CHAPTER - 2

Silicon Based Computing Devices
A computer is generally designed from various electronic devices made from semiconductor materials like Silicon and Germanium. The electronic devices that used mostly based on silicon material, so called Silicon Based computer.

2.1 Designing Aspects of Silicon Based Computer Devices

A computer can be defined as electronic machine which receives input data, stores data and extract/manipulates data or information, and gives output in a useful format [21]. The initial stage electronic computers were quite big in the size of a large room, consuming power as compared hundreds modern personal computers (PCs). Modern computers are based on integrated circuits and are millions to billions of times faster and occupy a very less space.

![Fig.2.1 General Architecture of a Computer]

Any general purpose computer has four main components: the arithmetic logic unit (ALU), the control unit, the memory, and the input and output devices (I/O). These components are generally interconnected by busses made of groups of wires. Each component consists of thousands to trillions of small electrical circuits which can be turned off or on by means of an electronic switch. Each circuit represented by a bit (binary digit) of information, it is “1” when the circuit is turned on and “0” when it is turned off. The circuits are arranged by logic gates so that state of one or more circuits can be controlled by one or more of the other circuits. The Central Processing Unit (CPU) consists of the Control Unit (CU), ALU (Arithmetic Logical Unit), Registers, and basic I/O. Early CPUs were composed of many separate components but since the mid-1970s
CPUs have typically been constructed on a single integrated circuit called a microprocessor [21].

Computer architecture consists of three main subcategories [22]: (A) Instruction set architecture (ISA): is the overall computing system includes the instruction sets, word size, memory address modes, processor registers, and address and data formats. (B) Micro architecture: also known as Computer organization gives more detailed description of the system which involves the study of interconnected and interoperates of the constituent parts of the system to implement in ISA. (C) System Design which includes hardware components of the computing system such as: buses and switches, memory controllers and hierarchies, CPU off-load mechanisms such as direct memory access (DMA) and issues like multiprocessing. The actual device is converted to designed hardware once both ISA and micro architecture get specified. This design process is called implementation. Implementation is usually not considered architectural definition, but rather hardware design engineering.

Implementation can be studied as

• Logic Implementation — designing of blocks as per micro architecture and gate levels.

• Circuit Implementation — transistor-level design of gate and ALUs.

• Physical Implementation—physical circuits are designed by placing the circuit in a chip floor plan.

A logical operation performed by a logic gate on one or more logic inputs and gives a single logic output. The logic performed is called as Boolean logic and commonly used in digital circuits. Logic gates are generally implemented electronically using diodes or transistors, but can also be constructed by using electromagnetic relay logic, fluidic logic, pneumatic logic, optics, molecules, or even mechanical elements.

In electronic logic, a logic level is represented by a voltage or current. Each logic gate requires power so that it can source and sink currents to achieve the correct output voltage.
2.2 The Basic Semiconductor Crystal Structure: N Type and P-type

To understand how diodes, transistors, and other semiconductor devices behave, it is necessary to understand the basic structure of all semiconductor devices. It can be designed from Germanium but Silicon is preferred one for day to day applications.

![Fig.2.2 (a) Crystal Structure of Silicon atoms](image)

The Fig.2.2 (a) is often used to represent the crystal structure of silicon. Silicon (and germanium) falls in column IVa of the Periodic Table. This is the carbon family of elements. The essential characteristic of these elements is that each atom has four electrons to share with adjacent atoms in forming bonds. The nature of a bond between two silicon atoms is such that each atom provides one electron to share with the other and is known as a covalent bond. This bond is very stable and requires a lot of energy to break this bond. There are no electrons available to move from place to place as an electrical current and we can say a pure silicon crystal is a good insulator [23].

Silicon crystal conducts electricity, when we allow some electrons to move from place to place within the crystal even though it has covalent bonds between atoms. The electron flow in the crystal is happened by adding an impurity such as Arsenic or Phosphorus. These elements are from column Va of the Periodic Table, and have five outer electrons to share with other atoms. In this application, four of these five electrons bond with adjacent silicon atoms as before, but the fifth electron is free. This electron can move with a small applied electrical voltage. It shows the crystal has an excess of current-carrying electrons, with a negative charge and known as "N-type" silicon. It exhibits some resistance in the flow of electricity and cannot be either called as conductor or an insulator, but as a semiconductor. Here one can not get actual control an electrical current.

[27]
If Aluminum (from column IIIa in the Periodic Table) introduced into the crystal, as shown in fig.2.2(c), the aluminum only has three electrons available to share with other atoms. These three electrons make three covalent bonds with adjacent silicon atoms, but the fourth bond can not formed and leaves a "hole" in the structure of the crystal. Experimentation shows that there should an empty place where an electron should logically go, and an electron should try to come to that space to fill it. However, the electron which is filling the hole leaves a covalent bond behind to fill the empty space, and therefore leaves another hole behind as it moves. Similarly another electron may move into that hole, leaving another hole behind, and so forth. The holes appear to move as positive charges in the crystal. So this type of semiconductor material is designated "P-type" silicon [23].

The effects of n-type and p-type can be studied when the two are combined in various ways, in a single crystal of silicon. The basic combination is a single crystal with an N-type region at one end and a P-type region at the other. This type of crystal with two regions is known as a semiconductor diode.

2.2.1 The PN Junction

A crystal with pure silicon can be a good electrical conductor by adding an impurity such as arsenic or phosphorus (for an N-type semiconductor) or aluminum or gallium (for a P-type semiconductor). An useful applications can happen only when a single semiconductor crystal contains both P-type and N-type regions. Here we will examine the properties of a single silicon crystal which is half N-type and half P-type.
Consider the silicon crystal shown in the below figure where Half is N-type and other half is P-type. The free electrons are in the N-type crystal and holes in the P-type crystal.

![N-Type and P-Type](image)

**Fig.2.2.1.(a) N-Type and P-Type half of a Single Semiconductor Crystal**

In real scenario, it is quite impossible to join two such crystals for necessary purpose, thus a practical PN junction can only be created by inserting impurities at different parts the crystal. The excess electrons in the N region strive to lose energy by filling the holes in the P region. This creates an empty zone, or depletion region, around the junction as shown **Fig.2.2.1.(b)**. This creates a small electrical imbalance in the crystal. The N region is losing some electrons so is positively charged. These electrons have migrated to fill holes in the P region, so have a negative charge. Therefore, at the junction, it may happen that the free electrons has to diffuse to the P-side and holes to the N-side and called as diffusion process.

As the electrons move from n-type to p-type, positive donor ions are uncovered; hence, a positive charge is built on the n-side of the junction. At the same time, the free holes cross the junction and uncover the negative acceptor ions by filling the holes. Therefore a net negative charge has to establish on p-side of the junction. Further diffusion is prevented if sufficient number of donor and acceptor ions is found. The positive charge on n-side repels the holes to cross from p-type to n-type and negative charge on p-side repels free electrons to enter from n-type to p-trpe. Thus it sets a barrier against further movement of charge carriers i.e. holes and electrons. This is called potential barrier or junction barrier $V_0$. The potential barrier is order of 0.1 to 0.3 volt. The potential distribution diagram is shown **Fig.2.2.1 (b)** below.
It is clear from the diagram that a potential barrier $V_0$ is set up which gives rise to electric field. The field prevents majority carriers to cross the barrier region. Outside this barrier, the material is still neutral. Inside the barrier, positive charge is on N-side and negative charge on P-side called as depletion layer. In this layer the mobile charge carriers (i.e. free electrons and holes) have depleted (i.e. emptied) [23].

2.2.1.1 Applying Voltage across PN Junction

In a germanium crystal, the electrical imbalance is about 0.3 volt, and about 0.65 to 0.7 volt in silicon crystal. This will vary somewhat depending on the concentration of the impurities on either side of the junction. It is not possible to explore the electrical imbalance as a power source, but we can apply an external voltage to the crystal and see the responses.

The potential difference across pn junction can be applied in two ways, namely; forward biasing and reverse biasing.
When external voltage applied to the junction in a direction which cancels the potential barrier and permits the current flow, it is called forward biasing. The forward bias is applied when positive terminal connect to P-type and negative terminal to N-type. With forward bias to pn junction, we can note that i) The potential barrier is reduced and at some forward voltage (0.1 to 0.3V), it is eliminated altogether. ii) The junction offers low resistance, called forward resistance to current flow and iii) Current flows in the circuit due to the establishment of low resistance path. The magnitude of current depends upon the applied forward voltage [23].

When the external voltage applied to the junction is in such a direction that potential barrier is increased, it is called reverse biasing. The reverse bias is applied when negative terminal is connected to P-type and positive terminal to N-type. With reverse bias to pn junction, we can note that i) The potential barrier is increased ii) The junction offers very high resistance, called reverse resistance to current flow and iii) No current flows in the circuit due to the establishment of high resistance path [23].

We can overview from the above that, with reverse bias to the junction, a high resistance path is established and hence no current flows occurs. On the other hand, with forward bias to the junction, a low resistance path is set up and hence current flows in the circuit. Because of this behavior, an electrical current can flow through the junction in the forward direction, but not in the reverse direction. This is the basic nature of an ordinary semiconductor diode.
2.2.1.2 Current Flow in a Forward Biased PN Junction

Under the influence of forward voltage, the free electrons in n-type move towards the junction (negative terminal of battery connects to n-type), leaving behind positively charged atoms. However, more electrons arrive from the negative battery terminal and enter the n-region to take up their places. As the free electrons reach the junction, they become valence electrons. A hole is in the co-valent bond. When free electron combines with a hole, it becomes a valence electron. As valence electrons, they move through the holes in the p-region. The valence electrons move forwards left in the p-region which is equivalent to the holes moving right. When the valence electrons reach the left end of the crystal, they flow into the positive terminal of the battery [23].

The mechanism of current flow in a forward biased pn junction can be summed up as [23]:

(i) The free electrons from the negative terminal continue to pour into the n-region while the free electrons in the n-region move towards the junction. (ii) The electrons travel through the n-region as free–electrons i.e. current in n-region is by free electrons. (iii) When these electrons reach the junction, they combine with holes and become valence electrons. (iv) The electrons travel through p-region as valence electrons i.e. current in the p-region is by holes and (v) When these valence electrons reach the left end of the crystal, they flow into the positive terminal of the battery.
One should observe that with reverse voltage increases, the depletion region continues to expand. If either end of the depletion region approaches its electrical contact too closely, the applied voltage has become high enough to generate an electrical arc straight through the crystal. This will destroy the diode. It is also possible to allow too much current to flow through the diode in the forward direction. The crystal is not a perfect conductor and exhibits some resistance. Heavy current flow will generate some heat within that resistance. If the resulting temperature gets too high, the semiconductor crystal will actually melt will not useful. That’s why manufacturers always specified the maximum specifications of a diode so that it can operate within the indicated limits.

2.2.1.3 Volt-Ampere (V-I) Characteristics of PN Junction

Volt-Ampere or V-I characteristics of a PN junction (also called crystal or semiconductor diode) is the curve between voltage across the junction and the circuit current. Usually, voltage is taken along x-axis and current along y-axis. The Figure.2.2.1.3 (a) below shows the circuit arrangement for determining the V-I characteristics of a pn junction. The characteristics get studied by zero external voltage, forward bias and reverse bias.

![Circuit arrangement of a PN junction](image)

Fig. 2.2.1.3.(a) Circuit arrangement of a PN junction
The V-I characteristics at different conditions are shown below in Fig.2.2.1.3 (b).

(i) Zero external voltage. When the external voltage is zero, i.e. circuit is open at K; the potential barrier at the junction does not permit current flow. Therefore, the circuit current is zero as indicated by point O in Fig. 2.2.1.3(b).

(ii) Forward bias: With forward bias to the pn junction i.e. p-type connected to positive terminal and n-type connected to negative terminal, the potential barrier is reduced. At some forward voltage (0.7v for Si and 0.3v for Ge), the potential barrier is altogether eliminated and current starts flowing in the circuit. After this the voltage, the current increases with the increase in forward voltage. Thus, arising curve OB is obtained with forward bias as shown in Fig. 2.2.1.3 (b). From the forward characteristic, it is seen that at first (region OA), the current increases very slowly and the curve is non-linear. It is because the external applied voltage is used up in overcoming the potential. Once the external voltage exceeds the potential barrier voltage, the pn junction behaves like an ordinary conductor. Therefore, the current rises very sharply with increase in external voltage (region AB on the curve). The curve is almost linear.
iii) Reverse bias: In a reverse bias p-n junction, the p-type connected to negative terminal and n-type connected to positive terminal results in increasing the potential barrier at the junction. This makes the junction resistance becomes very high and practically no current flows through the circuit. This is called reverse current due to the minority carriers. The undesirable free electrons in p-type and holes in n-type are called minority carriers. If reverse voltage is increased continuously, a single breakdown of the junction occurs; characterized by a sudden rise of reverse current and a sudden fall of the resistance of barrier region. This may destroy the junction permanently.

The forward current through a pn junction is due to the majority carriers produced by the impurity. However, reverse current is due to the minority carriers produced due to breaking of some covalent bonds at room temperature.

