads around the periphery of the chip, as indicated in Fig. 2.31. This process is rapid as all of the connections are made simultaneously; however, the bonding pads on the chip must be bumped to provide the clearance required by the presence of the polyimide carrier.

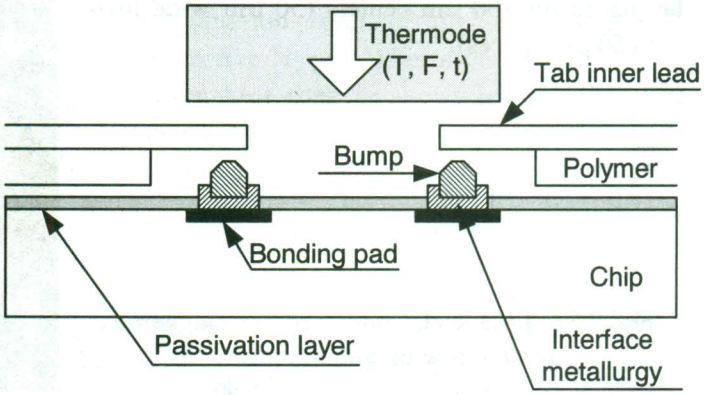


Fig. 2.31 Thermo-compression welding of copper leads to the bumped bonding pads about the periphery of a chip.

The process for bumping the chip bonding pads is illustrated in Fig. 2.32. After the chip pad have been coated with a layer of aluminum, the surface of the chip is coated with a passivation layer th is usually polyimide. The next step is to apply two or three layers of interface metallurgy (sometim called Under Bump Metallization (UBM). The interface metallurgy often consists of three layers metal, which include:

1. An adhesion layer to provide a strong, low-stress mechanical and electrical connection the bond pad metal and the adjacent passivation.
2. A diffusion barrier layer to constrain the diffusion of solder into the adhesion layer or th aluminum.
3. A solder wettable layer offers a wettable surface that is compatible with either gold o solder used to form the bumps.

As indicated in Fig. 2.32, photoresist is used to delineate the bonding pads after the application of th interface metallurgy. Either gold or silver bumps are applied by plating; however, high-lead solde bumps may be applied by evaporation. For TAB connections, the top of the bumps are finished with relatively flat top.

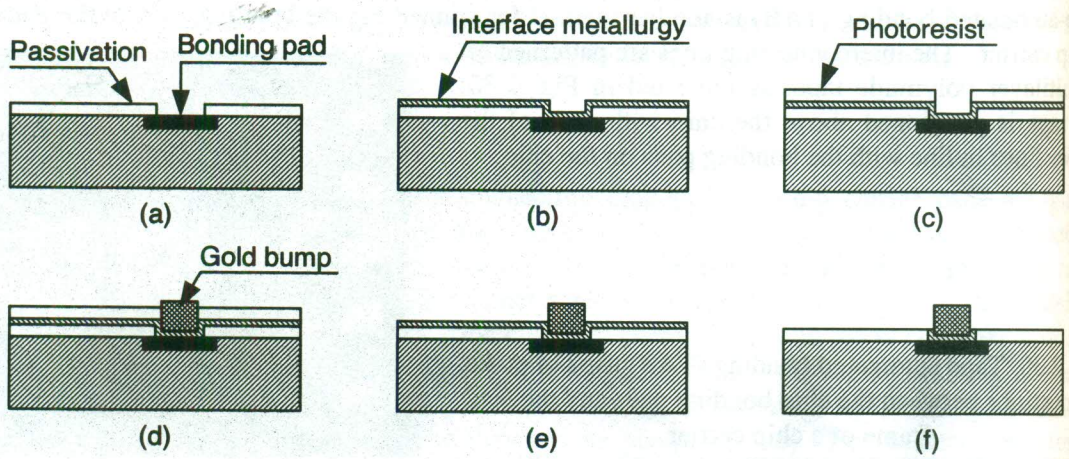


Fig. 2.32 Six steps in a process used for chip bumping.

An example of TAB connections to gold bumps is illustrated in Fig. 2.33. Note the staggered double row of solder bumps. This design permits additional I/O to be deployed about the perimeter of the chip. The TAB leads with the staggered I/O design can be placed on 100 μm centers (50 μm wide lines with 50 μm spaces).

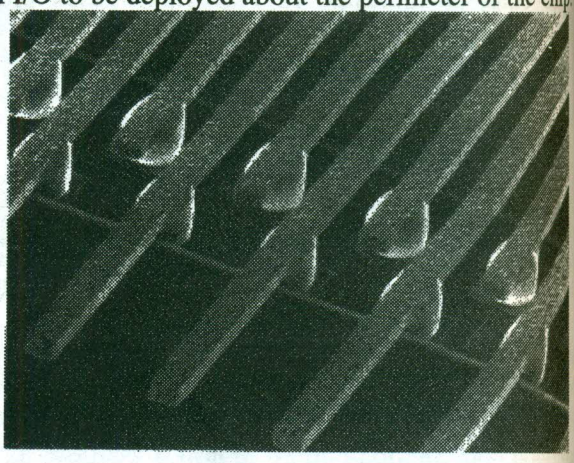


Fig. 2.33 TAB leads connected to a staggered double row of solder bumps.

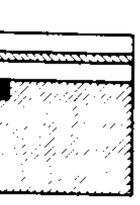
r the chip pads
ation layer that
gy (sometimes
three layers of

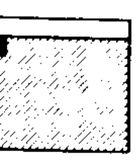
connection to

on layer or the

either gold or

olication of the
gh-lead solder
inished with a





ggered double
er of the chip.

Flip-Chip Bumping

There are several steps in fabricating flip-chip packages, which include bumping the wafer, cutting the dies from the wafer, attaching the bumped chip to a substrate and filling the remaining space under the die with an adhesive. The conductive bump, the attachment materials, and the processes employed differ with the types of flip-chip packages.

The bump serves four different functions in a flip-chip package. First, it provides the electrical and thermal conductive path from chip to substrate. Second, the bump serves to mechanically attach the chip to the substrate. Third, it acts as a spacer preventing electrical contact between the chip and substrate conductors. Finally, the bumps serve as very short leads that relieve the thermally induced shearing strains, which develop between board and substrate due to temperature changes.

The first step in solder bumping is preparing the bond pads located on the chip. This step involves the removal of the insulating aluminum oxide layer from the pads to prepare them for several layers of metals that are applied to enable good mechanical and electrical connection to the solder bump. Next a ball-limiting metal (BLM) is placed on the chip bond pads by either sputtering or plating. The BLM, which usually consists of several layers, defines and limits the solder-wetted area. The adhesion layer adheres well to the bond pad metal and the surrounding passivation, providing strong, low-stress mechanical and electrical connections. The diffusion barrier layer limits the diffusion of the bump material into the underlying material. The solder wettable layer offers an easily wettable surface to the molten solder during assembly to enhance bonding of the solder to the underlying metal. A protective layer may be used to prevent oxidation of this underlying layer.

There are several processes employed in applying metal to form the flip-chip bumps. Wet chemical processes to etch away the aluminum oxide and plate conductive nickel-gold bumps onto the chip bond pads. After plating the required thickness of nickel, a gold layer is added for corrosion protection. The gold stud bump process employs a slight modification of the wire bonding process. A gold ball is produced by melting the end of a gold wire forming a sphere. The gold ball is attached to the chip bond pad as the first part of a wire bond. The modified wire bonder breaks off the wire after attaching the ball to the chip bond pad. This gold ball provides a permanent connection through the aluminum oxide layer to the underlying metal.

High lead solder (97% Pb and 3% Sn) is one of the most popular materials for chip bumping. The IBM C4 process was based on evaporation of high lead solder to form the bumps. Conventional tin-lead eutectic solder cannot be evaporated because of the difference in vapor pressures of the two metals. Bump size and position is determined by the solder mask openings and spacing; however, the solder is subsequently reflowed forming spherical solder bumps.

Electroplating solder to form the bumps is less costly and this manufacturing technique is more flexible than evaporation. The BLM consists of an adhesion layer of titanium tungsten, a chrome/copper layer, a copper wetting layer and a gold protective layer[4]. After a photoresist layer is patterned to form the bump sites, solder[5] is electroplated over these sites, as shown in Fig. 2.34. The photoresist is then removed and the layers of BLM are etched away. The solder bumps are reflowed to form spheres that are attached to the chip pads, as illustrated in Fig. 2.35 and Fig. 2.36.

---

[4] A number of different metallurgies are employed in sputtering the BLM layers including—TiW/CrCu/Cu/Ni/Au, Cr/CrCu/Cu/Au, Ti/Cu/Ni/Au, Ti/Cu and Cr/Cu/Cu/Ni/Au.

[5] Several different solder compositions may be employed including—Pb/Sn (97/3), Pb/Sn (37/63) and the Sn/Ag/Cu family of lead free solders.

