



SRI SRI SRI MOOKAMBIKA
EDUCATIONAL SOCIETY

PRODDATUR



SRI SRI SRI MOOKAMBIKA
EDUCATIONAL SOCIETY

VAAGDEVI INSTITUTE OF TECHNOLOGY AND SCIENCE
Peddasettipilli (V), Proddatur-516360

Year/Semester/Branch: I B.Tech II SEM (ECE)

*Name of the Subject:
Electronic Devices & Circuits
(20A04302T)*

**Teaching Notes, Short Answer Questions, Multiple Choice
Questions, Long Answer Question**

**Prepared by,
S. Siddeswara Reddy, B.Tech, M.Tech.,(Ph.D),
Assitant Professor in ECE & HOD,
Dept. of ECE, VITS, Proddatur.**

ELECTRONIC DEVICES & CIRCUITS AND DESIGN (R20)

TEACHING NOTES

INDEX

UNIT NUMBER	Name of the Unit	Page No.
1	Semiconductors-Diodes	
2	Zener Diodes	
3	BJT'S	
4	MOSFET's (MOSField Effect Transistors)	
5	MOSFET Small Signal Operation Models	

(20A04101T) ELECTRONIC DEVICES & CIRCUITS

(Common to EEE and ECE)

Course Objectives:

- To understand the basic principles of all semiconductor devices.
- To be able to solve problems related to diode circuits, and amplifier circuits.
- To analyze diode circuits, various biasing and small signal equivalent circuits of amplifiers.
- To be able to compare the performance of BJTs and MOSFETs
- To design rectifier circuits and various amplifier circuits using BJTs and MOSFETs.

Unit – 1

Review of Semiconductors: Intrinsic semiconductors, Doped Semiconductors, Current Flow in Semiconductors, PN Junction with Open Circuit, PN Junction with Applied Voltage, Capacitive Effects in PN Junction.

Diodes: Introduction, The Ideal Diode – current voltage characteristic, rectifier, diode logic gates, Terminal Characteristics of Junction Diodes– forward bias, reverse bias, and breakdown regions, Modeling the Diode Forward Characteristics- exponential model, graphical analysis and Iterative analysis using the exponential model, constant voltage drop model, the small signal model.

Learning outcomes:

- Remember and understand the basic characteristics of semiconductor diode (L1)
- Understand iterative and graphical analysis of simple diode circuits (L1)

Unit – 2

Zener Diodes– Zenerdiode Characteristics, Voltage shunt regulator, Temperature Effects, Rectifier Circuits– half-wave, full-wave and bridge rectifier circuits, rectifier with a filter capacitor, C-L-C filter, Clipping and Clamping Circuits– limiter circuit, the clamped capacitor, voltage doubler, Special Diode Types– UJT, Schottkybarrier diode, Varactor diode, photo diode, light emitting diode(LED), Problem Solving.

Bipolar Junction Transistors(BJTs): Physical Operation - simplified structure and modes of operation, Operation of the npn, and pnp transistors: cutoff, active, and saturation modes, V-ICharacteristics- of different configurations - graphical representation of transistor characteristics, dependence of collector current on collector voltage, the Early Effect.

Learning outcomes:

- Understand principle of operation of Zener diode and other special semiconductor diodes (L1)
- Understand the V-I characteristics of BJT and its different configurations (L1)
- Analyze various applications of diode and special purpose diodes (L3)
- Design rectifier and voltage regulator circuits (L4)

Unit- 3

BJT circuits at DC, Applying the BJT in Amplifier Design- Voltage Amplifier, Voltage Transfer Characteristic (VTC), Small-Signal Voltage Gain, determining the VTC by Graphical Analysis, Qpoint, Small-signal operation and models- the transconductance, input resistance at the base, input resistance at the emitter, Voltage gain, separating the Signal and the DC Quantities, The Hybrid- π Model, the T Model, Basic BJT Amplifier Configurations - Common-Emitter (CE) amplifier without and with emitter resistance, Common-Base (CB) amplifier, Common-Collector (CC) amplifier or Emitter Follower, Biasing in BJT Amplifier Circuits- Fixed bias, Self bias, voltage divider bias circuits, biasing using a Constant-Current Source, CE amplifier – Small signal analysis and design, Transistor breakdown and Temperature Effects, Problem solving.

Learning outcomes:

- Solve problems on various biasing circuits using BJT (L2)
- Analyze BJT based biasing circuits (L3)
- Design an amplifier using BJT based on the given specifications (L4)

Unit – 4

MOS Field-Effect Transistors (MOSFETs): Introduction, Device Structure and Physical Operation – device structure, operation with zero gate voltage, creating a channel for current flow, operation for different drain to source voltages, the P-channel MOSFET, CMOS, V-I characteristics – i_D - v_{DS} characteristics, i_D - v_{GS} characteristics, finite output resistance in saturation, characteristics of the p- Channel MOSFET, MOSFET Circuits at DC, Applying the MOSFET in Amplifier Design – voltage transfer characteristics, biasing the MOSFET to obtain linear amplification, the small signal voltage gain, graphical analysis, the Q-point. Problem solving.

Learning outcomes:

- Understand principle of operation of various types of MOSFET devices (L1)
- Understand the V-I characteristics of MOSFET devices and their configurations (L1)

Unit – 5

MOSFET Small Signal Operation Models- the dc bias, separating the DC analysis and the signal analysis, Small signal equivalent circuit models, the trans conductance, the T equivalent circuit model, Basic MOSFET Amplifier Configurations- three basic configurations, characterizing amplifiers, common source(CS) amplifier without and with

source resistance, common gate (CG) amplifier, source follower, the amplifier frequency response,Biasing in MOSFET Amplifier Circuits– biasing by fixing VGS with and without source resistance, biasing using drain to gate feedback resistor, biasing using constant current source, Common Source Amplifier using MOSFETs – Small signal analysis and design, Body Effect, Problem Solving.

Learning outcomes:

- Solve problems on small signal equivalent of MOSFET devices (L2)
- Analyze various biasing circuits based on different types of MOSFETs (L3)
- Design an amplifier using BJT based on the given specifications (L4)

UNIT – I

SEMICONDUCTORS & DIODES

UNIT-1

1.1 Intrinsic Semiconductors

As their name implies, semiconductors are materials whose conductivity lies between that of

conductors, such as copper, and insulators, such as glass. There are two kinds of semiconductors: single-element semiconductors, such as germanium and silicon, which are in group IV in the periodic table; and compound semiconductors, such as gallium-arsenide, which are formed by combining elements from groups III and V or groups II and VI. Compound semiconductors are useful in special electronic circuit applications as well as in applications that involve light, such as light-emitting diodes (LEDs). Of the two elemental semiconductors, germanium was used in the fabrication of very early transistors (late 1940s, early 1950s). It was quickly supplanted, however, with silicon, on which today's integrated-circuit technology is almost entirely based. For this reason, we will deal mostly with silicon devices throughout this book.¹ A silicon atom has four valence electrons, and thus it requires another four to complete its outermost shell. This is achieved by sharing one of its valence electrons with each of its four neighboring atoms. Each pair of shared electrons forms a **covalent bond**. The result is that a crystal of pure or intrinsic silicon has a regular lattice structure, where the atoms are held in their position by the covalent bonds. Figure 1.1 shows a two-dimensional representation of such a structure.

At sufficiently low temperatures, approaching absolute zero (0 K), all the covalent bonds are intact and no electrons are available to conduct electric current. Thus, at such low temperatures, the intrinsic silicon crystal behaves as an insulator.

At room temperature, sufficient thermal energy exists to break some of the covalent bonds, a process known as thermal generation. As shown in Fig. 1.2, when a covalent bond is broken, an electron is freed. The **free electron** can wander away from its parent atom, and it becomes available to conduct electric current if an electric field is applied to the crystal. As the electron leaves its parent atom, it leaves behind a net positive charge, equal to the magnitude of the electron charge. Thus, an electron from a neighboring atom may be attracted to this positive charge, and leaves its parent atom. This action fills up the “hole” that existed in the ionized atom but creates a new hole in the other atom. This process may repeat itself, with the result that we effectively have a positively charged carrier, or **hole**, moving through the silicon crystal structure and being available to conduct electric current.

The charge of a hole is equal in magnitude to the charge of an electron. We can thus see that as temperature increases, more covalent bonds are broken and electron–hole pairs are generated. The increase in the numbers of free electrons and holes results in an increase in the conductivity of silicon.

An exception is the subject of gallium arsenide (GaAs) circuits, which though not covered in this edition

of the book, is studied in some detail in material provided on the text website.

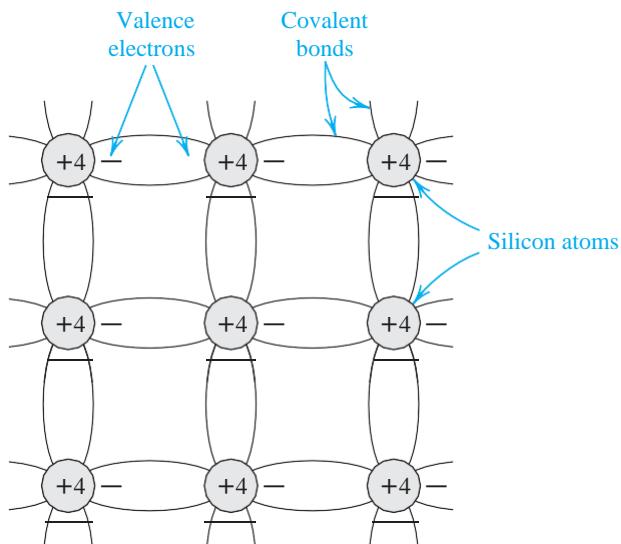


Figure 1.1 Two-dimensional representation of the silicon crystal. The circles represent the inner core of silicon atoms, with +4 indicating its positive charge of $+4q$, which is neutralized by the charge of the four valence electrons. Observe how the covalent bonds are formed by sharing of the valence electrons. At 0 K, all bonds are intact and no free electrons are available for current conduction.

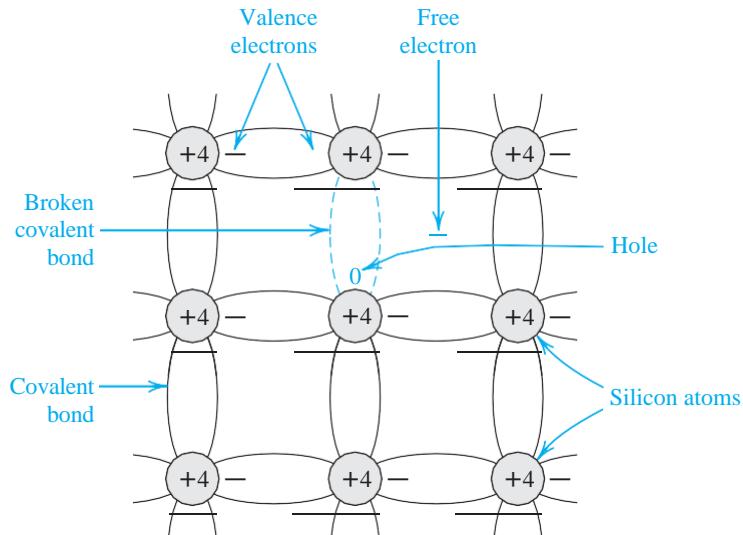


Figure 1.2 At room temperature, some of the covalent bonds are broken by thermal generation. Each broken bond gives rise to a free electron and a hole, both of which become available for current conduction.

Thermal generation results in free electrons and holes in equal numbers and hence equal concentrations, where concentration refers to the number of charge carriers per unit volume (cm^3). The free electrons and holes move randomly through the silicon crystal structure, and in the process some electrons may fill some of the holes. This process, called **recombination**, results in the disappearance of free electrons and holes. The recombination rate is proportional to the number of free electrons and holes, which in turn is determined by the thermal **generation** rate. The latter is a strong function of temperature. In thermal equilibrium, the recombination rate is equal to the generation rate, and one can conclude that the concentration of free electrons n is equal to the concentration of holes p ,

$$n = p = n_i \quad (1.1)$$

where n_i denotes the number of free electrons and holes in a unit volume (cm^3) of intrinsic silicon at a given temperature. Results from semiconductor physics gives n_i as

$$n_i = BT^{3/2}e^{-E_g/2kT} \quad (1.2)$$

Example 1.1

Calculate the value of n_i for silicon at room temperature ($T = 300 \text{ K}$).

Solution

Substituting the values given above in Eq. (3.2) provides

$$\begin{aligned} n_i &= 7.3 \times 10^{15} (300)^{3/2} e^{-1.12/2(8.62 \times 10^{-5} \times 300)} \\ &= 1.5 \times 10^{10} \text{ carriers/cm}^3 \end{aligned}$$

Although this number seems large, to place it into context note that silicon has 5×10^{22} atoms/ cm^3 . Thus at room temperature only one in about 5×10^{12} atoms is ionized and contributing a free electron and a hole!

where B is a material-dependent parameter that is $7.3 \times 10^{15} \text{ cm}^{-3}\text{K}^{-3/2}$ for silicon; T is the temperature in K ; E_g , a parameter known as the **bandgap energy**, is 1.12 electron volt (eV) for silicon²; and k is Boltzmann's constant (8.62×10^{-5} eV/K). It is interesting to know that the bandgap energy E_g is the minimum energy required to break a covalent bond and thus generate an electron-hole pair.

Finally, it is useful for future purposes to express the product of the hole and free-electron concentration as

$$pn = n_i^2 \quad (1.3)$$

where for silicon at room temperature, $n_i \times 1.5 \times 10^{10} / \text{cm}^3$. As¹⁰ will³ be seen shortly, this relationship extends to extrinsic or doped silicon as well.

Note that $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$.

LCDs, THE FACE OF ELECTRONICS:

The existence of liquid crystals whose color could be changed by means of an external heat source was first reported in 1888 by an Austrian botanical physiologist. The LC idea lay dormant until the late 1940s, however. Subsequent developments in the field of solid-state electronics provided the technology to harness the technique in display media, with the first LCDs being demonstrated by RCA beginning in 1962. Today, LCDs are an essential component in every mobile device as the interface to the world of electronics within. At the other end of the scale, large LCDs are used in flat-panel TVs, and very large LCDs are appearing as “dynamic” wallpaper in museum display settings.

1.2 Doped Semiconductors

The intrinsic silicon crystal described above has equal concentrations of free electrons and holes, generated by thermal generation. These concentrations are far too small for silicon to conduct appreciable current at room temperature. Also, the carrier concentrations and hence the conductivity are strong functions of temperature, not a desirable property in an electronic device. Fortunately, a method was developed to change the carrier concentration in a semiconductor crystal substantially and in a precisely controlled manner. This process is known as doping, and the resulting silicon is referred to as **doped silicon**.

Doping involves introducing impurity atoms into the silicon crystal in sufficient numbers to substantially increase the concentration of either free electrons or holes but with little or no change in the crystal properties of silicon. To increase the concentration of free electrons, n , silicon is doped with an element with a valence of 5, such as phosphorus. The resulting doped silicon is then said to be of ***n* type**. To increase the concentration of holes, p , silicon is doped with an element having a valence of 3, such as boron, and the resulting doped silicon is said to be of ***p* type**.

Figure 1.3 shows a silicon crystal doped with phosphorus impurity. The dopant (phosphorus) atoms replace some of the silicon atoms in the crystal structure. Since the phosphorus atom has five electrons in its outer shell, four of these electrons form covalent bonds with the neighboring atoms, and the fifth electron becomes a free electron. Thus each phosphorus atom *donates* a free electron to the silicon crystal, and the phosphorus impurity is called a **donor**. It should be clear, though, that no holes are generated by this process. The net positive charge associated with the phosphorus atom is a **bound charge** that does not move through the crystal.

If the concentration of donor atoms is N_D , where N_D is usually much greater than n_i , the concentration of free electrons in the n -type silicon will be $n_n \times N_D$

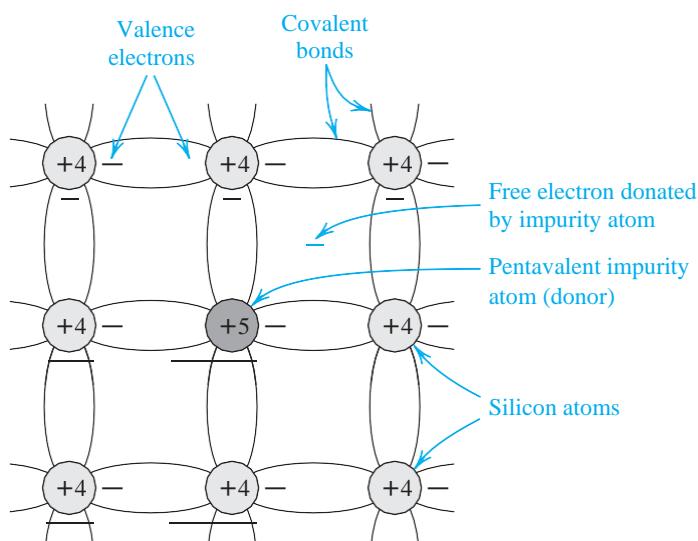


Figure 1.3 A silicon crystal doped by a pentavalent element. Each dopant atom donates a free electron and is thus called a donor. The doped semiconductor becomes n type.

where the subscript n denotes n -type silicon. Thus n_n is determined by the doping concentration and not by temperature. This is not the case, however, for the hole concentration. All the holes in the n -type silicon are those generated by thermal ionization. Their concentration p_n can be found by noting that the relationship in Eq. (1.3) applies equally well for doped silicon, provided thermal equilibrium is achieved. Thus for n -type silicon

$$p_n n_n = n_i$$

Substituting for n_n from Eq. (3.4), we obtain for p_n

Thus p_n will have the same dependence on temperature as that of n^2 . Finally, we note that in n -type silicon the concentration of free electrons n_n will be much larger than that of holes. Hence electrons are said to be the **majority** charge carriers and holes the **minority** charge carriers in n -type silicon.

To obtain p -type silicon in which holes are the majority charge carriers, a trivalent impurity such as boron is used. Figure 3.4 shows a silicon crystal doped with boron. Note that the boron atoms replace some of the silicon atoms in the silicon crystal structure. Since each boron atom has three electrons in its outer shell, it **accepts** an electron from a neighboring atom, thus forming covalent bonds. The result is a hole in the neighboring atom and a bound negative charge at the **acceptor** (boron) atom. It follows that each acceptor atom provides a hole. If the acceptor doping concentration is N_A , where $N_A n_i$, the hole concentration becomes

$$p_p \times N_A \quad (1.6)$$

where the subscript p denotes p -type silicon. Thus, here the majority carriers are holes and their concentration is determined by N_A . The concentration of minority electrons can be found

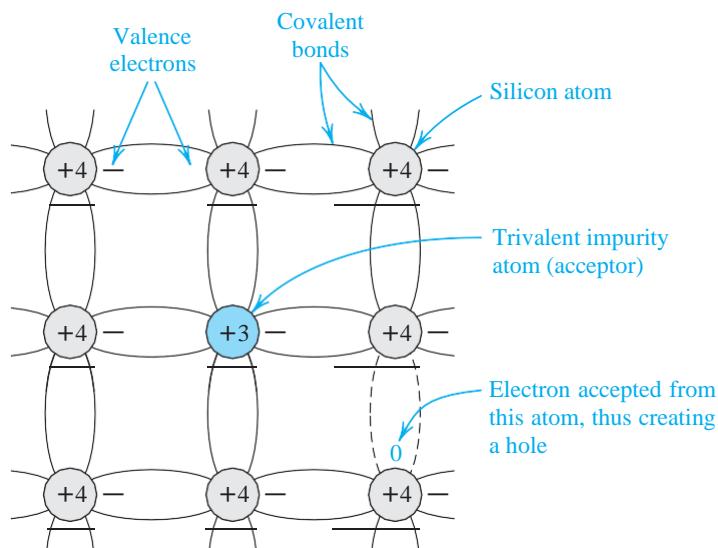


Figure 1.4 A silicon crystal doped with boron, a trivalent impurity. Each dopant atom gives rise to a hole, and the semiconductor becomes p type.

by using the relationship

$$p_p n_p = n_i$$

and substituting for p_p from Eq. (1.6),

$$n_p \times \frac{n_i^2}{N_A} \quad (1.7)$$

Thus, the concentration of the minority electrons will have the same temperature dependence as that of n^2 .

It should be emphasized that a piece of n -type or p -type silicon is electrically neutral; the charge of the majority free carriers (electrons in the n -type and holes in the p -type silicon) are neutralized by the bound charges associated with the impurity atoms.

Example 1.2

Consider an n -type silicon for which the dopant concentration $N = 10^{17}/\text{cm}^3$. Find the electron and hole concentrations at $T = 300 \text{ K}$.

Solution

The concentration of the majority electrons is

$$n_n \times N = 10^{17} / \text{cm}^3$$

Example 1.2 continued

The concentration of the minority holes is

$$p_n \times \frac{n_i^2}{N_D}$$

In Example 3.1 we found that at $T = 300 \text{ K}$, $n_i = 1.5 \times 10^{10}/\text{cm}^3$. Thus,

$$\begin{aligned} p_n &= \frac{1.5 \times 10^{10}}{10^{17}}^2 \\ &= 2.25 \times 10^3 / \text{cm}^3 \end{aligned}$$

Observe that $n_n \gg n_i$ and that n_n is vastly higher than p_n .

1.2 Current Flow in Semiconductors

There are two distinctly different mechanisms for the movement of charge carriers and hence for current flow in semiconductors: drift and diffusion.

1.2.1 Drift Current

When an electrical field E is established in a semiconductor crystal, holes are accelerated in the direction of E , and free electrons are accelerated in the direction opposite to that of E . This situation is illustrated in Fig. 1.5. The holes acquire a velocity v_p -drift given by

$$v_{p\text{-drift}} = \mu_p E \quad (1.8)$$

where μ_p is a constant called the **hole mobility**: It represents the degree of ease by which holes move through the silicon crystal in response to the electrical field E . Since velocity has the units of centimeters per second and E has the units of volts per centimeter, we see from Eq. (3.8) that the mobility μ_p must have the units of centimeters squared per volt-second ($\text{cm}^2/\text{V} \cdot \text{s}$). For intrinsic silicon $\mu_p = 480 \text{ cm}^2/\text{V} \cdot \text{s}$.

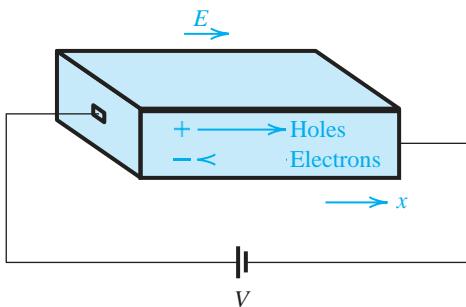


Figure 1.5 An electric field E established in a bar of silicon causes the holes to drift in the direction of E and the free electrons to drift in the opposite direction. Both the hole and electron drift currents are in the direction of E .

The free electrons acquire a drift velocity v_n -drift given by

$$v_{n\text{-drift}} = -\mu_n E \quad (1.9)$$

where the result is negative because the electrons move in the direction opposite to E . Here μ_n is the **electron mobility**, which for intrinsic silicon is about $1350 \text{ cm}^2/\text{V} \cdot \text{s}$. Note that μ_n is about 2.5 times μ_p , signifying that electrons move with much greater ease through the silicon crystal than do holes.

Let's now return to the single-crystal silicon bar shown in Fig. 1.5. Let the concentration of holes be p and that of free electrons n . We wish to calculate the current component due to the flow of holes. Consider a plane perpendicular to the x direction. In one second, the hole charge that crosses that plane will be $(Aqp\mathbf{v}_{p\text{-drift}})$ coulombs, where A is the cross-sectional area of the silicon bar and q is the magnitude of electron charge. This then must be the hole component of the drift current flowing through the bar,

$$I_p = Aqp\mathbf{v}_{p\text{-drift}} \quad (1.10)$$

Substituting for $\mathbf{v}_{p\text{-drift}}$ from Eq. (1.8), we obtain

$$I_p = Aqp\mu_p E$$

We are usually interested in the current density J_p , which is the current per unit cross-sectional area, the current component due to the drift of free electrons can be found in a similar manner. Note, however, that electrons drifting from right to left result in a current component from left to right. This is because of the convention of taking the direction of current flow as the direction of flow of positive charge and opposite to the direction of flow of negative charge. Thus,

$$I_n = -Aqn\mathbf{v}_{n\text{-drift}}$$

Substituting for $\mathbf{v}_{n\text{-drift}}$ from Eq. (1.9), we obtain the current density $J_n = I_n/A$ as

$$J_n = qn\mu_n E \quad (1.12)$$

The total drift current density can now be found by summing J_p and J_n from Eqs. (1.11) and (1.12),

$$J = J_p + J_n = q p\mu_p + n\mu_n E \quad (1.13)$$

This relationship can be written as

$$J = \sigma E \quad (1.14)$$

Observe that Eq. (1.15) is a form of Ohm's law and can be written alternately as

$$\rho = \frac{E}{J} \quad (1.15)$$

Thus the units of ρ are obtained from: $\frac{V/cm}{A/cm^2} \Delta \cdot cm$.

Example 1.3

Find the resistivity of (a) intrinsic silicon and (b) p -type silicon with $N = 10^{16}/cm^3$. Use $n = 1.5 \times 10^{10}/cm^3$, and assume that for intrinsic silicon $\mu_n = 1350 \text{ cm}^2/V \cdot s$ and $\mu_p = 480 \text{ cm}^2/V \cdot s$, and for the doped silicon $\mu_n = 1110 \text{ cm}^2/V \cdot s$ and $\mu_p = 400 \text{ cm}^2/V \cdot s$. (Note that doping results in reduced carrier mobilities.)

Solution

(a) For intrinsic silicon,

$$p = n = \eta = 1.5 \times 10^{10}/cm^3$$

Thus,

$$\begin{aligned}\rho &= \frac{1}{q p \mu_p + n \mu_n} \\ \rho &= \frac{1}{1.6 \times 10^{-19} \cdot 1.5 \times 10^{10} \times 480 + 1.5 \times 10^{10} \times 1350} \\ &= 2.28 \times 10^5 \Delta \cdot cm\end{aligned}$$

(b) For the p -type silicon

$$\begin{aligned}p_p \times N &= 10^{16}/cm^3 \\ n_p \times \frac{n^2}{N_A} &= \frac{1.5 \times 10^{10}}{10^{16}} = 2.25 \times 10^4/cm^3\end{aligned}$$

Thus,

$$\begin{aligned}\rho &= \frac{1}{q p \mu_p + n \mu_n} \\ &= \frac{1}{1.6 \times 10^{-19} \times 10^{16} \times 400 + 2.25 \times 10^4 \times 1110} \\ &\times \frac{1}{1.6 \times 10^{-19} \times 10^{16} \times 400} = 1.56 \Delta \cdot \text{cm}\end{aligned}$$

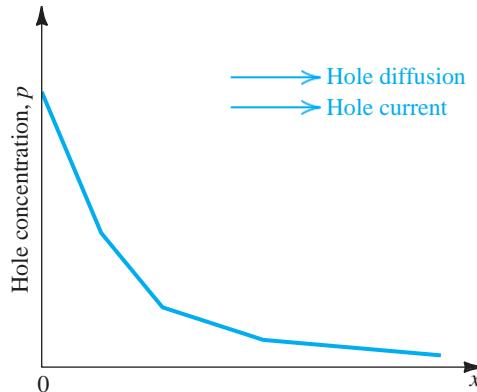
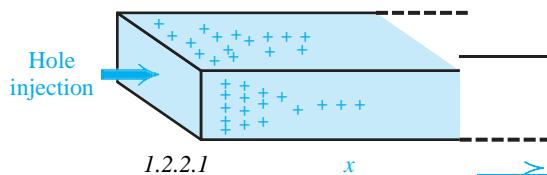
Observe that the resistivity of the *p*-type silicon is determined almost entirely by the doping concentration. Also observe that doping the silicon reduces its resistivity by a factor of about 10^4 , a truly remarkable change.