2.2.1.4 Limitations in the Operating Conditions of PN Junction

Every pn junction has limiting values of maximum forward current, peak inverse voltage and maximum power rating and perform well under these limitations [23] otherwise it may be destroyed due to excessive heat.

(i) Maximum forward current. It is the highest instantaneous forward current that a pn junction can conduct without damage to the junction. Manufacturers’ data sheet usually specifies this rating. If the forward current in a pn junction is more than this rating, the junction will be destroyed due to overheating. (ii) Peak inverse voltage (PIV). It is the maximum reverse voltage that can be applied to the pn junction without damage to the junction. If the reverse voltage across the junction exceeds its PIV, the junction may be destroyed due to excessive heat. (iii) Maximum power rating. It is the maximum power that can be dissipated at the junction without damaging it. The power dissipated at the junction is equal to the product of junction current and the voltage across the junction. This is very important consideration and is invariably specified by the manufacturer in the data sheet.

The electronic properties of atoms used in a semiconductor crystal can be analyzed by orbital structure of the atoms. The orbital structure of atoms helps in finding different energy levels of atoms which facilitates the flow of electrons in the crystal at different voltage of power supply. Below the discussion has made regarding the electronic orbital structures of the elements.
2.2.2 Electronic Orbital Structure of the Elements

The electronic properties of any element can study with help of wave function theory. In general there are four quantum numbers are required to define the wave function. The total energy, the orbital angular momentum, the component of the angular momentum along a fixed axis in space, and the electron spin are quantized. The four quantum numbers are identified as follows [28].

1. The principal quantum number \( n \) is an integer 1, 2, 3,.. and determines the total energy associated with a particular state. This number may be considered to define the size of the classical elliptical orbit, and it corresponds to the quantum number \( n \) of the Bohr atom.

2. The orbital angular momentum quantum number \( l \) have the values 0, 1, 2,.. \((n-1)\).This indicates the shape of the classical orbit. The magnitude of this angular momentum is

\[
\sqrt{(n^2 - l^2) + \frac{\hbar^2}{2m}}
\]

where \( \hbar = \text{Plank's constant} = 4.1 \times 10^{-15} \text{ eV.Sec} = 6.626 \times 10^{-34} \text{ Joule.Sec} \)

3. The orbital magnetic number \( m_l \) may have the values 0, ±1, ±2,......±l. This number gives the orientation of the classical orbit with respect to an applied magnetic field. The magnitude of the component of angular momentum along with direction of the magnetic field is \((\hbar/2\pi)m_s\).

4. Electron spin is basically the rotation of electron around its own axis. When an electron system is subjected to a magnetic field, the spin axis will orient itself either parallel or antiparallel to the direction field. The spin is quantized to one of two possible values of angular momentum is given by \( m_s (\hbar/2\pi) \), where the spin quantum number \( m_s \) maybe+1/2or-1/2.

Certain fundamental principles that atomic orbital follows in arranging electrons are:

2.2.2.1 The Pauli Exclusion Principle

This states that that no two electrons in an electronic system can have the same set of four quantum numbers \( n, l, m_l \) and \( m_s \)[24].
2.2.2.2 The Aufbau Principle

It is used to determine the electron configuration of an atom, molecule or ion. This principle postulates that an atom is "built up" by progressively adding electrons. As they are added, they assume their most stable conditions (electron orbitals) with respect to the nucleus and those electrons already there. According to the principle, electrons fill orbitals starting at the lowest available (possible) energy states before filling higher states (e.g. 1s before 2s). The number of electrons that can occupy each orbital is limited by the Pauli Exclusion Principle. If multiple orbitals of the same energy are available, Hund's rule says that unoccupied orbitals will be filled before occupied orbitals are reused (by electrons having different spins) [25].

2.2.2.3 Hund's rules-Rule of Maximum Multiplicity

It is an observational rule which states that a greater total spin state usually makes the resulting atom more stable. [26]

The stability of an atom increases with commonly manifested lower energy state as high-spin state forces the unpaired electrons to reside in different spatial orbitals. The increased stability is due to the decrease in the screening of electron-nuclear attractions. Total spin state is calculated as the total number of unpaired electrons + 1, or twice the total spin + 1 written as \(2S+1\).

As a result of Hund's rule, constraints are placed on the way atomic orbitals are filled using the Aufbau principle. It is general rule that if a group of \(n\) or fewer electrons occupy a set of \(n\) degenerate orbitals, they will spread themselves out among the orbitals and give \(n\) unpaired spins. This is Hund's Rule or The Rule of Maximum Multiplicity. The three rules are i) for a given electron configuration, the term with maximum multiplicity has the lowest energy or can say with maximum S (spin angular momentum) ii) For a given multiplicity, the term with the largest value of L has the lowest energy, where L is the orbital angular momentum and iii) for a given term, in an atom with outermost subshell half-filled or less, the level with the lowest value of lies lowest in energy. If the outermost shell is more than half-filled, the level with highest value of J is lowest in energy. J is the total angular momentum, \(J = S + L\).
These rules specify in a simple way how the usual energy interactions dictate the ground state term. The rules assume that the repulsion between the outer electrons is very much greater than the spin-orbit interaction which is in turn stronger than any other remaining interactions. This is referred to as the LS coupling regime [26].

It can be shown that the spin-orbit interaction can shift all the energy levels together. Thus the ordering of energy levels depends only on outer valence electrons.

### 2.2.2.4 Electron shells

All the electrons in an atom which have the same value of n are said to belong to the same electron shell. These shells are identified by the letters K,L,M,N,......,corresponding to n=1,2,3,4,......,respectively. A shell is subdivided to subshells corresponding to different values of l and identified as s,p,d,f,.....,corresponding to l=0,1,2,3,4,....., respectively. Taking in to account of the exclusion principle, the distribution of electrons in an atom among the shells and subshells is shown in below table [28].

- n=1, l=0, m_l=0, m_s=±1/2 - 1s states.
- n=2,l=0, m_l=0,m_s=±1/2 - 2s subshell. There in addition, six energy levels corresponding to n=2,l=1,m_l=-1,0 or +1, and m_s=±1. These designed as 2p subshell.

<table>
<thead>
<tr>
<th>Shell......</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>n...........</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>l.........</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subshell</td>
<td>s</td>
<td>s</td>
<td>p</td>
<td>s</td>
</tr>
<tr>
<td>Number of electrons</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

| 2 | 8 | 18 | 32 |

**Table 2.2.2.4 Shell and orbital**
The inner-shell electrons are very strongly bound to an atom, and cannot be easily removed. That is the electrons closest to the nucleus are the most tightly bound and lowest energy. The energy band theory of semiconductor crystal can be studied quantum mechanically for the array atoms it contains.

2.2.3 The Energy-Band Theory of Crystals

X-ray and other studies reveal that most metals and semiconductors are crystalline in structure. A crystal consists of a space array of atoms or molecules built up by regular repetition in three dimensions of some fundamental structural unit. The electronic energy levels can be studied from the time-dependent Schrödinger equation [28].

\[ \Psi^2 + \frac{\hbar^2}{2m} \left( \nabla^2 - U \right) \Psi = 0 \]  \hspace{1cm} (2.2.3)

Where \( \Psi \) is called wave function, and it must describe the behavior of particle. \( \Psi^2 \) gives the probability of finding the electron. The potential energy \( U \) characterizing the crystalline structure is a periodic function of space as a result of contributions from every atom. \( W \) is the total energy of the particle.

In a crystal, it is generally the levels of the outer-shell electrons of atom that are changed considerably as these electrons are shared by more than one atom in the crystal. The new energy levels of the outer electrons can be determined by coupling between outer-shell electrons of the atoms forming a band of closely spaced energy states. A qualitative discussion of this energy-band structure follows.

<table>
<thead>
<tr>
<th>Element</th>
<th>Atomic Number</th>
<th>Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>6</td>
<td>1s^22s^22p^6</td>
</tr>
<tr>
<td>Si</td>
<td>14</td>
<td>1s^22s^22p^63s^23p^6</td>
</tr>
<tr>
<td>Ge</td>
<td>32</td>
<td>1s^22s^22p^63s^23p^63d^{10}4s^24p^2</td>
</tr>
<tr>
<td>Sn</td>
<td>50</td>
<td>1s^22s^22p^63s^23p^63d^{10}4s^24p^64d^{10}5s^25p^2</td>
</tr>
</tbody>
</table>

Table 2.2.3 Electronic Configuration of basic atoms used in Semiconductor crystal
Consider a crystal consisting of N atoms of one element of above table. If the atoms in the crystal are far apart then the interaction between them is negligible so the energy levels will coincide with those of the isolated atom. The outer two sub-shells for each element in the table contain two s electrons and two p electrons. Hence, if we ignore the inner-shell levels, then as indicated to the extreme right in below Fig.2.2.3.(a), there are 2N electrons completely filling the 2N possible s levels, all at the same energy. Since the p atomic subshell has six possible states, our assumed crystal of widely spaced atoms has 2N electrons, which fill only one-third of the 6N possible p states, all at the same level.

![Fig.2.2.3.(a) Energy levels for a Crystal](image-url)

**Fig.2.2.3.(a) Energy levels for a Crystal**

**Fig.2.2.3. (b) Energy band for a Crystal**
On decreasing the interatomic spacing of the imaginary crystal, an atom will exert an electric force on its neighbours, so the coupling between atoms, the atomic-wave functions overlap, and the crystal becomes an electronic system which must obey the Pauli exclusion principle. Hence the $2N$ degenerate ‘s’ states must spread out in energy. The separation between levels is small, but since $N$ is very large ($\sim 10^{23}\text{cm}^{-3}$); the total spread between the minimum and maximum energy may be several electron volts if the interatomic distance is decreased sufficiently. This large number of discrete but closely spaced energy levels is called energy band, and is indicated schematically by the lower region in Fig.2.2.3.(a). The $2N$ states in this band are completely filled with $2N$ electrons. Similarly, the upper shaded region Fig.2.2.3.(a) is a band of $6N$ states which has only $2N$ of its levels occupied by electrons.

There is an energy gap (a forbidden band) between the two bands discussed above and that this gap decreases as the atomic spacing decreases. For small enough distance shown in Fig.2.2.3. (b), these bands will overlap. Under such circumstances the $6N$ upper states merge with the $2N$ lower states, giving a total $8N$ levels, half of which are occupied by the $2N+2N=4N$ available electrons. At this spacing each atom has given up four electrons to the band; these electrons can no longer be said to orbit in s or p sub shells of an isolated atom, but rather they belong to the crystal as a whole. In this sense the elements in table 2.2.3 are tetravalent, since they contribute four electrons each to the crystal. The band these electrons occupy is called the valence band.

Solutions of Schrödinger’s equation are complicated, and have been obtained approximately for only relatively few crystals. These solutions lead us to expect an energy-band diagram as in Fig.2.2.3(b). At the crystal–lattice spacing (the shaded sector), we find the valence band filled with $4N$ electrons separated by a forbidden band (no allowed energy states) of extent $E_G$ (gap energy) from an empty band consisting of $4N$ additional states. Thus upper vacant band is called conduction band.
CHAPTER- 2  
Silicon Based Computing Devices

Consider the Fig.2.2.3.(c) below. This figure shows the important range of electron energy bands. To understand this diagram, let's define a few terms.

**Fig.2.2.3.(c) Band Diagram**

**Conduction Band:** It is the range of electron energy where electrical conduction is possible. Electrons with this much energy is free of their parent atoms and can move through the medium in which they exist.

**Valence Band:** It is the range of electron energy where electrical conduction is not possible. Electrons with this much energy is bound into the atomic structure of the material, and is unavailable to conduct an electrical current.

**Forbidden Zone:** It is the energy range between the valence band and the conduction band. Electrons cannot remain within this range of energy; they must either gain or lose energy so as to attain either the conduction band or the valence band.

**Fermi Level:** The highest energy level in the crystal that can remain populated by electrons at a temperature of absolute zero. Electrons with greater energy than this may be available for conduction; electrons with less energy are bound to the crystal structure.

*Diagram A* above represents a *good conductor*, such as copper or silver. Here, at temperatures above absolute zero, electrons are always available to conduct electrical current, even with no applied energy. In metals, the valence and conduction bands actually overlap.

[42]
Diagram B shows a typical insulator, such as glass. All electrons are locked into the atomic structure, and are unavailable as current carriers. It will take a lot of energy to break any electrons loose for conduction.

Diagram C represents a crystal of N-type silicon (or germanium). The forbidden zone is still present, but much smaller than for an insulator. These are called as "semiconductor." With the crystal doped with N-type impurities, there are lots of electrons around with almost enough energy to roam freely, so the Fermi level gets pushed up close to the conduction band. If the doping level is heavy enough (large dosage of impurities), the Fermi level can actually enter the conduction band.

Diagram D represents a P-type semiconductor crystal. Here, the p-type impurities have left holes in the atomic structure, which tend to attract and hold free electrons. This pulls the Fermi level down until it gets close to the valence band. Similar to the highly-doped N-type crystal, a highly-doped P-type crystal will have its Fermi level within the valence band instead of just above it.