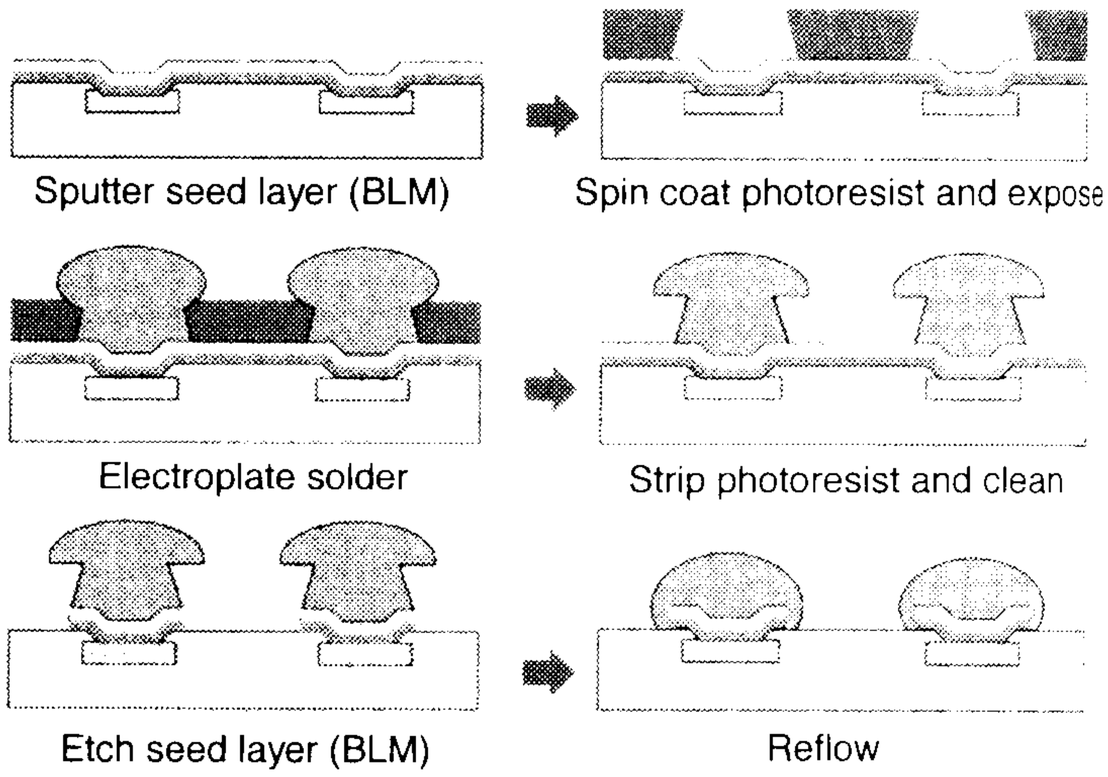


Fig. 2.34 Fabrication steps in producing solder ball bumps for flip-chip bonding.

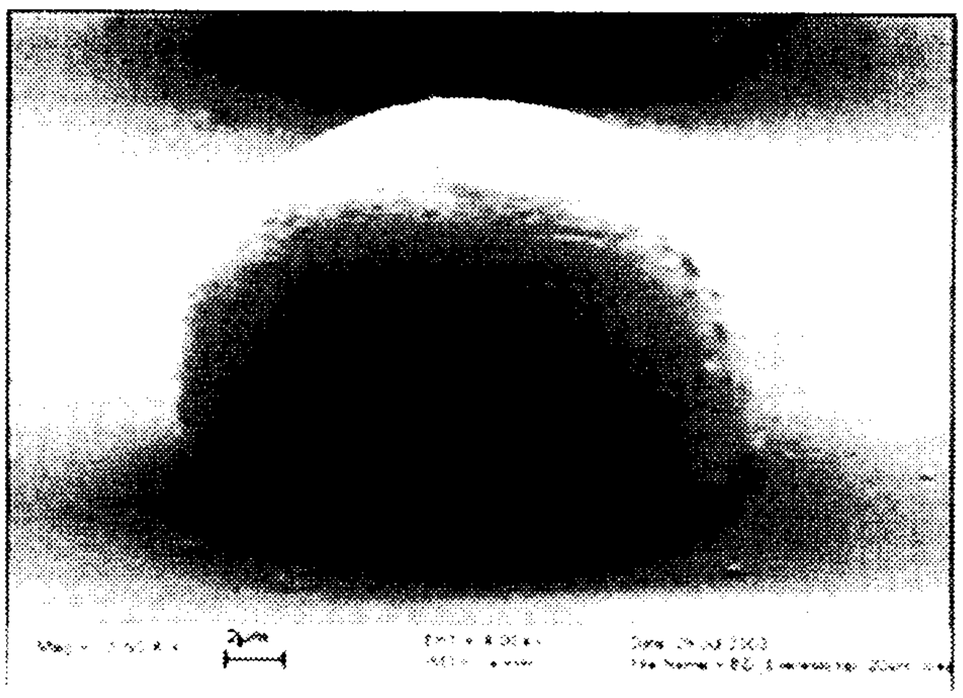


Fig. 2.35 Photomicrograph of a micro-solder bump 25 μm in diameter on a 50 μm pitch.

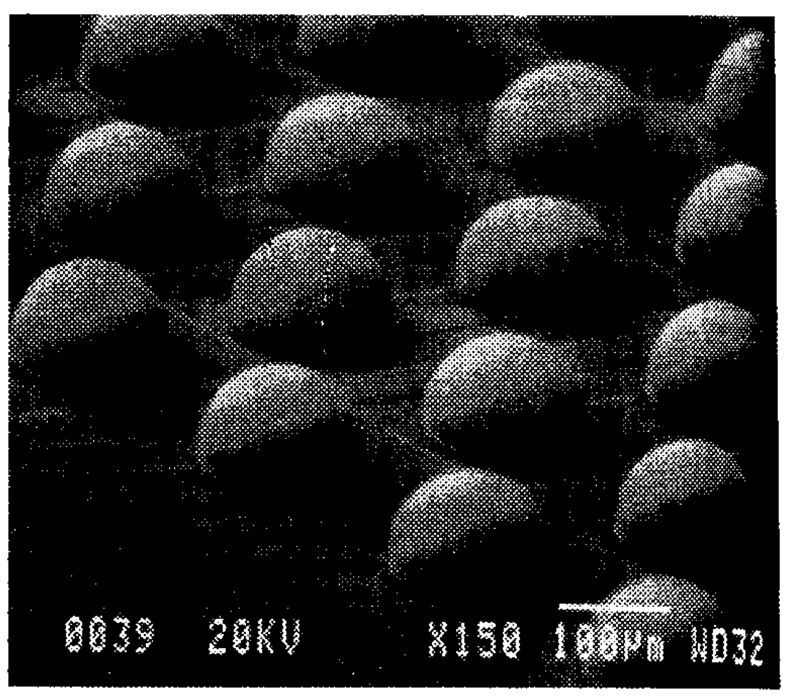


Fig. 2.36 Staggered array of solder balls on the back side of a flip-chip.

2.10.4 Advanced Chip Structures

As chips are designed with millions of transistors that operate at increasing frequencies the number of on-chip interconnections and the number of I/O are increasing dramatically. The cost and the development time for a line of chips are also increasing markedly. These facts are drivers that are changing both design and manufacturing methods used in the production of chips. One important advance has been the development of a process for using copper to replace the aluminum used to interconnect the transistors, gates and capacitors on integrated circuits. In the late 1990s, IBM introduced a technology that allows the use of copper lines, in place of the aluminum interconnects in chips. IBM has developed electroplating process for copper that is known as damascene. The process begins by coating the surface of the wafer with a seed layer, or plating base, whose function is to conduct the current from the anode located at the wafer edge to all points on the wafer where a deposit of copper is specified. One approach to accommodate the seed layer with complex inner connection structures is illustrated in Fig. 2.37. Damascene plating involves depositing the seed layer over a patterned material, which is the insulator that must remain in place because it is a functional part of the

circuit. After plating, copper covers the entire surface of the wafer. Complete coverage causes major problems because all of the wiring lines are shorted out. It is necessary to remove the excess copper by chemical-mechanical polishing (CMP) which planarizes the surface of the wafer.

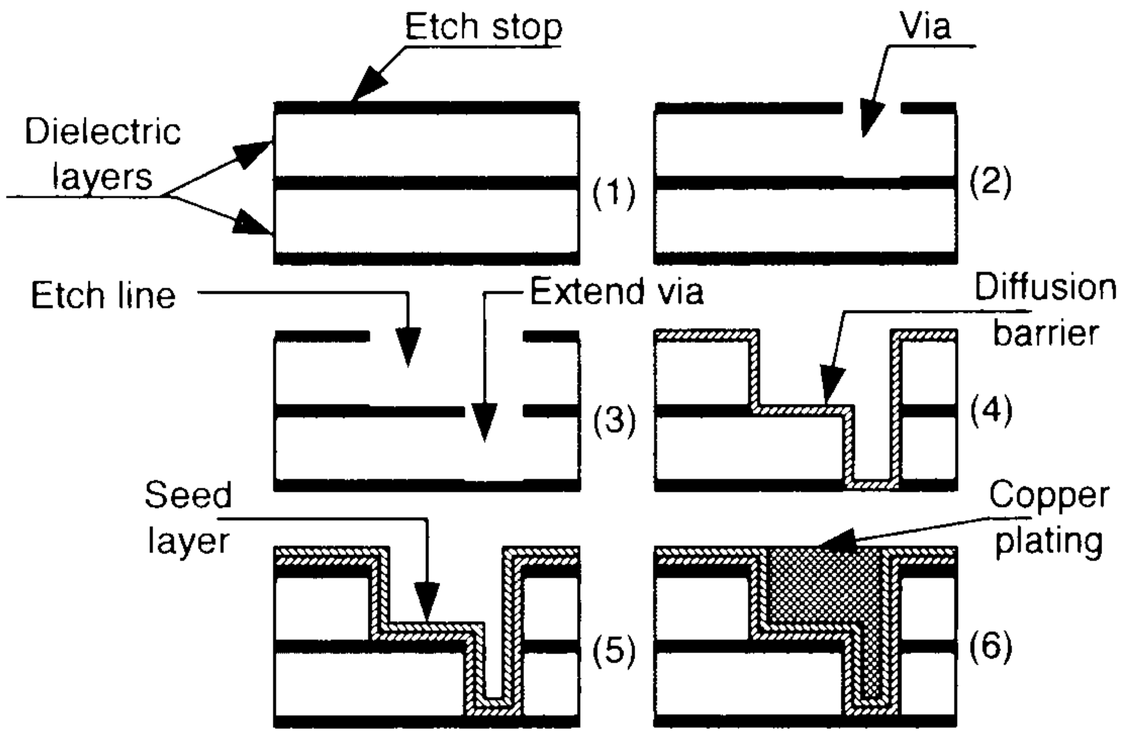


Fig. 2.37 Schematic illustration of several of the steps involved in the damascene plating process adapted to making multi-layer inner connections on chips.