1.2.2 Diffusion Current

Carrier diffusion occurs when the density of charge carriers in a piece of semiconductor is not uniform. For instance, if by some mechanism the concentration of, say, holes, is made higher in one part of a piece of silicon than in another, then holes will diffuse from the region of high concentration to the region of low concentration. Such a diffusion process is like that observed if one drops a few ink drops in a water-filled tank. The diffusion of charge carriers gives rise to a net flow of charge, or **diffusion current**.

As an example, consider the bar of silicon shown in Fig. 1.6(a): By some unspecified process, we have arranged to inject holes into its left side. This continuous hole injection gives rise to and maintains a hole **concentration profile** such as that shown in Fig. 1.6(b). This profile in turn causes holes to diffuse from left to right along the silicon bar, resulting in a hole current in the *x* direction. The magnitude of the current at any point is proportional to the slope of the concentration profile, or the **concentration gradient**, at that point,

$$J_p = -q D_p \frac{dp(x)}{dx} \quad (1.19)$$



1.2.2.2

Figure 1.6 A bar of silicon (a) into which holes are injected, thus creating the hole concentration profile along the x axis, shown in (b). The holes diffuse in the positive direction of x and give rise to a hole diffusion current in the same direction. Note that we are not showing the circuit to which the silicon bar is connected.

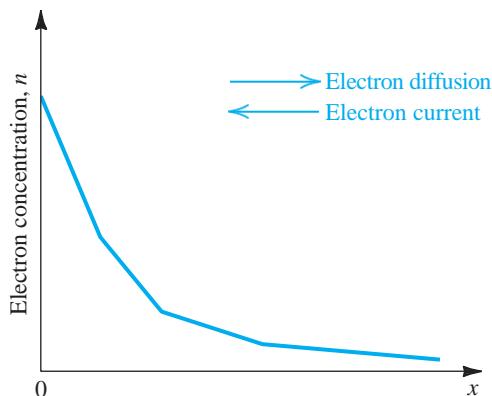


Figure 1.7 If the electron concentration profile shown is established in a bar of silicon, electrons diffuse in the x direction, giving rise to an electron diffusion current in the negative- x direction.

where J_p is the hole-current density (A/cm^2), q is the magnitude of electron charge, D_p is a constant called the **diffusion constant** or **diffusivity** of holes; and $p(x)$ is the hole concentration at point x . Note that the gradient (dp/dx) is negative, resulting in a positive current in the x direction, as should be expected.

In the case of electron diffusion resulting from an electron concentration gradient (see Fig. 1.7), a similar relationship applies, giving the electron-current density,

$$J_n = qD_n \frac{dn(x)}{dx} \quad (1.20)$$

where D_n is the diffusion constant or diffusivity of electrons. Observe that a negative (dn/dx) gives rise to a negative current, a result of the convention that the positive direction of current is taken to be that of the flow of positive charge (and opposite to that of the flow of negative).

For holes and electrons diffusing in intrinsic silicon, typical values for the diffusion constants are $D_p = 12 \text{ cm}^2/\text{s}$ and $D_n = 35 \text{ cm}^2/\text{s}$.

At this point the reader is probably wondering where the diffusion current in the silicon bar in Fig. .6(a) goes. A good question, as we are not showing how the right-side end of the bar is connected to the rest of the circuit. We will address this and related questions in detail in our discussion of the *pn* junction in later sections.

Example 1.4

Consider a bar of silicon in which a hole concentration profile described by

$$p(x) = p_0 e^{-x/L_p}$$

is established. Find the hole-current density at $x = 0$. Let $p_0 = 10^{16} \text{ cm}^{-3}$, $L_p = 1 \mu\text{m}$, and $D_p = 12 \text{ cm}^2/\text{s}$. If the cross-sectional area of the bar is $100 \mu\text{m}^2$, find the current I_p .

Solution

$$\begin{aligned} J_p &= -qD_p \frac{dp(x)}{dx} \\ &= -qD_p \frac{d}{dx} p_0 e^{-x/L_p} \\ &= q \frac{D_p}{L_p} p_0 e^{-x/L_p} \end{aligned}$$

Thus,

$$\begin{aligned} J_p(0) &= q \frac{D_p}{L_p} p_0 \\ &= 1.6 \times 10^{-19} \times \frac{12}{1 \times 10^{-4}} \times 10^{16} \\ &= 192 \text{ A/cm}^2 \end{aligned}$$

The current I_p can be found from

$$\begin{aligned} I_p &= J_p \times A \\ &= 192 \times 100 \times 10^{-8} \\ &= 192 \mu\text{A} \end{aligned}$$

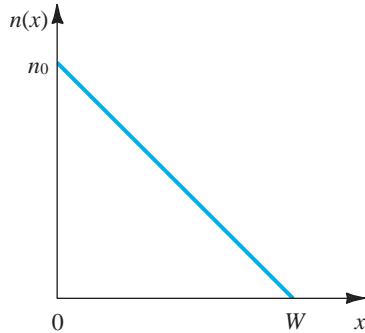


Figure E3.5

Ans. $56 \mu\text{A}/\mu\text{m}^2$; $18 \mu\text{m}^2$

1.2.3 Relationship between D and μ

A simple but powerful relationship ties the diffusion constant with the mobility,

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = V_T \quad (1.21)$$

where $V_T = kT/q$. The parameter V_T is known as the **thermal voltage**. At room temperature, $T = 300 \text{ K}$ and $V_T = 25.9 \text{ mV}$. We will encounter V_T repeatedly throughout this book. The relationship in Eq. (1.21) is known as the **Einstein relationship**.

1.3 The *pn* Junction with Applied voltage

Having learned important semiconductor concepts, we are now ready to consider our first practical semiconductor structure—the *pn* junction. As mentioned previously, the *pn* junction implements the diode (Chapter 4) and plays the dominant role in the structure and operation of the bipolar junction transistor (BJT, Chapter 6). As well, understanding *pn* junctions is very important to the study of the MOSFET operation (Chapter 5).

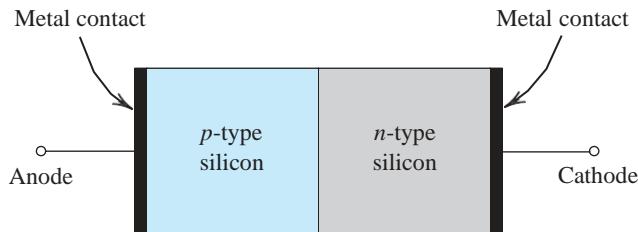


Figure 1.8 Simplified physical structure of the *pn* junction. (Actual geometries are given in Appendix A.) As the *pn* junction implements the junction diode, its terminals are labeled anode and cathode.

1.4.1 Physical Structure

Figure 1.8 shows a simplified physical structure of the *pn* junction. It consists of a *p*-type semiconductor (e.g., silicon) brought into close contact with an *n*-type semiconductor material (also silicon). In actual practice, both the *p* and *n* regions are part of the same silicon crystal; that is, the *pn* junction is formed within a single silicon crystal by creating regions of different dopings (*p* and *n* regions). Appendix A provides a description of the fabrication process of integrated circuits including *pn* junctions. As indicated in Fig. 3.8, external wire connections are made to the *p* and *n* regions through metal (aluminum) contacts. If the *pn* junction is used as a diode, these constitute the diode terminals and are therefore labeled “anode” and “cathode” in keeping with diode terminology.³

1.4.2 Operation with Open-Circuit Terminals

Figure 1.9 shows a *pn* junction under open-circuit conditions—that is, the external terminals are left open. The “+” signs in the *p*-type material denote the majority holes. The charge of these holes is neutralized by an equal amount of bound negative charge associated with the acceptor atoms. For simplicity, these bound charges are not shown in the diagram. Also not shown are the minority electrons generated in the *p*-type material by thermal ionization.

In the *n*-type material the majority electrons are indicated by “−” signs. Here also, the bound positive charge, which neutralizes the charge of the majority electrons, is not shown in order to keep the diagram simple. The *n*-type material also contains minority holes generated by thermal ionization but not shown in the diagram.

The Diffusion Current I_D Because the concentration of holes is high in the *p* region and low in the *n* region, holes diffuse across the junction from the *p* side to the *n* side. Similarly, electrons diffuse across the junction from the *n* side to the *p* side. These two current components add together to form the diffusion current I_D , whose direction is from the *p* side to the *n* side, as indicated in Fig. 1.9.

The Depletion Region The holes that diffuse across the junction into the n region quickly recombine with some of the majority electrons present there and thus disappear from the scene. This recombination process results also in the disappearance of some free electrons from This terminology in fact is a carryover from that used with vacuum-tube technology, which was the technology for making diodes and other electronic devices until the invention of the transistor in 1947. This event ushered in the era of solid-state electronics, which changed not only electronics, communications, and computers but indeed the world!

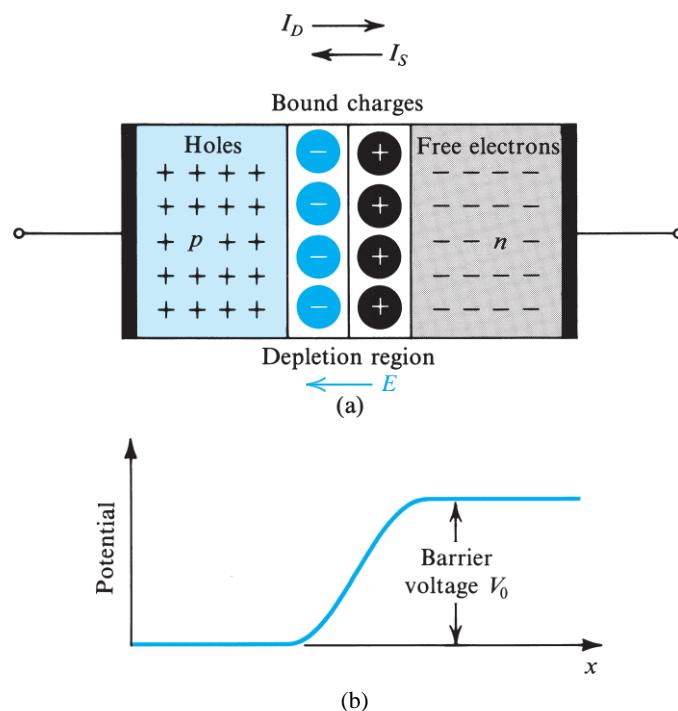


Figure 1.9 (a) The pn junction with no applied voltage (open-circuited terminals). (b) The potential distribution along an axis perpendicular to the junction.

n -type material. Thus some of the bound positive charge will no longer be neutralized by free electrons, and this charge is said to have been **uncovered**. Since recombination takes place close to the junction, there will be a region close to the junction that is *depleted of freeelectrons* and contains uncovered bound positive charge, as indicated in Fig. 1.9.

The electrons that diffuse across the junction into the p region quickly recombine with some of the majority holes there, and thus disappear from the scene. This results also in the disappearance of some majority holes, causing some of the bound negative charge to be uncovered (i.e., no longer

neutralized by holes). Thus, in the *p* material close to the junction, there will be a region *depleted of holes* and containing uncovered bound negative charge, as indicated in Fig. 1.9.

From the above it follows that a **carrier-depletion region** will exist on both sides of the junction, with the *n* side of this region positively charged and the *p* side negatively charged. This carrier-depletion region—or, simply, **depletion region**—is also called the **space-charge region**. The charges on both sides of the depletion region cause an electric field *E* to be established across the region in the direction indicated in Fig. 1.9. Hence a potential difference results across the depletion region, with the *n* side at a positive voltage relative to the *p* side, as shown in Fig. 1.9(b). Thus the resulting electric field opposes the diffusion of holes into the *n* region and electrons into the *p* region. In fact, the voltage drop across the depletion region acts as a **barrier** that has to be overcome for holes to diffuse into the *n* region and electrons to diffuse into the *p* region. The larger the barrier voltage, the smaller the number of carriers that will be able to overcome the barrier, and hence the lower the magnitude of diffusion current. Thus it is the appearance of the barrier voltage V_0 that limits the carrier diffusion process. It follows that the diffusion current I_D depends strongly on the voltage drop V_0 across the depletion region.

The Drift Current I_S and Equilibrium In addition to the current component I_D due to majority-carrier diffusion, a component due to minority-carrier drift exists across the junction. Specifically, some of the thermally generated holes in the *n* material move toward the junction and reach the edge of the depletion region. There, they experience the electric field in the depletion region, which sweeps them across that region into the *p* side. Similarly, some of the minority thermally generated electrons in the *p* material move to the edge of the depletion region and get swept by the electric field in the depletion region across that region into the *n* side. These two current components—electrons moved by drift from *p* to *n* and holes moved by drift from *n* to *p*—add together to form the drift current I_S , whose direction is from the *n* side to the *p* side of the junction, as indicated in Fig. 1.9. Since the current I_S is carried by thermally generated minority carriers, its value is strongly dependent on temperature; however, it is independent of the value of the depletion-layer voltage V_0 . This is due to the fact that the drift current is determined by the number of minority carriers that make it to the edge of the depletion region; any minority carriers that manage to get to the edge of the depletion region will be swept across by *E* irrespective of the value of *E* or, correspondingly, of V_0 .

Under open-circuit conditions (Fig. 1.9) no external current exists; thus the two opposite currents across the junction must be equal in magnitude:

$$I_D = I_S$$

This equilibrium condition⁴ is maintained by the barrier voltage V . Thus, if for some reason I_D exceeds I_S , then more bound charge will be uncovered on both sides of the junction, the depletion layer will widen, and the voltage across it (V_0) will increase. This in turn causes I_D to decrease until equilibrium is achieved with $I_D = I_S$. On the other hand, if I_S exceeds I_D , then the amount of uncovered charge will decrease, the depletion layer will narrow, and the voltage across it (V_0) will decrease. This causes I_D to increase until equilibrium is achieved with $I_D = I_S$.

The Junction Built-in Voltage With no external voltage applied, the barrier voltage V_0 across the pn junction can be shown to be given by⁵

$$V_0 = V \ln \frac{N_A N_D}{n_i^2} \quad (1.22)$$

where N_A and N_D are the doping concentrations of the p side and n side of the junction, respectively. Thus V_0 depends both on doping concentrations and on temperature. It is known as the **junction built-in voltage**. Typically, for silicon at room temperature, V_0 is in the range of 0.6 V to 0.9 V.

When the pn junction terminals are left open-circuited, the voltage measured between them will be zero. That is, the voltage V_0 across the depletion region *does not* appear between the junction terminals. This is because of the contact voltages existing at the metal–semiconductor junctions at the terminals, which counter and exactly balance the barrier voltage. If this were not the case, we would have been able to draw energy from the isolated pn junction, which would clearly violate the principle of conservation of energy.

Width of and Charge Stored in the Depletion Region Figure 1.10 provides further illustration of the situation that obtains in the pn junction when the junction is in equilibrium. In fact, in equilibrium the equality of drift and diffusion currents applies not just to the total currents but also to their individual components. That is, the hole drift current must equal the hole diffusion current and, similarly, the electron drift current must equal the electron diffusion current.

The derivation of this formula and of a number of others in this chapter can be found in textbooks dealing with devices, such as that by Streetman and Bannerjee (see the reading list in Appendix I).

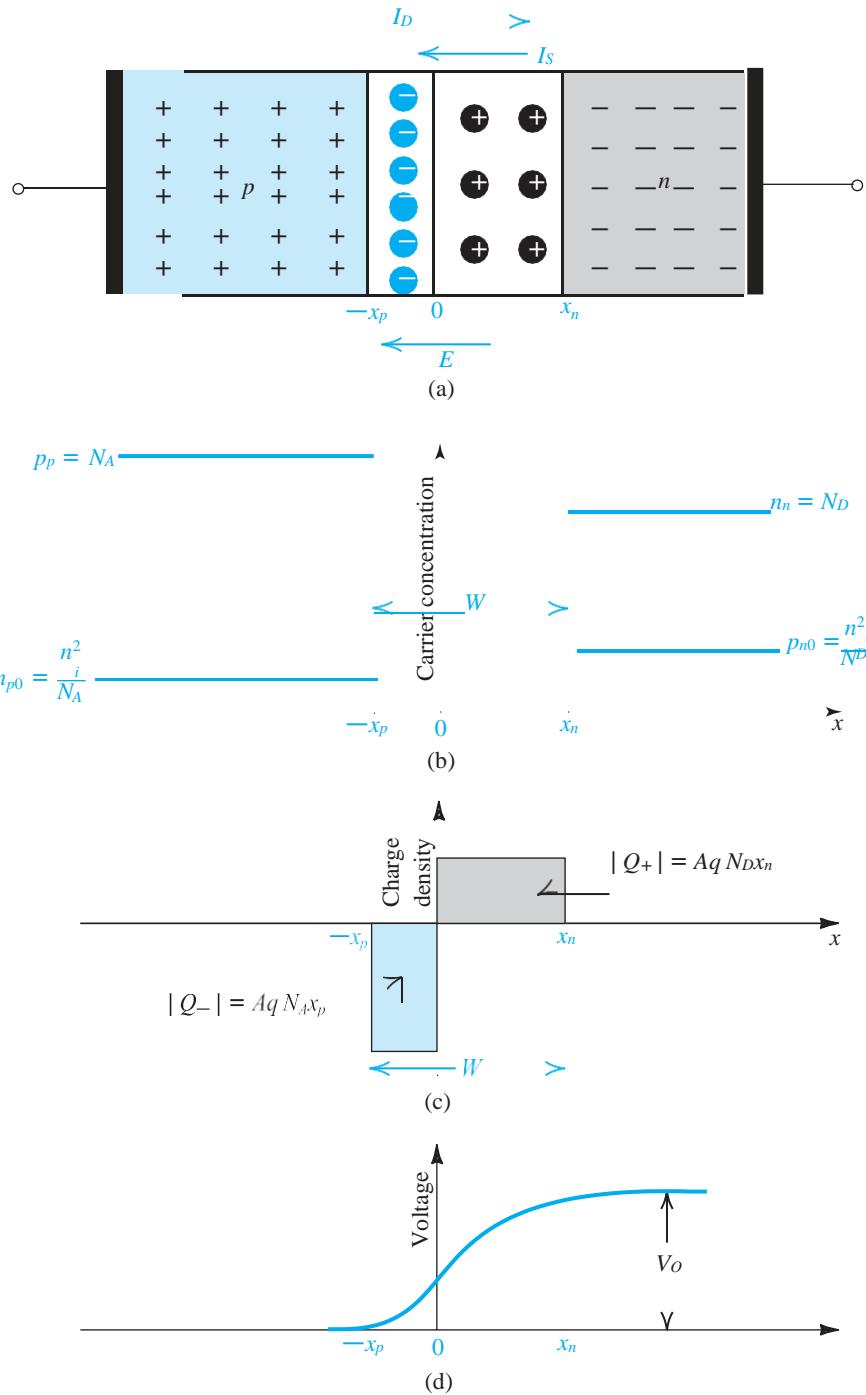


Figure 1.10 (a) A *pn* junction with the terminals open-circuited. (b) Carrier concentrations; note that $N_A > N_D$. (c) The charge stored in both sides of the depletion region; $Q_J = |Q_+| + |Q_-|$. (d) The built-in Voltage

In Fig. 1.10(a) we show a junction in which $N_A > N_D$, a typical situation in practice. This is borne out by the carrier concentration on both sides of the junction, as shown in Fig. 1.10(b). Note that we have denoted the minority-carrier concentrations in both sides by n_{p0} and p_{n0} , with the additional subscript “0” signifying equilibrium (i.e., before external voltages are applied, as will be seen in the next section). Observe that the depletion region extends in

both the p and n materials and that equal amounts of charge exist on both sides (Q_D and Q_N as the + - in Fig. 3.10c). However, since usually unequal A and N are used, in dopings N .

case illustrated in Fig. 3.10, the width of the depletion layer will not be the same on the two sides. Rather, to uncover the same amount of charge, the depletion layer will extend deeper into the more lightly doped material. Specifically, if we denote the width of the depletion region in the p side by x_p and in the n side by x_n , we can express the magnitude of the charge on the n side of the junction as

$$Q_+ = qAx_nN_D \quad (1.23)$$

and that on the p side of the junction as

$$Q_- = qAx_pN_A \quad (1.24)$$

where A is the cross-sectional area of the junction in the plane perpendicular to the page. The charge equality condition can now be written as

$$qAx_nN_D = qAx_pN_A$$

which can be rearranged to yield

$$\frac{x_n}{x_p} = \frac{N_A}{N_D} \quad (1.25)$$

In actual practice, it is usual for one side of the junction to be much more heavily doped than the other, with the result that the depletion region exists almost entirely on one side (the lightly doped side).

The width W of the depletion layer can be shown to be given by

$$W = x_n + x_p = \frac{2\epsilon_s}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) V_0 \quad (1.26)$$

where ϵ_s is the electrical permittivity of silicon = $11.7\epsilon_0 = 11.7 \times 8.85 \times 10^{-14}$ F/cm = 1.04×10^{-12} F/cm. Typically W is in the range 0.1 μm to 1 μm . Eqs. (1.25) and (1.26) can be used to obtain x_n and x_p in terms of W as

$$x_n = W \frac{N_A}{N_A + N_D} \quad (1.27)$$

$$x_p = W \frac{N_D}{N_A + N_D} \quad (1.28)$$

The charge stored on either side of the depletion region can be expressed in terms of W by utilizing

Eqs. (1.23) and (1.27) to obtain

$$Q_J = Q_+ - \frac{Q_-}{N_A N_D} W \quad (1.29)$$

Finally, we can substitute for W from Eq. (1.26) to obtain

$$Q_J = A q \frac{2e}{s} \frac{N_A N_D}{N_A + N_D} V_0 \quad (1.30)$$

These expressions for Q_J will prove useful in subsequent sections

1.4 The *pn* Junction with an Applied Voltage

Having studied the open-circuited *pn* junction in detail, we are now ready to apply a dc voltage between its two terminals to find its electrical conduction properties. If the voltage is applied so that the *p* side is made more positive than the *n* side, it is referred to as a forward-bias⁶ voltage. Conversely, if our applied dc voltage is such that it makes the *n* side more positive than the *p* side, it is said to be a reverse-bias voltage. As will be seen, the *pn* junction exhibits vastly different conduction properties in its forward and reverse directions.

Our plan is as follows. We begin by a simple qualitative description in Section 1.5.1 and then consider an analytical description of the *i*-*v* characteristic of the junction in Section 1.5.2.

1.5.1 Qualitative Description of Junction Operation

Figure 1.11 shows the *pn* junction under three different conditions: (a) the open-circuit or equilibrium condition studied in the previous section; (b) the reverse-bias condition, where a dc voltage V_R is applied; and (c) the forward-bias condition, where a dc voltage V_F is applied.

⁶For the time being, we take the term *bias* to refer simply to the application of a dc voltage. We will see in later chapters that it has a deeper meaning in the design of electronic circuits

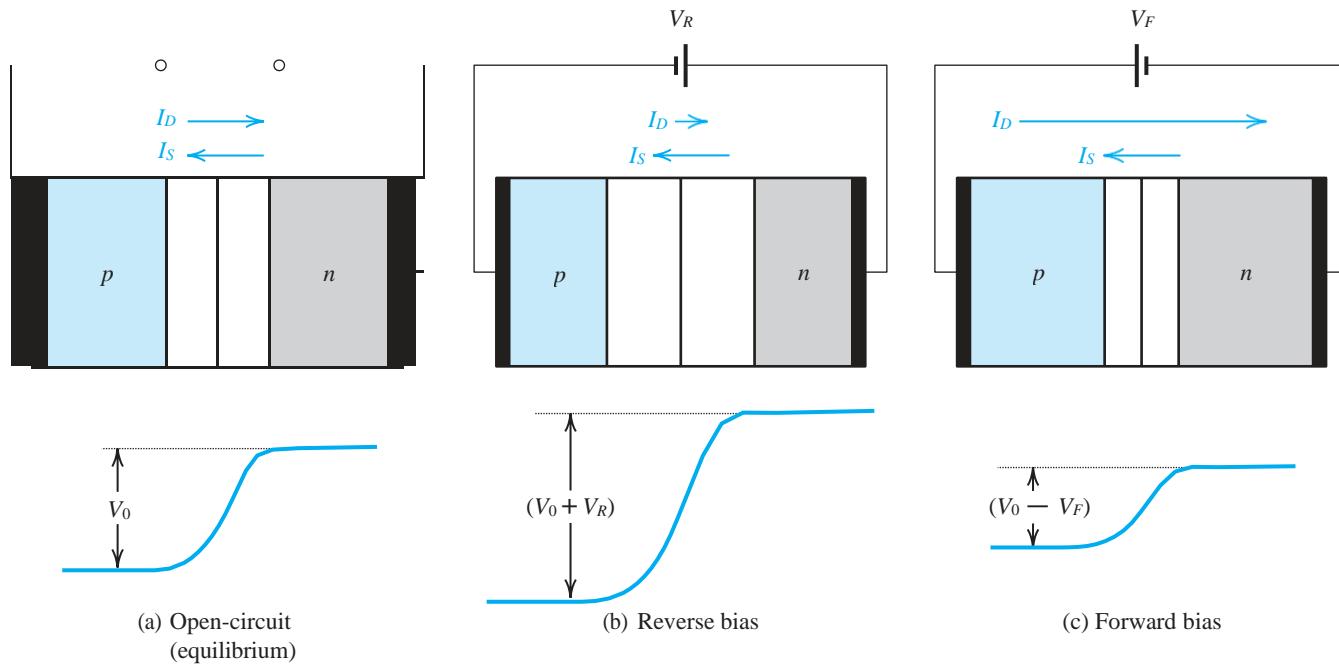


Figure 1.11 The *pn* junction in: (a) equilibrium; (b) reverse bias; (c) forward bias.

Observe that in the open-circuit case, a barrier voltage V_0 develops, making n more positive than p , and limiting the diffusion current I_D to a value exactly equal to the drift current I_S , thus resulting in a zero current at the junction terminals, as should be the case, since the terminals are open-circuited. Also, as mentioned previously, the barrier voltage V_0 , though it establishes the current equilibrium across the junction, does *not* in fact appear between the junction terminals.

Consider now the reverse-bias case in (b). The externally applied reverse-bias voltage V_R is in the direction to add to the barrier voltage, and it does, thus increasing the effective barrier voltage to $(V_0 + V_R)$ as shown. This reduces the number of holes that diffuse into the n region and the number of electrons that diffuse into the p region. The end result is that the diffusion current I_D is dramatically reduced. As will be seen shortly, a reverse-bias voltage of a volt or so is sufficient to cause $I_D \approx 0$, and the current across the junction and through the external circuit will be equal to I_S . Recalling that I_S is the current due to the drift across the depletion region of the thermally generated minority carriers, we expect I_S to be very small and to be strongly dependent on temperature. We will show this to be the case very shortly. We thus conclude that in the reverse direction, the pn junction conducts a very small and almost-constant current equal to I_S .