Diagram E shows the entire crystal as a whole has one Fermi level. The conduction and valence bands have different energy levels in the crystal making the N-type conduction band is very close in energy to the P-type valence band. The transition region corresponds to the depletion region within the crystal. This is a major factor in the operation of all semiconductor devices, and helps to get specific properties from a given device suitable for manufacturing the semiconductor crystal.
2.2.3.1 Insulators, Semiconductors and Metals

A very poor conductor of electricity is called an insulator; an excellent conductor is a metal; and a if the conductivity lies between these two is a semiconductor. A material may be placed in one of these three classes, depending upon its energy-band structure.

**Insulator:** The energy-band structure of Fig.2.2.3.1 at the normal lattice spacing is indicated schematically in Fig.2.2.3.(b) for a diamond (carbon) crystal the region containing no quantum states is several electron volts high ($E_G \sim 6\text{eV}$). This large forbidden band separates the filled valence region from the vacant conduction band. Since the electron cannot acquire externally applied energy, conduction is impossible, and hence diamond is an insulator.

**Semiconductor:** A substance for which the width of the forbidden energy region is relatively small ($\sim 1\text{ eV}$) is called a semiconductor. Graphite, a crystalline form of carbon but having a crystal symmetry which is different from diamond, has small values of $E_G$ of 0.785 and 1.21eV, respectively, at 0$^\circ$K. Energies of this magnitude normally cannot be acquired from an applied field. Hence the valence band remains full, the conduction band empty, and these materials are insulators at low temperature. However, the conductivity increases with temperature, as explained below, and for this reason these substances are known as intrinsic semiconductors.

[44]
As the temperature is increased, some of these valence electrons acquire thermal energy greater than $E_G$ and hence move into the conduction band. These are now free electrons as that they can move about under the influence of even a small applied field. These free, or conduction, electrons are indicated schematically by dots in Fig. 2.2.3.1(b). The insulator has now become slightly conducting; it is a semiconductor. The absence of an electron in the valence band is called a hole. The holes in a semiconductor refer to the empty energy levels in a filled valence band. This in fact considered as extrinsic (impurity) semiconductor.

As the band-gap energy of a crystal is a function of interatomic spacing, $E_G$ depends somehow on temperature. $E_G$ for silicon decreases with temperature at the rate of $3.60 \times 10^{-4}$ ev/°K. Hence for silicon, $E_G(T) = (1.21) - (3.60) \times 10^{-4} T$ and at room temperature ($300^°K$), $E_G = 1.1$ eV. Similarly, for germanium, $E_G(T) = 0.785 - 2.23 \times 10^{-4} T$ and at room tempertaure, $E_G = 0.72$ eV.

**Metal**: The band structure of a crystal may contain no forbidden energy region, so that the valence band merges into an empty band, as indicated in Fig 2.2.3.1(c). Under an applied electric field the electrons acquire additional energy and move into higher energy states. Since these mobile electrons constitute a current, this substance is a conductor, and the empty region is the conduction band. A metal is characterized by a band structure containing overlapping valence and conduction bands.

### 2.2.3.2 The Potential–Energy field in a metal: Multi Electron System

Considering a metal as a single crystal with a multi electron system, the resultant potential energy is the sum of potential energies produced by all the ions of the lattice. Here we do not have single nuclei where electrons are revolving around; rather it can be row of nuclei.

Consider the conditions that exist near the surface of the metal, the exact position of the surface cannot be defined. It is located at a small distance from the last nucleus in the row. As there is no nuclei exist to the right of last nucleus, there cannot be lowering and flattening of the potential-energy curve as exists between other nuclei. This leads to most important conclusion: a potential-energy hill, or barrier, exists at the surface of the metal.

[45]
The analysis is for the free electrons in the metal rather than in bound ones. A free electron is one having energy corresponding to an energy in the conduction band. Within the metal the total energy converted into potential energy. The velocity of electron is not zero inside the metal and the electron travels more or less freely throughout the body of the metal. But, when the electron reaches the surface of the metal, it collides with potential-energy barrier there.

2.2.3.3 Energy Density in a Metal

The distribution in energy of the electrons in a metal can be represented by \( \frac{d\hat{n}_E}{\rho E} \), where \( d\hat{n}_E \) represents the number of free electrons per cubic meter whose energies lie in the energy interval \( dE \),and \( \rho E \) gives the density of electrons in this interval.

The function \( \rho E \) may expressed as \( \rho E = f(E) N(E) \),where \( N(E) \) is the density of states(number of states per electron volt per cubic meter)in the conduction band, and \( f(E) \) is the probability that a quantum state with energy \( E \) is occupied by electron.

\[
N(E) = \tilde{\gamma} E^{1/2}, \quad \text{where} \quad \tilde{\gamma} = \frac{4\pi}{h^3} (2m)^{3/2} \left( 1.60 \times 10^{-19} \right)^{3/2} = 6.82 \times 10^{27} \text{ (m}^{-3} \text{ eV})^{3/2}
\]

Where \( m \) is the mass of the electron in kilograms ad \( h \) is Planck's constant in joule-seconds.

\( f(E) \) is called the Fermi-Dirac probability function, and specifies the fraction of all states at energy \( E \) (electron volt) occupied under conditions of thermal equilibrium.Fom quantum statistics it is found that \( f(E) = 1/1+e^{(E-E_F)/kT} \)

Where \( k \)=Boltzmann constant, eV/\( ^0\text{K} \)

\( T=\text{temperature}, \ 0^0\text{K} \)

\( E_F=\text{Fermi level, or characteristic energy, for the crystal, eV} \)

Deducing the condition (i) when \( E= E_F \), \( f(E) =1/2 \) for any temperature

(ii) If \( E> E_F \), \( f(E) =0 \), there is no probability of finding an occupied quantum state of energy greater than \( E_F \) at absolute zero.

(iii) If \( E< E_F \), \( f(E) =1 \), All quantum levels with energies less than \( E_F \) will be occupied at \( T=0^0\text{K} \).
\[ \rho_E = \begin{cases} \frac{1}{\sqrt{2\pi}} E^{1/2} & \text{for } E < E_F \\ 0 & \text{for } E > E_F \end{cases} \]

The relationship above is called **completely degenerate energy distribution function**. The Fermi level for \( E_F \) may obtain on the basis of the complete degenerate function.

\[ n = \int_{0}^{E_F} \frac{1}{\sqrt{2\pi}} E^{1/2} \, dE \]

\[ = \frac{2}{3} \sqrt{\frac{\pi}{2}} E_F^{3/2} \]

\[ = (3n/2\sqrt{\pi})^{2/3} = 3.64 \times 10^{-19} n^{2/3} \]

As the density 'n' varies from metal to metal, \( E_F \) will also vary among metals. If specific gravity, atomic weight and number of free electrons per atom are known, one can calculate

2.2.3.4 The Density of states: The Schrodinger Equation

A one-dimensional model of potential energy can be found out by Schrödinger Equation assuming potential-energy barrier at the surface is high so that no electron escape.

A one-dimensional model of the potential-energy diagram and two possible wave functions are shown below.

Fig. 2.2.3.4 A one dimensional problem for potential energy

The fig.2.2.3.4 shows a one dimensional problem at \( U=0 \) for a distance \( L \), but rises forward infinity at the boundaries \( x=0 \) and \( x=L \).

The Schrödinger equation is generally expressed as
Putting $U=0$, the general solution of this second-order linear differential equation has two arbitrary constants, $C_1$ and $C_2$, and in the interval $0 \leq x \leq 1$ is given by

$$\Psi = C_1 + \sin \alpha x + C_2 \cos \alpha x$$

Where $\alpha^2 = \frac{8 \pi^2 m W}{h^2}$-------------------Eq(2.2.3.4.2)

Since for $x=0$, $\Psi = 0$, then $C_2 = 0$, since for $x=L$, $\Psi = 0$, $\sin \alpha L = 0$, or $\alpha L = n_x \pi$, where $n_x$ is an integer.

Putting in above equation

$$W = \frac{n_x^2 \hbar^2}{8 m L^2}.$$ -------------------Eq(2.2.3.4.3)

The wave function is $\Psi = C_1 \sin \left(\frac{n_x \pi x}{L}\right)$. Since the probability of finding electron somewhere in the metal unity is unity, i.e $\Psi$ should be normalized.

$$\int_0^L \Psi^2 dx = \int_0^L C_1^2 \sin^2 \frac{n_x \pi x}{L} dx = 1 = C_1^2 L/2.$$ -------------------Eq(2.2.3.4.4)

Or $C_1 = (2/L)^{1/2}$, and $\Psi = (2/L)^{1/2} \sin \frac{n_x \pi x}{L}$

$n_x$ can not be zero since, if it were, $\Psi$ would vanish everywhere. For $n_x = 1$ the wave function is plotted in (b) and for $n_x = 2$, the wave function is plotted in (c).
2.2.3.5 The Uncertainty Principle

In the one-dimensional electronic problem discussed above, there is an inherent uncertainty $\Delta p_x$ in momentum because $n_x$ can have only integral values. The smallest value of $\Delta n_x=1$ and hence $\Delta p_x=h/2L$. Since the electron is somewhere between $x=0$ and $x=L$, the uncertainty in position $\Delta x=L$.

Therefore

$\Delta p_x \Delta x=h/2$  \hspace{1cm} Eq(2.2.3.5.1)

This equation is referred as the Heisenberg's uncertainty principle which postulates that, for all physical systems, there is always an uncertainty in the position and in the momentum of a particle and the product of these two uncertainties is of the order of magnitude of Planck's constant $h$[27].

2.2.3.6 Many Electron System and Density-Functional Theory

Density-functional theory (DFT) helps in calculating the electronic structures and band plots. This is a microscopic first-principles theory of condensed matter physics that explain the electron-electron many-body problem via the introduction of an exchange-correlation term in the functional of the electronic density. DFT-calculated bands are mostly found in agreement with experimentally measured bands like in determining band shape. But sometime the difference comes in systematic errors as compared to experiment results. In particular, DFT systematically underestimate about 30-40% of the band gap in insulators and semiconductors [29, 30, 31, 32, 33, 34, 35, 36].

DFT put an exact theory to reproduce and predict ground state properties (e.g., the total energy, the atomic structure, etc.). But it does not address the excited state properties of electrons injected or removed from the system.DFT appears in excitation states in terms of Kohn-Sham energies which have no physical interpretation at all .But Hartree-Fock energies, can be truly considered as an approximation for quasiparticle energies. Hence, DFT is not a band theory suitable for calculating bands and band-plots rather it is a quantum mechanical theory used in physics and chemistry to investigate the electronic structure in the ground state of many-body systems, in particular atoms, molecules, and the condensed phases. With DFT, the properties of a many-electron system can be determined by using functionals, i.e. functions of another function, which
is nothing but the spatially dependent electron density. Hence the name density functional theory comes from the use of functionals of the electron density. DFT is the most popular methods available in condensed-matter physics, computational physics, and computational chemistry. Despite recent improvements, there are still difficulties in using DFT for intermolecular interactions, van der Waals forces, charge transfer excitations, transition states, potential energy surfaces, and in calculations of the band gap in semiconductors.

2.2.3.7 Many Electron System and Green’s Function

Green’s function method is used to calculate the bands in electron-electron many-body effects. As improved to DFT, the Green’s function of a system provides both ground as well as excited state of the system. The poles of the Green’s function are the quasiparticle energies i.e the bands of a material. For real systems like solids, the self-energy is a very complex quantity and usually approximations are needed to solve the problem. One such approximation is the GW approximation, so called from the mathematical form the self-energy takes as the product \( \Sigma = GW \) of the Green’s function \( G \) and the dynamically screened interaction \( W \). This approach is more pertinent for calculation of band plots. The GW approximation gives more result in determining band gaps of insulators and semiconductors in agreement with experiment, and hence more accepted than DFT [37, 38, 39].

2.3 Conduction in Semiconductors

2.3.1 Electrons and Holes in an Intrinsic Semiconductor

The conductivity is generally expressed as
\[
\sigma = ne\mu \quad \text{Eq(2.3.1.1)}
\]
Where \( \mu \) (square meters per volt-second) is called mobility of the electrons and \( n \) is the concentration of free electrons per cubic meter and \( e \) is electron volt [28].

The conductivity is proportional to the concentration \( n \) of free electrons. For a good conductor \( n \) is very large (~\( 10^{28} \) electrons/m\(^3\)). For insulator, \( n \) is very small (~\( 10^7 \)); and for a semiconductor lies between these values. The electrons in a semiconductor is not wander around here and there, but rather trapped in a bond between two adjacent ions.
It has explained the crystal structure of two important semiconductor used in electronic devices in the introduction part of this chapter, section 2.2 in briefing the basic semiconductor crystal structure.