The production steps illustrated in Fig. 2.37 are described below:

1. Deposit insulation with a low dielectric constant and separate layers with an etch stop (silicon nitride).
2. Define vias and etch first layer of insulation.
3. Define lines and via extensions into second layer and etch.
4. Deposit barrier layer to prevent diffusion of copper into the thin dielectric layer.
5. Deposit seed layer of copper over the entire surface of the wafer.
6. Plate copper filling vias and lines.
7. Not shown in Fig. 2.37 is the chemical-mechanical polishing (CMP) of the wafer surface necessary to remove the copper from the surface. The only copper remain is in via holes and trenches for the wiring lines.

Damascene electroplating is ideally suited for the fabrication of multi-layer interconnection structures, because it is possible to apply copper to fill via holes and adjacent trenches to form wiring lines without voids or seams. This plating process is compatible with the requirement for a barrier layer between the seed layer and the insulating (dielectric) layer. The purpose of the barrier layer is to prevent interaction between the copper and the insulator.

Examples of copper multi-layer interconnection structures are presented in Fig. 2.38 and Fig. 2.39. In Fig. 2.38, it is clear that six layers of copper wiring lines have been applied over the initial layer of tungsten interconnections. The layers are applied in sequence with chemical-mechanical polishing (CMP) of the wafer surface after each layer of copper has been deposited by the damascene plating process. The photomicrograph shown in Fig. 2.39 provides a view of the copper layers after the dielectric insulation has been removed. The photomicrograph indicates the complexity of the wiring necessary to inner connect the dense high-end chips developed with feature sizes less than 180 nm. The primary reason for using copper metallization is the reduced cross sectional area of the wiring lines increases the parasitic resistance to an unacceptable level if aluminum lines are employed. Companies such as IBM and Intel are already using this process with six or seven layers of copper wiring for inner connections. It is believed that nearly all semiconductor manufacturers will employ copper interconnections in the near future.

usually very small, the packaging methods used to house them are similar to those used to house chip
While the production methods are similar for MEMS and micro-electronic devices, the applicatio
differ significantly. For this reason, we will not treat MEMS, other than this brief introduction.

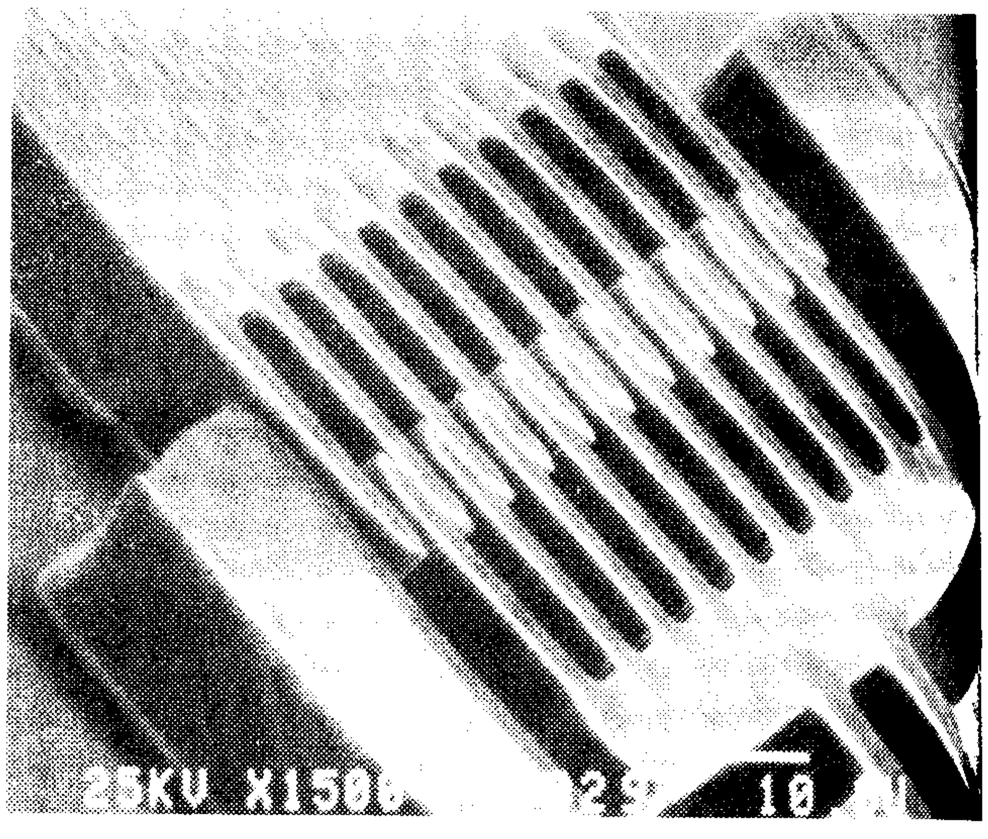


Fig. 2.41 Photomicrograph of a inner laced forks on a MEMS device.

2.11 SCALE OF INTEGRATION

In the 1970s Gordon Moore predicted the doubling of the number of transistors on integrated circuits (an excellent measure of chip performance) every 18 months. By the end of that decade, the predicition became known as **Moore's Law**. It permited the prediction of the maximim number of transistors on the most complex chips producted with the most modern technology. Because computing speed is related to the number of transistors on a chip Moore's law is also used to describe the continuing improvements in computing speed per unit cost.

Moore's law is empirical because it was formulated based on early observations. However, its global acceptance over the past 25 years has elevated its status. The law now serves as a goal for the semiconductor industry. Today semiconductor manufacturers invest significant amounts of money and engineering talent to follow Moore's Law because they know their competitors will be able to double their transistor count every 18 months and gain competitive advantage.

The implication of Moore's law for semiconductor manufacturers is immense. A typical major design project such as a new CPU takes between two to four years to reach production. Consider this developemnt time together with the implications of Moore's law. Moore's law is equivalent to a 1% per week improvement in performance for the industry. This fact implies that a short delay of a month or two in bring the new chip to market can mean the difference between a successful product one that is too large or too slow for customers to buy.

Recent semiconductor industry technology "roadmaps" predict that Moore's Law will continue for several chip generations. However the doubling time may be extended from 18 months to three years. Depending on the doubling time used in the preditions, Moore's law implies an increase in the number of transistors by a fctor from 10 to 100 in the coming decase.

2.11.1 Scaling of Feature Size

Since the introduction of integrated circuits in 1959, there has been a rapid increase in the number of components (transistors, resistors, capacitors and diodes) which can be placed on a single chip. The dramatic increase in the number of transistors, presented in Fig. 1.16, is due to a decrease in feature size and an increase in the size of the area used for the chip. Today, an advanced chip may support as many as 500 million transistors, with gate lengths of about 40 nm. Continued progress in reducing gate lengths to about 10 to 15 nm is anticipated in the coming decade unless limits on the lithographic process slow further development. The trend line for the marked reduction in feature size (gate length and on-chip wiring pitch) over the past four and a half decades is shown in Fig. 2.42.

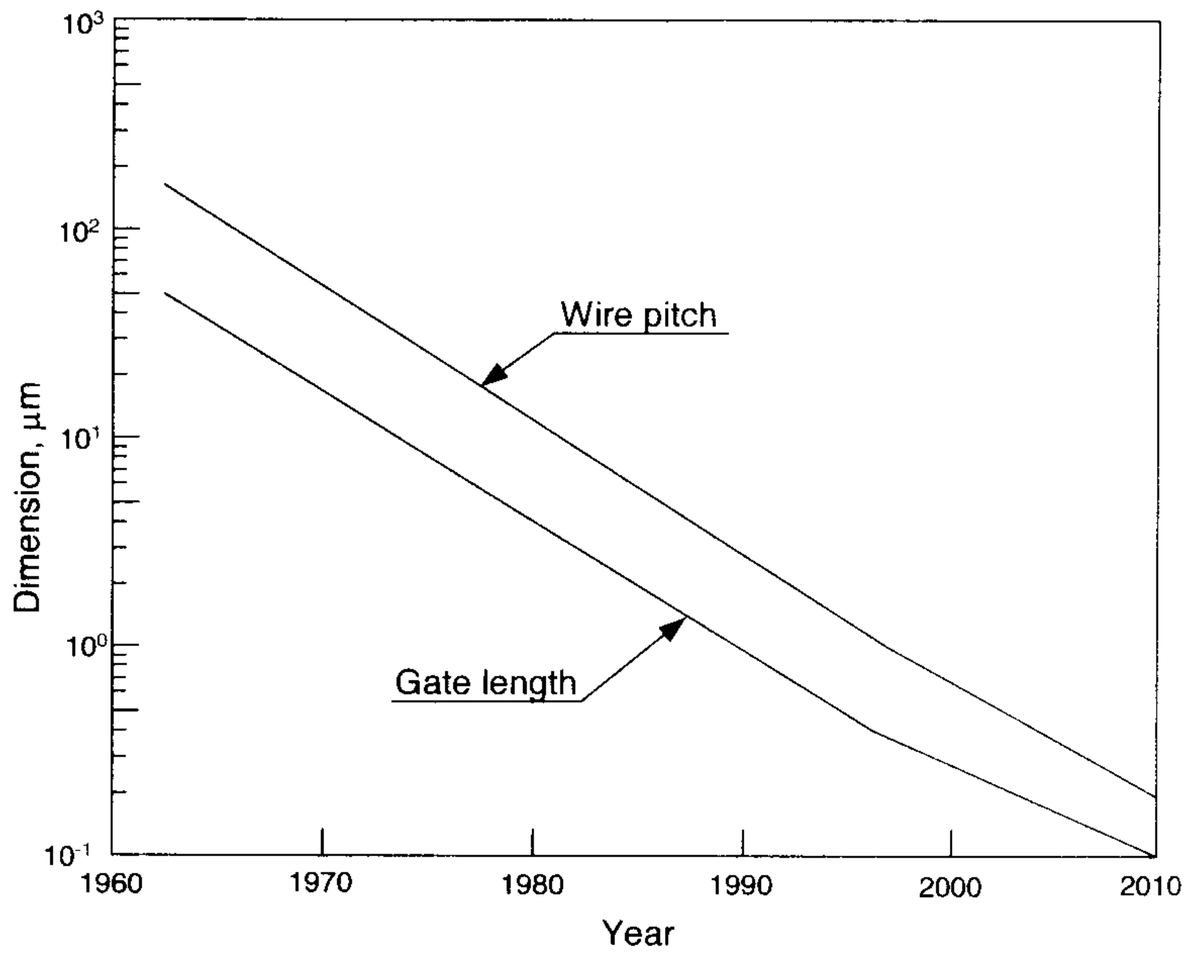


Fig. 2.42 Trend lines for scaling down the gate length and on-chip wiring pitch since 1960.