Before leaving the reverse-bias case, observe that the increase in barrier voltage will be accompanied by a corresponding increase in the stored uncovered charge on both sides of the depletion region. This in turn means a wider depletion region, needed to uncover the additional charge required to support the larger barrier voltage $(V_0 + V_R)$. Analytically, these results can be obtained easily by a simple extension of the results of the equilibrium case. Thus the width of the depletion region can be obtained by replacing V_0 in Eq. (3.26) by $(V_0 + V_R)$,

$$W = x_n + x_p = \frac{2e_s}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right) (V_0 + V_R) \quad (1.31)$$

and the magnitude of the charge stored on either side of the depletion region can be determined by replacing V_0 in Eq. (3.30) by $(V_0 + V_R)$,

$$Q_J = A_s \frac{2e q}{N_A + N_D} \left(\frac{N_A N_D}{V_0 + V_R} \right) \quad (1.32)$$

We next consider the forward-bias case shown in Fig. 3.11(c). Here the applied voltage V_F is in the direction that subtracts from the built-in voltage V_0 , resulting in a reduced barrier voltage $(V_0 - V_F)$ across

the depletion region. This reduced barrier voltage will be accompanied by reduced depletion-region charge and correspondingly narrower depletion-region width W . Most importantly, the lowering of the barrier voltage will enable more holes to diffuse from p to n and more electrons to diffuse from n to p . Thus the diffusion current I_D increases substantially and, as will be seen shortly, can become many orders of magnitude larger than the drift current I_S . The current I in the external circuit is of course the difference between I_D and I_S ,

$$I = I_D - I_S$$

and it flows in the forward direction of the junction, from p to n . We thus conclude that the pn junction can conduct a substantial current in the forward-bias region and that current is mostly a diffusion current whose value is determined by the forward-bias voltage V_F .

1.5.2 The Current-Voltage Relationship of the Junction

We are now ready to find an analytical expression that describes the current–voltage relationship of the pn junction. In the following we consider a junction operating with a forward applied voltage V and derive an expression for the current I that flows in the forward direction (from p to n). However, our derivation is general and will be seen to yield the reverse current when the applied voltage V is made negative.

From the qualitative description above we know that a forward-bias voltage V subtracts from the built-in voltage V_0 , thus resulting in a lower barrier voltage ($V_0 - V$). The lowered barrier in turn makes it possible for a greater number of holes to overcome the barrier and diffuse into the n region. A similar statement can be made about electrons from the n region diffusing into the p region.

Let us now consider the holes injected into the n region. The concentration of holes in the n region at the edge of the depletion region will increase considerably. In fact, an important result from device physics shows that the steady-state concentration at the edge of the depletion region will be

$$p_n(x_n) = p_{n0} e^{VV_T} \quad (1.33)$$

That is, the concentration of the minority holes increases from the equilibrium value of p_{n0} (see Fig. 1.10) to the much larger value determined by the value of V , given by Eq. (1.33).

We describe this situation as follows: The forward-bias voltage V results in an **excess concentration** of minority holes at $x = x_n$, given by

$$\begin{aligned} \text{Excess concentration} &= p_{n0} e^{VV_T} - p_{n0} \\ &= p_{n0} e^{VV_T} - 1 \end{aligned} \quad (1.34)$$

The increase in minority-carrier concentration in Eqs. (1.33) and (1.34) occurs at the edge of the depletion region ($x = x_n$). As the injected holes diffuse into the n material, some will recombine with the majority electrons and disappear. Thus, the excess hole concentration will decay exponentially with distance. As a result, the total hole concentration in the n material will be given by

$$p_n(x) = p_{n0} + (\text{Excess concentration}) e^{-(x-x_n)/L_p}$$

Substituting for the “Excess concentration” from Eq. (3.34) gives

$$p_n(x) = p_{n0} + p_{n0} e^{VV_T} - 1 e^{-(x-x_n)/L_p} \quad (1.35)$$

The exponential decay is characterized by the constant L_p , which is called the **diffusion length** of holes in the n material. The smaller the value of L_p , the faster the injected holes will recombine with the majority electrons, resulting in a steeper decay of minority-carrier concentration.

Figure 3.12 shows the steady-state minority-carrier concentration profiles on both sides of a pn junction in which $N_A \ll N_D$. Let's stay a little longer with the diffusion of holes into the n region. Note that the shaded region under the exponential represents the excess minority carriers (holes). From our study of diffusion in Section 3.3, we know that the establishment of a carrier concentration profile such as that in Fig. 3.12 is essential to support a steady-state diffusion current. In fact, we can now find the value of the hole-diffusion current density by applying Eq. (1.19),

$$J_p(x) = -qD_p \frac{dp_n(x)}{dx}$$

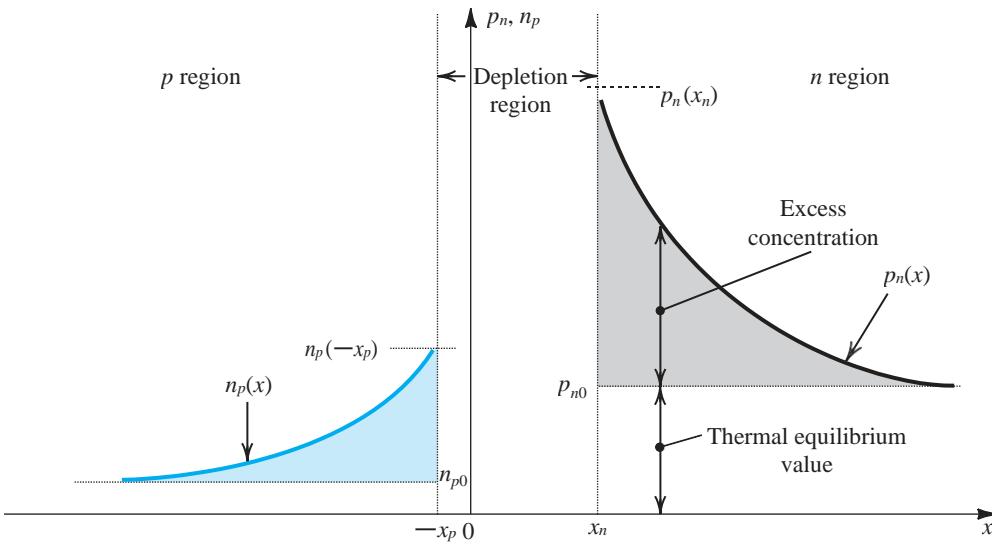


Figure 1.12 Minority-carrier distribution in a forward-biased *pn* junction. It is assumed that the *p* region is more heavily doped than the *n* region; $N_A > N_D$.

Substituting for $p_n(x)$ from Eq. (1.35) gives

$$J_p(x) = q \frac{D_p}{L_p} p_{n0} e^{VVT} - 1 e^{-(x-x_n)L_p} \quad (1.36)$$

As expected, $J_p(x)$ is highest at $x = x_n$,

$$J_p(x_n) = q \frac{D_p}{L_p} p_{n0} e^{VVT} - 1 \quad (1.37)$$

and decays exponentially for $x > x_n$, as the minority holes recombine with the majority electrons. This recombination, however, means that the majority electrons will have to be replenished by a current that injects electrons from the external circuit into the *n* region of the junction. This latter current component has the same direction as the hole current (because electrons moving from right to left give rise to current in the direction from left to right). It follows that as $J_p(x)$ decreases, the electron current component increases by exactly the same amount, making the total current in the *n* material constant at the value given by Eq. (1.37).

An exactly parallel development can be applied to the electrons that are injected from the *n* to the *p* region, resulting in an electron diffusion current given by a simple adaptation of Eq. (1.37),

$$J_n(-x_p) = q \frac{D_n}{L_n} n_{p0} e^{VVT} - 1 \quad (1.38)$$

Now, although the currents in Eqs. (1.37) and (1.38) are found at the two edges of the depletion region, their values do not change in the depletion region. Thus we can drop the location descriptors (x_n), $-x_p$, add the two current densities, and multiply by the junction area A to

obtain the total current I as

$$I = A J_p + J_n$$

$$I = Aq \frac{D_p}{L_p} p_{n0} + \frac{D_n}{L_n} n_{p0} e^{V/V_T} - 1$$

Substituting for $p_{n0} = \frac{n^2}{N_D}$ and for $n_{p0} = \frac{n^2}{N_A}$ gives

$$I = Aqn^2 \frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} e^{V/V_T} - 1 \quad (1.39)$$

From this equation we note that for a negative V (reverse bias) with a magnitude of a few times V_T (25.9 mV), the exponential term becomes essentially zero, and the current across the junction becomes negative and constant. From our qualitative description in Section 1.5.1, we know that this current must be I_S . Thus,

$$I = I_S e^{V/V_T} - 1 \quad (1.40)$$

where



$$I_S = Aqn^2 \frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \quad (1.41)$$

Figure 1.13 shows the I - V characteristic of the pn junction (Eq. 1.40). Observe that in the reverse direction the current saturates at a value equal to $-I_S$. For this reason, I_S is given the name **saturation current**. From Eq. (1.41) we see that I_S is directly proportional to the cross-sectional area A of the junction. Thus, another name for I_S , one we prefer to use in this book, is the junction **scale current**. Typical values for I_S , for junctions of various areas, range from 10^{-18} A to 10^{-12} A. Besides being proportional to the junction area A , the expression for I_S in Eq. (1.41) indicates that I_S is proportional to n , which is a very strong function of temperature (see Eq. 1.2).

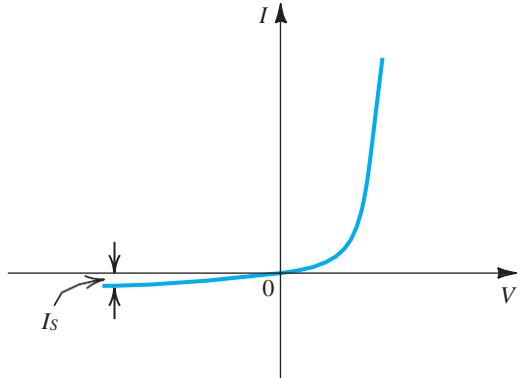


Figure 3.13 The pn junction I - V characteristic.

Example 1.6

For the pn junction considered in Example 3.5 for which $N_A = 10^{18}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$, $A = 10^{-4}\text{cm}^2$, and $n_i = 1.5 \times 10^{10}/\text{cm}^3$, let $L_p = 5 \mu\text{m}$, $L_n = 10 \mu\text{m}$, D_p (in the p region) = $10 \text{ cm}^2/\text{V}\cdot\text{s}$, and D_n (in the n region) = $18 \text{ cm}^2/\text{V}\cdot\text{s}$. The pn junction is forward biased and conducting a current $I = 0.1 \text{ mA}$. Calculate: (a) I_s ; (b) the forward-bias voltage V ; and (c) the component of the current I due to hole injection and that due to electron injection across the junction.

Solution

(a) Using Eq. (1.41), we find I_s as

$$I_s = 10^{-4} \times 1.6 \times 10^{-19} \times 1.5 \times 10^{10}^2 \\ \times \frac{10}{5 \times 10^{-4} \times 10^{16}} + \frac{18}{10 \times 10^{-4} \times 10^{18}} \\ = 7.3 \times 10^{-15} \text{ A}$$

(b) In the forward direction,

$$I_s e^{\frac{V}{V_T}} = I e^{\frac{V}{V_T}}$$

Thus,

$$V = V \ln \frac{I}{I_s}$$

For $I = 0.1 \text{ mA}$,

$$V = 25.9 \times 10^{-3} \ln \frac{0.1 \times 10^{-3}}{7.3 \times 10^{-15}} \\ = 0.605 \text{ V}$$

(c) The hole-injection component of I can be found using Eq.(1.37)

$$I_p = A q \frac{D_p}{L_p} n_0 e^{\frac{V}{V_T}} - 1 \\ A q \frac{D n^2}{L_p N_D} e^{\frac{V}{V_T}} - 1$$

Similarly, I_n can be found using Eq. (1.39),

$$I_n = A q \frac{n^2}{L N} e^{\frac{V}{V_T}} - 1$$

Thus,

$$\frac{I_p}{I_n} = \frac{D_p}{D_n} \frac{L_n}{L_p} \frac{N_A}{N_D}$$

For our case,

$$\frac{I_p}{I_n} = \frac{10}{18} \times \frac{10}{5} \times \frac{10^{18}}{10^{16}} = 1.11 \times 10^2 = 111$$

Example 1.6 continued

Thus most of the current is conducted by holes injected into the n region.

Specifically,

$$I_p = \frac{111}{112} \times 0.1 = 0.0991 \text{ mA}$$

$$I_n = \frac{1}{112} \times 0.1 = 0.0009 \text{ mA}$$

This stands to reason, since the p material has a doping concentration 100 times that of the n material.

1.5.3 Reverse Breakdown

The description of the operation of the pn junction in the reverse direction, and the I – V relationship of the junction in Eq. (3.40), indicate that at a reverse-bias voltage $-V$, with $V \gg V_T$, the reverse current that flows across the junction is approximately equal to I_S and thus is very small. However, as the magnitude of the reverse-bias voltage V is increased, a value is reached at which a very large reverse current flows as shown in Fig. 1.14. Observe that as V reaches the value $-V_Z$, the dramatic increase in reverse current is accompanied by a very small increase in the reverse voltage; that is, the reverse voltage across the junction.

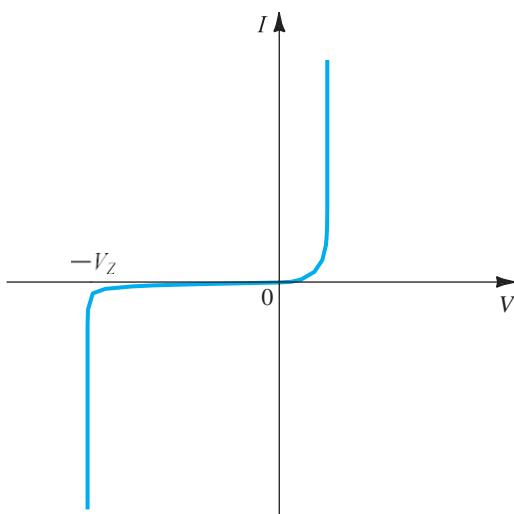


Figure 3.14 The I – V characteristic of the pn junction showing the rapid increase in reverse current in the breakdown region.

remains very close to the value V_Z . The phenomenon that occurs at $V < V_Z$ is known as **junction breakdown**. It is not a destructive phenomenon. That is, the *pn* junction can be repeatedly operated in the breakdown region without a permanent effect on its characteristics. This, however, is predicated on the assumption that the magnitude of the reverse-breakdown current is limited by the external circuit to a “safe” value. The “safe” value is one that results in the limitation of the power dissipated in the junction to a safe, allowable level.

There are two possible mechanisms for *pn* junction breakdown: the **zener effect**⁷ and the **avalanche effect**. If a *pn* junction breaks down with a breakdown voltage $V_Z < 5$ V, the breakdown mechanism is usually the zener effect. Avalanche breakdown occurs when V_Z is greater than approximately 7 V. For junctions that break down between 5 V and 7 V, the breakdown mechanism can be either the zener or the avalanche effect or a combination of the two.

Zener breakdown occurs when the electric field in the depletion layer increases to the point of breaking covalent bonds and generating electron–hole pairs. The electrons generated in this way will be swept by the electric field into the *n* side and the holes into the *p* side. Thus these electrons and holes constitute a reverse current across the junction. Once the zener effect starts, a large number of carriers can be generated, with a negligible increase in the junction voltage. Thus the reverse current in the breakdown region will be large and its value must be determined by the external circuit, while the reverse voltage appearing between the diode terminals will remain close to the specified breakdown voltage V_Z .

The other breakdown mechanism, avalanche breakdown, occurs when the minority carriers that cross the depletion region under the influence of the electric field gain sufficient kinetic energy to be able to break covalent bonds in atoms with which they collide. The carriers liberated by this process may have sufficiently high energy to be able to cause other carriers to be liberated in another ionizing collision. This process keeps repeating in the fashion of an avalanche, with the result that many carriers are created that are able to support any value of

Named after an early worker in the area. Note that the subscript *Z* in V denotes *zener*. We will use V_z to denote the breakdown voltage whether the breakdown mechanism is the zener effect or the avalanche effect.

reverse current, as determined by the external circuit, with a negligible change in the voltagedrop across the junction.

As will be seen in Chapter 4, some *pn* junction diodes are fabricated to operate specifically in the breakdown region, where use is made of the nearly constant voltage V_Z .

1.5 Capacitive Effects in the *pn* Junction

There are two charge-storage mechanisms in the *pn* junction. One is associated with the charge stored in the depletion region, and the other is associated with the minority-carrier charge stored in the *n* and *p* materials as a result of the concentration profiles established bycarrier injection. While the first is easier to see when the *pn* junction is reverse biased, the second is in effect only when the junction is forward biased.

1.6.1 Depletion or Junction Capacitance

When a *pn* junction is reverse biased with a voltage V_R , the charge stored on either side of the depletion region is given by Eq. (1.32),

$$Q_J = A \frac{2e q \frac{N_A N_D}{N_A + N_D} (V_0 + V_R)}{s}$$

Thus, for a given *pn* junction,

$$Q_J = \alpha \frac{V_0 + V_R}{s} \quad (1.42)$$

where α is given by

$$\alpha = A \frac{\frac{N_A N_D}{N_A + N_D}}{2e q s} \quad (1.43)$$

Thus Q_J is nonlinearly related to V_R , as shown in Fig. 1.15. This nonlinear relationship makes it difficult to define a capacitance that accounts for the need to change Q_J whenever V_R is

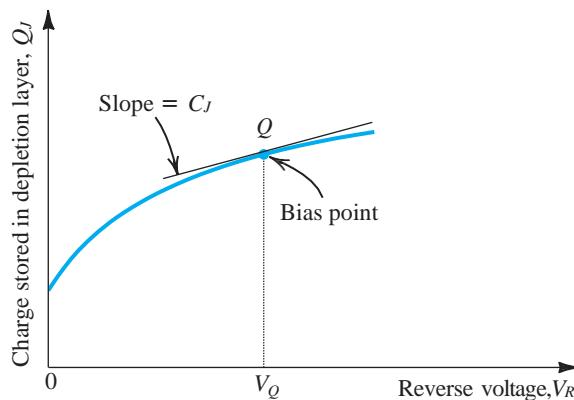


Figure 1.15 The charge stored on either side of the depletion layer as a function of the reverse voltage V_R .

changed. We can, however, assume that the junction is operating at a point such as Q , as indicated in Fig. 1.15, and define a capacitance C_j that relates the change in the charge Q_j to a change in the voltage V_R ,

$$C_j = \frac{dQ_j}{dV_R} \Big|_{V_R=V_Q} \quad (1.44)$$

This incremental-capacitance approach turns out to be quite useful in electronic circuit design, as we shall see throughout this book.

Using Eq. (1.44) together with Eq. (1.42) yields

$$C_j = \frac{\alpha}{2(V_0 + V_R)} \quad (1.45)$$

The value of C_j at zero reverse bias can be obtained from Eq. (1.45) as

$$C_{j0} = \frac{\alpha}{2V_0} \quad (1.46)$$

which enables us to express C_j as

$$C_j = \frac{C_{j0}}{1 + \frac{V_R}{V_0}} \quad (1.47)$$

where C_{j0} is given by Eq. (1.46) or alternatively if we substitute for α from Eq. (1.43) by

$$C_{j0} = A \frac{\frac{e_s}{q}}{\frac{N_{ND}}{x_r - x_r} + \frac{1}{V_r}} \quad (1.48)$$

Before leaving the subject of depletion-region or junction capacitance we point out that in the *pn* junction we have been studying, the doping concentration is made to change abruptly at the junction boundary. Such a junction is known as an **abrupt junction**. There is another type of *pn* junction in which the carrier concentration is made to change gradually from one side of the junction to the other. To allow for such a **graded junction**, the formula for the junction capacitance (Eq. 1.47) can be written in the more general form

$$C_j = \frac{C_{j0}}{\frac{V_R}{V_0}^m} \quad (1.49)$$

where m is a constant called the **grading coefficient**, whose value ranges from 1/3 to 1/2 depending on the manner in which the concentration changes from the *p* to the *n* side.

1.6.1 Diffusion Capacitance

Consider a forward-biased *pn* junction. In steady state, minority-carrier distributions in the *p* and *n* materials are established, as shown in Fig. 1.12. Thus a certain amount of excess minority-carrier charge is stored in each of the *p* and *n* bulk regions (outside the depletion region). If the terminal voltage V changes, this charge will have to change before a new steady-state is achieved. This charge-storage phenomenon gives rise to another capacitive effect, distinctly different from that due to charge storage in the depletion region.

To calculate the excess minority-carrier charge, refer to Fig. 1.12. The excess hole charge stored in the *n* region can be found from the shaded area under the exponential as follows.

$$\begin{aligned} Q_p &= Aq \times \text{shaded area under the } p_n(x) \text{ curve} \\ &= Aq[p_n(x_n) - p_{n0}]L_p \end{aligned}$$

Substituting for $p_n(x_n)$ from Eq. (1.33) and using Eq. (1.37) enables us to express Q_p as

$$Q_p = \frac{L_p^2}{\mu_p I} \quad (1.50)$$

The factor L^2/D that relates Q to I is a useful device parameter that has the dimension of time (s) and is denoted τ_p



$$\tau_p = \frac{L_p^2}{D_p} \quad (1.51)$$

Thus,

$$Q_p = \tau_p I_p \quad (1.52)$$

The time constant τ_p is known as the excess **minority-carrier (hole) lifetime**. It is the average time it takes for a hole injected into the n region to recombine with a majority electron. This definition of τ_p implies that the entire charge Q_p disappears and has to be replenished every τ_p seconds. The current that accomplishes the replenishing is $I_p Q_p / \tau_p$. This is an alternate derivation for Eq. (1.52).

⁸Recall that the area under an exponential curve $Ae^{-x/B}$ is equal to AB .

A relationship similar to that in Eq. (3.52) can be developed for the electron charge stored in the p region,

$$Q_n = \tau_n I_n \quad (1.53)$$

where τ_n is the electron lifetime in the p region. The total excess minority-carrier charge can be obtained by adding together Q_p and Q_n ,

$$Q = \tau_p I_p + \tau_n I_n \quad (1.54)$$

This charge can be expressed in terms of the diode current $I = I_p + I_n$ as

$$Q = \tau_T I \quad (1.55)$$

where τ_T is called the **mean transit time** of the junction. Obviously, τ_T is related to τ_p and τ_n . Furthermore, for most practical devices, one side of the junction is much more heavily doped than the other. For instance, if $N_A N_D$, one can show that $I_p \ll I_n$. For example, if $N_A \gg N_D$, then $I_p \ll I_n$.

$$I_n, I \quad I_p, Q_p \quad Q_n, Q \quad Q_p, \text{ and thus } \tau_p \ll \tau_n.$$

For small changes around a bias point, we can define an **incremental diffusion capacitance** C_d as

and can show that

$$C_d = \frac{dQ}{dV} \quad (1.56)$$

$$C_d = \frac{\tau_T}{V_T} I \quad (1.57)$$

where I is the forward-bias current. Note that C_d is directly proportional to the forward current I and thus is negligibly small when the diode is reverse biased. Also note that to keep C_d small, the transit time τ_T must be made small, an important requirement for a pn junction intended for high-speed or high-frequency operation.

analysis of diode circuits. The latter task involves the important subject of device modeling. Our study of modeling the diode characteristics will lay the foundation for our study of modeling transistor operation in the next three chapters.

Of the many applications of diodes, their use in the design of rectifiers (which convert ac to dc) is the most common. Therefore we shall study rectifier circuits in some detail and briefly look at a number of other diode applications. Further nonlinear circuits that utilize diodes and other devices will be found throughout the book, but particularly in Chapter 18.

The junction diode is nothing more than the pn junction we studied in Chapter 3, and most of this chapter is concerned with the study of silicon pn -junction diodes. In the last section, however, we briefly consider some specialized diode types, including the photodiode and the light-emitting diod

1.6 The Ideal Diode

1.7.1 Current-Voltage Characteristic

The ideal diode may be considered to be the most fundamental nonlinear circuit element. It is a two-terminal device having the circuit symbol of Fig. 4.1(a) and the $i-v$ characteristic shown in Fig. 1..1(b). The terminal characteristic of the ideal diode can be interpreted as follows: If a negative voltage (relative to the reference direction indicated in Fig. 1.1a) is applied to the diode, no current flows and the diode behaves as an open circuit (Fig. 1.1c). Diodes operated in this mode are said to be **reverse biased**, or operated in the reverse direction. An ideal diode has zero current when operated in the reverse direction and is said to be **cut off**, or simply **off**.

On the other hand, if a positive current (relative to the reference direction indicated in Fig. 1.1(a) is applied to the ideal diode, zero voltage drop appears across the diode. In other words, the ideal diode behaves as a short circuit in the *forward* direction (Fig. 1.1d); it passes any current with zero voltage drop. A **forward-biased** diode is said to be **turned on**, or simply **on**.

From the above description it should be noted that the external circuit must be designed to limit the forward current through a conducting diode, and the reverse voltage across a cutoff diode, to predetermined values. Figure 4.2 shows two diode circuits that illustrate this point. In the circuit of Fig. 1.2(a) the diode is obviously conducting. Thus its voltage drop will be zero, and the current through it will be determined by the 10-V supply and the 1-k Ω resistor as 10 mA. The diode in the circuit of Fig. 4.2(b) is obviously cut off, and thus its current will be zero, which in turn means that the entire 10-V supply will appear as reverse bias across the diode.

The positive terminal of the diode is called the **anode** and the negative terminal the **cathode**, a carryover from the days of vacuum-tube diodes. The $i-v$ characteristic of the ideal diode (conducting in one direction and not in the other) should explain the choice of its arrow-like circuit symbol.

As should be evident from the preceding description, the $i-v$ characteristic of the ideal diode is highly nonlinear; although it consists of two straight-line segments, they are at 90° to one another. A nonlinear curve that consists of straight-line segments is said to be **piecewise linear**. If a device having a piecewise-linear characteristic is used in a particular application in such a way that the signal across its terminals swings along only one of the linear segments, then the device can be considered a linear circuit element as far as that particular circuit

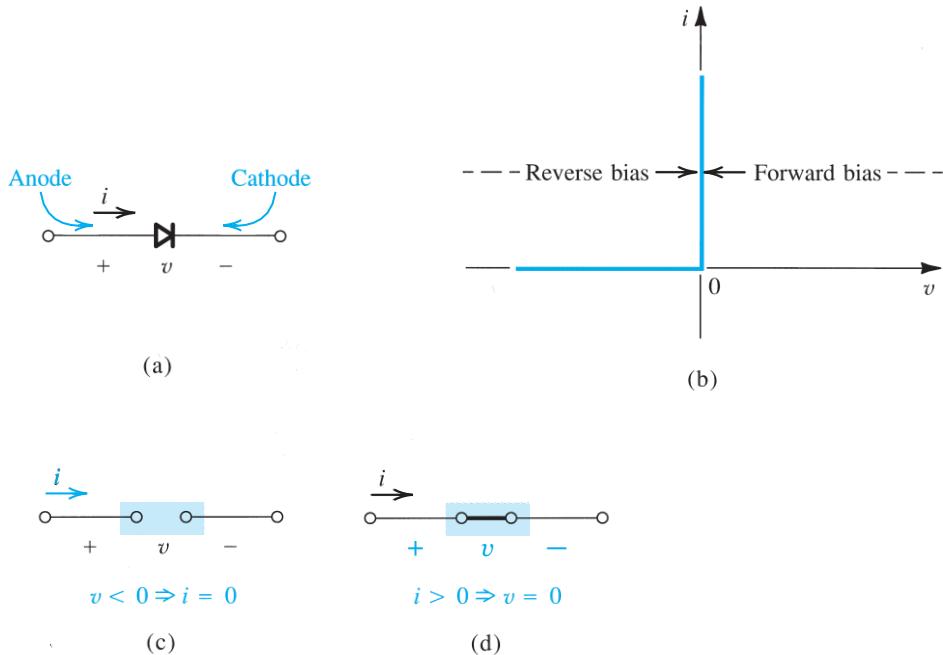


Figure 1.7.1 The ideal diode: (a) diode circuit symbol; (b) i - v characteristic; (c) equivalent circuit in the reverse direction; (d) equivalent circuit in the forward direction.