### 2.3.1.1 Conductivity of a Semiconductor

Two charge-carrying particles are formed by each hole-electron pair. One is negative (free electron) of mobility $\mu_n$ and other positive (holes) of mobility $\mu_p$. These particles move in opposite directions in an electric field $\epsilon$, but since they are opposite sign, the current of each is in the same direction. Hence the current density $J$ is given by

$$J = (n\mu_n + p\mu_p) e \epsilon$$

Where $n=$ magnitude of free electrons concentration

$p=$ magnitude of holes concentration

$\sigma =$ conductivity.

Hence

$$\sigma = (n\mu_n + p\mu_p)$$

--- Eq(2.3.1.1.2)

For a pure (intrinsic) semiconductor considered here, $n=p=n_i$

Where $n_i=$ intrinsic concentration

In a pure germanium at room temperature there is about one hole-electron pair for every $2 \times 10^9$ germanium atoms. With increasing temperature, the density of hole-electron pairs increases, and correspondingly, the conductivity increases.

The density is given by

$$n_i^2 = (2.33 \times 1043) (m_m m_p / m^2)^{1/2} T^3 \epsilon^E / kT$$

--- Eq(2.3.1.1.3)

The intrinsic concentration $n_i$, varies with temperature in accordance with the relationship

$$n_i^2 = A_0 T^3 \epsilon^E / kT$$

--- Eq(2.3.1.1.4)

The conductivity of germanium (silicon) is found to be increases approximately 6(8) percent per degree increase in temperature. *Such a large change in conductivity with temperature places a limitation on the use of semiconductor devices in some circuits.* On the other hand it is this property which is also considered as an advantage to make circuit.

As it is explained in the case of a metal, the number of conduction electrons per cubic meter ($d_n$) whose energy lies between $E$ and $E+dE$ is given by
\[ d_n = \rho E \cdot dE = f(E) \cdot N(E) dE \]

In a semiconductor the lowest energy in the conduction band is \( E_c \), is given by

\[ N(E) = \mathcal{Y} (E - E_c^{1/2}) \]

So the fermi equation is given by

\[ f(E) = \frac{1}{1 + \varepsilon (E - E_F)/kT} \]

--- Eq(2.3.1.5)---

**Fig. 2.3.1.1 Fermi Dirac distribution of energy at 0\(^0\)K, 300\(^0\)K and 1000\(^0\)K**

As shown in Fig.2.3.1.1, at room temperature some electrons are excited to higher energies and some states near the bottom of the conduction band \( E_c \). Similarly near the top of valence band \( E_v \), the probability of occupancy is decreased from unity since some electrons have escaped from their covalent bond and are now in the conduction band.

Since hole signifies an empty energy level, the Fermi function will be \( 1-f(E) \).

So \( 1-f(E) = 1-[1 + \varepsilon (E - E_F)/kT] \sim \varepsilon - (E_F - E)/kT \), where \( E_F - E >> kT \) for \( E <= E_v \).

The number of holes per cubic meter in the valence band is

\[ p = N_v \varepsilon (E_F - E_v)/kT \]

where \( N_v = 2(2\pi m_h kT/\hbar^2)^{3/2} \)

The Fermi-level for intrinsic can be calculated both for electron and holes as

\[ N_c \varepsilon (E_c - E_F)/kT = N_v \varepsilon (E_F - E_v)/kT \]

--- Eq(2.3.1.6)---

Taking logarithm on both side and evaluating,

\[ E_F = (E_c + E_v)/2 - (kT/2) \log \left( \frac{N_c}{N_v} \right) \]

If effective masses of a hole and a free electron are the same, \( N_c = N_v \), then
EF = (Ec + Ev)/2, and the Fermi level lies in the center of the forbidden energy band, as shown in Fig2.3.1.1.

Taking product of electron and hole

\[ np = \frac{N_c N_v \varepsilon^{-\left(E_c - E_v\right)/kT}}{N_c N_v \varepsilon^{-E_G/kT}} \]

This product is independent of the fermi level, but does depend upon the temperature and the energy gap \( E_G = E_c - E_v \).

The above equation valid for both extrinsic and intrinsic material.

So \( n = n_i \) and \( p = p_i = n_i \)

\[ np = n_i^2 = \left(2.33 \times 10^{43}\right) \left(\frac{m_e m_p}{m^2}\right)^{3/2} T^3 \varepsilon^{-E_G/kT} \]

The energy gap decreases linearly with temperature, so that

\[ E_G = E_G^0 - \beta T \]

Where \( E_G^0 \) is the magnitude of the energy gap at 0\(^0\)K. Substituting this relationship in above equation,

\[ n_i^2 = A_0 T^3 \varepsilon^{-E_G^0/kT} \]

This result also has verified experimentally [40, 41]

2.3.2 Donor and Acceptor Impurities

When donor impurities are added to a semiconductor, allowable energy levels are introduced a very small distance below conduction band. These new levels are discrete level as added impurity atoms are far apart in the crystal structure, and their interaction is small.

If the intrinsic semiconductor is doped with n-type impurities, not only does the number of electrons increases, but the number of holes decreases below that which would be available in the intrinsic semiconductor. The reason for the decrease in the number of holes is that the larger number of electrons present increases the rate of recombination of electrons with holes [28].

[53]
Fig. 2.3.2 (a) Energy Band –N-type semiconductor (b) Energy Band –P-type semiconductor

When acceptor or p-type, impurities are added to the intrinsic semiconductor, they produce an allowable discrete energy levels which is just above the valence band. Since very small amount of energy required for an electron to leave the valence band and occupy the acceptor energy level, it follows that the holes generated in the valence band by these electrons constitute the largest number of carriers in the semiconductor material [28].

The important result comes here is that the doping of an intrinsic semiconductor not only increases the conductivity, but also serves to act as a conductor in which the electric carriers are either predominantly holes or predominantly electrons. In a N-type semiconductor, the electrons are called majority carries and holes are minority. In a P-type material, the holes are majority and electrons are minority.

If we take \( N_D = \) donor ions and \( N_A = \) acceptor ions and as the semiconductor is electrically neutral,

\[
N_D + p = N_A + n
\]

Considering a n-type material \( N_A = 0 \),

\[
n = N_D \quad \text{--- Eq (2.3.2.1)}
\]

Similarly for p-type material \( N_D = 0 \),

\[
n = N_A \quad \text{--- Eq (2.3.2.2)}
\]

So the concentration of \( n \) holes in n-type semiconductor

\[
p = \frac{n^2}{N_D} \quad \text{--- Eq (2.3.2.3)}
\]

Similarly for p-type semiconductor,

\[
n_p = \frac{n_p^2}{N_A} \quad \text{--- Eq (2.3.2.4)}
\]

The positions of Fermi level with impurities in n-type material and p-type material can be deduced to:

\[
E_F = E_C - kT \log \left( \frac{N_D}{N_A} \right) \quad \text{and} \quad E_F = E_V - kT \log \left( \frac{N_A}{N_D} \right) \quad \text{--- Eq (2.3.2.5)}
\]
2.3.3 Diffusion

In addition to a conduction current, the transport of charges in a semiconductor can be studied by a mechanism called diffusion. Under non uniform concentration of particles in a semiconductor, the concentration $p$ of holes varies with distance $x$ in the semiconductor, and the concentration gradient represents as $(dp/dx)$ in the density of carriers. The existence of gradient implies that, if an imaginary surface drawn in the semiconductor, the holes are in random motion. It may happen that in a given time interval, more holes will cross the surface from one side of greater density to side of smaller density and the net transport of charges across the surface gives flow of current.

The diffusion **hole-current density** $J_p$ (amperes per square meter) is proportional to the concentration gradient and is given by

$$J_p = -e D_p dp/dx \quad \text{Eq(2.3.3.1)}$$

where $D_p$ (square meters per second) is called the diffusion constant for holes.

Similarly for diffusion **electron-current density** $J_n$ is given by

$$J_n = e D_n dn/dx \quad \text{Eq(2.3.3.2)}$$

Since the diffusion and mobility are statistical phenomenon and $\mu$ are not independent and the relationship will be given by

$$D_p/\mu_p=D_n/\mu_n = V_T, \text{ where } V_T=\text{Volt equivalent of temperature}=k^'/T/e =T(k^'/e)=T/11,600$$

Where $k^'$ is the Boltzmann constant in joules per degree Kelvin.

2.3.4 Carrier Lifetime

Due to thermal agitation, new hole-electron pairs continue to produce while other hole-electron pair disappears. On an average, a hole (an electron) will exist for $\tau_p(\tau_n)$ sec for recombination. Considering a bar of n-type silicon illuminated by light of proper frequency, the concentration of holes and electrons will increase.

If $p_{n0}$ and $n_{n0}$ are the equilibrium concentrations of holes and electrons in the n-type specimen. We have

$$p'_{n0}p_{n0} = n'_{n0}n_{n0}, \text{ where } p'_{n0} \text{ and } n'_{n0} \text{ are represent carrier concentrations during steady irradiation.}$$

If we turn off the light, the carriers will return to their equilibrium.

$$p_{n0} = (p'_{n0}p_{n0}e^{tr}$$

$$n_{n0} = (n'_{n0}n_{n0}) e^{tr}$$
The rate of concentration change will be
\[ \frac{dp_n}{dt} = -\frac{(p_n-p_{n0})}{\tau} = \frac{d}{dt} (p_n-p_{n0}) \]  \hspace{1cm} \text{Eq}(2.3.4.1)

Where the quantity \((p_n-p_{n0})\) represents the injected or excess carrier density.

The minus sign indicates that the change is a decrease in the case of recombination and an increase when the concentration is recovering from a temporary depletion. Recombination is affected not only by volume impurities, but also by surface imperfections in the crystal [42, 43]. Gold is extensively used as a recombination agent by semiconductor device manufacturers and the device engineer can obtain desired carrier lifetimes by introducing gold into silicon under controlled condition [44, 45]

2.4 Semiconductor-Diode Characteristics

As the thesis has analyzed the design of various logic gates designed from diodes we have specifies in more detail the semiconductor diode characteristics in this section. It has already discussed in the previous discussions above that pn junction conducts current easily when forward biased and practically no current flows when it is reverse biased. Semiconductor diodes are smaller in size, cheaper and robust and usually operate with greater efficiency. So in this section the focus is more on applications of semiconductors diodes and the quantitative and qualitative theory lies in designing the semiconductor diodes.

A pn junction is known as a semiconductor or crystal diode. It is generally represented by the schematic symbol as below.

Fig. 2.4 A schematic Symbol of a Crystal diode

The outstanding property of a crystal diode to conduct current in one direction only permits it to be used as a rectifier. A crystal diode has two terminals depend on how it is used as forward or reverse biased. If the external circuit is trying to push the conventional current in the direction of arrow, the diode is forward biased. If it is opposite to arrowhead, the diode is reverse biased [23].
2.4.1 Rectifying Diode

![Diode Circuit Diagram]

**Fig.2.4.1 Rectifying Action of a Crystal Diode**

As shown in Fig 2.4.1, the rectification of a.c input voltage it has connected to the diode and load $R_L$ in series. The d.c output is obtained across the load. During positive half-cycle of input voltage, the arrowhead becomes positive w.r.t bar. Therefore diode is forward biased and conducts current in the circuit. The result is that positive half –cycle of input voltage appears across $R_L$. During the negative cycle of input a.c voltage, the diode becomes reversed biased because arrowhead will negative w.r.t bar. So under this condition it will not conduct and there is no voltage across $R_L$.

The behavior of diode is like switch. When the diode is forward biased, it behaves like a **closed switch** and connects the a.c. Supply to the load $R_L$. When the diode is **reversed biased**; it behaves like an open switch and disconnects the a.c. Supply from the load $R_L$.

**2.4.1.1 Resistance of Crystal Diode**

There two types of resistance generally observed in a crystal diode i.e. forward resistance and reverse resistance. The resistance of a forward bias diode is small compared to a reversed one [23].

In a Forward resistance the resistance is not the same for the same flow of direct current or as per the changing current. Accordingly this resistance exists in two types, namely d.c. forward resistance and a.c forward resistance.

The **d.c forward resistance** is the opposition offered by the diode to the direct current. It is measured by the ratio of d.c voltage across the diode to the resulting d.c current through it.
The a.c forward resistance is the opposition offered by the diode to the changing forward current. It is measured by the ratio of change in voltage across diode to the resulting change in current through it.

![Fig.2.4.1.1 (a) d.c.Forward Resistance](image)

![Fig.2.4.1.1 (b) a.c Forward resistance](image)

The d.c. forward resistance is given by, \( R_F = \frac{OA}{OB} \)

The a.c.forward resistance= (change in voltage across diode) / (corresponding change in current through diode) = \( \frac{oc-oa}{of-od} = \frac{ac}{df} \)

The forward resistance of a crystal diode is very small, ranging from 1 to 25Ohm.

On the other hand, the reverse resistance is the resistance offered by the diode to the reverse bias is known as reverse resistance. It can be d.c reverse resistance or a.c reverse resistance depending upon whether the reverse bias is direct or changing voltage. Ideally the reverse resistance of a diode is infinite. However practically for any value of reverse bias; there exists a small leakage current. In germanium diodes, the ratio of reverse to forward resistance is 40000:1 while for silicon 1000000:1[23].