The increase in the number of components per chip reflects the gate length, the on-chip wiring pitch and the size of the silicon chip (die). The size of the silicon dies has also been increasing with continuous improvements in chip manufacturing methods that enable larger size chips while maintaining satisfactory yields. The trend line for increasing chip size with time is presented in Fig. 2.43.

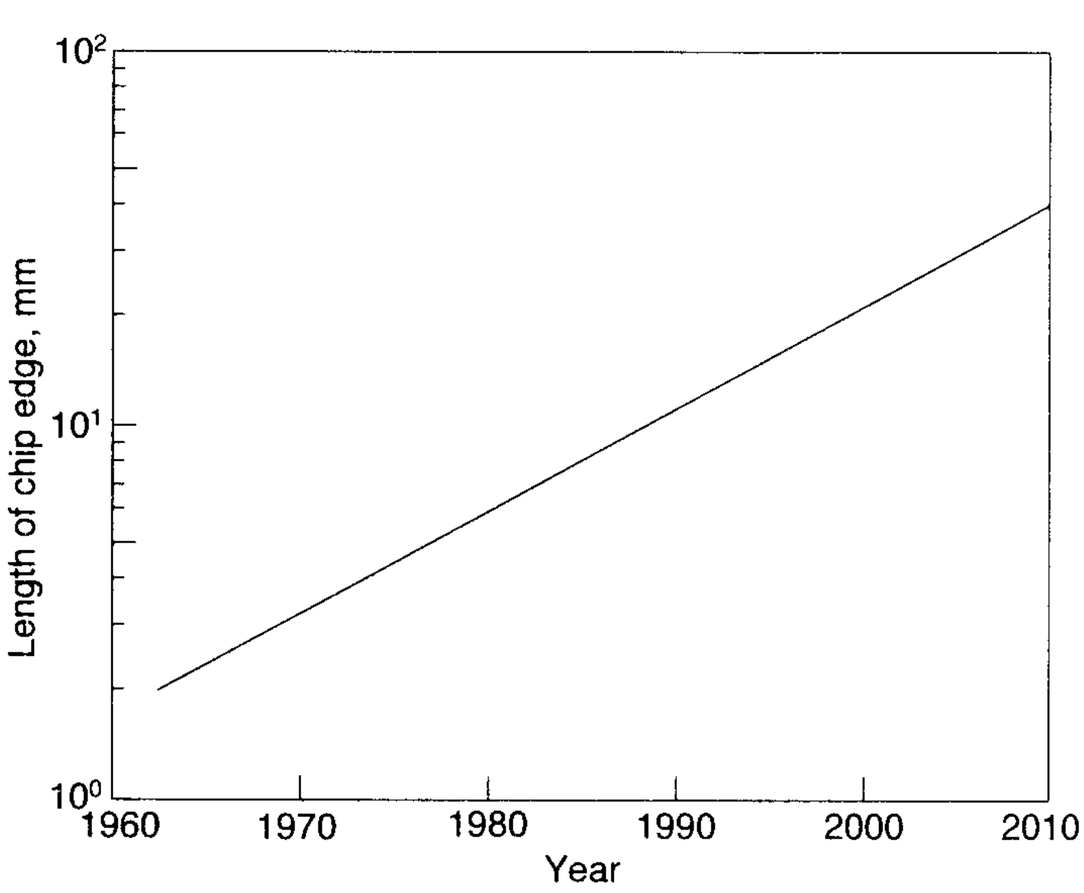


Fig. 2.43 Trend line for increasing chip size since 1960.

The trends presented in Fig. 2.42 and Fig. 2.43 yield an increase in the number of wiring tra and the number of wiring grids[6]. The trend lines showing this increase is presented in Fig. 2.44. T number of grids on a chip is an indication of the amount of functionality that can be achieved designing a complex integrated circuit. Modern chips have about 10^9 grid areas and by 2010 it expected that they will have 10^{10} wiring grid areas.

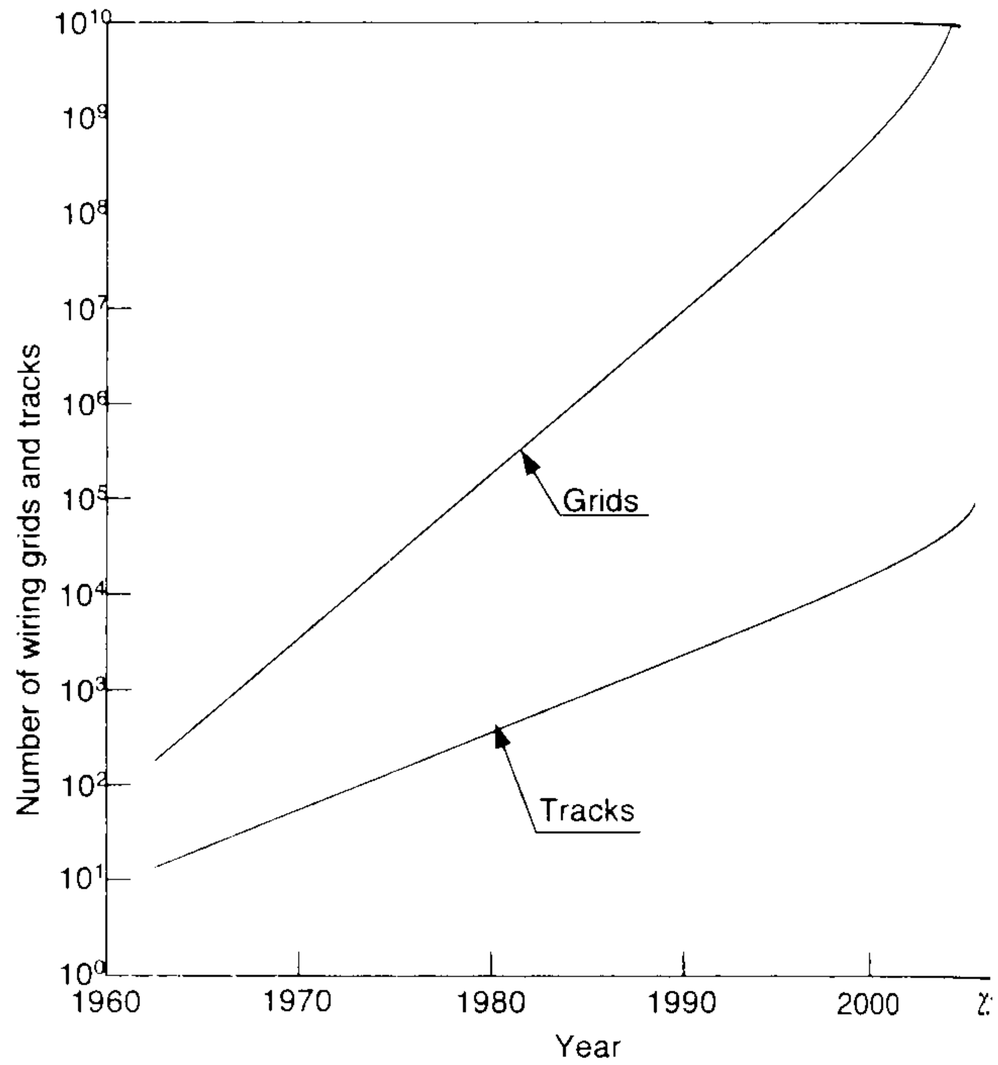


Fig. 2.44 Trend lines for increasing wiring grids and tracks since 1960.

The speed of on-chip functions is increasing as a consequence of decreasing gate length a illustrated in Fig. 2.45. The delay time of typical AND gate has decreased from tens of nanoseconds i the early 1960s to about 0.1 nanoseconds today and it is anticipated that it will decrease to a few tens o picoseconds by 2010.