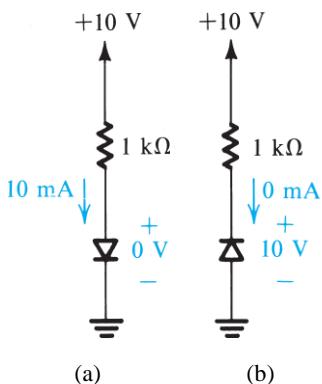


Figure 1.1.2 The two modes of operation of ideal diodes and the use of an external circuit to limit (a) the forward current and (b) the reverse voltage.

application is concerned. On the other hand, if signals swing past one or more of the break points in the characteristic, linear analysis is no longer possible.

1.7.1 A Simple Application: The Rectifier

A fundamental application of the diode, one that makes use of its severely nonlinear i - v curve, is the rectifier circuit shown in Fig. 1.3(a). The circuit consists of the series connection of a diode D and a resistor R . Let the input voltage v_I be the sinusoid shown in Fig. 1.3(b), and assume the diode to be ideal. During the positive half-cycles of the input sinusoid, the positive

v_I will cause current to flow through the diode in its forward direction. It follows that the diode voltage v_D will be very small—ideally zero. Thus the circuit will have the equivalent shown in Fig. 1.3(c), and the output voltage v_O will be equal to the input voltage v_I . On the other hand, during the negative half-cycles of v_I , the diode will not conduct. Thus the circuit will have the equivalent shown in Fig. 1.3(d), and v_O will be zero. Thus the output voltage will have the waveform shown in Fig. 1.3(e). Note that while v_I alternates in polarity and has a zero average value, v_O is unidirectional and has a finite average value or a *dc component*. Thus the circuit of Fig. 1.3(a) **rectifies** the signal and hence is called a **rectifier**. It can be used to generate dc from ac. We will study rectifier circuits in Section 4.5.

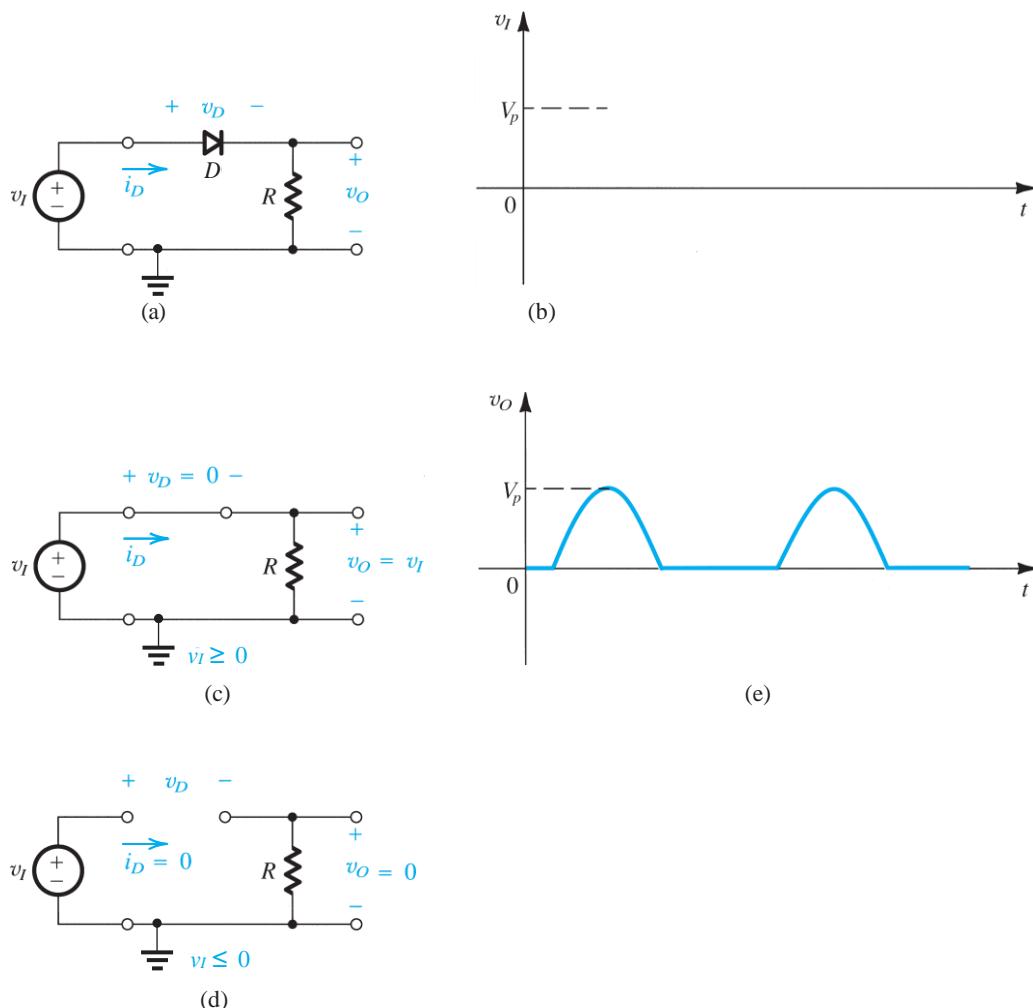


Figure 1.3 (a) Rectifier circuit. (b) Input waveform. (c) Equivalent circuit when $v_I \geq 0$. (d) Equivalent circuit when $v_I \leq 0$. (e) Output waveform.

Example 1.1

Figure 4.4(a) shows a circuit for charging a 12-V battery. If v_s is a sinusoid with 24-V peak amplitude, find the fraction of each cycle during which the diode conducts. Also, find the peak value of the diode current and the maximum reverse-bias voltage that appears across the diode.

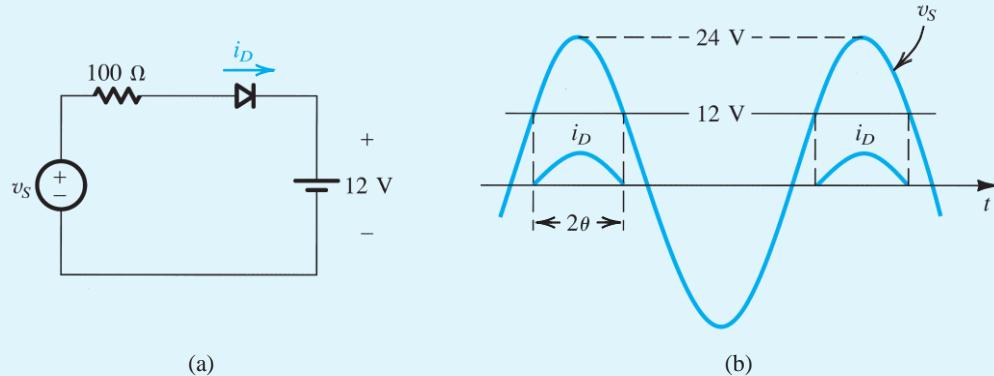


Figure 4.4 Circuit and waveforms for Example 4.1.

Solution

The diode conducts when v_s exceeds 12 V, as shown in Fig. 4.4(b). The conduction angle is 2θ , where θ is given by

$$24 \cos \theta = 12$$

Thus $\theta = 60^\circ$ and the conduction angle is 120° , or one-third of a cycle.

The peak value of the diode current is given by

$$I_d = \frac{24 - 12}{100} = 0.12 \text{ A}$$

The maximum reverse voltage across the diode occurs when v_s is at its negative peak and is equal to $24 + 12 = 36 \text{ V}$.

1.7.1 Another Application: Diode Logic Gates

Diodes together with resistors can be used to implement digital logic functions. Figure 1.5 shows two diode logic gates. To see how these circuits function, consider a positive-logic system in which voltage values close to 0 V correspond to logic 0 (or low) and voltage values close to 5 V correspond to logic 1 (or high). The circuit in Fig. 1.5(a) has three inputs, v_A , v_B , and v_C . It is easy to see that diodes connected to 5-V inputs will conduct, thus clamping the output v_Y to a value equal to 5 V. This positive voltage at the output will keep the diodes whose inputs are low (around 0 V) cut off.

Thus the output will be high if one or more of the inputs are high. The circuit therefore implements the **logic OR function**, which in Boolean

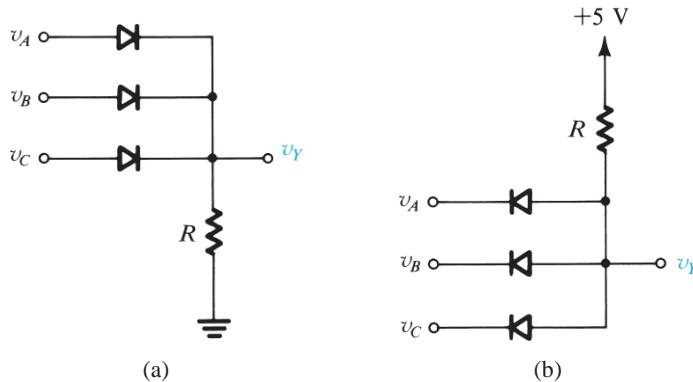


Figure 1.5 Diode logic gates: (a) OR gate; (b) AND gate (in a positive-logic system).

notation is expressed as

$$Y = A + B + C$$

Similarly, the reader is encouraged to show that using the same logic system mentioned above, the circuit of Fig. 1.5(b) implements the **logic AND function**,

Example 1.2

Assuming the diodes to be ideal, find the values of I and V in the circuits of Fig. 4.6.

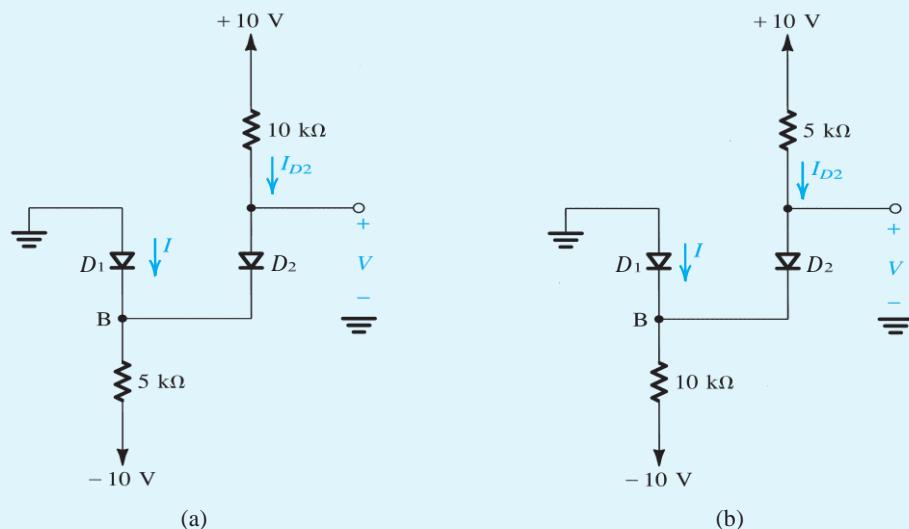


Figure 1.6 Circuits for Example 4.2.

$$Y = A \cdot B \cdot C$$

Example 1.2 continued

Solution

In these circuits it might not be obvious at first sight whether none, one, or both diodes are conducting. In such a case, *we make a plausible assumption, proceed with the analysis, and then check whether we end up with a consistent solution*. For the circuit in Fig. 4.6(a), we shall assume that both diodes are conducting. It follows that $V_B = 0$ and $V = 0$. The current through D_2 can now be determined from

$$I_{D2} = \frac{10 - 0}{10} = 1 \text{ mA}$$

Writing a node equation at B,

$$I + 1 = \frac{0 - (-10)}{5}$$

results in $I = 1$ mA. Thus D_1 is conducting as originally assumed, and the final result is $I = 1$ mA and $V = 0$ V.

For the circuit in Fig. 4.6(b), if we assume that both diodes are conducting, then $V_B = 0$ and $V = 0$. The current in D_2 is obtained from

$$I_{D2} = \frac{10 - 0}{5} = 2 \text{ mA}$$

The node equation at B is

$$I + 2 = \frac{0 - (-10)}{10}$$

which yields $I = -1$ mA. Since this is not possible, our original assumption is *not* correct. We start again, assuming that D_1 is off and D_2 is on. The current I_{D2} is given by

$$I_{D2} = \frac{10 - (-10)}{15} = 1.33 \text{ mA}$$

and the voltage at node B is

$$V_B = -10 + 10 \times 1.33 = +3.3 \text{ V}$$

Thus D_1 is reverse biased as assumed, and the final result is $I = 0$ and $V = 3.3$ V.

1.7 Terminal Characteristics of Junction Diodes

The most common implementation of the diode utilizes a *pn* junction. We have studied the physics of the *pn* junction and derived its *i*-*v* characteristic in Chapter 3. That the *pn* junction is used to implement the diode function should come as no surprise: the *pn* junction can conduct substantial current in the forward direction and almost no current in the reverse direction. In this section we study the *i*-*v* characteristic of the *pn* junction diode in detail in order to prepare ourselves for diode circuit applications.

Figure 1.7 shows the *i*-*v* characteristic of a silicon junction diode. The same characteristic is shown in Fig. 4.8 with some scales expanded and others compressed to reveal details. Note that the scale changes have resulted in the apparent discontinuity at the origin.

As indicated, the characteristic curve consists of three distinct regions:

1. The forward-bias region, determined by $v > 0$
2. The reverse-bias region, determined by $v < 0$
3. The breakdown region, determined by $v < -V_{ZK}$

These three regions of operation are described in the following sections.

1.8.1 The Forward-Bias Region

The forward-bias—or simply forward—region of operation is entered when the terminal voltage v is positive. In the forward region the *i*-*v* relationship is closely approximated by

$$i = I_S e^{v/V_T} - 1 \quad (1.1)$$

In this equation¹ I is a constant for a given diode at a given temperature. A formula for I in terms of the diode's physical parameters and temperature was given in Eq. (1.41). The current

Equation (1.1), the diode equation, is sometimes written to include a constant n in the exponential,

$$i = I (e^{v/nV_T} - 1)$$

with n having a value between 1 and 2, depending on the material and the physical structure of the diode. Diodes using the standard integrated-circuit fabrication process exhibit $n = 1$ when operated under normal conditions. For simplicity, we shall use $n = 1$ throughout this book, unless otherwise specified.

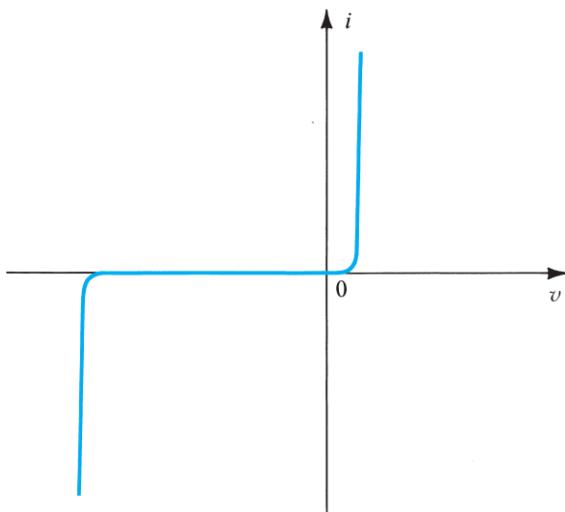


Figure 1.7 The i - v characteristic of a silicon junction diode.

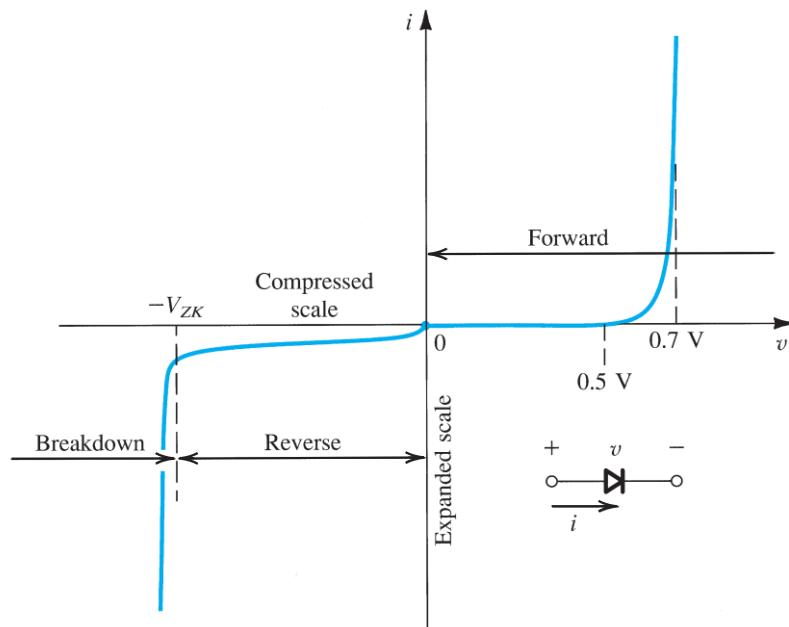


Figure 1.8 The diode i - v relationship with some scales expanded and others compressed in order to reveal details.

I_S is usually called the **saturation current** (for reasons that will become apparent shortly). Another name for I_S , and one that we will occasionally use, is the **scale current**. This name arises from the fact that I_S is directly proportional to the cross-sectional area of the diode. Thus doubling of the junction area results in a diode with double the value of I_S and, as the diode equation indicates, double the value of current i for a given forward voltage v . For “small-signal” diodes, which are small-size diodes intended for low-power applications, I_S is

on the order of 10^{-15} A. The value of I is, however, a very strong function of temperature.

As a rule of thumb, I_S doubles in value for every 5°C rise in temperature.

The voltage V_T in Eq. (1.1) is a constant called the **thermal voltage** and is given by

$$V_T = \frac{kT}{q} \quad (1.2)$$

where

k = Boltzmann's constant = 8.62×10^{-5} eV/K = 1.38×10^{-23} joules/kelvin

T = the absolute temperature in kelvins = $273 +$ temperature in $^\circ\text{C}$

q = the magnitude of electronic charge = 1.60×10^{-19} coulomb

Substituting $k = 8.62 \times 10^{-5}$ eV/K into Eq. (4.2) gives

$$V_T = 0.0862T, \text{ mV} \quad (1.2a)$$

Thus, at room temperature (20°C) the value of V_T is 25.3 mV. In rapid approximate circuit analysis we shall use V_T 25 mV at room temperature.

For appreciable current i in the forward direction, specifically for $i > I_S$, Eq. (1.1) can be approximated by the exponential relationship

$$i = I_S e^{v/V_T} \quad (1.3)$$

This relationship can be expressed alternatively in the logarithmic form

$$v = V \ln \frac{i}{I_S} \quad (1.4)$$

where \ln denotes the natural (base e) logarithm. The exponential relationship of the current i to the voltage v holds over many decades of current (a span of as many as seven decades—i.e., a factor of 10^7 —can be found). This is quite a remarkable property of junction diodes, one that is also found in bipolar junction transistors and that has been exploited in many interesting applications.

Let us consider the forward i - v relationship in Eq. (1.3) and evaluate the current I_1 corresponding to a diode voltage V_1 :

$$I_1 = I_s e^{V_1/V_T}$$

Similarly, if the voltage is V_2 , the diode current I_2 will be

$$I_2 = I_s e^{V_2/V_T}$$

A slightly higher ambient temperature (25°C or so) is usually assumed for electronic equipment operating inside a cabinet. At this temperature, $VT \approx 25.8 \text{ mV}$. Nevertheless, for the sake of simplicity and to promote rapid circuit analysis, we shall use the more arithmetically convenient value of $VT \approx 25 \text{ mV}$ throughout this book.

These two equations can be combined to produce

$$\frac{I_2}{I_1} = e^{(V_2 - V_1)/V_T}$$

This equation simply states that for a decade (factor of 10) change in current, the diode voltage drop changes by $2.3V_T$, which is approximately 60 mV. This also suggests that the diode i - v relationship is most conveniently plotted on semilog paper. Using the vertical, linear axis for v and the horizontal, log axis for i , one obtains a straight line with a slope of 60 mV per decade of current.

A glance at the i - v characteristic in the forward region (Fig. 4.8) reveals that the current is negligibly small for v smaller than about 0.5 V. This value is usually referred to as the **cut-in voltage**. It should be emphasized, however, that this apparent threshold in the characteristic is simply a consequence of the exponential relationship. Another consequence of this relationship is the rapid increase of i . Thus, for a “fully conducting” diode, the voltage drop lies in a narrow range, approximately 0.6 V to 0.8 V. This gives rise to a simple “model” for the diode where it is assumed that a conducting diode has approximately a 0.7-V drop across it. Diodes with different current ratings (i.e., different areas and correspondingly different I_s) will exhibit the 0.7-V drop at different currents. For instance, a small-signal diode may be considered to have a 0.7-V drop at $i = 1 \text{ mA}$, while a higher-power diode may have a 0.7-V drop at $i = 1 \text{ A}$. We will study the topics of diode-circuit analysis and diode models in the next section.

Example 1.3

A silicon diode said to be a 1-mA device displays a forward voltage of 0.7 V at a current of 1 mA. Evaluate the junction scaling constant I_S . What scaling constants would apply for a 1-A diode of the same manufacture that conducts 1 A at 0.7 V?

Solution

Since

$$i = I_S e^{V/V_T}$$

then

$$I_S = i e^{-V/V_T}$$

Example 1.3 continued

For the 1-mA diode:

$$I_S = 10^{-3} e^{-700/25} = 6.9 \times 10^{-16} \text{ A}$$

The diode conducting 1 A at 0.7 V corresponds to one-thousand 1-mA diodes in parallel with a total junction area 1000 times greater. Thus I_S is also 1000 times greater,

$$I_S = 6.9 \times 10^{13} \text{ A}$$

Since both I_S and V_T are functions of temperature, the forward $i-v$ characteristic varies with temperature, as illustrated in Fig. 4.9. At a given constant diode current, the voltage drop across the diode decreases by approximately 2 mV for every 1°C increase in temperature. The change in diode voltage with temperature has been exploited in the design of electronic thermometers.

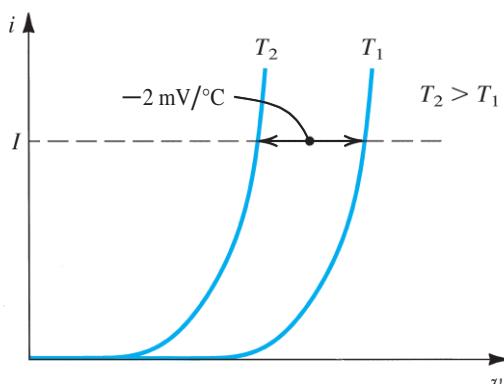


Figure 1.9 Temperature dependence of the diode forward characteristic. At a constant current, the voltage drop decreases by approximately 2 mV for every 1°C increase in temperature.

1.8.2 The Reverse-Bias Region

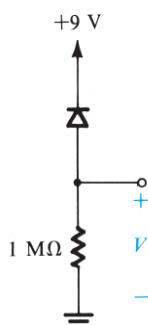
The reverse-bias region of operation is entered when the diode voltage v is made negative. Equation (1.1) predicts that if v is negative and a few times larger than V_T (25 mV) in magnitude, the exponential term becomes negligibly small compared to unity, and the diode current becomes

$$i = -I_S$$

That is, the current in the reverse direction is constant and equal to I_S . This constancy is the reason behind the term *saturation current*.

Real diodes exhibit reverse currents that, though quite small, are much larger than I_S .^{14 – 15} For instance, a small-signal diode whose I_S is on the order of 10^{-15} A to 10^{-14} A could show a reverse current on the order of 1 nA. The reverse current also increases somewhat with the increase in magnitude of the reverse voltage. Note that because of the very small magnitude of the current, these details are not clearly evident on the diode i - v characteristic of Fig. 1.8.

A large part of the reverse current is due to leakage effects. These leakage currents are proportional to the junction area, just as I_S is. Their dependence on temperature, however, is different from that of I_S . Thus, whereas I_S doubles for every 5°C rise in temperature, the corresponding rule of thumb for the temperature dependence of the reverse current is that it doubles for every 10°C rise in temperature.



1.8.3 The Breakdown Region

The third distinct region of diode operation is the breakdown region, which can be easily identified on the diode i - v characteristic in Fig. 4.8. The breakdown region is entered when the magnitude of the reverse voltage exceeds a threshold value that is specific to the particular diode, called the **breakdown voltage**. This is the voltage at the “knee” of the i - v curve in Fig. 4.8 and is denoted V_{ZK} , where the subscript Z stands for zener (see Section 3.5.3) and K denotes knee.

As can be seen from Fig. 1.8, in the breakdown region the reverse current increases rapidly, with the associated increase in voltage drop being very small. Diode breakdown is normally not destructive, provided the power dissipated in the diode is limited by external circuitry to a “safe” level. This safe value is normally specified on the device data sheets. It therefore is necessary to limit the reverse current in the breakdown region to a value consistent with the permissible power dissipation.

The fact that the diode i - v characteristic in breakdown is almost a vertical line enables it to be used in voltage regulation. This subject will be studied in Section 1.5.

1.8 Modeling the Diode Forward Characteristic

Having studied the diode terminal characteristics we are now ready to consider the analysis of circuits employing forward-conducting diodes. Figure 1.10 shows such a circuit. It consists of a dc source V_{DD} , a resistor R , and a diode. We wish to analyze this circuit to determine the diode voltage V_D and current I_D . To aid in our analysis, we need to represent the diode with a model. There are a variety of diode models, of which we now know two: the ideal-diode model and the exponential model. In the following discussion we shall assess the suitability of these two models in various analysis situations. Also, we shall develop and comment on other models. This material, besides being useful in the analysis and design of diode circuits, establishes a foundation for the modeling of transistor operation that we will study in the next three chapters.

1.9.1 The Exponential Model

The most accurate description of the diode operation in the forward region is provided by the exponential model. Unfortunately, however, its severely nonlinear nature makes this model the most difficult to use. To illustrate, let’s analyze the circuit in Fig. 4.10 using the exponential diode model.

Assuming that V_{DD} is greater than 0.5 V or so, the diode current will be much greater than I_S , and we can represent the diode i - v characteristic by the exponential relationship,

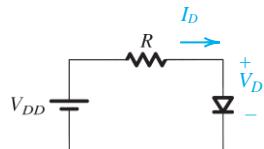


Figure 1.10 A simple circuit used to illustrate the analysis of circuits in which the diode is forward conducting.

resulting in

$$I_D = I_S e^{V_D/V_T} \quad (1.6)$$

The other equation that governs circuit operation is obtained by writing a Kirchhoff loop equation, resulting in

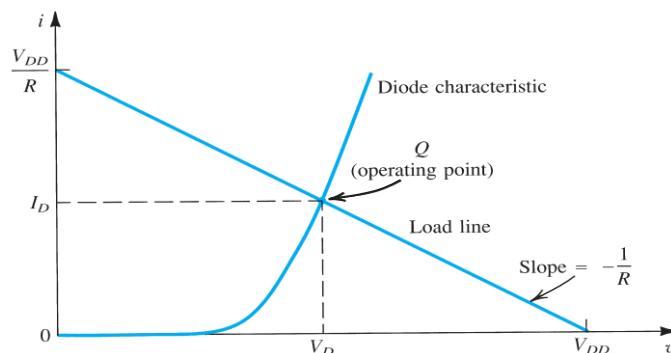
$$I_D \frac{V_{DD} - V_D}{R} \quad (1.7)$$

Assuming that the diode parameter I_S is known, Eqs. (4.6) and (1.7) are two equations in the two unknown quantities I_D and V_D . Two alternative ways for obtaining the solution are graphical analysis and iterative analysis.