### 2.4.1.2 Equivalent Circuit of Crystal Diode

When the forward voltage \( V_F \) is applied across a diode, it will not conduct till the potential barrier \( v_0 \) at the junction is overcome. When the forward voltage exceeds the potential barrier voltage, the diode starts conducting as shown Fig.2.4.1.2 (a). The forward current \( I_F \) flowing through the diode causes a voltage drop in its internal resistance \( r_f \). Therefore, the forward voltage \( V_F \) applied across the actual diode has to overcome

(i) Potential barrier \( V_0 \)

(ii) Internal drop \( I_f r_f \)

So \( V_F = V_0 + I_f r_f \)
Therefore, equivalent circuit for a crystal diode is a switch in series with a battery $V_0$ and internal resistance $r_f$ as shown Fig. 2.4.1.2(b). This equivalent circuit of diode is very helpful in studying the performance of a diode in a circuit.

For a silicon diode, $V_0 = 0.7V$ whereas for germanium diode $V_0 = 0.3V$.

**Ideal diode:** An ideal diode is one which behaves as a perfect conductor when forward biased and as a perfect insulator when reversed biased. Therefore in hypothetical situation $r_f = 0$ and potential barrier $V_0$ is considered negligible.

**Forward Current:** It is the current flowing through a forward biased diode. Every diode has a maximum value of forward current which it can safely carry. If this value exceeded, the diode may be destroyed due to excessive heat. That's why manufacturer always mention maximum forward current that a diode can handle.

**Peak Inverse voltage:** It is maximum reverse voltage that a diode can withstand without destroying the junction. If the reverse voltage across a diode exceeds this value, the reverse current increases sharply and breakdown the junction due to excessive heat. Peak inverse voltage is extremely important when diode is used as a rectifier.

**Reverse Current or leakage current:** It is the current that flows through a reverse biased diode. This current is due to the minority carriers. Under normal operating voltages, the reverse current is quite small. Its value is extremely small ($<1\mu A$) for silicon diodes but it is appreciable ($\approx 100 \mu A$) for germanium diodes.

There different types of crystal diode rectifier used for d.c supply (i) half-Wave rectifier (ii) Full-wave Rectifier-(a) Center –tap full-wave rectifier (b) Full-wave bridge rectifier.
2.4.1.3 Band Structure of an Open-Circuited p-n Junction

We have already mentioned that placing p- and n-type material in intimate contact on an atomic scale forms a p-n junction. Under these conditions the Fermi level must be constant throughout the specimen at equilibrium. If it is not, electrons on one side of the junction would have an average energy higher than those on the other side, and there would be a transfer of electrons and energy until the Fermi levels in the two sides did line up. In previous discussion it is verified that the Fermi level \( E_F \) is closer to the conduction band edge \( E_{Cn} \) in the n-type material and closer to the valence band edge \( E_{Vp} \) in the p side. So, the conduction band edge \( E_{Cp} \) in the p material can not be at the same level as \( E_{Cn} \), nor can the valence band edge \( E_{Vn} \) in the n side line up with \( E_{Vp} \). Hence the energy-band diagram for a p-n junction appears as shown in below fig2.4.1.4.1, where a shift in energy levels \( E_0 \). This represents potential energy for electrons. The width of the forbidden gap is \( E_g \) in electron volts.

![Band Diagram for an Open-Circuited p-n Junction](image)

**Fig. 2.4.1.3.(a) Band Diagram for s p-n junction in Open-Circuit**

[60]
E₀ = Ecₚ – Ecₙ = Evₚ - Evₙ = E₁ + E₂

This energy E₀ represents the potential energy of the electrons at the junction Fig. 2.4.1.3. (b).

From the Fig. 2.4.1.3(a), we can see the contact difference of potential as

E_F - Eᵥₚ = 1/2Eᵢ-E₁ and Eᵥₙ - E_F = 1/2Eᵢ-E₂

Adding these two equations, we obtain

E₀ = E₁ + E₂ = Eᵢ - (Eᵥₙ - E_F) - (E_F - Eᵥₚ)
### Silicon Based Computing Devices

#### Acceptor | Junction | Donor
---|---|---
Holes | Electron | P-type

#### 0.5cm

**Space charge region**

\[
d^2V/dx^2 = -\rho/\varepsilon
\]

\[
\rho/\varepsilon
\]

\[
\varepsilon = -dV/dx = \int \varepsilon \cdot dx
\]

**Electric field intensity**

\[
\psi = -\int \varepsilon \cdot dx
\]

**Electrostatic potential**

\[
V_0
\]

**Potential energy barrier**

---

Fig. 2.4.1.3(b) P-N Junction with Charge Density, Electric-Field Intensity and Potential-Energy Barrier

[62]
From equation 2.3.1.1.7 and 2.3.1.1.8,

\[ E_G = kT \ln \frac{N_c N_v}{n_i^2} \]

From equation 2.3.2.1, \( E_{c_n} - E_F = kT \ln \frac{N_c}{N_D} \)

From equation 2.3.2.1, \( E_F - E_{v_p} = kT \ln \frac{N_v}{N_A} \)

Now \( E_0 = kT \left( \ln \frac{N_c N_v}{n_i^2} - \ln \frac{N_c}{N_D} - \ln \frac{N_v}{N_A} \right) = kT \left( \frac{N_c N_v}{n_i^2} \cdot \frac{N_D}{N_c} \cdot \frac{N_A}{N_v} \right) \)

\[ = kT \ln \left( \frac{N_D N_A}{n_i^2} \right) \quad \text{Eq}(2.4.1.3.1) \]

\( E_0 \)'s are expressed in electron volt and \( k \) has the dimensions of electron volts per degree Kelvin. The contact difference in potential \( V_0 \) is expressed in volts and equivalent numerically to \( E_0 \). \( V_0 \) only depends on the equilibrium concentrations and not at all on the charge density in the transition region.

Other expression of \( E_0 \) are obtained by substituting the concentration for a n-type semiconductor as \( n_n \approx N_D \) (eq.2.3.2.1), \( p_n = n_i^2 / N_D \) (eq.2.3.2.3). Similarly for p-type semiconductor \( p_p \approx N_A \) (Eq.2.3.2.2), \( n_p = n_i^2 / N_A \) (2.3.2.4) in equation 2.4.1.3.1.

\[ E_0 = kT \ln \frac{p_p}{p_n} = kT \ln \frac{n_n}{n_p} \quad \text{Eq}(2.4.1.4.2) \]

Using reasonable values \( p_p = 10^{16} \text{ cm}^{-3} \), \( p_n = 10^4 \text{ cm}^{-3} \), and \( kT = 0.026 \text{ eV} \). At room temperature, we obtain \( E_0 \approx 0.5 \text{ eV} \).

Under equilibrium condition of zero resultant hole current, the hole current density is zero, the negative of the hole diffusion current must equal the hole drift current.

\[ \frac{D_p}{p} = -\frac{dV}{V_T}, \text{if this equation is integrated between limits which extend across the junction from the p material, where the equilibrium hole concentration is } p_{p_0}, \text{to the n side, where the whole density is } p_{n_0}, \text{the result is} \]

\[ p_{p_0} = p_{n_0} e^{V_0 / V_T}, \text{since } V_0 / V_T = E_0 / kT, \]

It will be \( E_0 = kT \ln \frac{p_{p_0}}{p_{n_0}} \quad \text{Eq}(2.4.1.3.3) \)
2.4.1.4 The Current Components in a p-n Diode

When a forward bias is applied to a diode, holes are injected into n side and electrons into p side. The number of these injected minority carriers falls off exponentially with distance from the junction. Since the diffusion current of minority carriers is proportional to the concentration gradient, this current must also vary exponentially with distance. There are two minority currents, $I_{pn}$ and $I_{np}$, and are described in below fig.6-5.

![Fig. 2.4.1.4 Current Components in a p-n diode](image)

$I_{pn}(x)$ represents hole current in the n material and $I_{np}(x)$ represents electron current in the p side as a function of x. Electrons crossing the junction at x=0 from right to left constitute a current in the same direction as holes crossing the junction from left to right.

Hence total current $I$ at x=0 is

$$I = I_{pn}(0) + I_{np}(0)$$

The hole current in the p side $I_{pp}$ (a majority carrier current) is given by

$$I_{pp}(x) = I - I_{np}(x)$$

As the holes approach the junction, some of them recombine with the electrons, which are injected into the p side from the n side. Hence part of the current $I_{pp}$ becomes a negative current just equal in magnitude to the diffusion current $I_{np}$. The current $I_{pp}$ thus decreases toward the junction. What remains of $I_{pp}$ at the junction enters the n side and becomes the whole diffusion current $I_{pn}$. Similar remarks can be done with $I_{nn}$.

[64]
The current in a p-n diode is bipolar in character since it is made up of both positive and negative carriers of electricity. The total current is constant throughout the device, but the proportion due to holes and that due to electrons varies with distance.

### 2.4.1.4.1 Quantitative theory of the p-n Diode Currents

Assuming the barrier width is zero, in a forward bias diode, the holes are injected from the p side into the n material. The concentration $p_n$ of holes in the n side is increased above its thermal-equilibrium value $p_{n0}$ and is given by

$$p_n(x) = p_{n0} + P_n(0) e^{-x/L_p}$$

Where the parameter $L_p$ is called the diffusion length for holes in the n material, and the injected, or excess, concentration at $x=0$ is

$$P_n(0) = p_n(0) - p_{n0}$$

The below fig. 2.4.1.4.1 shows the exponential decrease of the density $p_n(x)$ with distance $x$ into n material.

![Fig. 2.4.1.4.1 Quantitative Theory of the p-n Diode Currents](image)

The diffusion hole current in the n side is given by

$$I_{pn} = -AeD_p \left( \frac{dp_n}{dx} \right)$$

Taking derivative of $p_n(x)$ and substitute in above equation.

$$I_{pn}(x) = AeD_p \frac{P_n(0)}{L_p} e^{-x/L_p}$$
This equation verifies that the hole current decrease exponentially with distance. The dependence of $I_{pn}$ upon applied voltage is contained implicitly in the factor $P_n(0)$ because the injected concentrations are a function of voltage. We now find the dependence of $P_n(0)$ upon $V$.

### 2.4.1.4.2 The Law of Junction

If the hole concentrations at the edges of the space-charge region are $p_p$ and $p_n$ in the $p$ and $n$ materials, respectively, and if the barrier potential across this depletion layer is $V_B$, then

$$p_p = p_n \varepsilon \frac{V_B}{V_T} \quad \text{Eq(2.4.1.4.2.1)}$$

If we apply this to the case of open-circuited $p$-$n$ junction, then $p_p = p_{p0}$, $p_n = p_{n0}$ and $V_B = V_0$. Substituting these values in above equation,

It will be $p_{p0} = p_{n0} \varepsilon \frac{V_0}{V_T}$.

Considering now a junction biased in the forward direction by an applied voltage $V$. Then the barrier voltage $V_B$ is decreased from its equilibrium value $V_0$ by the amount $V$, or $V_B = V_0 - V$. Now $p_p = p_{p0}$.

At the edge of depletion layer, $x=0$, $p_n = p_n(0)$.

So, $p_{p0} = p_n(0) \varepsilon (V_0 - V) / V_T \quad \text{Eq(2.4.1.4.2.2)}$

$$p_{n0} \varepsilon \frac{V_0}{V_T} = p_n(0) \varepsilon (V_0 - V) / V_T, \text{ which implies}$$

$$p_n(0) = p_{n0} \varepsilon \frac{V_0}{V_T} \quad \text{Eq(2.4.1.4.2.3)}$$

This boundary condition is called law of junction.

It indicates that for a forward bias ($V > 0$), the hole concentration $p_n(0)$ at the junction is greater than the equilibrium value $p_{n0}$. A similar workout can be done for electrons by interchanging $p$ and $n$.

The hole concentration now will be $P_n(0) = p_n(0) - p_{n0} \varepsilon \frac{V_0}{V_T} - p_{n0} = p_{n0} (\varepsilon^{\frac{V_0}{V_T}} - 1)$
2.4.1.4.3 The Forward Current

The hole current $I_{pn}(0)$ crossing the junction into the n side at $x=0$ is given by

$$I_{pn}(0) = (AeD_p n_0) / L_p (\varepsilon V / VT - 1)$$  \text{Eq (2.4.1.5.3.1)}

The electron current

$$I_{np}(0) = (AeD_n p_0) / L_n (\varepsilon V / VT - 1)$$  \text{Eq (2.4.1.5.3.2)}

The total diode current $I$ is sum of $I_{pn}(0)$ and $I_{np}(0)$, or

$$I = I_0 (\varepsilon V / VT - 1)$$  \text{Eq (2.4.1.5.3.3)}

Where $I_0 = (AeD_p n_0) / L_p + (AeD_n p_0) / L_n$  \text{Eq (2.4.1.5.3.4)}

2.4.1.4.4 The Reverse Saturation Current

For a reverse bias whose magnitude is large compared with $V_T$ (~26mV at room temperature), $I \rightarrow I_0$. Hence $I_0$ is called the reverse saturation current.

$$I_0 = Ae (D_p / L_p N_D + D_n / L_n N_A) n_i^2$$  \text{Eq (2.4.1.4.4.1)}

Where $n_i^2 = A_0 T^{3.5} \varepsilon^{-E_{GO}/kT} = A_0 T^{3} \varepsilon^{-V_{GO}/kT}$  \text{Eq (2.4.1.4.4.2)}

Where $V_{GO}$ is a voltage which is numerically equal to the forbidden-gap energy $E_{GO}$ in electron volts, and $V_T$ is the volt equivalent of temperature.