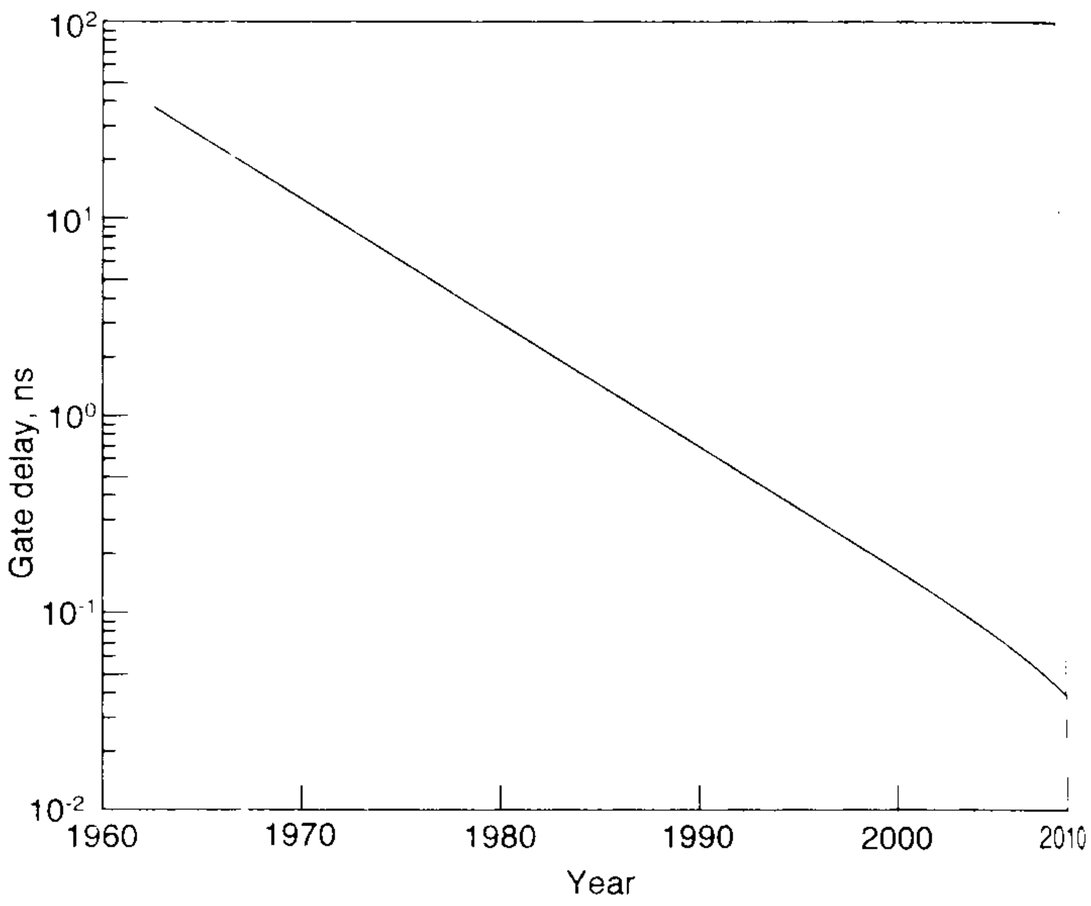


Fig. 2.45 Trend line for decreasing gate delay.

[6] Square wiring grids are superimposed on a chip and the wiring tracks extend along a sequence of these grids in either the x or y direction. The number of grid squares defined by the wiring tracks increases as the square of the number of wiring tracts.

A final gra shows the enhance silicon chip with ti chip depends on th carries and the sp Hence, a measure the number of wir gate delay. Th indicate that chip from about 4 in t 100 billion today.

Fig. 2.46 Trend capa

2.11.2 Scaling

The scaling dow electrical compo factor[7] K_w for th Wire pi area A scales a respectively. T

[7] **The scale fa**

A final graph showing the benefit of reducing feature size is presented in Fig. 2.46. This graph shows the enhanced capability of a typical silicon chip with time. The capability of a chip depends on the number of functions it carries and the speed of these functions. Hence, a measure of capability is given by the number of wiring grids divided by the gate delay. The results of Fig. 2.46 indicate that chip capability has increased from about 4 in the early 1960s to nearly 100 billion today.

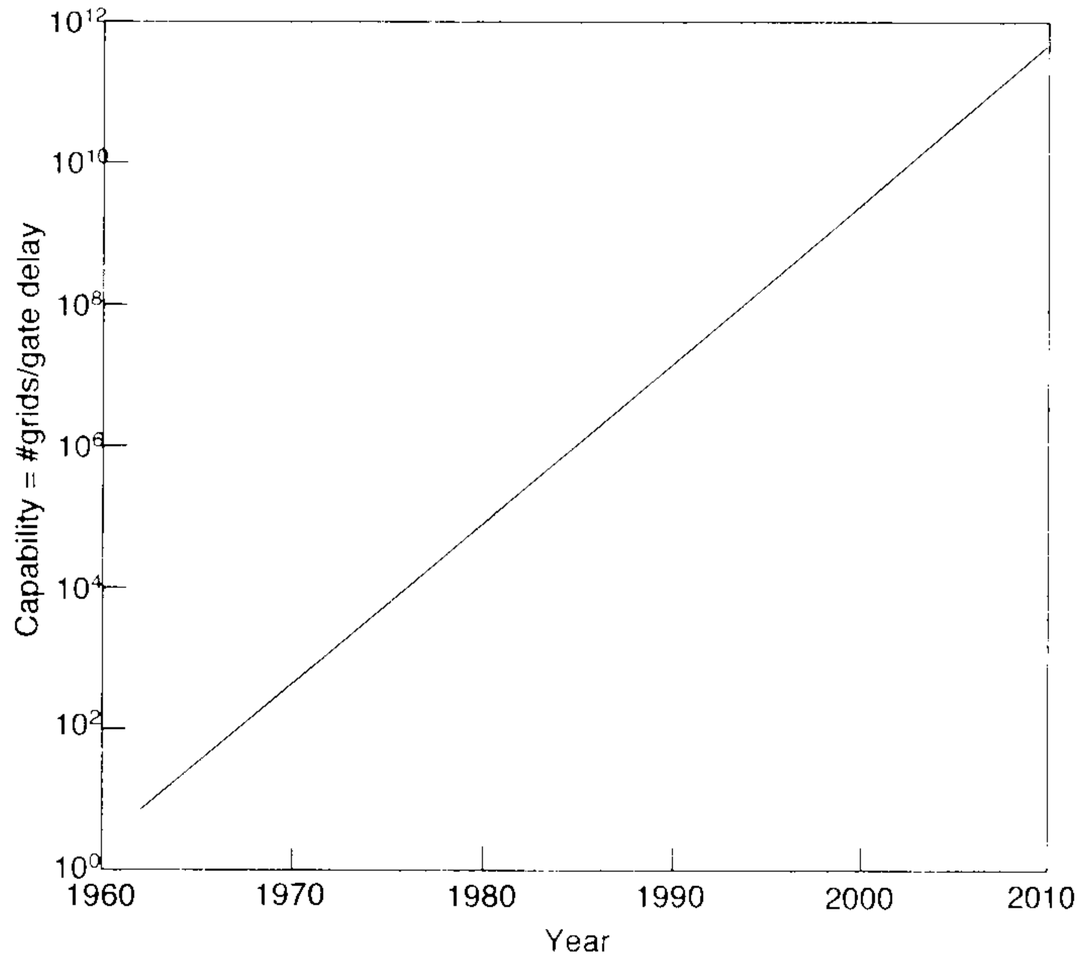


Fig. 2.46 Trend line for increasing chip capacity with time.

2.11.2 Scaling of Effects on Feature Size

The scaling down the on-chip feature size affects many electrical characteristics of transistors and other electrical components. The relationship of the change in the electrical parameters due to the scaling factor[7] K_w for the wire pitch and K_e for the chip edge dimension is given in Table 2.5.

Wire pitch and chip edge dimensions both scale linearly with K_w and K_e respectively. The chip area A scales as K_e^2 and the number of tracks and the grids on a chip scale as K_e/K_w and $(K_e/K_w)^2$, respectively. To maintain constant field strength, the voltage is also scaled by K_w.

Table 2.5
Influence of scaling factors on-chip electrical parameters.

Parameter	Symbol	Relationship	Scaling factor
Wire pitch	K_w		
Chip edge	K_e		
Area	A		K_e^2
Tracts/chip	N_T		K_e/K_w
Grids/chip	N_G	N_T^2	$(K_e/K_w)^2$
Voltage	V		K_w
Wire capacitance	C_w		K_w
Wire resistance	R_w		$1/K_w$
Switching time	t_{sw}		K_w
Switching energy	E_{sw}	C_wV^2	K_w^3
Current	I_w	C_wV/t_{sw}	K_w
	I_A	I_w/K_w^2	$1/K_w$
	I_{Chip}	AI_A	K_e^2/K_w

[7] The scale factors are approximately $K_w = 0.90$ and $K_e = 1.05$.

The area A of an on chip capacitor decreases as a function of K_w^2 but the thickness h of the dielectric decreases with K_w; hence, the capacitance ($C = kA/h$) decreases with K_w. The resistance of an on chip wire ($R = \rho L/A$) scales as $1/K_w$ because its length scales with K_w and its cross sectional area scales with K_w^2.

The switching energy $E_{sw} = CV^2$ scales with K_w^3. The current flowing along a wire I_w scales as K_w, because the frequency of switching has increased while the voltage has decreased. Note also that the current flow per unit area on the chip I_A scales as $1/K_w$ and the total current flowing into a chip I_c scales as K_e^2/K_w.

To illustrate the use of these scaling factors in predicting future requirements consider the following example.

EXAMPLE 2.1

Determine the current I_{chip} required to drive a chip that is to be developed four years from today. The average today for the current supplied to this type of chip is 10 A. The scale factors for the wire pitch $K_w = 0.9$ and for the chip edge dimension $K_e = 1.05$ represent annual changes anticipated for the feature size.

$$\left[I_{chip}\right]_4 = \left[I_{chip}\right]_0 \left(\frac{K_e^2}{K_w}\right)^4 = 10\left(\frac{(1.05)^2}{0.9}\right)^4 = 22.52 \text{ A}$$

2.11.3 Effects of Scaling the Wires

As the feature sizes continue to decrease, the scaling of the wires becomes more important in terms of power, delay and density. To show the effect of the wire size on electrical parameters including power distribution and on-chip communications, consider the data presented in Table 2.6. This table shows the scaling relationship for a transistor length wire, a fixed length wire (1 μm) and a length of wire across the entire chip.