1.9.2 Graphical Analysis Using the Exponential Model

Graphical analysis is performed by plotting the relationships of Eqs. (1.6) and (1.7) on the i - v plane. The solution can then be obtained as the coordinates of the point of intersection of the two graphs. A sketch of the graphical construction is shown in Fig. 1.11. The curve represents the exponential diode equation (Eq. 1.6), and the straight line represents Eq. (1.7). Such a straight line is known as the **load line**, a name that will become more meaningful in later chapters. The load line intersects the diode curve at point Q , which represents the **operating point** of the circuit. Its coordinates give the values of I_D and V_D .

Graphical analysis aids in the visualization of circuit operation. However, the effort involved in performing such an analysis, particularly for complex circuits, is too great to be justified in practice.



1.9.2 Iterative Analysis Using the Exponential Model

Equations (1.6) and (1.7) can be solved using a simple iterative procedure, as illustrated in the following example.

Example 1.4

Determine the current I_D and the diode voltage V_D for the circuit in Fig. 4.10 with $V_{DD} = 5$ V and $R = 1$ k Ω . Assume that the diode has a current of 1 mA at a voltage of 0.7 V.

Solution

To begin the iteration, we assume that $V_D = 0.7$ V and use Eq. (4.7) to determine the current,

$$\begin{aligned} I_D &= \frac{V_{DD} - V_D}{R} \\ &= \frac{5 - 0.7}{1} = 4.3 \text{ mA} \end{aligned}$$

We then use the diode equation to obtain a better estimate for V_D . This can be done by employing Eq. (4.5), namely,

$$V_2 - V_1 = 2.3V_T \log \frac{I_2}{I_1}$$

Substituting $2.3V_T = 60$ mV, we have

$$V_2 = V_1 + 0.06 \log \frac{I_2}{I_1}$$

Substituting $V_1 = 0.7$ V, $I_1 = 1$ mA, and $I_2 = 4.3$ mA results in $V_2 = 0.738$ V. Thus the results of the first iteration are $I_D = 4.3$ mA and $V_D = 0.738$ V. The second iteration proceeds in a similar manner:

$$\begin{aligned} I_D &= \frac{V_2 - V_D}{R} = 4.262 \text{ mA} \\ V_2 &= 0.738 + 0.06 \log \frac{4.262}{4.3} \\ &= 0.738 \text{ V} \end{aligned}$$

Thus the second iteration yields $I_D = 4.262$ mA and $V_D = 0.738$ V. Since these values are very close to the values obtained after the first iteration, no further iterations are necessary, and the solution is $I_D = 4.262$ mA and $V_D = 0.738$ V.

1.9.3 The Need for Rapid Analysis

The iterative analysis procedure utilized in the example above is simple and yields accurate results after two or three iterations. Nevertheless, there are situations in which the effort and time required are still greater than can be justified. Specifically, if one is doing a pencil-and-paper design of a relatively complex circuit, rapid circuit analysis is a necessity.

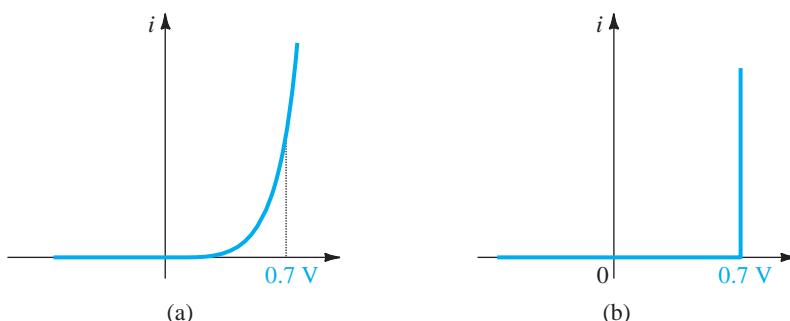
Through quick analysis, the designer is able to evaluate various possibilities before deciding on a suitable circuit design. To speed up the analysis process, one must be content with less precise results. This, however, is seldom a problem, because the more accurate analysis can be postponed until a final or almost-final design is obtained. Accurate analysis of the almost-final design can be performed with the aid of a computer circuit-analysis program such as SPICE (see Appendix B and the website). The results of such an analysis can then be used to furtherrefine or “fine-tune” the design.

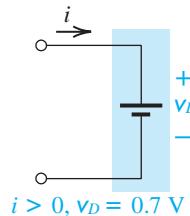
To speed up the analysis process, we must find a simpler model for the diode forward characteristic.

1.9.4 The Constant-Voltage-Drop Model

The simplest and most widely used diode model is the constant-voltage-drop model. This model is based on the observation that a forward-conducting diode has a voltage drop that varies in a relatively narrow range, say, 0.6 to 0.8 V. The model assumes this voltage to be constant at a value, say, 0.7 V. This development is illustrated in Fig. 1.12.

The constant-voltage-drop model is the one most frequently employed in the initial phases of analysis and design. This is especially true if at these stages one does not have detailed information about the diode characteristics, which is often the case.





(c)

Figure 1.12 Development of the diode constant-voltage-drop model: (a) the exponential characteristic; (b) approximating the exponential characteristic by a constant voltage, usually about 0.7 V; (c) the resulting model of the forward-conducting diodes.

Finally, note that if we employ the constant-voltage-drop model to solve the problem in Example 4.4, we obtain

$$V_D = 0.7 \text{ V}$$

and

$$\begin{aligned} I_D &= \frac{V_{DD} - 0.7}{R} \\ &= \frac{5 - 0.7}{1} = 4.3 \text{ mA} \end{aligned}$$

which are not very different from the values obtained before with the more elaborate exponential model.

1.9.5 The Ideal-Diode Model

In applications that involve voltages much greater than the diode voltage drop (0.6 V–0.8 V), we may neglect the diode voltage drop altogether while calculating the diode current. The result is the ideal-diode model, which we studied in Section 4.1. For the circuit in Example 4.4 (i.e., Fig. 1.10 with $V_{DD} 5 \text{ V}$ and $R 1 \text{ k}\Omega$), utilization of the ideal-diode model leads to

$$\begin{aligned} V_D &= 0 \text{ V} \\ I &= \frac{5 - 0}{1} = 5 \text{ mA} \end{aligned}$$

which for a very quick analysis would not be bad as a gross estimate. However, with almost no additional work, the 0.7-V-drop model yields much more realistic results. We note, however, that the greatest utility of the ideal-diode model is in determining which diodes are on and which are off in a multidiode circuit, such as those considered in Section 4.1.

1.9.6 The Small-Signal Model

Consider the situation in Fig. 1.13(a), where a dc voltage V_{DD} establishes a dc current I_D through the series combination of a resistance R and a diode D . The resulting diode voltage is denoted V_D . As mentioned above, values of I_D and V_D can be obtained by solving the circuit using the diode exponential characteristic or, much more quickly, approximate values can be found using the diode constant-voltage-drop model.

Next, consider the situation of V_{DD} undergoing a small change $\pm \Delta V_{DD}$, as shown in Fig. 1.13(b). As indicated, the current I_D changes by an increment ΔI_D , and the diode voltage V_D changes by an increment ΔV_D . We wish to find a quick way to determine the values of these incremental changes. Toward that end, we develop a “small-signal” model for the diode.

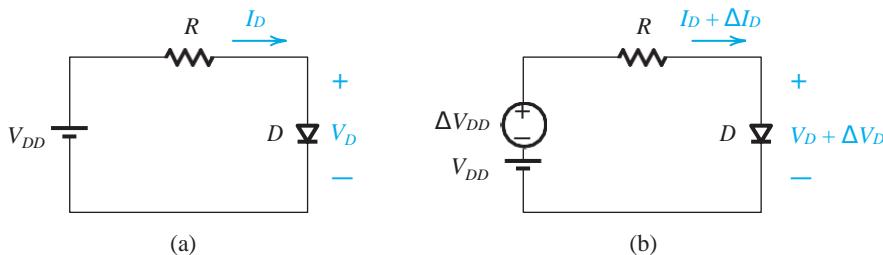


Figure 1.13 (a) A simple diode circuit; (b) the situation when V_{DD} changes by ΔV_{DD} .

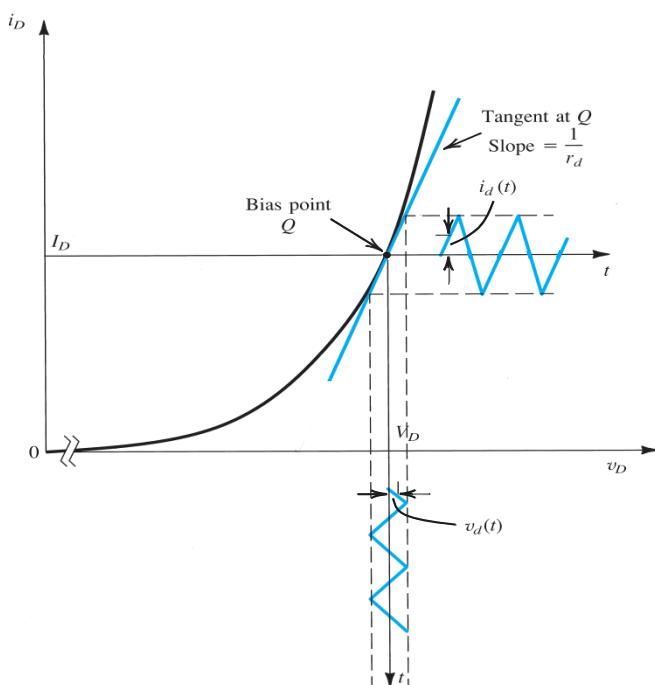


Figure 1.14 Development of the diode small-signal model.

Here the word *signal* emphasizes that in general, V_{DD} can be a time-varying quantity. The qualifier “small” indicates that this diode model applies only when V_D is kept sufficiently small, with “sufficiently” to be quantified shortly.

To develop the diode small-signal model, refer to Fig. 1.14. We express the voltage across the diode as the sum of the dc voltage V_D and the time-varying signal $v_d(t)$,

$$v_D(t) = V_D + v_d(t) \quad (1.8)$$

Correspondingly, the total instantaneous diode current $i_D(t)$ will be

$$i_D(t) = I_S e^{V_D/V_T} \quad (1.9)$$

Substituting for v_D from Eq. (1.8) gives

$$i_D(t) = I_S e^{(V_D+v_d)/V_T}$$

which can be rewritten

$$i_D(t) = I_S e^{V_D/V_T} e^{v_d/V_T} \quad (1.10)$$

In the absence of the signal $v_d(t)$, the diode voltage is equal to V_D , and the diode current is I_D , given by

$$I_D = I_S e^{V_D/V_T} \quad (1.11)$$

Thus, $i_D(t)$ in Eq. (1.10) can be expressed as

$$i_D(t) = I_S e^{v_d/V_T} \quad (1.12)$$

Now if the amplitude of the signal $v_d(t)$ is kept sufficiently small such that

$$\frac{v_d}{V_T} \ll 1 \quad (1.13)$$

then we may expand the exponential of Eq. (1.12) in a series and truncate the series after the first two terms to obtain the approximate expression

$$i_D(t) \approx I_D \left[1 + \frac{v_d}{V_T} \right] \quad (1.14)$$

This is the **small-signal approximation**. It is valid for signals whose amplitudes are smaller than about 5 mV (see Eq. 4.13, and recall that $V_T = 25$ mV).

V_T

From Eq. (4.14) we have

Thus, superimposed on the dc current I_D , we have a signal current component directly proportional to the signal voltage v_d . That is,

$$i_D = I_D + i_d \quad (1.16)$$

where

$$i_d = \frac{I_D}{V_T} v_d \quad (1.17)$$

The quantity relating the signal current i_d to the signal voltage v_d has the dimensions of conductance, mhos (V), and is called the **diode small-signal conductance**. The inverse of this parameter is the **diode small-signal resistance**, or **incremental resistance**, r_d ,

$$r_d = \frac{V_T}{I_D} \quad (1.18)$$

Note that the value of r_d is inversely proportional to the bias current I_D .

³ For $v^{\frac{1}{2}} = 5$ mV, $v^{\frac{1}{2}}/V = 0.2$. Thus the next term in the series expansion of the exponential will be $1 \times 0.2^2 = 0.02$, a factor of 10 lower than the linear term we kept.

Additional insight into the small-signal approximation and the small-signal diode model can be obtained by considering again the graphical construction in Fig. 1.14. Here the diode is seen to be operating at a dc bias point Q characterized by the dc voltage V_D and the corresponding dc current I_D . Superimposed on V_D we have a signal $v_d(t)$, assumed (arbitrarily) to have a triangular waveform.

It is easy to see that using the small-signal approximation is equivalent to assuming that *the signal amplitude is sufficiently small such that the excursion along the i - v curve is limited to a short almost-linear segment*. The slope of this segment, which is equal to the slope of the tangent to the i - v curve at the operating point Q , is equal to the small-signal conductance. The reader is encouraged to prove that the slope of the i - v curve at $i = I_D$ is equal to I_D/V_T , which is $1/r_d$; that is,

$$r_d = \left. \frac{\partial i}{\partial v} \right|_{i_D = I_D} \quad (1.19)$$

From the preceding we conclude that superimposed on the quantities V_D and I_D that define the dc bias

point, or **quiescent point**, of the diode will be the small-signal quantities $v_d(t)$ and $i_d(t)$, which are related by the diode small-signal resistance r_d evaluated at the bias point (Eq. 1.18). Thus the small-signal analysis can be performed separately from the dc bias analysis, a great convenience that results from the linearization of the diode characteristics inherent in the small-signal approximation. Specifically, after the dc analysis is performed, the small-signal equivalent circuit is obtained by eliminating all dc sources (i.e., short-circuiting dc voltage sources and open-circuiting dc current sources) and replacing the diode by its small-signal resistance. Thus, for the circuit in Fig. 1.13(b), the dc analysis is obtained by using the circuit in Fig. 1.13(a), while the incremental quantities ΔI_D and ΔV_D can be determined by using the small-signal equivalent circuit shown in Fig. 1.15. The following example should further illustrate the application of the small-signal model.

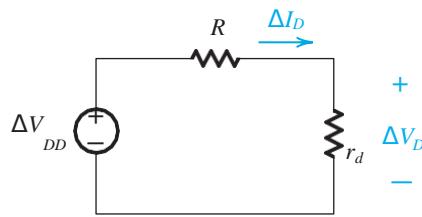


Figure 1.15 Circuit for determining the incremental quantities ΔI_D and ΔV_D for the circuit in Figure 1.13(b). Note that replacing the diode with its small-signal resistance r_d results in a linear circuit.

Example 1.5

Consider the circuit shown in Fig. 4.16(a) for the case in which $R = 10 \text{ k}\Omega$. The power supply V^+ has a dc value of 10 V on which is superimposed a 60-Hz sinusoid of 1-V peak amplitude. (This “signal” component of the power-supply voltage is an imperfection in the power-supply design. It is known as the **power-supply ripple**. More on this later.) Calculate both the dc voltage of the diode and the amplitude of the sine-wave signal appearing across it. Assume the diode to have a 0.7-V drop at 1-mA current.

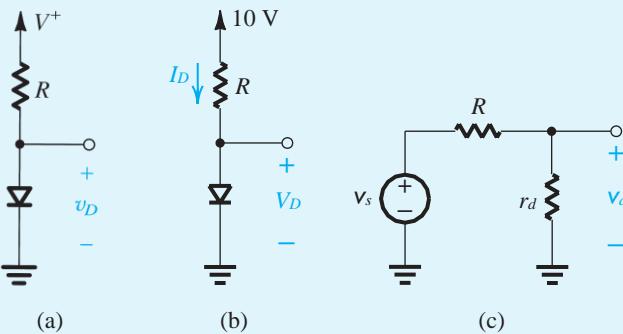


Figure 1.16 (a) Circuit for Example 4.5. (b) Circuit for calculating the dc operating point. (c) Small-signal equivalent circuit.

Solution

Considering dc quantities only, we assume $V_D = 0.7$ V and calculate the diode dc current

$$I_D = \frac{10 - 0.7}{10} = 0.93 \text{ mA}$$

Since this value is very close to 1 mA, the diode voltage will be very close to the assumed value of 0.7 V. At this operating point, the diode incremental resistance r_d is

$$r_d = \frac{V_T}{I_D} = \frac{25}{0.93} = 26.9 \text{ } \Delta$$

The signal voltage across the diode can be found from the small-signal equivalent circuit in Fig. 4.16(c). Here v_s denotes the 60-Hz 1-V peak sinusoidal component of V^+ , and v_d is the corresponding signal across the diode. Using the voltage divider rule provides the peak amplitude of v_d as follows:

$$\begin{aligned} v_d(\text{peak}) &= \hat{v}_s \frac{r_d}{R + r_d} \\ &= 1 \frac{0.0269}{10 + 0.0269} = 2.68 \text{ mV} \end{aligned}$$

Finally, we note that since this value is quite small, our use of the small-signal model of the diode is justified.

From the above we see that for a diode circuit that involves both dc and signal quantities, a small-signal equivalent circuit can be obtained by eliminating the dc sources and replacing each diode with its small-signal resistance r_d . Such a circuit is linear and can be solved using linear circuit analysis.

Finally, we note that while r_d models the small-signal operation of the diode at low frequencies, its dynamic operation is modeled by the capacitances C_j and C_d , which we

studied in Section 3.6 and which also are small-signal parameters. A complete model of the diode includes C_j and C_d in parallel with r_d .

1.9.7 Use of the Diode Forward Drop in Voltage Regulation

A further application of the diode small-signal model is found in a popular diode application, namely, the use of diodes to create a regulated voltage. A **voltage regulator** is a circuit whose purpose is to provide a constant dc voltage between its output terminals. The output voltage is required to remain as constant as possible in spite of (a) changes in the load current drawn from the regulator output terminal and (b) changes in the dc power-supply voltage that feeds the regulator circuit. Since the forward-voltage drop of the diode remains almost constant at approximately $0.7 \pm \frac{1}{4}$ V while the current through it varies by relatively large amounts, a forward-biased diode can make a simple voltage regulator. For instance, we have seen in Example 4.5 that while the 10-V dc supply voltage had a ripple of 2 V peak-to-peak (a 10% variation), the corresponding ripple in the diode voltage was only about 2.7 mV (a 0.4% variation). Regulated voltages greater than 0.7 V can be obtained by connecting a number of diodes in series. For example, the use of three forward-biased diodes in series provides a voltage of about 2 V. One such circuit is investigated in the following example, which utilizes the diode small-signal model to quantify the efficacy of the voltage regulator that is realized.

Example 4.6

Consider the circuit shown in Fig. 4.17. A string of three diodes is used to provide a constant voltage of about 2.1 V. We want to calculate the percentage change in this regulated voltage caused by (a) a $\pm 10\%$ change in the power-supply voltage, and (b) connection of a 1-k Ω load resistance.

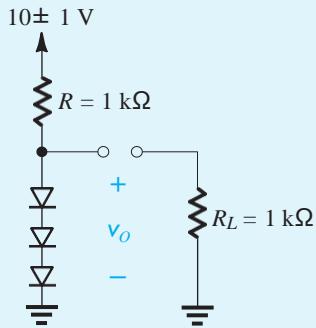


Figure 4.17 Circuit for Example 4.6.

Solution

With no load, the nominal value of the current in the diode string is given by

$$I = \frac{10 - 2.1}{1} = 7.9 \text{ mA}$$

2-Marks Questions

1. Define drift current?

Drift current is developed due to potential gradient and the Phenomenon found both in semiconductors & metals.

2. Define diffusion current?

What is meant by diffusion current in semiconductor?

In addition to the drift there may exist an additional current due to the transport of charges in a semiconductor. It is developed due to charge concentration gradient.

Such an additional current is due to the phenomenon called diffusion

3. Define the term transition capacitance?

When a diode is reversed biased, the width of the depletion region increases. Due to this the p region and n region act like a plates of capacitor while the depletion region acts as dielectric. Thus there exists a capacitance called transition capacitance.

4. What is avalanche breakdown?

Breaking of covalent bonds is due to collision of accelerated charge carriers having large velocities and kinetic energy with adjacent atoms. The process is called carrier multiplication. This occurs for zener diodes with VZ greater than 6V.

5. What is PN junction diode?

The two type of materials namely p-type and n-type are chemically combined with a special fabrication technique to form p-n junction. This p-n junction forms a semiconductor device called p-n junction diode.

6. Differentiate drift and diffusion current?.

Drift current:

Developed due to potential gradient.

Phenomenon found both in semiconductors & metals.

Diffusion current:

Developed due to charge concentration gradient.

Only in semiconductors.

8. What is meant by depletion region?

In PN junction, the diffusion of holes and electrons start initially. Near the junction, holes recombine in N-region to form immobile positive ions. Similarly electrons recombine in P-region to form immobile negative ions. With sufficient accumulation of such immobile ions on both sides, the diffusion stops. So near the junction, there exists a region in which immobile positive and negative charge reside while mobile charge carriers in this region get completely depleted. This region is called depletion region.

9. What is meant by Knee or Zener Knee?

At a certain reverse voltage, current through Zener diode increases rapidly. The change from low value to large value of current is very sharp and well defined. Such a sharp change in the reverse characteristics is called Knee or Zener Knee of the curve.

10. What is switching characteristics of diode?

When diode is switched from forward biased to the reverse biased state or vice versa, it takes finite time to attain a steady state. This time consists of Transient and an interval of time before the diode attains a steady state. The Behavior of the diode during this time is called switching characteristics of the diode.

11. Define and explain peak inverse voltage (PIV) of a Diode?

Peak inverse voltage 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 to its peak inverse voltage, the junction may be destroyed due to excessive heat.

12 .Define cut in voltage?

This is the voltage at which the forward bias curve abruptly increases from the smaller value. The cut in voltage of Ge is 0.3V & for Si is 0.7V.

13. Why Si is preferred over Ge in the manufacture of semiconductor devices?

The Si is preferred over Ge in the manufacture of semiconductor device because, Silicon has low leakage current .

Silicon has higher temperature stability The peak inverse voltage for Si is higher.

14. Explain the terms knee voltage & break down voltage with respect to diodes?

Knee voltage:It is the forward voltage of a PN diode at which the current through the junction starts increasing rapidly.

Breakdown Voltage:It is the reverse voltage of a PN junction diode at which the junction breaks down with sudden rise in the reverse current.

15. Name Some Donor And Acceptor Impurities?

Donor--arsenic, phosphorus, nitrogen

Acceptor--boron, aluminum, gallium

16. Define recovery time of a diode?

The recovery time is the time difference between the 10 percent point of the diode voltage and the time when this voltage reaches and remains within 10 percent of its final value.

17. Define forbidden energy gap?

Forbidden gap is the energy gap between valence and conduction bands. For insulators, the gap will be more and for conductors it is nil. For Si, the gap is 1.1eV and for Ge it is 0.7eV.

18. Write the diode current equation?

$$I_d = I_s [e^{V_d / \eta V_T} - 1]$$

I_d = diode current

I_s = reverse saturation current in amperes

V_d = applied voltage

η = 1 for germanium diode, 2 for silicon diode

V_T = Thermal voltage

19. . What are holes and electrons in a PN-diode?

An electron is negatively charged. The electrons in the outer shell determine the electrical characteristics of each particular type of atom.

A hole is defined as an absence of an electron in a shell where one could exist.

20. What is the difference between diffusion and drift current?

Drift Current	Diffusion current
1. Developed due to potential gradient.	1. Developed due to charge concentration gradient.
2. Phenomenon found both in Semiconductors and metals.	2. Only in semiconductors.

Essay Questions:

1. Define static and dynamic resistance of diode?

Draw the circuit diagram

Explanation about static and dynamic resistance.

2. With neat diagram, explain the operation of p-n junction diode considering different basing conditions ?

Draw the circuit diagram

Explanation about operation of p-n junction diode.

3. Draw the energy band diagrams of metals, semiconductors and Insulators?

Draw the circuit diagram

Explanation about different energy bands mentioned.

4. Explain about p-type and n-type semiconductors?

Explanation about p & n type semiconductors with diagrams.

5. Draw the energy band diagram of a p-n junction under open circuit condition and derive the expression for contact potential?

Draw the diagram,

Derive the expressions for contact potential.

6. Define transition capacitance in a diode and derive the expression for it?

Transition capacitance,

Derive the expression

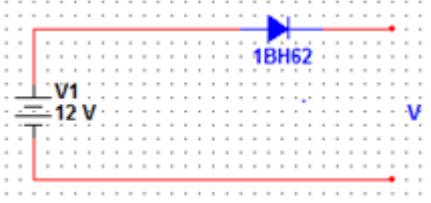
MCQ's:

- 1. How many junction/s do a diode consist?** (b)
a) 0 b) 1 c) 2 d) 3
- 2. If the positive terminal of the battery is connected to the anode of the diode, then it is known as** (a)
a) Forward biased
b) Reverse biased
c) Equilibrium
d) Schottky barrier
- 3. During reverse bias, a small current develops known as** (c)
a) Forward current
b) Reverse current
c) Reverse saturation current
d) Active current
- 4. If the voltage of the potential barrier is V_0 . A voltage V is applied to the input, at what moment will the barrier disappear?** (b)
a) $V < V_0$
b) $V = V_0$
c) $V > V_0$
d) $V \ll V_0$
- 5. What is the maximum electric field when $V_{bi}=2V$, $V_R=5V$ and width of the semiconductor is 7cm?** (b)
a)-100V/m
b) -200V/m
c) 100V/m
d) 200V/m
- 6. During the reverse biased of the diode, the back resistance decrease with the increase of the temperature. Is it true or false?** (a)
a) True
b) False
- 7. When the diode is reverse biased with a voltage of 6V and $V_{bi}=0.63V$. Calculate the total potential.** (b)
a) 6V
b) 6.63V
c) 5.27V
d) 0.63V

8. It is possible to measure the voltage across the potential barrier through a voltmeter? (b)

- a) True
- b) False

9. What will be the output of the following circuit? (Assume 0.7V drop across the diode) (c)



- a) 12V
- b) 12.7V
- c) 11.3V
- d) 0V

10. Which of the following formula represents the correct formula for width of the depletion region? (a)

- a) $W = \left\{ 2 \in \left(\frac{V_{bi} + V_R}{\epsilon} \right) \left[\frac{Na + Nd}{NaNd} \right] \right\}^{0.5}$
- b) $W = \left\{ 2 \in \left(\frac{V_{bi} - V_R}{\epsilon} \right) \left[\frac{Na + Nd}{NaNd} \right] \right\}^{0.5}$
- c) $W = \left\{ 2 \in \left(\frac{V_{bi} + V_R}{\epsilon} \right) \left[\frac{Na - Nd}{NaNd} \right] \right\}^{0.5}$
- d) $W = \left\{ 2 \in \left(\frac{V_{bi} - V_R}{\epsilon} \right) \left[\frac{Na - Nd}{NaNd} \right] \right\}^{0.5}$

UNIT – II

ZENER DIODES

UNIT - II

2.1 Zener Diodes

The very steep I-v curve that the diode exhibits in the breakdown region and the almost- constant voltage drop that this indicates suggest that diodes operating in the breakdown region can be used in the design of voltage regulators. From the previous section, the reader will recall that voltage regulators are circuits that provide a constant dc output voltage in the face of changes in their load current and in the system power-supply voltage. This in fact turns out to be an important application of diodes operating in the reverse breakdown region, and special diodes are manufactured to operate specifically in the breakdown region. Such diodes are called breakdown diodes or, more commonly, as noted earlier, Zener diodes. Figure 2.1 shows the circuit symbol of the Zener diode. In normal applications of Zener diodes, current flows into the cathode, and the cathode is positive with respect to the anode.