We have neglected carrier generation and recombination in the space-charge region for germanium. But for silicon, the diffusion current is negligible compared with the transition-layer charge-generation current, which is given by

$$I = I_0 (\varepsilon V / VT - 1)$$  \text{Eq (2.4.1.4.4.3)}

Where $\eta \approx 2$ for small currents and $\eta \approx 1$ for large currents. Also $I_0$ is now found to be proportional to $n_i$ instead of $n_i^2$. Hence if $K_2$ is a constant,

$$I_0 = K_2 T^{1.5} \varepsilon^{-\frac{V_{GO}}{2V_T}}$$  \text{Eq (2.4.1.4.4.4)}
2.4.1.5 The Volt-Ampere Characteristic

For p-n junction, the current $I$ is related to the voltage by the equation

$$I = I_0 (e^{\frac{V}{V_T}} - 1) \quad \text{Eq(2.4.1.5.1)}$$

A positive value of $I$ means current flows from the p to the n side. The diode is forward biased if $V$ is positive, indicating that the p side of the junction is positive with respective to the n side. The symbol $\eta$ is unity for germanium and approximately 2 from silicon.

$V_T$ stands for volt equivalent of temperature, and is given by

$$V_T = \frac{T}{11,600} \quad \text{Eq(2.4.1.5.2)}$$

At room temperature ($T=300^0K$), $V_T = 0.026V = 26mV$. The form of volt-ampere characteristics is shown below fig. 2.4.1.5

![Volt-ampere characteristic of ideal p-n diode](image)

**Fig. 2.4.1.5 Volt-ampere characteristic of ideal p-n diode**

The current increases exponential if $V$ is positive and several times of $V_T$. When the diode is reverse-biased and $|V|$ is several times $V_T$, $I \approx -I_0$. The reverse current is therefore constant, independent of the applied reverse bias. Consequently $I_0$ is referred as reverse saturation current.

In order to display forward and reverse characteristics conveniently, it is necessary to use two different current scales. The volt-ampere characteristics shown in the fig has a forward current scale in mill amperes and a reverse scale in microamperes.
The dashed portion of the curve indicates that a reverse biasing voltage $V_Z$, the diode characteristic exhibits an abrupt and marked departure. At this critical voltage a large reverse current flows, and the diode is said to be in the breakdown region.

### 2.4.1.6 Space-Charge (Depletion Region)

Considering a junction in which there is an abrupt change from acceptor ions on one side to donor ions on the other side. Such a junction is formed experimentally called alloy, or fusion or junction. It is necessary that the concentration $N_A$ of acceptor ions equal the concentration $N_D$ of donor impurities. As a matter of fact, it is often advantageous to have an unsymmetrical junction [28].

\[
e_{N_A}W_p = e_{N_D}W_n \quad \text{Eq}(2.4.1.6.1)
\]
If \( N_A << N_D \), then \( W_p >> W_n \). For simplicity, we neglect \( W_n \) and assume that the entire barrier potential \( \psi_B \) appears across the uncovered acceptor ions. The relationship between potential and charge density is given by Poisson’s equation,

\[
\frac{d^2V}{dx^2} = \frac{eN_A}{\varepsilon}, \quad \text{Eq (2.4.1.6.2)},
\]

where \( \varepsilon \) is the permittivity of the semiconductor.

If \( \varepsilon_r \) is the relative dielectric constant and \( \varepsilon_0 \) is the permittivity of free space, then \( \varepsilon = \varepsilon_r \ast \varepsilon_0 \). The electric lines of flux start on the positive donor ions and terminate on the negative acceptor ions. Hence there are no flux lines to the left of the boundary \( x=0 \) in Fig.2.4.1.6 and \( \varepsilon = -dV/dx = 0 \) at \( x=0 \). Also, since the zero of potential is arbitrary, we chose \( V=0 \) at \( x=0 \).

Integrating the above equation, subject to the boundary condition yields

\[
V = \frac{eN_A}{2 \varepsilon} \frac{x^2}{2} \quad \text{Eq (2.4.1.6.3)}
\]

At \( x=W_p=W \), \( V=V_B \), the barrier height.

Thus

\[
V_B = \left( \frac{eN_A}{2 \varepsilon} \right) W^2 \quad \text{Eq (2.4.1.6.4)}
\]

If we now reverse the symbol \( V \) for the applied bias, then \( V_B = V_0 - V \), where \( V \) is a negative number for an applied reverse bias and \( V_0 \) is the contact potential. This equation confirms our qualitative conclusion that the thickness of the depletion layer increases with applied voltage. We now see that \( W \) varies as \( V_B^{1/2} \).

If \( A \) is the area of the junction, the charge in the distance \( W \) is

\[
Q = eN_A W A
\]

The transition capacitance \( C_T \) is

\[
C_T = \frac{dQ/dV}{eN_A A} dW/dV = \frac{\varepsilon A}{W} \quad \text{Eq (2.4.1.6.5)}
\]

This formula is exactly the expression which is obtained for a parallel-plate capacitor of area \( A \) and plate separation \( W \) containing a material of permittivity \( \varepsilon \).

[70]
2.4.2 Tunneling Diode

A p-n junction diode of the type discussed above section has an impurity concentration of about 1 part in $10^8$. With this amount of doping, the width of the depletion layer, which constitutes a potential barrier at the junction, is of the order 5 microns ($5 \times 10^{-4}$ cm). This potential barrier restrains the flow of carriers from the side of the junction where they constitute majority carriers to the side where they constitute minority carriers. If the concentration of impurity atoms is greatly increased, say to 1 part in $10^3$ (corresponding to a density in excess of $10^{19}$ cm$^{-3}$), the device characteristic are completely changed. This new diode was announced in 1958 by Esaki who also gave the correct theoretical explanation for its volt-ampere characteristic [28], depicted in Fig. 2.4.2.

![Fig. 2.4.2 Volt-Ampere Characteristic of Tunnel Diode](image-url)
2.4.2.1 The Tunneling Phenomenon

A tunnel diode or Esaki diode is a type of semiconductor diode which is capable of very fast operation, well into the microwave frequency region, by using quantum mechanical effects. These diodes have a heavily doped p–n junction only some 10 nm (100 Å) wide. The heavy doping results in a broken bandgap, where conduction band electron states on the n-side are more or less aligned with valence band hole states on the p-side. The width of the junction varies inversely with square root of impurity concentration [28].

As depicted in Eq (2.4.1.6.4), $V_b = \frac{eN_A}{2\varepsilon} W^2$ or $W = \sqrt{\frac{(2V_b\varepsilon/e)^*1/N_A}}$

From above equation it implies that width is reduced from 5 microns to less than 100A⁰ (10⁻⁶ cm). This thickness is only about one-fifth the wavelength of visible light. Classically, a particle must have an energy at least equal to the height of a potential-energy barrier if it is to move from one side of the barrier to the other. However, for barriers as thin as those estimated above in the Esaki diode, the Schrödinger equation indicates that there is a large probability that an electron will penetrate through the barrier. This quantum-mechanical behavior is referred to as tunneling, and hence these high-impurity-density p-n junction devices are called tunnel diodes. In an one-dimensional problem: An electron of total energy $W$ (joules) moves in the region 1, where the potential energy may be taken as zero, $U=0$. At $x=0$, there is a potential-energy barrier of height $U_0 > W$, and indicated in Fig. 2.4.2.1, the potential energy remains constant in region 2 for $x>0$.

![Fig. 2.4.2.1 Potential Energy Hill](image-url)
CHAPTER- 2  Silicon Based Computing Devices

Region 1

The Schrodinger equation is represented as
\[ \frac{d^2 \Psi}{dx^2} + \frac{8\pi^2 m}{\hbar^2} W \Psi = 0 \]

The solution Schrodinger equation is in the form
\[ \Psi = C e^{i \frac{8\pi^2 m}{\hbar^2} W \frac{x^2}{2}} \]  

Where \( C \) is a constant. The electronic wave function \( \Phi = e^{i\omega t} \Psi \) represents a travelling wave.

We discussed earlier the product of \( \Psi \) and its complex conjugate \( \Psi^* \) is interpreted as giving the probability of finding an electron between \( x \) and \( x+dx \) (in a one-dimensional space). Since \( \Psi \Psi^* = |\Psi|^2 = C^2 = \text{const} \), the electron has an equal probability of being found anywhere in a region of zero potential energy.

Region 2

The Schrodinger equation for \( x > 0 \) is represented as
\[ \frac{D^2 \Psi}{dx^2} - \frac{8\pi^2 m}{\hbar^2} (U_0 - W) \Psi = 0 \]

Since \( U_0 > W \), this equation has a solution of the form
\[ \Psi = A e^{i \frac{8\pi^2 m}{\hbar^2} (U_0 - W) \frac{x^2}{2}} \]

Where \( A \) is a constant and
\[ \omega = \frac{1}{2} \left( \frac{\hbar^2}{(2m(2m(U_0 - W))^2} \right)^{1/2} \]

The probability of finding electron between \( x \) and \( x+dx \) in region 2 is
\[ |\Psi|^2 dx = A^2 e^{2 \omega x} \]
From this we can see that an electron can penetrate a potential energy barrier and that this probability decreases exponentially with distance into the barrier region. A calculation of $d_0$ for $U_0-W=1.60 \times 10^{-20} \text{J}$ (corresponding to 0.1eV) yields $d_0 \approx 3 \text{Å}$. For impurity densities in excess of those indicated above ($10^{19} \text{cm}^{-3}$), the barrier depth $d$ approaches $d_0$, and $A^2 e^{-d/d_0}$ becomes large enough to represent an appreciable number of electrons which have tunneled through the hill[28].

2.4.2.2 Energy-Band Structure of a high Doped p-n Diode

The condition that $d'$ be of the same order of magnitude as $d_0$ is a necessary but not a sufficient condition for tunneling. It is also required that occupied energy states exist on the side from which the electron tunnels and that allowed empty states exist on the other side at the same energy level. Hence we must now consider the energy-band picture when the impurity concentration is very high. Earlier we discussed for a lightly doped p-n diode, the Fermi level $E_F$ lies inside the forbidden gap. We shall now demonstrate that, for a diode which is doped heavily enough to make tunneling possible, $E_F$ lies outside the forbidden band.

We know $E_F = E_C - kT \ln N_C/N_D$

For a lightly doped semiconductor, $N_D < N_C$ so that $\ln (N_C/N_D)$ is a positive number. Hence $E_F < E_C$, and the Fermi level lies inside the forbidden band.

Since $N_C \approx 10^{19} \text{cm}^{-3}$, then for donor concentrations in excess of this amount ($N_D > 10^{19} \text{cm}^{-3}$, corresponding to a doping in excess of 1 part in $10^3$), $\ln (N_C/N_D)$ is negative. Hence $E_F > E_C$, and the Fermi level in the n-type material lies in the conduction band. By similar reasoning we conclude that, for a heavily doped p region, $N_A > N_V$, and Fermi level lies in the valence band. Hence $E_F > E_G$, so that the contact difference of potential energy $E_0$ now exceeds the forbidden-energy-gap voltage $E_G$. Hence under open conditions, the band structure of a heavily doped p-n junction must be as shown below Fig. 2.4.2.2.
Fig. 2.4.2.2 Energy Band in heavily doped p-n-Diode

The Fermi-level $E_F$ in the p side is at the same energy as the Fermi level $E_F$ in the n side. Note that there are no filled states on one side of the junction which are at the same energy as empty allowed states on the other side. Hence there can be no flow of charge in either direction across the junction, and the current is zero, an obviously correct conclusion for an open-circuited diode.

### 2.4.2.3 The Volt-Ampere Characteristics

With the aid of the energy-band picture of the above fig. 2.4.2.2 and the concept of quantum-mechanical tunneling, the tunnel-diode characteristics of Fig-2.4.2 may be explained. Let us consider that the p material is grounded and that a voltage applied across the diode shifts the n side with respect to the p-side. For example, if a reverse-bias voltage is applied, we know that the height of the barrier is increased above the open-circuited value $E_0$, hence the n-side levels must shift downward with respect to the p-side levels, as indicated in Fig 2.4.2.2 (b). We now observe that there are some energy
states (the heavily shaded region) in the valence band of the p side which lie at the same level as allowed empty states in the conduction band of the n side. Hence these electrons will tunnel from the p to the n side, giving rise to a reverse diode current. As the magnitude of reverse bias increases, the heavily shaded area grows in size, causing the reverse current to increases, as shown section 1 of Fig. 2.4.2.3.(a)

Fig. 2.4.2.3(a) The Tunneling Current (solid line)
Fig. 2.4.2.3 Energy-Band picture in heavily doped p-n diode under Forward Bias.