Table 2.6
Influence of scaling wire size on electrical parameters, power distribution and on-chip communication.

Parameter	Transistor length	Fixed 1 μm	Chip length
C	K_w	1	K_e
R	$1/K_w$	$1/K_w^2$	K_e/K_w^2
I	K_w	1	K_e
IR	1	$1/K_w^2$	$(K_w/K_w)^2$
IR/V	$1/K_w$	$1/K_w^3$	K_e^2/K_w^3
RC	1	$1/K_w^2$	$(K_w/K_w)^2$
RC/t_{sw}	$1/K_w$	$1/K_w^3$	K_e^2/K_w^3

The first two data rows show how the electrical parameters capacitance C and resistance R change with K_w and K_e. Note that the results differ depending on the purpose of the wire—whether it is a fixed length, device length, or if it extends across the chip.

The future difficulties in current and power distribution are characterized by the rows identified as I, IR and IR/V. The data in these three rows represent the current flowing through a single supply wire. The current I decreases with K_w if the wire is local (device length), but increases with K_e if the wire crosses the chip. The voltage drop over the length of a wire IR does not change for device length wires, but the IR loss increases with $1/K_w^2$ for fixed length wires and as $(K_e/K_w)^2$ for wires that cross the chip. The more important parameter is IR/V, which is the voltage drop relative to the supply voltage. For a chip length wire, this parameter is increasing as K_e^2/K_w^3 or 51% per year for $K_w = 0.9$ and $K_e = 1.05$. The scaling factor K_e^2/K_w^3 represents a serious problem in future developments because the voltage drop relative to the supply voltage for long on-chip wiring is increasing at an alarming rate.

The problem associated with on-chip communication is represented by the data presented in the RC and RC/t_{sw} of Table 2.6. The time delay of signal propagation is related to the RC constant which increases as $1/K_w^2$ for a fixed length wire and as $(K_e/K_w)^2$ for a cross-chip length wire. The parameter RC/t_{sw} represents the wire delay time relative to the gate delay time (switching time). Note that RC/t_{sw} is increasing as $1/K_w^3$ for fixed length wires and as K_e^2/K_w^3 for cross-chip length wires. For $K_w = 0.90$ and $K_e = 1.05$, $K_e^2/K_w^3 = 1.51$, which represents an annual increase of 51% in wire delay time relative to the gate delay. It is apparent that wiring delays will become more important than gate delay unless the chip architecture minimizes wire lengths.

2.12 I/O COUNT AND RENT'S RULE

Bonding pads are provided around the perimeter of the chip to permit input and output signals to enter or leave the chip. Bonding pads are also provided for the supply voltages and the ground lines which provide the power to the chip. These bonding pads, shown in Fig. 2.47, are terminals for the interconnections required to wire a group of chips together to provide digital logic functions or memory required in a complete electronic system. In more modern high density chips the bonding pads are distributed of the entire surface of the chip. In these cases, the chip is connected to the chip carrier by using controlled collapsible chip connections (C4), which will be described later in Chapter 4.

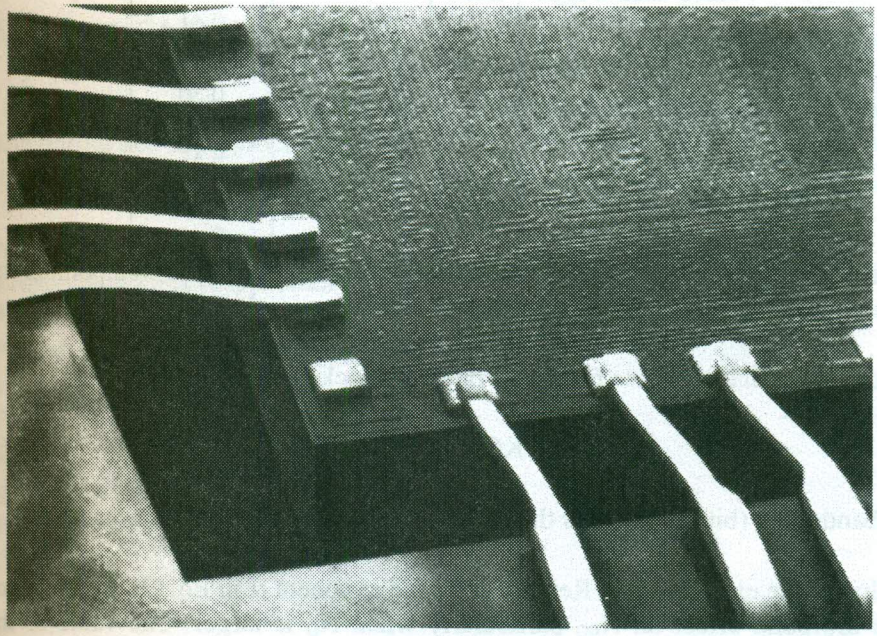


Fig. 2.47 Top surface of a chip showing perimeter bonding pads and leads for the first level package.

The number of I/O increases with the number of gates and it is important to estimate the numb required on the chip, on the circuit board or the sub assembly which consists of several circuit boar and a back panel. A common method for estimating the I/O count is to utilize Rent's rule [1, 7], whic is an empirical equation developed in the early 1960's[8]. Rent's rule relates gate count and I/O count fo random logic connections as:

$$N_{I/O} = kN_G^{p} \tag{2.6}$$

where $N_{I/O}$ is the number of I/O, N_G is the number of gates involved, k is a constant depending on th function of the card and p the exponent, which also depends on the function of the card, varies betwee 0.1 and 0.80. Some values of Rent's constant k and Rent's exponent p are given in Table 2.7 fo modern card designs with different functions.

Table 2.7
Values of the Rent constant p and the Rent exponent k for select functions

Design Type	Date	Rent's coefficient	Rent's exponent
SRAM	1990	6	0.12
Gate arrays	1990	1.9	0.50
Chip and Module	1990	1.4	0.63
Microprocessors	1990	0.82	0.45
Microprocessors	1995	2.09	0.36
ASIC control logic IFU*	2004	0.8	0.69
ASIC control logic FPU*	2004	2.2	0.66
ASIC control logic FXU*	2004	4.4	0.61
ASIC control logic IDU*	2004	20.5	0.30
ASIC control logic ISU*	2004	23.3	0.31
ASIC control logic LSU*	2004	7.3	0.46

*IFU is an instruction-fetch unit; FPU is a floating-point unit; FXU is a fixed point unit; IDU is an instruction-decode unit; ISU is an instruction-sequence unit; LSU is a load store unit.

Another similar relation that accounts for signal rate and clock cycle is given by:

$$B = KC^{\alpha} \tag{2.7}$$

where B is the bandwidth (bits/s) and C is the capability (number of gates × frequency).

The value of the exponent p in Rent's rule or the exponent α in Eq. (2.7) is very important as the exponent has a profound effect on $N_{I/O}$ particularly when N_G is large. The value of the exponent k

[8] The original memorandum written by E. F. Rent is shown in a recent publication by M. Y. Lanzerotti et al in reference [7].

required for a given system depends strongly on performance and the digital function. It was possible to build systems with Rent's exponent of about k = 0.7 in the 1960s soon after this relationship was developed; however, reduced feature size of the on-chip components has made it impossible to maintain this exponent. Today lower exponents are often achieved as indicated in Table 2.7.

The impressive scaling down of feature size in modern ULSI designs has acerbated the I/O problem. The number of bonding pads that can be placed on a chip increases as the perimeter of the chip multiplied by the linear density of the pads. This statement is valid even for area bonding (flip-chip) because the pads must be routed out from under the chip to its perimeter at the next level of packaging. As indicated in Fig. 2.48, the pin count on a chip is increasing at about 12% per year. This increase in pin count is due to increased perimeter dimensions (6%/year) as well as improved bonding pad densities (6% per year).

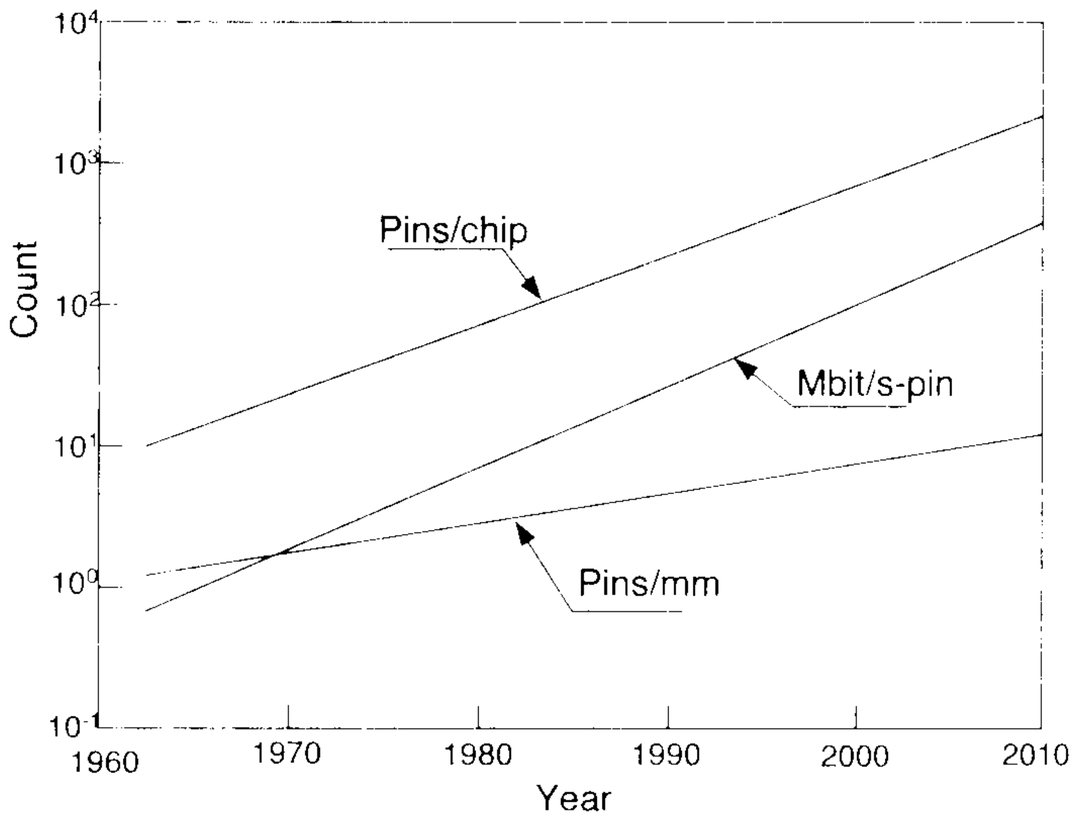


Fig. 2.48 Trend lines for increasing pin count, pin density and signal rate time.