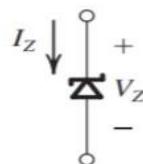


Figure 2.1 Circuit symbol for a Zener diode

2.1.1 Zener diode characteristics:

Figure 2.2 shows details of the diode i-v characteristic in the breakdown region. We observe that for currents greater than the knee current I_{ZK} (specified on the data sheet of the zener diode), the i-v characteristic is almost a straight line. The manufacturer usually specifies the voltage across the Zener diode V_Z at a specified test current, I_{ZT} . We

have indicated these parameters in Fig. 2.2 as the coordinates of the point labelled Q. Thus a 6.8-V zener diode will exhibit a 6.8-V drop at a specified test current of, say, 10 mA. As the current through the zener deviates from I_{ZT} , the voltage across it will change, though only slightly. Figure 2.2 shows that corresponding to current change the zener voltage changes by ΔV , which is related to ΔI by

$$\Delta V = r_z \Delta I$$

where r_z is the inverse of the slope of the almost-linear i-v curve at point Q. Resistance r_z is the incremental resistance of the zener diode at operating point Q. It is also known as the dynamic resistance of the zener, and its value is specified on the device data sheet. Typically, r_z is in the range of a few ohms to a few tens of ohms. Obviously, the lower the value of r_z is, the more constant the zener voltage remains as its current varies, and thus the more ideal its performance becomes in the design of voltage regulators. In this regard, we observe from Fig. 2.2 that while r_z remains low and almost constant over a wide range of current, its value

increases considerably in the vicinity of the knee. Therefore, as a general design guideline, one should avoid operating the zener in this low-current region.

Zener diodes are fabricated with voltages V_Z in the range of a few volts to a few hundred volts. In addition to specifying V_Z (at a particular current I_{ZT}), r_z , and I_{ZK} , the manufacturer

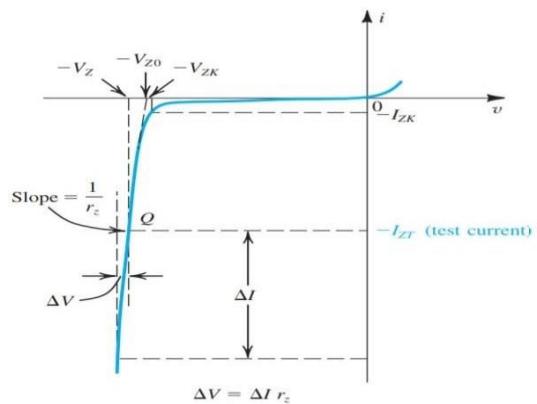


Figure 2.2 The diode i - v characteristic with the breakdown region shown in some detail

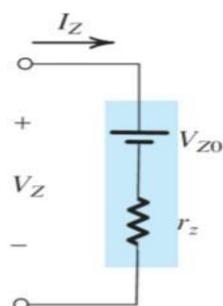


Figure 2.3 Model for Zener diode

also specifies the maximum power that the device can safely dissipate. Thus a 0.5-W, 6.8-V zener diode can operate safely at currents up to a maximum of about 70 mA. The almost-linear i - v characteristic of the zener diode suggests that the device can be modeled as indicated in Fig.

2.3. Here V_{Z0} denotes the point at which the straight line of slope $1/r_z$ intersects the voltage axis (refer to Fig. 2.2). Although V_{Z0} is shown in Fig. 2.2 to be slightly different from the knee voltage V_{ZK} , in practice their values are almost equal. The equivalent circuit model of Fig. 2.3 can be analytically described by

$$V_Z = V_{Z0} + r_z I_Z$$

and it applies for $I_Z > I_{ZK}$ and, obviously, $V_Z > V_{Z0}$.

2.1.2 Voltage Shunt Regulator:

We now illustrate, by way of an example, the use of zener diodes in the design of shunt regulators, so named because the regulator circuit appears in parallel (shunt) with the load.

2.1.3 Temperature Effects:

The dependence of the zener voltage V_z on temperature is specified in terms of its temperature coefficient T_C , or temco as it is commonly known, which is usually expressed in $\text{mV}/^\circ\text{C}$. The value of T_C depends on the zener voltage, and for a given diode the T_C varies with the operating current. Zener diodes whose V_z are lower than about 5 V exhibit a negative T_C . On the other hand, Zener diodes with higher voltages exhibit a positive T_C . The T_C of a zener diode with a V_z of about 5 V can be made zero by operating the diode at a specified current. Another commonly used technique for obtaining a reference voltage with low temperature coefficient is to connect a zener diode with a positive temperature coefficient of about $2 \text{ mV}/^\circ\text{C}$ in series with a forward-conducting diode. Since the forward-conducting diode has a voltage drop of

0.7 V and a T_C of about $-2 \text{ mV}/^\circ\text{C}$, the series combination will provide a voltage of $(V_z + 0.7)$ with a T_C of about zero.

2.2 Rectifier Circuits

One of the most important applications of diodes is in the design of rectifier circuits. A diode rectifier forms an essential building block of the dc power supplies required to power electronic equipment. A block diagram of such a power supply is shown in Fig. 2.4 As indicated, the power supply is fed from the 120-V (rms) 60-Hz ac line, and it delivers a dc voltage V_o (usually in the range of 4 V to 20 V) to an electronic circuit represented by the load block.

The dc voltage V_o is required to be as constant as possible in spite of variations in the ac line voltage and in the current drawn by the load. The first block in a dc power supply is the power transformer. It consists of two separate coils wound around an iron core that magnetically

couples the two windings. The primary winding, having N_1 turns, is connected to the 120-V ac supply, and the secondary winding, having N_2 turns, is connected to the circuit of the dc power supply. Thus, an ac voltage v_s of $120(N_2/N_1)$ V (rms) develops between the two terminals of the secondary winding.

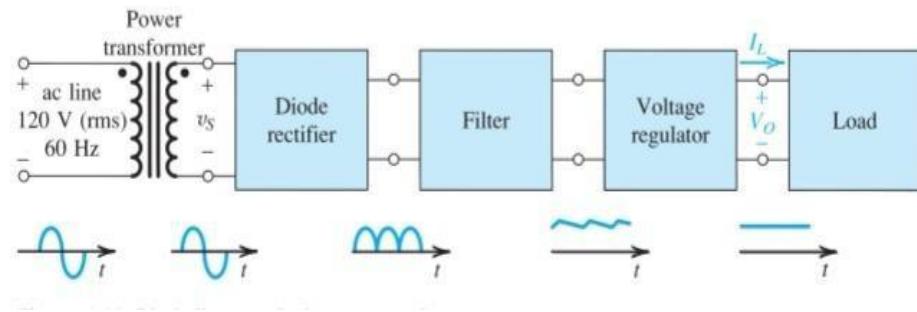


Figure 2.4 Block diagram of a dc power supply

By selecting an appropriate turns ratio (N_1/N_2) for the transformer, the designer can step the line voltage down to the value required to yield the particular dc voltage output of the supply. For instance, a secondary voltage of 8-V rms may be appropriate for a dc output of 5 V. This can be achieved with a 15:1 turns ratio.

In addition to providing the appropriate sinusoidal amplitude for the dc power supply, the power transformer provides electrical isolation between the electronic equipment and the power-line circuit. This isolation minimizes the risk of electric shock to the equipment user. The diode rectifier converts the input sinusoid v_s to a unipolar output, which can have the pulsating waveform indicated in Fig. 2.4. Although this waveform has a nonzero average or a dc component, its pulsating nature makes it unsuitable as a dc source for electronic circuits, hence the need for a filter. The variations in the magnitude of the rectifier output are considerably reduced by the filter block in Fig. 2.4.

In this section we shall study a number of rectifier circuits and a simple implementation of the output filter. The output of the rectifier filter, though much more constant than without the filter, still contains a time-dependent component, known as ripple. To reduce the ripple and to stabilize the magnitude of the dc output voltage against variations caused by changes in load current, a voltage regulator is employed. Such a regulator can be implemented using the zener shunt regulator configuration. Alternatively, and much more commonly at present, an integrated-circuit regulator can be used.

2.2.1 The half wave rectifier:

The half-wave rectifier utilizes alternate half-cycles of the input sinusoid. Figure 2.5(a) shows the circuit of a half-wave rectifier. This circuit was analysed assuming an ideal diode.

Using the more realistic constant-voltage-drop diode model, we obtain

$$v_o = 0, \quad v_s <$$

$$V_D \quad v_o = v_s - V_D,$$

$$v_s \geq$$

$$V_D$$

The transfer characteristic represented by these equations is sketched in Fig. 2.5(b), where $V_D = 0.7$ V or 0.8 V. Figure 2.5(c) shows the output voltage obtained when the input v_s is a sinusoid. In selecting diodes for rectifier design, two important parameters must be specified: the current-handling capability required of the diode, determined by the largest current the diode is expected to conduct, and the peak inverse voltage (PIV) that the diode must be able

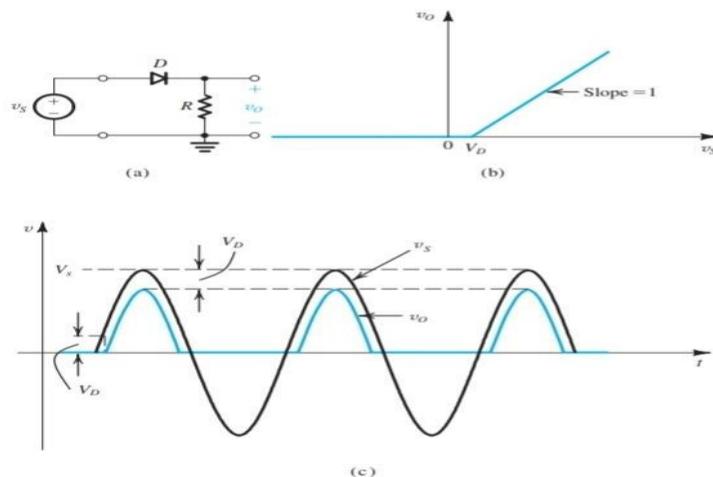


Figure 2.5 (a) Half-wave rectifier. (b) Transfer characteristic of the rectifier circuit. (c) Input and output waveforms.

to withstand without breakdown, determined by the largest reverse voltage that is expected to appear across the diode. In the rectifier circuit of Fig. 2.5(a), we observe that when v_s is negative the diode will be cut off and v_o will be zero. It follows that the PIV is equal to the peak of v_s ,

$$\text{PIV} = V_s$$

It is usually prudent, however, to select a diode that has a reverse breakdown voltage at least 50% greater than the expected PIV. Before leaving the half-wave rectifier, the reader should note two points. First, it is possible to use the diode exponential characteristic to determine the exact transfer characteristic of the rectifier. However, the amount of work involved is usually too great to be justified in practice. Of course, such an analysis can be easily done using a computer circuit-analysis program such as SPICE. Second, whether we analyze the circuit accurately or not, it should be obvious that this circuit does not function properly when the input signal is small. For instance, this circuit cannot be used to rectify an input sinusoid of 100-mV amplitude. For such an application one resorts to a so-called precision rectifier, a circuit utilizing diodes in conjunction with op amps.

2.2.2 The Full wave rectifier:

The full-wave rectifier utilizes both halves of the input sinusoid. To provide a unipolar output, it inverts the negative halves of the sine wave. One possible implementation is shown in Fig. 2.6(a). Here the transformer secondary winding is centre-tapped to provide two equal voltages v_s across the two halves of the secondary winding with the polarities indicated. Note that when the input line voltage (feeding the primary) is positive, both of the signals labelled v_s will be positive. In this case D1 will conduct and D2 will be reverse biased. The current through D1 will flow through R and back to the centre tap of the secondary. The circuit then behaves like a half-wave rectifier, and the output during the positive half-cycles when D1 conducts will be identical to that produced by the half-wave rectifier. Now, during the negative half-cycle of the ac line voltage, both of the voltages labelled v_s will be negative.

Thus, D1 will be cut off while D2 will conduct. The current conducted by D2 will flow through R and back to the centre tap. It follows that during the negative half-cycles while D2 conducts, the circuit behaves again as a half-wave rectifier. The important point, however, is that the current through R always flows in the same direction, and thus v_o will be unipolar, as indicated in Fig. 2.6(c). The output waveform shown is obtained by assuming that a conducting diode has a constant voltage drop V_D . Thus, the transfer characteristic of the full-wave rectifier takes the shape shown in Fig. 2.6(b). The full-wave rectifier obviously produces a more "energetic" waveform than that provided by the half-wave rectifier. In almost all rectifier applications, one opts for a full-wave type of some kind. To find the PIV of the diodes in the full-wave rectifier circuit, consider the situation during the positive half-cycles. Diode D1 is conducting, and D2 is cut off. The voltage at the cathode of D2 is v_o , and that at its anode is $-v_s$. Thus, the reverse voltage across D2 will be $(v_o + v_s)$, which will reach its maximum when v_o is at its peak value of $(V_s - V_D)$, and v_s is at its peak value of V_s ; thus,

$$PIV = 2V_s - V_D$$

which is approximately twice that for the case of the half-wave rectifier.

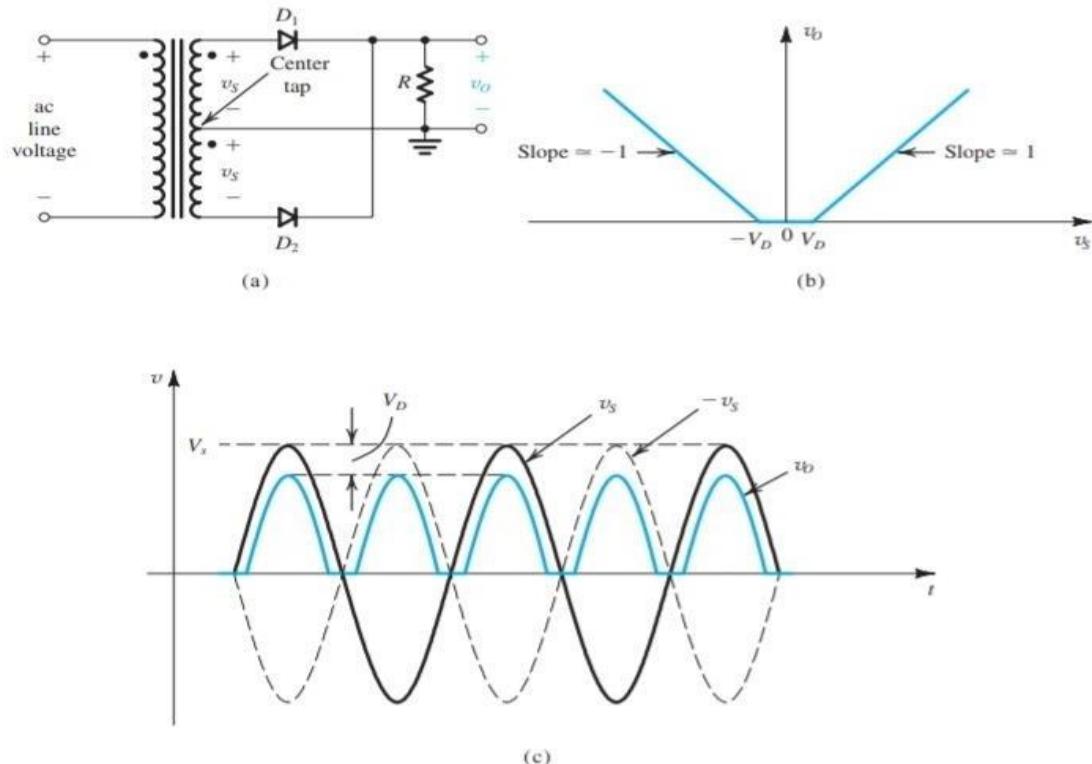


Figure 2.6 Full-wave rectifier utilizing a transformer with a centre-tapped secondary winding: (a) circuit; (b) transfer characteristic assuming a constant-voltage-drop model for the diodes; (c) input and output waveforms.

2.2.3 The Bridge rectifier:

An alternative implementation of the full-wave rectifier is shown in Fig. 2.7(a). This circuit, known as the bridge rectifier because of the similarity of its configuration to that of

7 | Pa

the Wheatstone bridge, does not require a centre-tapped transformer, a distinct advantage over the full-wave rectifier circuit of Fig. 2.7. The bridge rectifier, however, requires four diodes as compared to two in the previous circuit. This is not much of a disadvantage, because diodes are inexpensive and one can buy a diode bridge in one package.

The bridge-rectifier circuit operates as follows: During the positive half-cycles of the input voltage, v_s is positive, and thus current is conducted through diode D1, resistor R, and diode D2. Meanwhile, diodes D3 and D4 will be reverse biased. Observe that there are two diodes in series in the conduction path, and thus v_o will be lower than v_s by two diode drops (compared to one drop in the circuit previously discussed). This is somewhat of a disadvantage of the bridge rectifier. Next, consider the situation during the negative half-cycles of the input voltage. The secondary voltage v_s will be negative, and thus $-v_s$ will be positive, forcing current through D3, R, and D4. Meanwhile, diodes D1 and D2 will be reverse biased. The important point to note, though, is that during both half-cycles, current flows through R in the same direction (from right to left), and thus v_o will always be positive, as indicated in Fig. 2.7(b). To determine the peak inverse voltage (PIV) of each diode, consider the circuit during the positive half-cycles. The reverse voltage across D3 can be determined

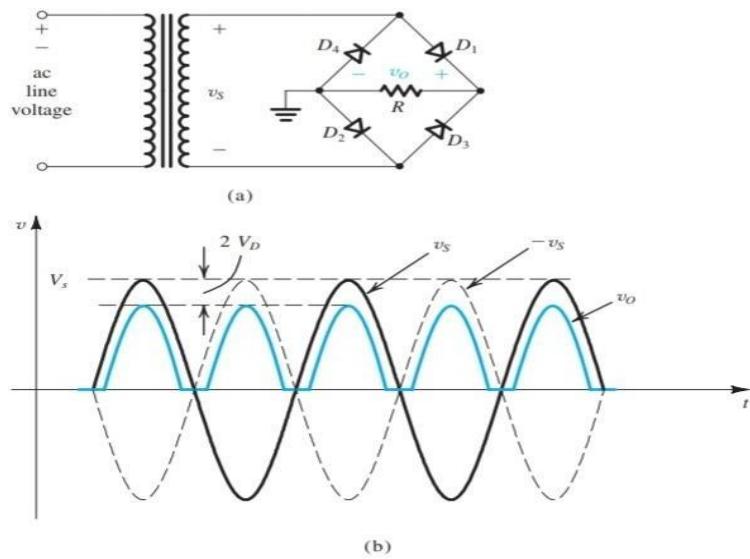


Figure 2.7 The bridge rectifier: (a) circuit; (b) input and output waveforms.

from the loop formed by D_3 , R , and D_2 as

$$v_{D3}(\text{reverse}) = v_O + v_{D2}(\text{forward})$$

Thus, the maximum value of v_{D3} occurs at the peak of v_O and is given by

$$\text{PIV} = V_s - 2V_D + V_D = V_s - V_D$$

Observe that here the PIV is about half the value for the full-wave rectifier with a centre-tapped transformer. This is another advantage of the bridge rectifier. Yet one more advantage of the bridge-rectifier circuit over that utilizing a centre-tapped transformer is that only about half as many turns are required for the secondary winding of the transformer. Another way of looking at this point can be obtained by observing that each half of the

secondary winding of the centre-tapped transformer is utilized for only half the time. These advantages have made the bridge rectifier the most popular rectifier circuit configuration.

2.2.4 The rectifier with a filter capacitor – Peak Rectifier:

The pulsating nature of the output voltage produced by the rectifier circuits discussed above makes it unsuitable as a dc supply for electronic circuits. A simple way to reduce the variation of the output voltage is to place a capacitor across the load resistor. It will be shown that this filter capacitor serves to reduce substantially the variations in the rectifier output voltage.

To see how the rectifier circuit with a filter capacitor works, consider first the simple circuit shown in Fig. 2.8. Let the input v_i be a sinusoid with a peak value V_p , and assume the diode to be ideal. As v_i goes positive, the diode conducts and the capacitor is charged so that $v_o = v_i$. This situation continues until v_i reaches its peak value V_p . Beyond the peak, as v_i decreases, the diode becomes reverse biased and the output voltage remains constant at the value V_p . In fact, theoretically speaking, the capacitor will retain its charge and hence its voltage indefinitely, because there is no way for the capacitor to discharge. Thus, the circuit provides a dc voltage output equal to the peak of the input sine wave. This is a very encouraging result in view of our desire to produce a dc output.

Next, we consider the more practical situation where a load resistance R is connected across the capacitor C , as depicted in Fig. 2.8(a). However, we will continue to assume the diode to be ideal. As before, for a sinusoidal input, the capacitor charges to the peak of the input V_p . Then the diode cuts off, and the capacitor discharges through the load resistance R . The capacitor discharge will continue for almost the entire cycle, until the time at which v_i

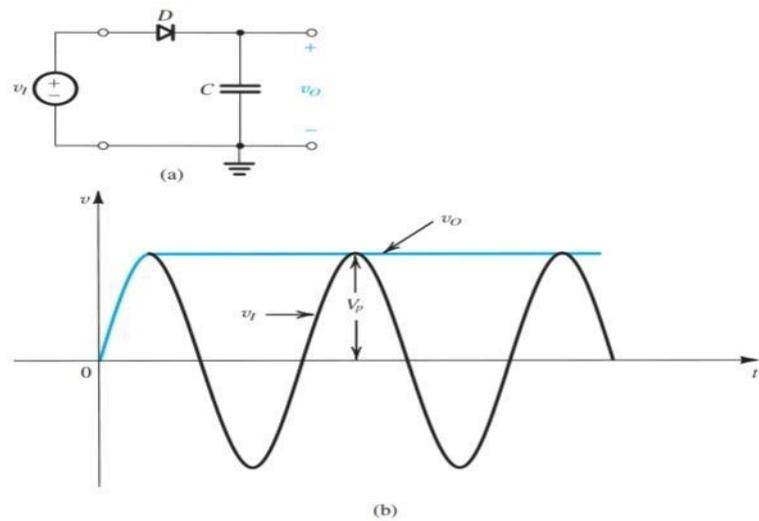


Figure 2.8 (a) A simple circuit used to illustrate the effect of a filter capacitor. (b) Input and output waveforms assuming an ideal diode. Note that the circuit provides a dc voltage equal to the peak of the input sine wave. The circuit is therefore known as a peak rectifier or a peak detector.

exceeds the capacitor voltage. Then the diode turns on again and charges the capacitor up to the peak of v_I , and the process repeats itself. Observe that to keep the output voltage from

decreasing too much during capacitor discharge, one selects a value for C so that the time constant CR is much greater than the discharge interval. We are now ready to analyze the circuit in detail. Figure 2.9(b) shows the steady-state input and output voltage waveforms under the assumption that $CR \gg T$, where T is the period of the input sinusoid. The waveforms of the load current

$$i_L = V_0/R$$

and of the diode current (when it is conducting)

$$\begin{aligned} i_D &= i_C + i_L \\ &= C dv_i / dt + i_L \end{aligned}$$

are shown in Fig. 2.9(c). The following observations are in order:

1. The diode conducts for a brief interval, Δt , near the peak of the input sinusoid and supplies the capacitor with charge equal to that lost during the much longer discharge interval. The latter is approximately equal to the period T .

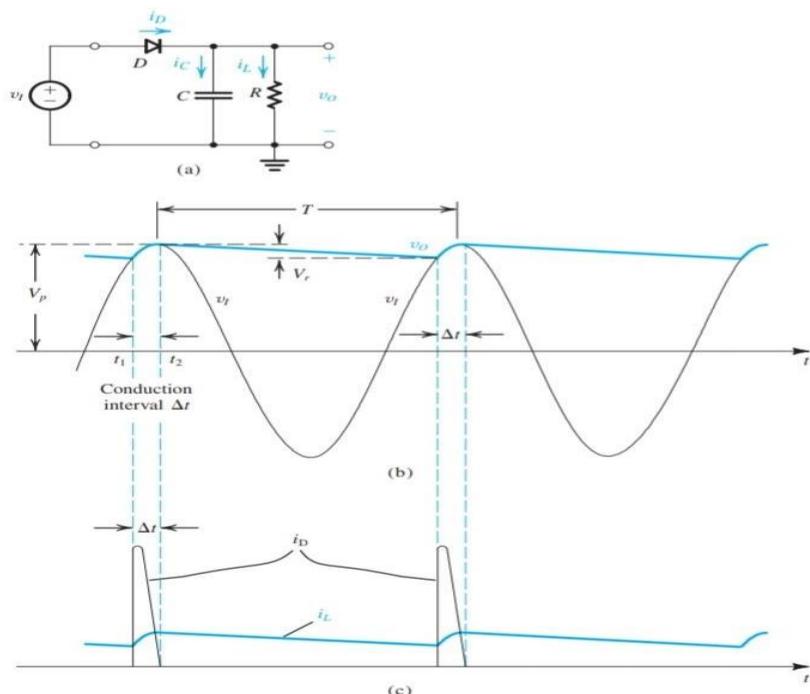


Figure 2.9 Voltage and current waveforms in the peak-rectifier circuit with $CR \gg T$. The diode is assumed ideal.

2. Assuming an ideal diode, the diode conduction begins at time t_1 , at which the input v_i equals the exponentially decaying output v_o . Conduction stops at t_2 shortly after the peak of v_i ; the exact value of t_2 can be determined by setting $i_D = 0$.
3. During the diode-off interval, the capacitor C discharges through R , and thus v_o decays exponentially with a time constant CR . The discharge interval begins just past the peak of v_i . At the end of the discharge interval, which lasts for almost the entire period T , $v_o = V_p - V_r$, where V_r is the peak-to-peak ripple voltage. When $CR \gg T$, the value of V_r is small.

4. When V_r is small, v_o is almost constant and equal to the peak value of v_i . Thus, the dc output voltage is approximately equal to V_p . Similarly, the current i_L is almost constant, and its dc component I_L is given by

$$I_L = V_p / R$$

If desired, a more accurate expression for the output dc voltage can be obtained by taking the average of the extreme values of v_o ,

$$V_o = V_p - 1/2V_r$$

With these observations in hand, we now derive expressions for V_r and for the average and peak values of the diode current. During the diode-off interval, v_o can be expressed as

$$v_o = V_p e^{-t/CR}$$

At the end of the discharge interval we have

$$V_p - V_r V_p e^{-T/CR}$$

Now, since $CR \gg T$, we can use the approximation $e^{-T/CR} \approx 1 - T/CR$ to obtain

$$V_r \approx V_p T / CR$$

We observe that to keep V_r small we must select a capacitance C so that $CR \gg T$. The ripple voltage V_r can be expressed in terms of the frequency $f = 1/T$ as

$$V_r = V_p / fCR$$

Using Eq., we can express V_r by the alternate expression

$$V_r = I_L / fC$$

Note that an alternative interpretation of the approximation made above is that the capacitor discharges by means of a constant current $I_L = V_p/R$. This approximation is valid as long as $V_r \ll V_p$.