When forward bias is applied to the diode, the potential barrier is decreased below $E_0$, hence the n-side levels must shift upward with respect to those on the p side, and the energy-band picture for this situation is indicated in Fig. 2.4.2.3 (b). It predicts that there are occupied states in the conduction band of the n material which are at the same energy as allowed empty states (holes) in the valence band of the p side. Hence electrons will tunnel from the n to the p material, giving rise to the forward current.
As the forward bias is increased further, the condition shown in Fig. 2.4.2.3.(c) is reached. Now the maximum number of electrons can leave occupied states on the right side of the junction, and tunnel through the barrier to empty states on the left side, giving rise to the peak current $I_P$ in Fig. 2.4.2.3 (a). If still more forward bias is applied, the situation in Fig. 2.4.2.3. (d) is obtained, and the tunneling current decreases, giving rise to Section 3 of Fig. 2.4.2.3 (a). Finally at an even larger forward bias, the band structure of Fig. 2.4.2.3. (e) is valid. Since now there are no empty allowed states on one side of the junction at the same energy as occupied states on the other side, the tunneling current must drop to zero. The dashed line in section-4 is regular p-n junction injection current.

### 2.4.2.4 Characteristics of a Tunneling Diode

The tunneling diode is an excellent conductor in the reverse direction (the p side of the junction negative with respect to the n side). Also, for small forward voltage, the resistance remains small. At the peak current $I_P$ correspond to the voltage $V_P$, then the current decreases. As a consequence, the dynamic conductance $g=dI/dV$ is negative. The tunneling diode exhibits a negative-resistance characteristic between the peak current $I_P$ and the minimum value $I_V$, called the valley current. At the valley voltage $V_V$ at which $I=I_V$, the conductance is gain zero, and beyond this point the resistance becomes and remains positive. At the so-called peak forward voltage the current increases beyond this value. For currents whose values are between $I_V$ and $I_P$, the curve is triple valued, because each current can be obtained at three different applied voltages. It is this multivalued feature which makes the tunneling diode useful in pulse and digital circuit. The standard circuit symbol shown below Fig. 2.4.2.4.

![Tunneling Diode and Small-signal model for negative-resistance](image)

Fig. 2.4.2.4 Tunneling Diode and Small-signal model for negative-resistance
2.5 Combinatorial Logic Devices and their Properties

The designing of a digital logic circuits generally happens in two ways.

i) Combinational logic design

ii) Sequential logic design.

A combinational logic is the one in which there are only inputs and outputs, but there is no feedback, but a sequential logic involves feedback along with inputs and outputs.

A bipolar circuit involves resistors, diodes and bipolar junction transistors. The simplest logic circuit developed with the diodes and resistors called diode logic. A semiconductor diode is fabricated with two types of semiconductor materials called P-type and N-Type.

---

![Fig. 2.5.(a) A semiconductor Diode](image)

This diode does not allow current i.e. open circuited when +ve voltage is given and short circuited when a –ve voltage is given as shown below fig. thus acts as a switch. (Fig.2.5.2).

---

![Fig.2.5.(b) Diode acts as switch](image)
2.5.1 Diode Logic

With the above specification various types of basic logic gates that can be made from diodes are AND, NAND, NOT, OR, NOR, XOR and XNOR. Generally, NAND gates were easier to construct from MOS (Metal Oxide Semiconductor) technology and thus NAND gates served as the first pillar of boolean logic in electronic computation. Boolean functions may be practically implemented by using electronic gates. The following basic things are needed for designing electronic gates. They are (i) a power supply (ii) Gate INPUTS are driven by voltages having two nominal values, e.g. 0V and 5V representing logic 0 and logic 1 respectively. (iii) The OUTPUT of a gate provides two nominal values of voltage only, e.g. 0V and 5V representing logic 0 and logic 1 respectively. In general, there is only one output to a logic gate except in some special cases. (iv) There is always a time delay between an input being applied and the output responding Diode.

2.5.1.1 AND Gate

Fig.2.5.1.1 Symbolic Representation of AND Gate

The AND gate is an electronic circuit that gives a high output (1) only if all its inputs are high. A dot (.) is used to show the AND operation i.e. A.B. Bear in mind that this dot is sometimes omitted i.e. AB.

2.5.1.2 OR Gate

Fig.2.5.1.2 Symbolic Representation of OR Gate

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A+B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
CHAPTER- 2 Silicon Based Computing Devices

The OR gate is an electronic circuit that gives a high output (1) if one or more of its inputs are high. A plus (+) is used to show the OR operation.

2.5.1.3 NOT Gate

The NOT gate is an electronic circuit that produces an inverted version of the input at its output. It is also known as an inverter. If the input variable is A, the inverted output is known as NOT A. This is also shown as A', or A with a bar over the top, as shown at the outputs. The diagrams below show two ways that the NAND logic gate can be configured to produce a NOT gate. It can also be done using NOR logic gates in the same way.

Fig. 2.5.1.3 (a) Symbolic Representation of NOT Gate from AND type

Fig. 2.5.1.3 (b) Symbolic Representation of NOT Gate from OR type

2.5.1.4 NAND Gate

Fig.2.5.1.4 Symbolic Representation of NAND Gate

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
This is a NOT-AND gate which is equal to an AND gate followed by a NOT gate. The outputs of all NAND gates are high if any of the inputs are low. The symbol is an AND gate with a small circle on the output. The small circle represents inversion.

2.5.1.5 NOR Gate

![Symbolic Representation of NOR Gate](image)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A+B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig.2.5.1.5 Symbolic Representation of NOR Gate

This is a NOT-OR gate which is equal to an OR gate followed by a NOT gate. The outputs of all NOR gates are low if any of the inputs are high. The symbol is an OR gate with a small circle on the output. The small circle represents inversion.

2.5.1.6 XOR gate

![Symbolic Representation of XOR Gate](image)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A⊕B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig.2.5.1.6 Symbolic Representation of XOR Gate

The 'Exclusive-OR' gate is a circuit which will give a high output if either, but not both, of its two inputs are high.
2.5.1.7 XNOR gate

The ‘Exclusive-NOR’ gate circuit does the opposite to the XOR gate. It will give a low output if either, but not both, of its two inputs are high. The symbol is an XOR gate with a small circle on the output. The small circle represents inversion.

The NAND and NOR gates are called universal functions since with either one the AND and OR functions and NOT can be generated.

Note: A function in sum of products form can be implemented using NAND gates by replacing all AND and OR gates by NAND gates. A neither function in product of sums form can be implemented using NOR gates by replacing all AND and OR gates by NOR gates.

<table>
<thead>
<tr>
<th>2 Input</th>
<th>XNOR gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig.2.5.1.7 Symbolic Representation of XNOR Gate
All logic gate symbols can summaries as

![Logic Gate Symbols](image)

<table>
<thead>
<tr>
<th>NOT gate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

### Table 2.5.1: Logic Gates representation using the Truth table

<table>
<thead>
<tr>
<th>INPUTS</th>
<th>OUTPUTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

#### 2.5.2 MULTI Input Logic Gates

Above we have shown all logic gates have taken 2 inputs. But we can use also more than 2 inputs as shown below: Ex-AND gate.

#### 2.5.2.1 Two Input AND Gate

Here is an example of a two input gate as we have already seen. It is an AND gate and the truth table for this gate can be seen to the right of it.
2.5.2.2 Three Input AND Gate

Here is an example of a three input AND gate. Notice that the truth table for the three input gate is similar to the truth table for the two input gate. It works on the same principle, this time all three inputs need to be high (1) to get a high output.
2.5.2.3 Four Input AND Gate

Here is an example of a four input AND gate. It also works on the same principle, all four inputs need to be high (1) to get a high output. The same principles apply to 5, 6, ..., n input gates.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>A.B.C.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table: 2.5.2.3 (4 input AND gate)

2.5.3 Binary Adder

Binary adders combine binary values to obtain a sum using the combinations of logic gates. These are classified according to their ability to accept and combine the digits. The two types of adder generally considered are half-adder and full-adder[46].
2.5.3.1 Half Adder

A half adder is a logical circuit that performs an addition operation on two binary digits. The half adder produces a sum and a carry value which are both binary digits.

\[
\begin{array}{c|c|c}
A & B & S \\
0 & 0 & 0 \\
1 & 0 & 1 \\
0 & 1 & 1 \\
1 & 1 & 0 \\
\end{array}
\]

**Sum (S) = XOR**

\[
\begin{array}{c|c|c}
A & B & C \\
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 1 & 1 \\
\end{array}
\]

**Carry (C) = AND**

\[
S = (A \text{ XOR } B) \text{ XOR } C_i
\]

\[
C_0 = (A \text{ AND } B) \text{ or } (C_i \text{ AND } (A \text{ XOR } B))
\]

\[
= (A \text{ and } B) \text{ or } (B \text{ and } C_i) \text{ or } (C_i \text{ and } A)
\]
2.6 Sequential Logic Design

Sequential switching circuits are those whose output levels at any instant of time are dependent not only on the levels present at that time, but also on the prior input level conditions. It means that sequential switching circuits have memory. Sequential circuits are made of combinatorial circuits and memory elements. The basic building blocks of sequential circuits are the flip-flop, bistable, and latch just as the basic building block of the combinational circuit is the gate [46].

Sequential circuits are very important for computer hardware because they are needed to implement registers, counters, and shifters, all of which are fundamental to the operation of the central processing unit in a computer [46].

2.6.1 Flip-Flop

The most important memory element is the flip-flop which is an assembly of logic gates. There are several different gate arrangements that are used to construct flip-flops in a wide variety of ways. Each type of flip-flop has several special features or characteristics necessary for particular applications. The flip-flops serve as a storage device. Flip-flops are the fundamental components of shift registers and counters. There different types of Flip-Flop available are (i) SR Flip Flop (ii) J-K Flip Flop and (iii) D-Flip Flop and (iv) T-Flip Flop. The term 'latch' is used for certain non-clocked flip-flops.

A flip-flop can have one or more inputs. These inputs are used to cause the flip-flop to switch back and forth between its possible output states. A flip-flop input has to be pulsed momentarily to cause a change in the flip-flop output, and the output will remain
in that new state even after the input pulse has been removed. This gives a memory characteristic to flip-flop[46].

2.6.1.1 S R Latch

It has two outputs labeled Q and Q’ and two inputs labeled as S and R. The state of the latch corresponds to level Q (HIGH or LOW, 1 or 0) and Q’ is complement of that state.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Output</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0</td>
<td>Q0</td>
<td>No change</td>
</tr>
<tr>
<td>0 1</td>
<td>0</td>
<td>RESET</td>
</tr>
<tr>
<td>1 0</td>
<td>1</td>
<td>SET</td>
</tr>
<tr>
<td>1 1</td>
<td>?</td>
<td>Not allowed</td>
</tr>
</tbody>
</table>

Fig. 2.6.1.1 Active-HIGH S-R Latch

2.6.1.2 NAND Gate S-R Latch

An active –LOW S-R latch can be constructed using cross-coupled NAND gates. Below shows the logic diagram and truth table of an active-LOW S-R Latch.

The analysis can be shown as (a) If S=0 and R=0, both Q and Q’ will be equal to 1 creating an invalid condition. (b) If S=1 and R=0, Q will be 0 and Q will be 1, Flip-Flop
resets. (c) If $S=0$ and $R=1$, $Q$ will be 1 and $\bar{Q}$ will be 0, Flip-Flop sets and (d) If $S=1$ and $R=1$, No change of state takes place, the output $Q$ and $\bar{Q}$ remains same as before [46].

2.6.1.3 The D Latch and the D flip-flop

We can construct a latch by single input $S$ and obtain $R$ by inverting it. This single input is labeled $D$. It differs from SR latch is that it has only one input.

![Diagram of D latch and D flip-flop]

**Fig.2.6.1.3 The D latch and the D Flip-Flop**
2.6.1.4 J-K Flip Flop

The J-K flip flop is the most versatile flip-flop, and the most commonly used flip flop when discrete devices are used to implement arbitrary state machines. Like the RS flip-flop, it has two data inputs, J and K, and a clock input.

It is generally represented by below figures.

These are the following characteristics of a Flip-Flop.

a) When J=0 and K=0, no change takes place even clock pulse applied.

b) When J=0 and K=1, the flip-flop resets

c) When J=1 and K=0, the flip-flop sets

d) When J=1 and K=1, the flip-flop toggles
2.6.1.5 T Flip-Flop

It is a simplified version of the JK flip-flop. It is not usually found as an IC chip but used in many circuits like counter and dividers. Its only function is that it toggles itself with every clock pulse and constructed from the RS flip-flop.

![T Flip-Flop Diagram](image)

Fig.2.6.1.5 T Flip-Flop

2.6.2 Register

As a flip-flop (FF) can store only one bit of data, a 0 or a 1, it is referred to as a single bit register. When more bits of data are to be stored, a number of FFs are used. A register is a set of FFs used to store binary data[46]. The storage capacity of a register is the number of bits (1s and 0s) of digital data it can retain.