Today large high-performance chips[9] contain more than a thousand bonding pads. It is clear that the new chip designs will have a much higher I/O count than the logic chips of the past. Design-for-performance chips will require chip carriers and circuit boards to accommodate pin counts of 1,000 to 2,000 pins at signal rates of 200 to 300 Mb/(s-pin).

2.13 MEMORY DEVICES

Digital logic circuits process extremely large amounts of information and it is necessary to store large amounts of data during and after processing to fully utilize the capabilities of a digital system. A large number of different storage devices are employed in this process including:

Memory chips
- RAM—random access memory
- ROM—read only memory
- PROM—programmable read only memory

Magnetic disks
- Floppy
- Hard

Compact disks
Digital video disks
Magnetic tapes

[9] The IBM Z990 processor employs a 16.8 × 16.8 mm signal control chip with 1,666 signal connections. The 16 chips that constitute the processors are mounted on a multi-chip-module (MCM) that contains a total of 5,184 pins of which 2930 are for signal connections.

Storage of data in memory requires the capability to write or to transfer information to the memory devices and to read or retrieve this information from memory at some later time. All of the memory devices listed above have the capability of reading and writing. However, the ROM and PROM are basically used only for reading as the information stored in these devices is essentially permanent and is written to memory in initializing the digital system.

The important characteristics of memory devices include storage capacity, access time, portability, reliability and cost. Access time and cost of the various devices trade off. The lower cost storage methods include the magnetic tape and the CD-ROM. The access time for the tape is very long ranging from about 10 to 100 s, because of the time required to position the tape relative to the read or write heads. The CD-ROM provides much better access time at a lower cost than the magnetic tape. Most magnetic tapes are used for back-up systems, where the entire contents of a computer are periodically stored in the event of a failure of the memory in the computer.

Hard disk drives provide significantly better access time about 9 to 10 ms while retaining very high capacity and very low cost. The improved performance is achieved by using flying read-write heads which are positioned rapidly over the entire surface of the rotating disks. Disk memories use a hard disk for very high capacities 500 GB or more with short access times.

Zip disks, memory sticks, compact flash cards and floppy disks are used for portable storage where files are to be transfer from one system to another. Zip disks are superior to floppy disks because the capacity of the zip disk (up to 35 GB) is much larger than the capacity of the floppy disk (1.44 MB). The access time of the zip disk is also less than that for the floppy disk. Memory sticks and compact flash cards have the advantage of small size and speed of recording data.

For higher speed applications, it is not possible to use electro-mechanical means (moving magnetic surfaces with read-write heads) to store information, because access time is so long that the computer's performance is degraded. For these applications, the information is stored electronically with access times which range from 10 to 100 ns. The dynamic random access memory (DRAM) is the most common way to read and write data with relatively low cost. DRAM consists of an array of storage cells as illustrated in Fig. 2.49. Each cell provides a memory location where either 0 or 1 may be stored. A particular cell is located by a row and a column address, and is activated only when both the row and column lines are high (1). The switching of a unit cell is illustrated in Fig. 2.50, where a MOS transistor is shown connected to row and column selection lines.

The data is stored in the unit cell by charging a small capacitor (10-15 fF). If the capacitor is charged the unit cell is high and storing 1, but if there is no charge the unit cell is storing 0. Because the unit cell loses charge, due either to leakage or a voltage loss which occurs during reading, it must be recharged periodically (every 1 to 2 ms the entire array of cell is replenished). The memory array is controlled by a regeneration circuit, which periodically samples each cell and replenishes the capacitors as required. After replenishment, the regeneration circuit permits reading and writing to the cell as identified by the row and column selection lines.

The capacity of dynamic RAM's has increased rapidly while its price/bit has dropped dramatically over the past years. In 2006 the price of 1 GB DDR-2-667 MHz module with a single 1 GB DRAM chip cost from about \$100 to \$110, depending on the supplier. The price of memory[10] for this popular module was slightly more than 1×10^{-6} cents per bit. The chip carrier for the DRAM is almost always a J leaded plastic chip carrier, because the number of I/O required for even large capacity DRAM is relatively small. Heat generation is also minimal as only one cell is active (replenish, read or write) at any instant. The primary requirement in packaging memory chips is to achieve high density in a location in close proximity to the logic circuits which address the DRAM chips.

---

[10] The price of DRAM chips dropped about 40% in 2005 due to over capacity of the semiconductor manufactures and relatively weak demand for this type of chip..

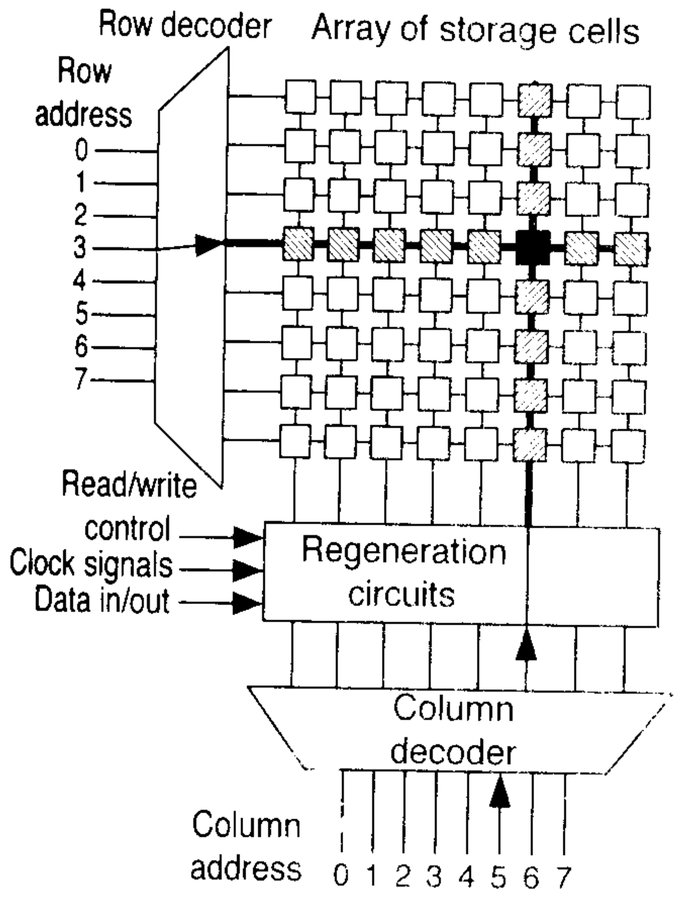


Fig. 2.49 Array of storage cells for DRAM.

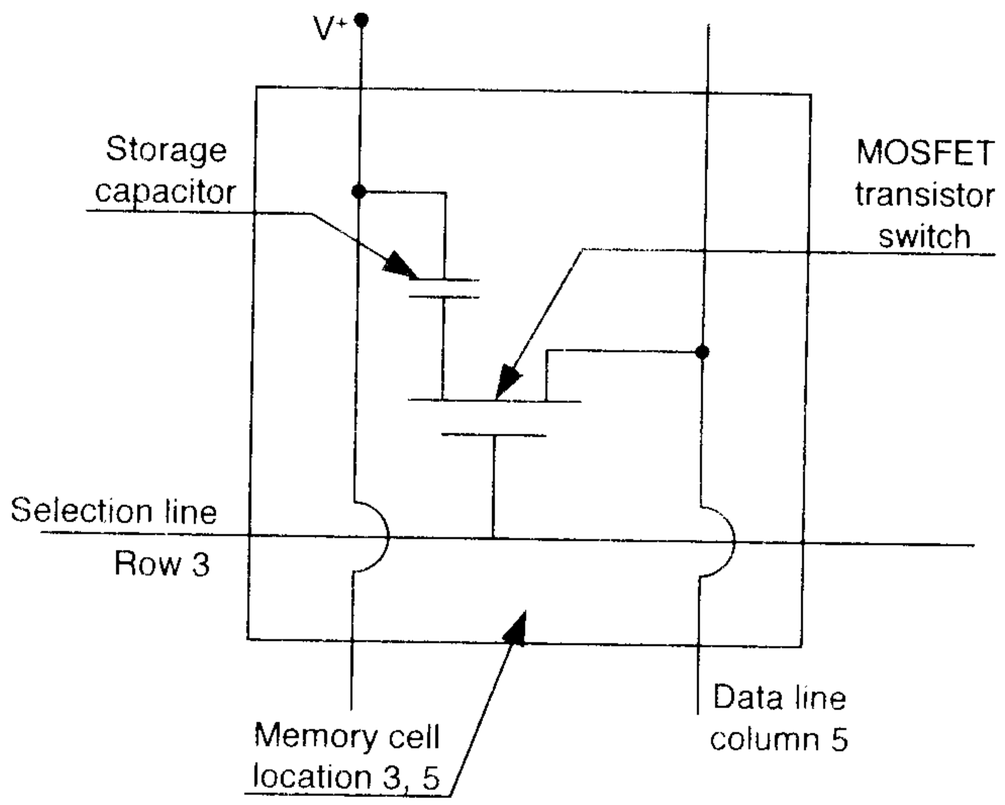


Fig. 2.50 A unit storage cell for DRAM.