Assuming that diode conduction ceases almost at the peak of v_i , we can determine the conduction interval Δt from

$$V_p \cos(\omega\Delta t) = V_p - V_r$$

where $\omega = 2\pi f = 2\pi/T$ is the angular frequency of v_i . Since $(\omega\Delta t)$ is a small angle, we can employ the approximation $\cos(\omega\Delta t) \approx 1 - 1/2 (\omega\Delta t)^2$ to obtain

$$\omega\Delta t \sqrt{(2V_r/V_p)}$$

We note that when $V_r \ll V_p$, the conduction angle $\omega\Delta t$ will be small, as assumed.

To determine the average diode current during conduction, i_{Dav} , we equate the charge that the diode supplies to the capacitor,

$$Q_{supplied} = i_{Cav}\Delta t$$

Where from

$$i_{Cav} = i_{Dav} - I_L$$

to the charge that the capacitor loses during the discharge interval,

$$Q_{lost} = CV_r$$

to obtain, using above Eqs,

$$i_{Dav} = I_L (1 + \pi \sqrt{(2V_p/V_r)})$$

Observe that when $V_r \ll V_p$, the average diode current during conduction is much greater than the dc load current. This is not surprising, since the diode conducts for a very short interval and must replenish the charge lost by the capacitor during the much longer interval in which it is discharged by i_L .

The peak value of the diode current, $i_{D\max}$, can be determined by evaluating the expression at the onset of diode conduction—that is, at $t = t_1 = -\Delta t$ (where $t = 0$ is at the peak). Assuming that i_L is almost constant at the value, we obtain

$$i_{D\max} = I_L 1 + 2\pi 2V_p/V_r$$

we see that for $V_r \ll V_p$, $i_{D\max} \approx 2i_{Dav}$, which correlates with the fact that the waveform of i_D is almost a right-angle triangle.

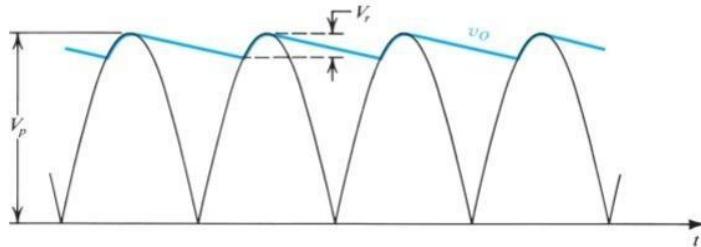


Figure 2.10 Waveforms in the full-wave peak rectifier

The circuit of Fig. 2.10(a) is known as a half-wave peak rectifier. The full-wave rectifier circuits of Figs. 2.6(a) and 2.7(a) can be converted to peak rectifiers by including a capacitor across the load resistor. As in the half-wave case, the output dc voltage will be almost equal to the peak value of the input sine wave. The ripple frequency, however, will be twice that of the input. The peak-to-peak ripple voltage, for this case, can be derived using a procedure identical to that above but with the discharge period T replaced by $T/2$, resulting in

$$V_r = V_p / 2fCR$$

While the diode conduction interval, Δt , the average and peak currents in each of the diodes will be given by

$$i_{Dav} = I_L (1 + \pi$$

$$\sqrt{V_p/2V_r}) i_{Dmax} = I_L$$

$$(1 + 2\pi\sqrt{V_p/2V_r})$$

Comparing these expressions with the corresponding ones for the half-wave case, we note that for the same values of V_p , f , R , and V_r (and thus the same I_L), we need a capacitor half the size of that required in the half-wave rectifier. Also, the current in each diode in the full-wave rectifier is approximately half that which flows in the diode of the half-wave circuit.

The analysis above assumed ideal diodes. The accuracy of the results can be improved by taking the diode voltage drop into account. This can be easily done by replacing the peak voltage V_p to which the capacitor charges with $(V_p - V_D)$ for the half-wave circuit and the full-

wave circuit using a centre-tapped transformer and with $(V_p - 2V_D)$ for the bridge-rectifier case.

We conclude this section by noting that peak-rectifier circuits find application in signal-processing systems where it is required to detect the peak of an input signal. In such a case, the circuit is referred to as a peak detector. A particularly popular application of the peak detector is in the design of a demodulator for amplitude-modulated (AM) signals. We shall not discuss this application further here.

2.3 Clipping and Clamping circuits

In this section, we shall present additional nonlinear circuit applications of diodes.

2.3.1 Limiter circuits:

Figure 2.11 shows the general transfer characteristic of a limiter circuit. As indicated, for inputs in a certain range, $L_-/K \leq v_i \leq L_+/K$, the limiter acts as a linear circuit, providing an output proportional to the input, $v_o = K v_i$. Although in general K can be greater than 1, the

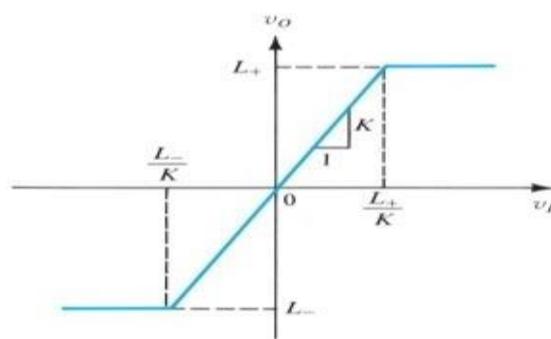


Figure 2.11 General transfer characteristic for a limiter circuit

circuits discussed in this section have $K \leq 1$ and are known as passive limiters. If v_i exceeds the upper threshold (L_+/K), the output voltage is limited or clamped to

the upper limiting level L_+ . On the other hand, if v_i is reduced below the lower limiting threshold L_-/K , the output voltage v_o is limited to the lower limiting level L_- .

The general transfer characteristic of Fig. 2.11 describes a double limiter—that is, a limiter that works on both the positive and negative peaks of an input waveform. Single limiters, of course, exist. Finally, note that if an input waveform such as that shown in Fig. 2.12 is fed to a double limiter, its two peaks will be clipped off. Limiters therefore are sometimes referred to as clippers.

The limiter whose characteristics are depicted in Fig. 2.11 is described as a hard limiter. Soft limiting is characterized by smoother transitions between the linear region and the saturation regions and a slope greater than zero in the saturation regions, as illustrated in Fig. 2.13. Depending on the application, either hard or soft limiting may be preferred.

Limiters find application in a variety of signal-processing systems. One of their simplest applications is in limiting the voltage between the two input terminals of an op amp to a value lower than the breakdown voltage of the transistors that make up the input stage of the op-amp circuit. We will have more to say on this and other limiter applications at later points in this book.

Diodes can be combined with resistors to provide simple realizations of the limiter function. A number of examples are depicted in Fig. 2.14. In each part of the figure both the circuit and its transfer characteristic are given. The transfer characteristics are obtained using the constant-voltage-drop ($V_D = 0.7$ V) diode model but assuming a smooth transition between the linear and saturation regions of the transfer characteristic. The circuit in Fig. 2.14(a) is that of the half-wave rectifier except that here the output is taken across the diode. For $v_i < 0.5$ V, the diode is cut off, no current flows, and the voltage drop across R is zero; thus $v_o = v_i$. As v_i exceeds 0.5 V, the diode turns on, eventually limiting

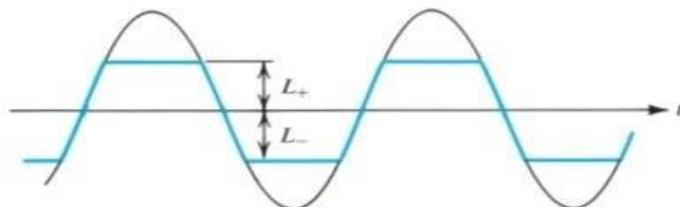


Figure 2.12 Applying a sine wave to a limiter can result in clipping off its two peaks

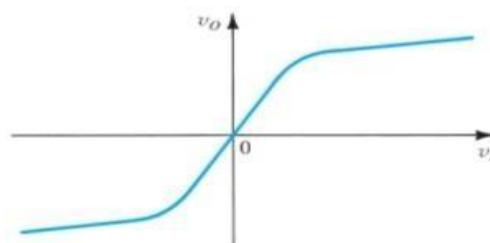


Figure 2.13 Soft Limiting

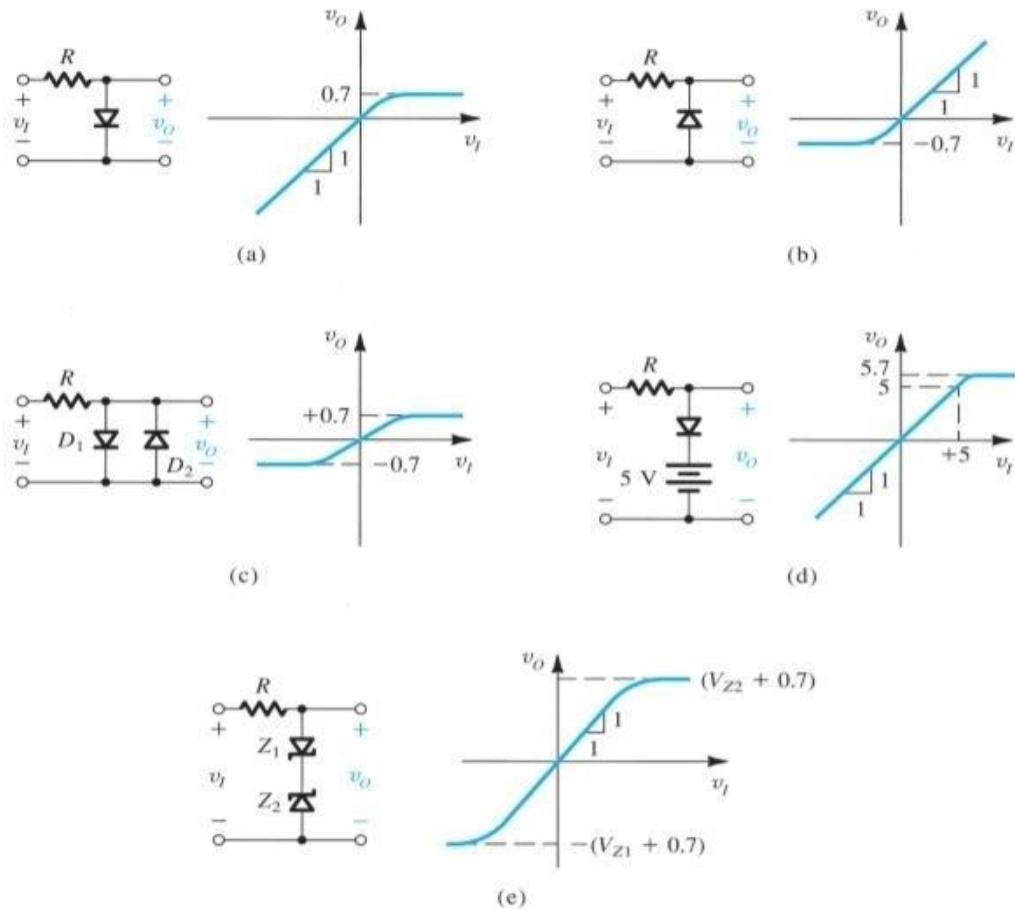


Figure 2.14 A variety of basic Limiting circuits

v_o to one diode drop (0.7 V). The circuit of Fig. 2.14(b) is similar to that in Fig. 2.14(a) except that the diode is reversed.

Double limiting can be implemented by placing two diodes of opposite polarity in parallel, as shown in Fig. 2.14(c). Here the linear region of the characteristic is obtained for

$-0.5 \text{ V} \leq v_I \leq 0.5 \text{ V}$. For this range of v_I , both diodes are off and $v_o = v_I$. As v_I exceeds 0.5 V , D_1 turns on and eventually limits v_o to $+0.7 \text{ V}$. Similarly, as v_I goes more negative than -0.5 V , D_2 turns on and eventually limits v_o to -0.7 V .

The thresholds and saturation levels of diode limiters can be

controlled by using strings of diodes and/or by connecting a dc voltage in series with the diode(s). The latter idea is illustrated in Fig. 2.14(d). Finally, rather than strings of diodes, we may use two Zener diodes in series, as shown in Fig. 2.14(e). In this circuit, limiting occurs in the positive direction at a voltage of $V_{Z2} + 0.7$, where 0.7 V represents the voltage drop across zener diode Z1 when conducting in the forward direction. For negative inputs, Z1 acts as a zener, while Z2 conducts in the forward direction. It should be mentioned that pairs of zener diodes connected in series are available commercially for applications of this type under the name double-anode zener. More flexible limiter circuits are possible if op amps are combined with diodes and resistors.

2.3.2 The Clamped capacitor or DC Restorer:

If in the basic peak-rectifier circuit, the output is taken across the diode rather than across the capacitor, an interesting circuit with important applications results. The circuit, called a dc restorer, is shown in Fig. 2.15 fed with a square wave. Because of the polarity in which the diode is connected, the capacitor will charge to a voltage v_C with the polarity indicated in Fig. 2.15 and equal to the magnitude of the most negative peak of the input signal. Subsequently, the diode turns off and the capacitor retains its voltage indefinitely.

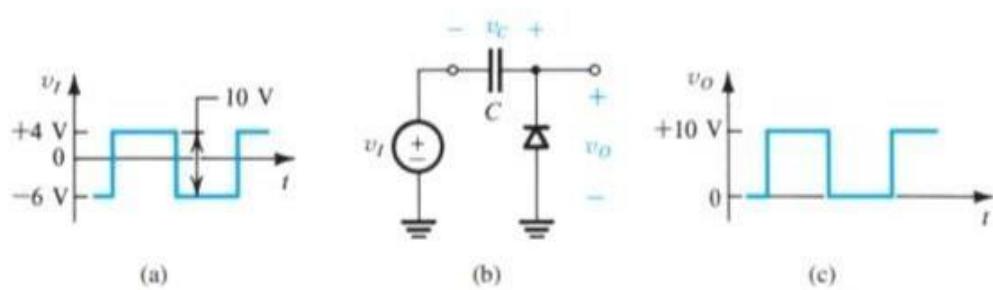


Figure 2.15 The clamped capacitor or dc restorer with a square-wave input and no load

If, for instance, the input square wave has the arbitrary levels -6 V and $+4\text{ V}$, then v_C will be equal to 6 V . Now, since the output voltage v_O is given by

$$v_O = v_I + v_C$$

it follows that the output waveform will be identical to that of the input, except that it is shifted upward by v_C volts. In our example the output will thus be a square wave with levels of 0 V and $+10\text{ V}$.

Another way of visualizing the operation of the circuit in Fig. 2.15 is to note that because the diode is connected across the output with the polarity shown, it prevents the output voltage from going below 0 V (by

conducting and charging up the capacitor, thus causing the output to rise to 0 V), but this connection will not constrain the positive excursion of v_o . The output waveform will therefore have its lowest peak clamped to 0 V, which is why the circuit is called a clamped capacitor. It should be obvious that reversing the diode polarity will provide an output waveform whose highest peak is clamped to 0 V. In either case, the output waveform will have a finite average value or dc component. This dc component is entirely unrelated to the average value of the input waveform. As an application, consider a pulse signal being transmitted through a capacitively coupled or ac-coupled system. The capacitive coupling will cause the pulse train to lose whatever dc component it originally had. Feeding the resulting pulse waveform to a clamping circuit provides it with a well-determined dc component, a process known as dc restoration. This is why the circuit is also called a dc restorer.

Restoring dc is useful because the dc component or average value of a pulse waveform is an effective measure of its duty cycle.⁵ The duty cycle of a pulse waveform can be modulated (in a process called pulse width modulation) and made to carry information. In

such a system, detection or demodulation could be achieved simply by feeding the received pulse waveform to a dc restorer and then using a simple RC low-pass filter to separate the average of the output waveform from the superimposed pulses.

When a load resistance R is connected across the diode in a clamping circuit, as shown in Fig. 2.16, the situation changes significantly. While the output is above ground, a current must flow in R . Since at this time the diode is off, this current obviously comes from the capacitor, thus causing the capacitor to discharge and the output voltage to fall.

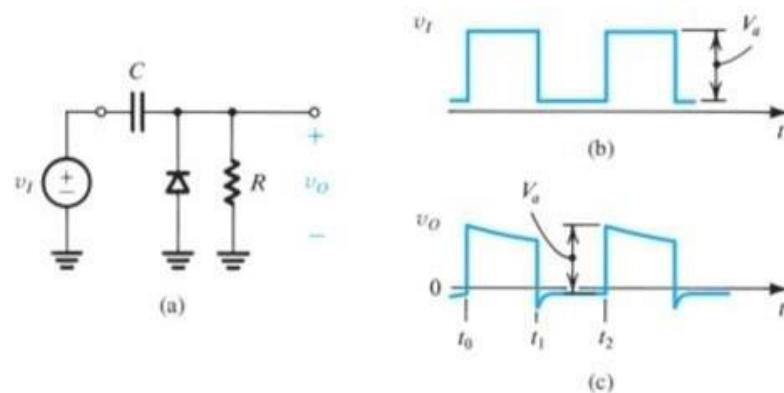


Figure 2.16 The clamped capacitor with a load resistor R

This is shown in Fig. 2.16 for a square-wave input. During the interval t_0 to t_1 , the output voltage falls exponentially with time constant CR . At t_1 the input decreases by V_a volts, and the output attempts to follow. This causes the diode to conduct heavily and to quickly charge the capacitor. At the end of the interval t_1 to t_2 , the output voltage would normally be a few tenths of a volt negative (e.g., -0.5 V). Then, as the input rises by V_a volts (at t_2), the output follows, and the cycle repeats itself. In the steady state the charge lost by the capacitor during the interval t_0 to t_1 is recovered during the interval t_1 to t_2 . This charge

equilibrium enables us to calculate the average diode current as well as the details of the output waveform.

2.3.3 Voltage Doubler:

Figure 2.17(a) shows a circuit composed of two sections in cascade: a clamped capacitor formed by C_1 and D_1 , and a peak rectifier formed by D_2 and C_2 . When excited by a sinusoid of amplitude V_p the clamping section provides the voltage waveform v_{D1} shown, assuming ideal diodes, in Fig. 2.17(b). Note that while the positive peaks are clamped to 0 V, the negative peak reaches $-2V_p$. In response to this waveform, the peak-detector section provides across capacitor C_2 a dc voltage equal to the negative peak of v_{D1} , that is, $-2V_p$. Because the output voltage is double the input peak, the circuit is known as a voltage doubler. The technique can be extended to provide output dc voltages that are higher multiples of V_p .

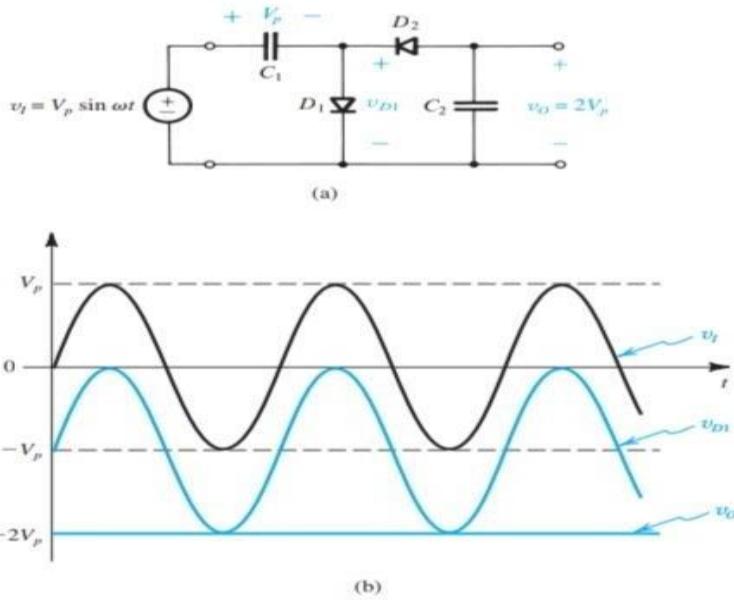


Figure 2.17 Voltage doubler: (a) circuit; (b) waveforms of the input voltage, the voltage across D_1 , and the output voltage $v_o = -2V_p$

2.4 Special diode types

In this section, we discuss briefly some important special types of diodes.

2.4.1 UJT:

A Uni Junction Transistor (UJT) is a three-lead electronic semiconductor device with only one junction that acts exclusively as an electrically controlled switch.

The UJT is not used as a linear amplifier. It is used in free-running oscillators, synchronized or triggered oscillators, and pulse generation circuits at low to moderate frequencies (hundreds of kilo Hertz). It is widely used in the triggering circuits for silicon-controlled rectifiers. In the 1960s, the low cost per unit, combined with its unique characteristic, warranted its use in a wide variety of applications like oscillators, pulse generators, saw-tooth generators, triggering circuits, phase control, timing circuits, and voltage- or current-regulated supplies. The original unijunction transistor types are now considered obsolete, but a later multi-layer device, the programmable unijunction transistor, is still widely available.

2.4.2 Schottky barrier diode:

The Schottky-barrier diode (SBD) is formed by bringing metal into contact with a moderately doped n-type semiconductor material. The resulting metal-semiconductor junction behaves like a diode, conducting current in one direction (from the metal anode to the semiconductor cathode) and acting as an open circuit in the other, and is known as the Schottky-barrier diode or simply the Schottky diode. In fact, the current-voltage characteristic of the SBD is remarkably similar to that of a PN-junction diode, with two important exceptions:

1. In the SBD, current is conducted by majority carriers (electrons). Thus, the SBD does not exhibit the minority-carrier charge-storage effects found in forward-biased pn junctions. As a result, Schottky diodes can be switched from on to off, and vice versa, much faster than is possible with PN-junction diodes.

2. The forward voltage drop of a conducting SBD is lower than that of a PN-junction diode. For example, an SBD made of silicon exhibits a forward voltage drop of 0.3 V to 0.5 V, compared to the 0.6 V to 0.8 V found in silicon PN-junction diodes. SBDs can also be made of gallium arsenide (GaAs) and, in fact, play an important role in the design of GaAs circuits.⁶ Gallium-arsenide SBDs exhibit forward voltage drops of about 0.7 V.

Apart from GaAs circuits, Schottky diodes find application in the design of a special form of bipolar-transistor logic circuits, known as Schottky-TTL, where TTL stands for transistor-transistor logic.

Before leaving the subject of Schottky-barrier diodes, it is important to note that not every metal-semiconductor contact is a diode. In fact, metal is commonly deposited on the semiconductor surface in order to make terminals for the semiconductor devices and to connect different devices in an integrated-circuit chip. Such metal-semiconductor contacts are known as ohmic contacts to distinguish them from the rectifying contacts that result in SBDs. Ohmic contacts are usually made by depositing metal on very heavily doped (and thus low-resistivity) semiconductor regions. (Recall that SBDs use moderately doped material.)

2.4.3 Varactor diode:

Previously we learned that reverse-biased PN junctions exhibit a charge-storage effect that is modelled with the depletion-layer or junction capacitance C_j . As indicates, C_j is a function of the reverse-bias voltage V_R . This dependence turns out to be useful in a number of applications, such as the automatic tuning of radio receivers. Special diodes are therefore fabricated to be used as voltage-variable capacitors known as varactors. These devices are optimized to make the capacitance a strong function of voltage by arranging the grading coefficient m is 3 or 4.

2.4.4 Photo diode:

If a reverse-biased pn junction is illuminated—that is, exposed to incident light—the photons impacting the junction cause covalent bonds to break,

and thus electron-hole pairs are generated in the depletion layer. The electric field in the depletion region then sweeps the liberated electrons to the n side and the holes to the p side, giving rise to a reverse current across the junction. This current, known as photocurrent, is proportional to the intensity of the incident light. Such a diode, called a photodiode, can be used to convert light signals into electrical signals.

Photodiodes are usually fabricated using a compound semiconductor such as gallium arsenide. The photodiode is an important component of a growing family of circuits known as optoelectronics or photonics. As the name implies, such circuits utilize an optimum combination of electronics and optics for signal processing, storage, and transmission. Usually, electronics is the preferred means for signal processing, whereas optics is most suited for transmission and storage. Examples include fibre-optic transmission of telephone and television signals and the use of optical storage in CD-ROM computer discs. Optical transmission provides very wide bandwidths and low signal attenuation. Optical storage allows vast amounts of data to be stored reliably in a small space.

Finally, we should note that without reverse bias, the illuminated photodiode functions as a solar cell. Usually fabricated from low-cost silicon, a solar cell converts light to electrical energy.

2.4.5 Light Emitting diode (LED):

The light-emitting diode (LED) performs the inverse of the function of the photodiode; it converts a forward current into light. In a forward-biased pn junction, minority carriers are injected across the junction and diffuse into the p and n regions. The diffusing minority carriers then recombine with the majority carriers. Such recombination can be made to give rise to light emission. This can be done by fabricating the pn junction using a semiconductor of the type known as direct-bandgap materials. Gallium arsenide belongs to this group and can thus be used to fabricate light-emitting diodes.

The light emitted by an LED is proportional to the number of recombinations that take place, which in turn is proportional to the forward current in the diode.

LEDs are very popular devices. They find application in the design of numerous types of displays, including the displays of laboratory instruments such as digital voltmeters. They can be made to produce light in a variety of colours. Furthermore, LEDs can be designed so as to produce coherent light with a very narrow bandwidth. The resulting device is a laser diode. Laser diodes find application in optical communication systems and in DVD players, among other things.

Combining an LED with a photodiode in the same package results in a device known as an optoisolator. The LED converts an electrical signal applied to the optoisolator into light, which the photodiode detects and converts back to an electrical signal at the output of the optoisolator. Use of the optoisolator provides complete electrical isolation between the electrical circuit that is connected to the isolator's input and the circuit that is connected to its output. Such isolation can be useful in reducing the effect of electrical interference on signal transmission within a system, and thus optoisolators are frequently employed in the design of digital systems. They can also be used in the design of medical instruments to reduce the risk of electrical shock to patients. Note that the optical coupling between an LED and a photodiode need not be accomplished inside a small

package. Indeed, it can be implemented over a long distance using an optical fibre, as is done in fibre-optic communication links.

Bipolar Junction Transistors (BJT)

Three-terminal devices are far more useful than two-terminal ones, because they can be used in a multitude of applications, ranging from signal amplification to the design of digital logic and memory circuits. The basic principle involved is the use of the voltage between two terminals to control the current flowing in the third terminal. In this way, a three-terminal device can be used to realize a controlled source, is the basis for amplifier design. Also, in the extreme, the control signal can be used to cause the current in the third terminal to change from zero to a large value, thus allowing the device to act as a switch.

The switch is the basis for the realization of the logic inverter, the basic element of digital circuits. The invention of the BJT in 1948 at the Bell Telephone Laboratories ushered in the era of solid-state circuits. The result was not just the replacement of vacuum tubes by transistors in radios and television sets but the eruption of an electronics revolution that led to major changes in the way we work, play, and indeed, live. The invention of the transistor also eventually led to the dominance of information technology and the emergence of the knowledge-based economy.

The bipolar transistor enjoyed nearly three decades as the device of choice in the design of both discrete and integrated circuits. Although the MOSFET had been known very early on, it was not until the 1970s and 1980s that it became a serious competitor to the BJT. By 2014, the MOSFET was undoubtedly the most widely used electronic device, and CMOS technology the technology of choice in the design of integrated circuits. Nevertheless, the BJT remains a significant device that excels in certain applications.