Loading a register means setting or resetting the individual FFs, i.e. inputting data into the register so that their states correspond to the bits of data to be stored. Loading may be serial or parallel. In a serial loading data is transferred into the register in serial form, i.e. 1 bit at a time, where as in parallel loading, the data is transferred into the register in parallel form meaning that all the FFs are triggered into their new states at the same time. Parallel input requires that the SET and/or RESET controls of every FF be accessible. A register may output data either in serial form or in parallel form. Serial means one bit at a time serially. Parallel means entire data available transferred out at the same time.
2.6.2.1 Data Register/Buffer Register

The simplest form of register is Buffer Register or Data register which is used for the temporary storage of a "word" of data. In its simplest form, it consists of a set of N D flip flops, all sharing a common clock. All of the digits in the N bit data word are connected to the data register by an N line "data bus". Below shows a 4-bit data register.

![Fig.2.6.2.1 4-bit Data Register](image)

The binary word to be stored is applied to the data terminals. On application of clock pulse, the output word becomes the same as the word applied at the input terminals i.e. the input word loaded into the register by the application of clock pulse.

2.6.2.2 Shift Register

It is simply a set of flip flops (usually D latches or RS flip-flops) connected together so that the output of one becomes the input of the next, and so on in series. It is called a shift register because the data is shifted through the register by one bit position on each clock pulse. Below shows a four bit shift register, implemented with D flip flops.

![Fig. 2.6.2.2 (a) Shift Register (Serial-In-Serial-Out)](image)
Shift registers are named by their method of input and output; either serial or parallel. They are (i) Serial-In-Serial Out (SISO) (ii) Serial-In-Parallel-Out (SIPO) (iii) Parallel-In-Serial Out (PISO) and (iv) Parallel-In-Parallel Out (PIPO). There are also bi-directional shift registers which allow shifting in both directions: L→R or R→L.

Fig. 2.6.2.2(b) Serial-In-Parallel Out

Fig. 2.6.2.2 (c) Parallel-In-Serial Out
Fig. 2.6.2.2 (d) Parallel-In-Parallel Out

2.6.3 Counters

A digital counter is a set of flip-flops whose states change in response to pulses applied at the input to the counter. The flip-flops are interconnected such that their combined state at anytime is the binary equivalent of the total number of pulses that have occurred up to that time [46]. A counter can also be used as frequency divider to obtain waveforms with frequencies that are specific fractions of clock frequency. Counters can be asynchronous counters or synchronous counters. Asynchronous counters are also called ripple counters. It is called ripple counter because when the counter for example goes from 1111 to 0000, the first stage causes the second to flip, the second causes third to flip and so on. The ripple counter is the simplest type of counter, the easiest to design and requires the least amount of hardware. In ripple counters, the FFs within the counter are not made to change the states at exactly the same time. This is because the
FFs are not triggered simultaneously. The clock does not directly control the time at which every stage changes state. An asynchronous counter uses T FFs to perform a counting function. The actual hardware used in asynchronous is usually J-K flip-flops. Ripple counters are also called serial or series counters.

Synchronous counters are clocked such that each FF in the counter is triggered at the same time. This is done by connecting the clock line at each stage of the counter. Synchronous counters are faster than asynchronous counters, because the propagation delay involved is less.

A counter may be up-counter or a down counter. An up-counter is a counter which counts in the upward direction i.e. 0,1,2,3,…N. A down-counter is a counter which counts in the downward direction, i.e N,N-1,N-2,N-3,…….,1,0. Each of the count of the counter is called as the state of the counter. The number of states through which the counter passes before returning to the starting state is called the modulus of the counter. Hence the modulus of a counter is equal to the total number of distinct states (counts) including zero that a counter can store. In general an n-bit counter will have n FFs and $2^n$ states, and divides the input frequency by $2^n$, hence; it is divide-by-$2^n$ counter or mod-$2^n$ counter.

A counter is shortened modulus which does not utilize all the possible states. In an asynchronous counter, the invalid states are bypassed by providing suitable feedback; where as in a synchronous counter, the invalid states are taken as don’t care excitation.
Fig.2.6.3 (a) (2 bit Asynchronous (Ripple) Counter)

Fig.2.6.3 (b) (3 bit Asynchronous (Ripple) Counter)
Fig.2.6.3(c) (4-bit Asynchronous Counter)(Decade Counter)
Synchronous counters have the advantages of high speed and less severe decoding problems but the disadvantage of having more circuitry than that of asynchronous counters.

---

**Fig.2.6.3(d) (4 bit Synchronous Counter)**

Synchronous counters have the advantages of high speed and less severe decoding problems but the disadvantage of having more circuitry than that of asynchronous counters.
2.7 Other Silicon Devices and Technology

2.7.1 Transistors

A transistor consists of two pn junctions formed by sandwiching either p-type or n-type semiconductor between a pair of opposite types. There are two types of transistors namely (i) n-p-n transistor and (ii) p-n-p transistor. A transistor may be regarded as a combination of two diodes connected back to back. The three sections used in both types of transistors are generally represented both side parts as emitter and collector and the middle one as base. These devices are generally bipolar as both p-type and n-type play part in conduction [23].

2.7.2 FET (Field Effect Transistor)

A FET is generally represented by two sides as Source and Drain. The middle part represented as Gate. Unlike the ordinary transistor, these transistors carry charge as either holes or electrons and regarded as unipolar transistor. These device gives high impedance in compare to ordinary transistors. FET is essential regarded as voltage-drive source. There is also no junction for FET. The high input impedance and low output impedance and low noise level make DET far superior to the bipolar transistor. These are two types (i) JFET (Junction Field Effect Transistor) used in linear circuits and (ii) MOS (Metal Oxide Semiconductor) used in digital circuits [47].

2.7.3 MOS (Metal Oxide Semiconductor)

A basic MOS consisting of three layers. The top layer is a conductive metal electrode, the middle layer is an insulator of glass or silicon dioxide, and the bottom layer is another conductive electrode made out of crystal silicon. This layer is a semiconductor whose conductivity changes with either doping or temperature [48].
2.7.4 MOSFET

The metal–oxide–semiconductor field-effect transistor (MOSFET, MOS-FET, or MOSFET) is a device used for amplifying or switching electronic signals. In MOSFETs, a voltage on the oxide-insulated gate electrode can induce a conducting channel between the two other contacts called source and drain. The channel can be of n-type or p-type, and is accordingly called an nMOSFET or a pMOSFET (also commonly nMOS, pMOS). It is by far the most common transistor in both digital and analog circuits, though the bipolar junction transistor was at one time much more common[48,49].

2.7.5 CMOS (Complementary Metal Oxide Semiconductor)

Complementary metal-oxide-semiconductor (CMOS) is a technology for constructing integrated circuits. CMOS technology is used in microprocessors, microcontrollers, static RAM, and other digital logic circuits. CMOS technology is also used for several analog circuits such as image sensors, data converters, and highly integrated transceivers for many types of communication. CMOS is also sometimes referred to as complementary-symmetry metal-oxide-semiconductor or COS-MOS. The words "complementary-symmetry" refer to the fact that the typical digital design style with CMOS uses complementary and symmetrical pairs of p-type and n-type metal oxide semiconductor field effect transistors (MOSFETs) for logic functions[47,50]. These devices are high noise immunity and low static power consumption. Significant power is only drawn while the transistors in the CMOS device are switching between on and off states. The phrase "metal-oxide-semiconductor" is a reference to the physical structure of certain field-effect transistors, having a metal gate electrode placed on top of an oxide insulator, which in turn is on top of a semiconductor material [46]. The CMOS is being used in digital circuitry which consists of only logic levels i.e. ‘0’ and ‘1’.

2.7.6 SCR (Silicon Controlled Rectifiers)

The silicon controlled rectifier is a three terminal semiconductor switching device and most important device after Diode and Transistor. SCR can be used for controlling rectification, inversion and regulation of power flow. It appears in market in various forms such as Thyristor and thyrode. This basically consists of Anode, Cathode and a Gate. This basically an ordinary rectifier (pn) and a junction (nnpn) combined in one unit to form pnppn device[46].
2.7.7 IC (Integrated Circuit)

An integrated circuit has various components such as resistors, capacitors, diodes, transistors etc. fabricated on a small semiconductor chip and components can not removed. The size of ICs are so small that, typically having the dimensions of 0.2 mm*0.2 mm*0.001 mm. Modern digital systems utilize integrated circuits in their design. The basic types of constructions are employed in the manufacture of integrated circuit, namely (i) mono-lithic (ii) thin-film (iii) thick-film and (iv) hybrid. Monolithic ICs are by far the most common type used in practice. A monolithic IC is an electronic circuit, that is constructed entirely on single piece of semiconductor material (usually silicon) called substrate, which is commonly referred to as a chip. ICs have the advantage of low cost, low power, smaller size and high reliability over discrete circuitry. ICs may be classified as analog (linear) and digital.

The various level of digital ICs are [46]

SSI (Small Scale Integration): The least complex digital ICs with less than 12 gate circuits on a single chip. Logic gates and flip-flops belong to this category.

MSI (medium Scale Integration): With 12 to 99 gate circuits on a single chip, the more complex logic circuits such as encoders, decoders, counters, registers, multiplexers, arithmetic circuits, etc. belong to this category.

LSI (Large Scale Integration): With 100 to 9999 gate circuits on a single chip, small memories and small microprocessors fall in this category.

VLSI (Very large Scale Integration): ICs with complexities ranging from 10,000 to 99,999 gate circuits per chip fall in this category. Large memories and large microprocessor systems etc. come in this category.

ULSI (Ultra large Scale Integration): With complexities of over 100000 gate circuits per chip, very large memories and microprocessor systems and single-chip computers come in this category.
2.8 Current Scenario and Challenges

The computer industry has been going through a period of fundamental change. Computer manufacturers have been relied on the manifestation of Moore's Law which predicts the reliable tendency for the number of transistors on a chip to double every 18 to 24 months [3, 53]. Making doubling means the transistors can be made smaller to smaller and hence the chips become faster and help in getting faster computers. But industry experts say the trend defined by Moore's Law will start to slow[3]. This has now brought semiconductor industry a fundamental limit. While the search continues for novel structures and materials to extend conventional technology as long as possible, other techniques may play a greater role in continuing the improvements in computer performance. Present-day logic and memory chips are based primarily on CMOS (complementary metal-oxide semiconductor) transistor technology. According to the latest CMOS technology "roadmap," published in November 1999 by the Semiconductor Industry Association, device designers started to begin using a 1.5 nanometer insulator as early as 2001, although it has not reached still [3]. The channel length is defined as the distance between the source and drain, the two points between which a current flows during the operation of a transistor. The shorter the distance, the less time the transistor takes to switch, but various kinds of so-called short-channel effects suggest that the channel length cannot be much shorter than 25 nanometers, a size that could be attained by 2010 at the current rate of scaling[3,51,53]. The power-supply voltage level is perhaps an even more critical value. "The electronic nature of silicon requires a minimum level of about 1 volt. Appreciably less than that, and the transistor cannot produce enough current to switch on and off properly. At present, the voltage level is between 1.2 and 1.5 volts, which leaves room for about one more round of reduction. These limits mean that future transistors will not be able to be made smaller and faster according to the time-tested scaling laws [51, 52].

The questions that have been triggering researchers and manufacturers is that whether computer performance continues to improve rapidly even without the help of faster transistors or entirely new technologies that can replace or supplant to silicon transistors. The first approach that could come up, closest to existing circuits, is based on novel materials and devices [51, 52, 53].
Summary

This chapter overviews the basic silicon computer design and the various associate devices needed for it e.g. input, memory unit, control unit, arithmetic & logic unit and output. The functionality of a computer depend on the microprocessor made from various logic gates. The various semiconductor materials described which help in design these semiconductor logic gates available as n-type (excess electrons) and p-type (excess holes). The basic PN junction described to create space charge or depletion region to study the resistance to current flow depending on the polarity of applied voltage. The semiconductor diode and its characteristics also described, as diode is the primary element taken for designing the various logic gates. The various Boolean gates AND, NOT, NAND, OR, NOR, XOR and XNOR described made from rectifying diode. The diode designed from silicon materials described and underlying atomic principles of charge carriers and energy barriers have described. The tunneling diode described which creates a resonant field of potential or junction. The various quantitative theories of charge density and transformation along with quantum mechanical application have described. The chapter also describes the memory and computational devices by specifying flip-flop, registers and counters. Last part of the chapter described various other devices pertaining to semiconductor scenario e.g transistor, Integrated circuits etc. The current industry standard of semiconductor minimaturation and future challenges have discussed.

The basic idea of this chapter is to verify the designing of logic gates from semiconductor materials and electron flow properties. The other approach is to study designing principles of rectifying diodes and tunneling didoes. The chapter also aim at other aspects to study challenges and limitation of semiconductor devices, whether molecule based computer is a solution for it and can we design the molecular gates in same underlying principle of silicon based device!