REFERENCES

1. Landman, B., and R. L. Russo, "On a Pin vs. Block relationship for Partitioning of Logic graphs," IEEE Transactions on Computers, Vol. C-20, No. 12, pp. 1469-1479, 1971.
2. Winkle, T.-M., et al., "First- and Second-Level Packaging of the z990 Processor Cage, IBM Journal of Research and Development, Vol. 48 No. 3/4, 2004.
3. Select chapters in Microelectronics, Scientific America, 1977 which include: R. N. Noyce, "Microelectronics" p. 2-11; J. D. Meindl, "Microelectronic Circuit Elements" p. 12-25; W. C. Holton, "The Large Scale Integration of Microelectronic Circuits" p. 26-39; D. A. Hodges, "Microelectronic Memories" p. 54-65.
4. Taylor, J., Zafiratos, C. and M. A. Dubson, Modern Physics for Scientists and Engineers, 2nd Edition, Prentice Hall, New York, NY, 2005.
5. Shackelford, J. F., Introduction to Material Science for Engineers, 6th Edition, Prentice Hall, New York, NY, 2005.
6. Hill, J. W., Petrucci, R. H., McCreary, T. W. and S. S. Perry, General Chemistry, 4th Edition, Prentice Hall, New York, NY, 2005.
7. Lanzerotti, M. Y., Fiorenza, G. and R. A. Rand, "Microminature Packaging and Integrated Circuitry: The work of E. F. Rent, with an Application to on-chip Interconnection Requirements", IBM Journal for Research and Development, Vol. 49, No. 4/5, 2005.

EXERCISES

2.1 Determine the resistance of a No. 22 gage copper wire of length (a) 3m, (b) 6 in. (c) 10 ft. (d) 10 meters.

2.2 A conductor 6 mil wide and 2.5 in. long is produced by photo-etching a laminated circuit board with a copper cladding having a thickness of 1.0 ounces/ft^2. Determine the resistance of the line.

2.3 Prepare a drawing of the atomic structure of insulating materials such as: (a) Al_2O_3, (b) MgO and (c) Si_3N_4. Draw all of the shells in each atom and identify the shared electrons involved in the covalent bonding.

2.4 Prepare a drawing of the atomic structure of germanium showing the shells in adjacent atoms and identifying the electrons involved in the covalent bonding evident in this structure.

2.5 Prepare a drawing of the atomic structure of gallium arsenide GaAs, which is also used as a semiconductor. Identify the shells and describe the electrons which serve to covalently bond the two atoms.

2.6 Describe the difference between intrinsic and extrinsic conduction.

2.7 Reference a periodic table and list the elements which are considered as: (a) Type III elements and (b) Type V elements.

2.8 Prepare a drawing of the lattice structure of silicon with a dopant atom boron B included in the lattice. Describe the conducting characteristics of this new material. Will the resistivity of the new material depend on the number of boron atoms included in the silicon lattice?

2.9 Repeat Exercise 2.8 but use arsenic as the dopant.

2.10 A batch of P type silicon is to be formulated with a resistivity of 10 Ω-cm. If the total batch has a volume 30 liters, determine the volume of silicon and boron which are mixed together prior to melting.

2.11 Repeat Exercise 2.10 but change the resistivity to 5 ohm-cm.

2.12 For a P-N junction describe the depletion layer. Indicate how the thickness of the depletion layer is adjusted in a semiconductor device.

2.13 What is the mechanical equivalent to a diode?

2.14 Reference Fig. 2.8 and note the bias voltages applied to the electrodes. Explain the direction of hole, electron and current flow shown in this illustration.

2.15 For a P-N junction forming a diode, define forward bias and reverse bias. Leakage current is observed in a diode subjected to reverse bias. Why? How large are the leakage currents in comparison to the forward currents?

2.16 Consider the transistor switch shown in Fig. 2.11 and use Ohms law to explain why the output voltage is either high or low when the switch is off or on.

2.17 A wafer 300 mm in diameter is used to fabricate chips with a die size of 10 by 12 mm. Determine the number of dice which can be processed on each wafer. If the yield from the process is 62%, estimate the number of good dies which can be obtained from each wafer. If 20 wafers are processed in a boat of wafers, find the number of good dies expected. Comment on your results.

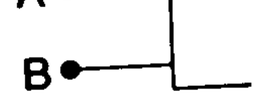
2.18 Prepare a table showing the voltage (high or low) on the source, gate and drain and the resulting switch state for a NMOS transistor of: (a) enhancement mode and (b) depletion mode.

2.19 Repeat Exercise 2.18 for a PMOS transistor.

2.20 In Fig. 2.17 a section view of a CMOS transistor which utilizes series connected NMOS and PMOS is shown. Both of these transistors operate in the enhancement mode. Prepare an equivalent drawing for a CMOS transistor fabricated from NMOS and PMOS transistors operating in the depletion mode.

2.21 Perform a trade-off analysis citing the relative advantages and disadvantages of bipolar, NMOS and CMOS transistors.

2.22 The EX



F

2.23 A logic
determi
2.24 A binary
100, 10
2.25 Describe
2.26 Describe
2.27 Describe
2.28 List the
2.29 Describe
bondin
2.30 Describe
high-er
2.31 Explain
2.32 What is
2.33 Estimat
you ma
2.34 Determi
to be r
1.05 re
2.35 Explain
2.36 Followi
Take k
2.37 For the
to 0.7.

2.22 The EXCLUSIVE-OR gate is shown in Fig. E2.22. Construct a truth table for this logic element.

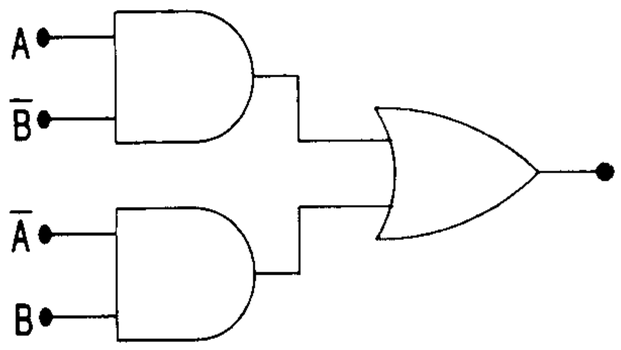


Fig. E2.22

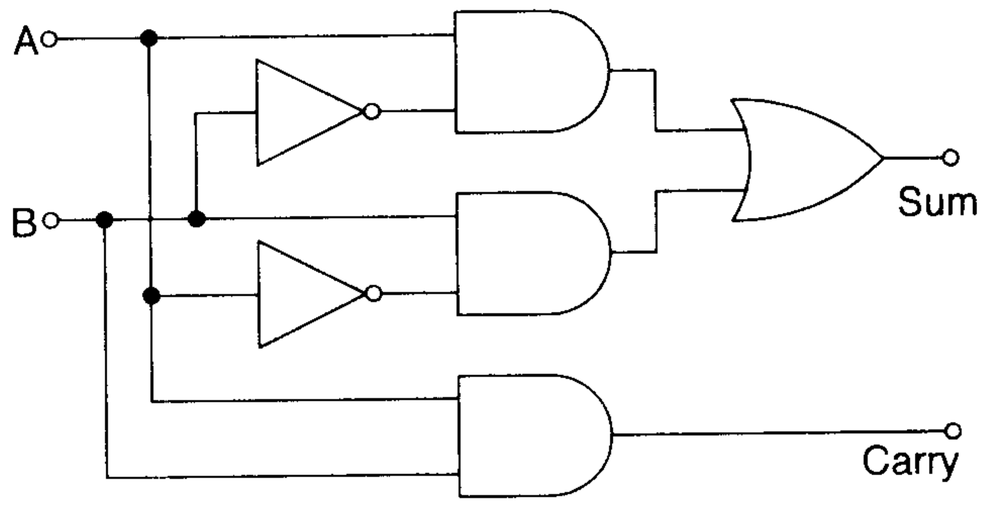


Fig. E2.23

2.23 A logic gate half-adder is shown in Fig. E2.23. Trace inputs AB of 00, 01, 10, and 11 and determine the corresponding outputs on the sum and carry terminals.

2.24 A binary adder shown in Fig. 2.22 has three inputs A, B, and C. For inputs of 000, 001, 010, 011, 100, 101, 110 and 111 trace through the circuit and indicate the output on the sum and carry bits.

2.25 Describe the advantages and disadvantages of CMOS used in the construction of logic circuits.

2.26 Describe the process for producing silicon ingots used in semiconductor manufacturing.

2.27 Describe the process for removing wafers from the single crystal silicon ingots.

2.28 List the various steps in the lithographic process used to implant the P and N features on a chip.

2.29 Describe the preparation of the pads for wire bonding; tape automated bonding; and flip-chip bonding.

2.30 Describe the damascene plating process and indicate why it is so important in the production of high-end chips.

2.31 Explain what is meant by the term "Back End of Line" (BEOL).

2.32 What is the definition of MEMS?

2.33 Estimate when the components placed on a single chip will reach 10^{12}. What assumptions have you made in your prediction?

2.34 Determine the current to drive a chip that is to be produced four years from today. The chip that is to be replaced requires a current of 12A. The scale factors for wire pitch and edge size are 0.9 and 1.05 respectively.

2.35 Explain the problem of on-chip communication as the feature size on chips continues to decrease.

2.36 Following Rent's rule estimate the number of I/O required on a circuit card with 10,000 logic gates. Take k = 2 and p = 0.45.

2.37 For the same circuit card as described in Exercise 2.36, plot the I/O required if p is varied from 0.3 to 0.7. Retain k = 2 in this exercise.