The BJT remains popular in discrete-circuit design, where it is used together with other discrete components such as resistors and capacitors to implement circuits that are assembled on printed-circuit boards (PCBs). Here we note the availability of a very wide selection of BJT types that fit nearly every conceivable application. As well, the BJT is still the preferred device in some very demanding analog and digital integrated-circuit applications. This is especially true in very-high-frequency and high-speed circuits. In particular, a very-high-speed digital logic-circuit family based on bipolar transistors, namely, emitter-coupled logic, is still in use (Chapter 15). Finally, bipolar transistors can be combined with MOSFETs to create innovative circuits that take advantage of the high-input-impedance and low-power operation of MOSFETs and the very-high-frequency operation and high-current-driving capability of bipolar transistors. The resulting technology is known as BiCMOS, and it is finding increasingly larger areas of application.

We shall start with a description of the physical operation of the BJT. Though simple, this physical description provides considerable insight regarding the performance of the transistor as a circuit element. We then quickly move from describing current flow in terms of electrons and holes to a study of the transistor terminal characteristics. Circuit models for transistor operation in different modes will be developed and utilized in the analysis and design of transistor circuits. The main objective of this chapter is to develop in the reader a high degree of familiarity with the BJT. Thus, it lays the foundation for the use of the BJT in amplifier design.

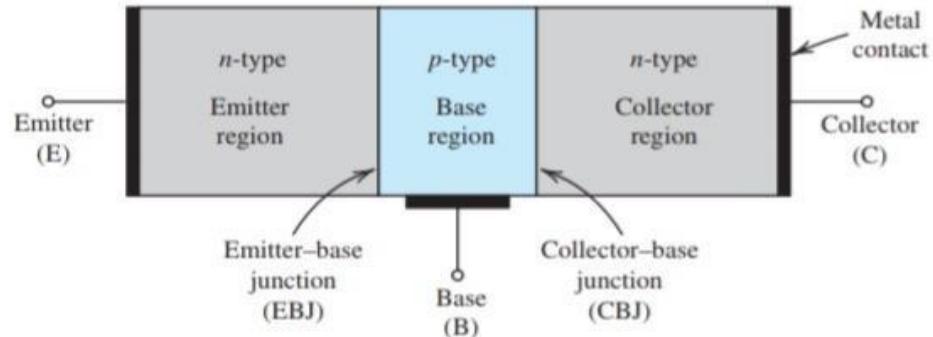
2.5 Physical Operation

2.5.1 Simplified structure and modes of operation:

Figure 2.18 shows a simplified structure for the BJT. A practical transistor structure will be shown later. As shown in Fig. 2.18, the BJT consists of three semiconductor regions: the emitter region (n type), the base region (p type), and the collector region (n type). Such a transistor is called npn transistor. Another transistor, a dual of the npn as shown in Fig. 2.19, has a p-type emitter, an n-type base, and a p-type collector, and is appropriately called a pnp transistor.

A terminal is connected to each of the three semiconductor regions of the transistor, with the terminals labelled emitter (E), base (B), and collector (C). The transistor consists of two pn junctions, the emitter-base junction (EBJ) and the collector-base junction (CBJ). Depending on the bias condition (forward or reverse) of each of these junctions, different modes of operation of the BJT are obtained, as shown in Table. The active mode is the one used if the transistor is to operate as an amplifier. Switching applications (e.g., logic circuits)

utilize both the cut-off mode and the saturation mode. As the name implies, in the cut-off mode no current flows because both junctions are reverse biased. As we will see shortly, charge carriers of both polarities—



that is, electrons and holes—participate in the current-conduction process in a bipolar transistor, which is the reason for the name bipolar.

Figure 2.18 A simplified structure of NPN transistor

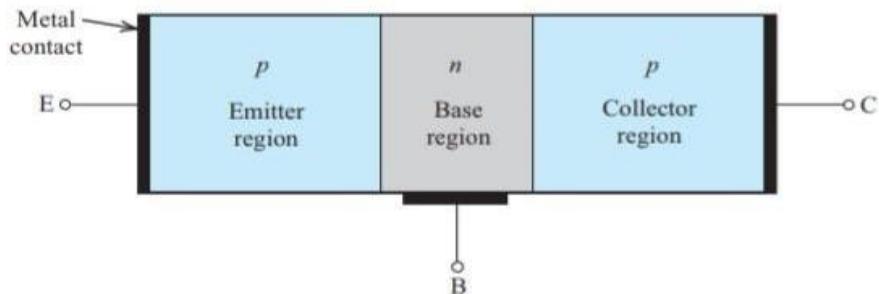


Figure 2.19 A simplified structure of PNP transistor

Table 2.1 BJT Modes of Operation

Mode	EBJ	CBJ
Cutoff	Reverse	Reverse
Active	Forward	Reverse
Saturation	Forward	Forward

2.5.2 Operation of the npn transistor in the active mode:

Of the three modes of operation of the BJT, the active mode is the most important. Therefore, we begin our study of the BJT by considering its physical operation in the active mode.² This situation is illustrated in Fig. 2.20 for the npn transistor. Two external voltage sources (shown as batteries) are used to establish the required bias conditions for active-mode operation. The voltage V_{BE} causes the p-type base to be higher in potential than the n-type emitter, thus forward biasing the emitter-base junction. The collector-base voltage V_{CB}

causes the n-type collector to be at a higher potential than the p-type base, thus reverse biasing the collector-base junction.

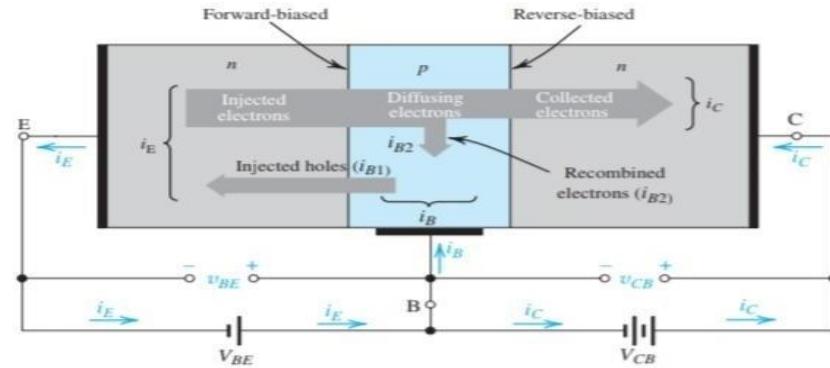


Figure 2.20 Current flow in npn transistor biased to operate in the active mode. (Reverse current components due to drift of thermally generated minority carriers are not shown.)

Current Flow The forward bias on the emitter-base junction will cause current to flow across this junction. Current will consist of two components: electrons injected from the emitter into the base, and holes injected from the base into the emitter. As will become apparent shortly, it is highly desirable to have the first component (electrons from emitter to base) be much larger than the second component (holes from base to emitter). This can be accomplished by fabricating the device with a heavily doped emitter and a lightly doped base; that is, the device is designed to have a high density of electrons in the emitter and a low density of holes in the base.

The current that flows across the emitter-base junction will constitute the emitter current i_E , as indicated in Fig. 2.20. The direction of i_E is “out of” the emitter lead, which, following the usual conventions, is in the direction of the positive-charge flow (hole current) and opposite to the direction of the negative-charge flow (electron current), with the emitter current i_E being equal to the sum of these two components. However, since the electron component is much larger than the hole

component, the emitter current will be dominated by the electron component.

From our study of the current flow across a forward-biased pn junction, we know that the magnitude of both the electron component and the hole component of i_E will be proportional to $e V_{BE} / V_T$, where V_{BE} is the forward voltage across the base-emitter junction and V_T is the thermal voltage (approximately 25 mV at room temperature).

Let's now focus our attention on the first current component, namely, that carried by electrons injected from the emitter into the base. These electrons will be minority carriers in the p-type base region. Because their concentration will be highest at the emitter side of the base, the injected electrons will diffuse through the base region toward the collector. In their journey across the base, some of the electrons will combine with holes, which are majority carriers in the base. However, since the base is usually very thin and, as mentioned earlier, lightly doped, the proportion of electrons that are "lost" through this recombination process will be quite small. Thus, most of the diffusing electrons will reach the boundary of the

collector-base depletion region. Because the collector is more positive than the base (by the reverse-bias voltage V_{CB}), these successful electrons will be swept across the CBJ depletion region into the collector. They will thus get collected and constitute the collector current i_c .

The Collector Current From the foregoing statements, we see that the collector current is carried by the electrons that reach the collector region. Its direction will be opposite to that of the flow of electrons, and thus into the collector terminal. Its magnitude will be proportional to e^{V_{CB}/V_T} , thus

$$i_c = I_s e^{V_{BE}/V_T}$$

where the constant of proportionality I_s , as in the case of the diode, is called the saturation current and is a transistor parameter. We will have more to say about I_s shortly.

An important observation to make here is that i_c is independent of the value of V_{CB} . That is, as long as the collector is positive with respect to the base, the electrons that reach the collector side of the base region will be swept into the collector and will register as collector current.

The Base Current Reference to Fig. 2.20. shows that the base current i_B is composed of two components. The first component i_{B1} is due to the holes injected from the base region into the emitter region. This current component is proportional to e^{V_{BE}/V_T} . The second component of base current, i_{B2} , is due to holes that have to be supplied by the external circuit in order to replace the holes lost from the base through the recombination process. Because i_{B2} is proportional to the number of electrons injected into the base, it also will be proportional to e^{V_{BE}/V_T} . Thus the total base current, $i_B = i_{B1} + i_{B2}$, will be proportional to e^{V_{BE}/V_T} , and can be expressed as a fraction of the collector current i_c as follows:

$$i_B = i_c / \beta$$

That is,

$$i_B = (I_S / \beta) e^{V_{BE} / V_T}$$

where β is a transistor parameter.

For modern *n-p-n* transistors, β is in the range 50 to 200, but it can be as high as 1000 for special devices. For reasons that will become clear later, the parameter β is called the *common-emitter current gain*.

The above description indicates that the value of β is highly influenced by two factors: the width of the base region, W , and the relative dopings of the base region and the emitter region, N_A/N_D . To obtain a high β (which is highly desirable since β represents a gain parameter) the base should be thin (W small) and lightly doped and the emitter heavily doped (making N_A/N_D small). For modern integrated circuit fabrication technologies, W is in the nanometer range. The Emitter Current Since the current that enters a transistor must leave it, it can be seen from Fig. 2.20 that the emitter current i_E is equal to the sum of the collector current i_C and the base current i_B ; that is,

$$i_E = i_C + i_B$$

$$i_E = (\beta + 1 / \beta) i_C$$

That
is,

$$i_E = (\beta + 1 / \beta) I_S e^{V_{BE} / VT}$$

Alternatively, we can express eq in the form

$$i_C = \alpha i_E$$

where the constant α is related to β by

$$\alpha = \beta / \beta + 1$$

Thus the emitter current in Eq. can be written

$$i_E = (I_S / \alpha) e^{V_{BE} / VT}$$

Finally, we can use Eq. to express β in terms of α , that is,

$$\beta = \alpha / 1 - \alpha$$

It can be seen from Eq. that α is a constant (for a particular transistor) that is less than but very close to unity. For instance, if $\beta = 100$, then $\alpha = 0.99$. Equation reveals an important fact: small changes in α correspond to very large changes in β . This mathematical observation manifests itself physically, with the result that transistors of the same type may have widely different values of β . For reasons that will become apparent later, α is called the common-base current gain.

Minority-Carrier Distribution Our understanding of the physical operation of the BJT can be enhanced by considering the distribution of minority charge carriers in the base and the emitter. Figure 2.21 shows the profiles of the concentration of electrons in the base and holes in the emitter of an npn transistor operating in the active mode. Observe that since the doping concentration in the emitter, N_D , is much higher than the doping concentration in the base, N_A , the concentration of electrons injected from emitter to base, $n_p(O)$, is much higher than the concentration of holes injected from the base to the emitter, $p_n(O)$. Both

quantities are proportional to e^{V_{BE}/V_T} , thus

$$n_p(O) = n_{po} e^{V_{BE}/V_T}$$

where n_{po} is the thermal-equilibrium value of the minority-carrier (electron) concentration in the base region.

Next, observe that because the base is very thin, the concentration of excess electrons decays almost linearly (as opposed to the usual exponential decay, as observed for the excess holes in the emitter region). Furthermore, the reverse bias on the collector-base junction causes the concentration of excess electrons at the collector side of the base to be zero. (Recall that electrons that reach that point are swept into the collector.)

The tapered minority-carrier concentration profile (Fig. 2.21) causes the electrons injected into the base to diffuse through the base region toward the collector. This electron

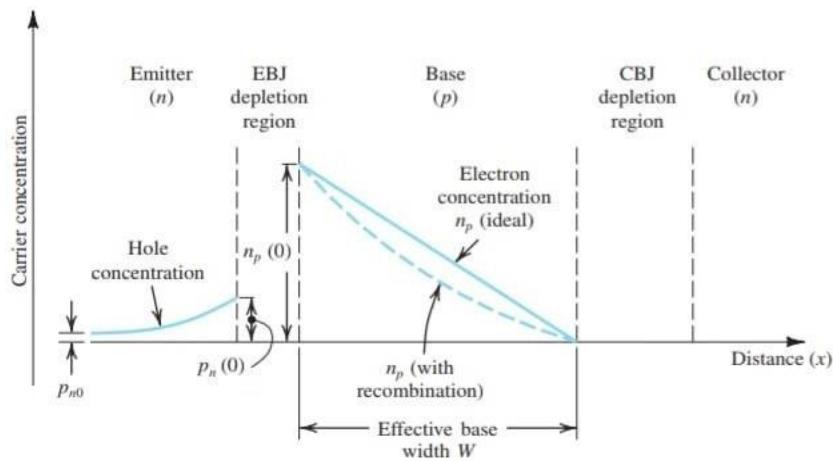


Figure 2.21 Profiles of minority-carrier concentrations in the base and in the emitter of npn transistor operating in the active mode: $v_{BE} > 0$ and $v_{CB} \geq 0$

current I_n is directly proportional to the slope of the straight-line concentration profile,

$$\begin{aligned} I_n &= A_E q D_n (dn_p(x) / dx) \\ &= A_E q D_n (-np(0) / W) \end{aligned}$$

where A_E is the cross-sectional area of the base-emitter junction (in the direction perpendicular to the page), q is the magnitude of the electron charge, D_n is the electron diffusivity in the base, and W is the effective width of the base. Observe that the negative slope of the minority-carrier concentration results in a negative current I_n across the base; that is, I_n flows from right to left (in the negative direction of x), which corresponds to the usual convention, namely, opposite to the direction of electron flow.

The recombination in the base region, though slight, causes the excess minority-carrier concentration profile to deviate from a straight line and take the slightly concave shape indicated by the broken line in Fig. 2.21. The slope of the concentration profile at the EBJ is slightly higher than that at the CBJ, with the difference accounting for the small number of electrons lost in the base region through recombination.

Finally, we have the collector current $i_C = I_n$, which will yield a negative value for i_C , indicating that i_C flows in the negative direction of the x axis (i.e., from right to left). Since we will take this to be the positive direction of i_C , we can drop the negative sign in Eq. Doing this and substituting for $n_p(0)$ from Eq., we can thus express the collector current i_C as

$$i_C = I_s e^{V_{BE}/VT}$$

where the saturation current I_s is given by

$$I_s = A_E q D_n n_{p0} / W$$

Substituting $n_{p0} = n_i^2 / N_A$, where n_i is the intrinsic carrier density and N_A is the doping concentration in the base, we can express I_s as

$$I_s = A_E q D_n n_i^2 i N_A W$$

The saturation current I_S is inversely proportional to the base width W and is directly proportional to the area of the EBJ. Typically I_S is in the range of 10^{-12} A to 10^{-18} A (depending on the size of the device). Because I_S is proportional to n^2_i , it is a strong function of temperature, approximately doubling for every 5°C rise in temperature. (For the dependence of n^2_i on temperature, refer to Eq.) Since I_S is directly proportional to the junction area (i.e., the device size), it will also be referred to as the scale current. Two transistors that are identical except that one has an EBJ area, say, twice that of the other will have saturation currents with that same ratio (i.e., 2). Thus for the same value of v_{BE} the larger device will have a collector current twice that in the smaller device. This concept is frequently employed in integrated-circuit design.

2.5.3 Operation in Saturation mode:

As mentioned above, for the BJT to operate in the active mode, the CBJ must be reverse biased. Thus far, we have stated this condition for the npn transistor as $v_{CB} \geq 0$. However, we know that a pn junction does not effectively become forward biased until the forward voltage across it exceeds approximately 0.4 V. It follows that one can maintain active-mode operation of an npn transistor for negative v_{CB} down to approximately -0.4 V. This is illustrated in Fig. 2.22, which is a sketch of i_C versus v_{CB} for a npn transistor operated with a constant emitter current I_E . As expected, i_C is independent of v_{CB} in the active mode, a situation that extends

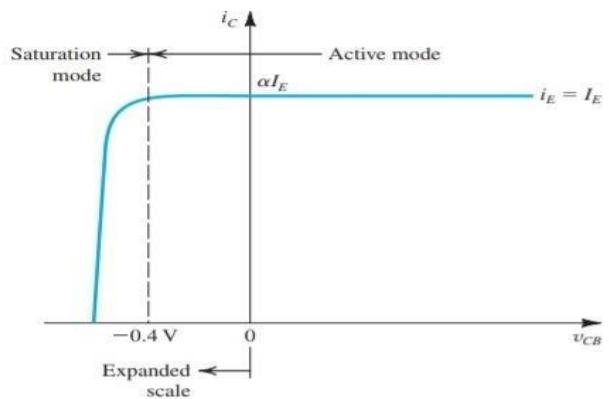


Figure 2.22 The i_C-v_{CB} characteristic of an npn transistor fed with a constant emitter current I_E . The transistor enters the saturation mode of operation for $v_{CB} < -0.4$ V, and the collector current diminishes

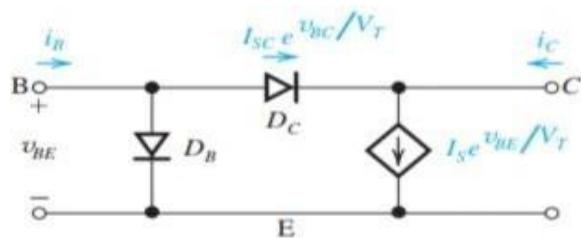


Figure 2.23 Modelling the operation of an npn transistor in saturation by augmenting the model with a forward-conducting diode D_C . Note that the current through D_C increases i_B and reduces i_C .

for v_{CB} going negative to approximately -0.4 V. Below this value of v_{CB} , the CBJ begins to conduct sufficiently that the transistor leaves the active mode and enters the saturation mode of operation, where i_c decreases. To see why i_c decreases in saturation, we can construct a model for the saturated npn transistor as follows. We augment the model with the forward-conducting CBJ diode DC, as shown in Fig. 2.23. Observe that the current i_{BC} will subtract from the controlled-source current, resulting in the reduced collector current i_c given by

$$i_c = I_s e^{v_{BE}/VT} - I_{SC} e^{v_{BC}/VT}$$

where I_{SC} is the saturation current for DC and is related to I_s by the ratio of the areas of the CBJ and the EBJ. The second term in Eq. will play an increasing role as v_{BC} exceeds 0.4 V or so, causing i_c to decrease and eventually reach zero. Figure 2.23 also indicates that in saturation the base current will increase to the value

$$i_B = (I_s/\beta) e^{v_{BE}/VT} + I_{SC} e^{v_{BC}/VT}$$

Above equations can be combined to obtain the ratio i_c/i_B for a saturated transistor. We observe that this ratio will be lower than the value of β . Furthermore, the ratio will decrease as v_{BC} is increased and the transistor is driven deeper into saturation. Because i_c/i_B of a saturated transistor can be set to any desired value lower than β by adjusting v_{BC} , this ratio is known as forced β and denoted β_{forced} ,

$$\beta_{forced} = (i_c/i_B) \mid_{saturation} \leq \beta$$

As will be shown later, in analyzing a circuit we can determine whether the BJT is in the saturation mode by either of the following two tests:

1. Is the CBJ forward biased by more than 0.4 V?
2. Is the ratio i_C/i_B lower than β ?

The collector-to-emitter voltage v_{CE} of a saturated transistor can be found from Fig. 2.33 as the difference between the forward-bias voltages

of the EBJ and the CBJ,

$$V_{CEsat} = V_{BE} - V_{BC}$$

Recalling that the CBJ has a much larger area than the EBJ, V_{BC} will be smaller than V_{BE} by 0.1 to 0.3 V. Thus,

$$V_{CEsat} \quad 0.1 \text{to } 0.3 \text{ V}$$

Typically we will assume that a transistor at the edge of saturation has $V_{CEsat} = 0.3$ V, while a transistor deep in saturation has $V_{CEsat} = 0.2$ V.

2.5.4 *The PNP transistor:*

The pnp transistor operates in a manner similar to that of the npn device described above. Figure 2.24 shows a pnp transistor biased to operate in the active mode. Here the voltage V_{EB} causes the p-type emitter to be higher in potential than the n-type base, thus forward biasing the emitter-base junction. The collector-base junction is reverse biased by the voltage V_{BC} , which keeps the p-type collector lower in potential than the n-type base.

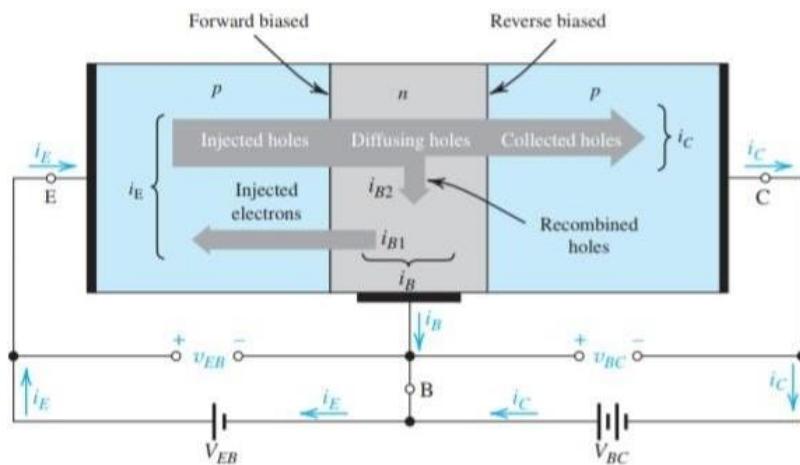


Figure 2.24 Current flow in a pnp transistor biased to operate in the active mode

Unlike the npn transistor, current in the pnp device is mainly conducted by holes injected from the emitter into the base as a result of the forward-bias voltage V_{EB} . Since the component of emitter current contributed by electrons injected from base to emitter is kept small by using a lightly doped base, most of the emitter current will be due to holes. The electrons injected from base to emitter give rise to the first component of base current, i_{B1} . Also, a number of the holes injected into the base will recombine with the majority carriers in the base (electrons) and will thus be lost. The disappearing base electrons will have to be replaced from the external circuit, giving rise to the second component of base current, i_{B2} . The holes that succeed in reaching the boundary of the depletion region of the collector-base junction will be attracted by the negative voltage on the collector. Thus these holes will be swept across the depletion region into the collector and appear as collector current. It can easily be seen from the above description that the current-voltage relationship of the pnp transistor will be identical to that of the npn transistor except that v_{BE} has to be replaced by v_{EB} . Also, the large-signal, active-mode operation of the pnp transistor can be modelled by any of four equivalent circuits similar to those for the npn transistor. Two of

these four circuits are shown in Fig. 2.25. Finally, we note that the pnp transistor can operate in the saturation mode in a manner analogous to that described for the npn device.

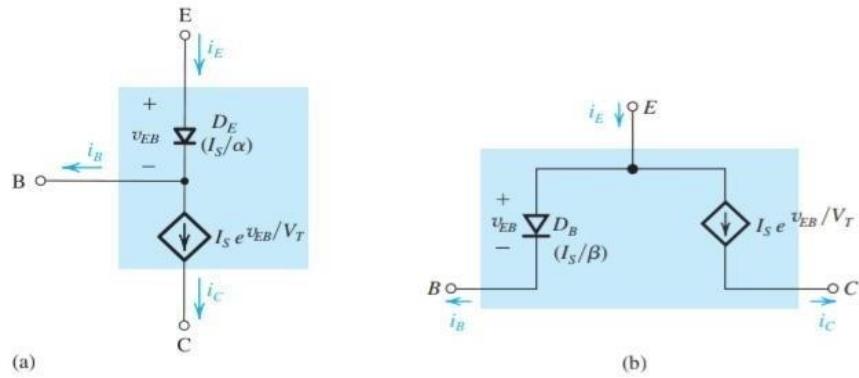


Figure 2.25 Two large-signal models for the pnp transistor operating in the active mode

2.6 V-I characteristics of different configurations

2.6.1 Graphical representation of transistor characteristics:

It is sometimes useful to describe the transistor i-v characteristics graphically. Figure 2.26 shows the $i_C - v_{BE}$ characteristic, which is the exponential relationship

$$i_C = I_s e^{v_{BE}/V_T}$$

which is identical to the diode i-v relationship. The $i_E - v_{BE}$ and $i_B - v_{BE}$ characteristics are also exponential but with different scale currents: I_s/α for i_E , and I_s/β for i_B . Since the constant of the exponential characteristic, $1/V_T$, is quite high (40), the curve rises very sharply. For v_{BE} smaller than about 0.5 V, the current is negligibly small. Also, over most of the normal current range v_{BE} lies in the range of 0.6 V to 0.8 V. In performing rapid first-order dc calculations, we normally will assume that $v_{BE} \approx 0.7$ V, which is similar to the approach used in the analysis of diode circuits. For a pnp transistor, the $i_C - v_{EB}$ characteristic will look identical to that of Fig.

2.26 with v_{BE} replaced with v_{EB} .

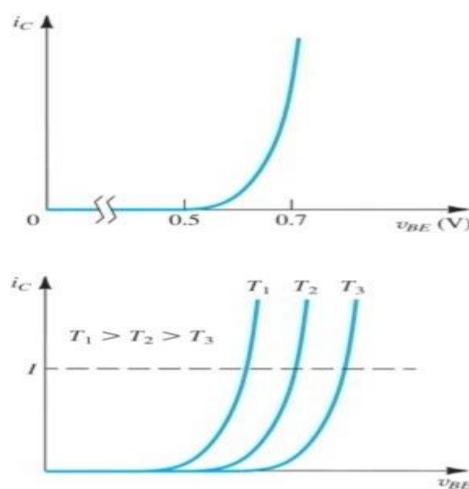


Figure 2.26 The $i_C - v_{BE}$ characteristic of npn transistor and Effect of temperature on the $i_C - v_{BE}$ characteristic. At a constant emitter current (broken line), v_{BE} changes by $-2\text{mV}/{}^\circ\text{C}$

As in silicon diodes, the voltage across the emitter-base junction decreases by about 2 mV for each rise of 1°C in temperature, provided the junction is operating at a constant current. Figure 2.26 illustrates this temperature dependence by depicting $i_{\text{C}}-v_{\text{BE}}$ curves for an npn transistor at three different temperatures.

2.6.2 Dependence on I_{C} on the collector voltage - The Early effect:

When operated in the active region, practical BJTs show some dependence of the collector current on the collector voltage, with the result that, unlike the graph shown in Fig. 2.27, their $i_{\text{C}}-v_{\text{CB}}$ characteristics are not perfectly horizontal straight lines. To see this dependence more

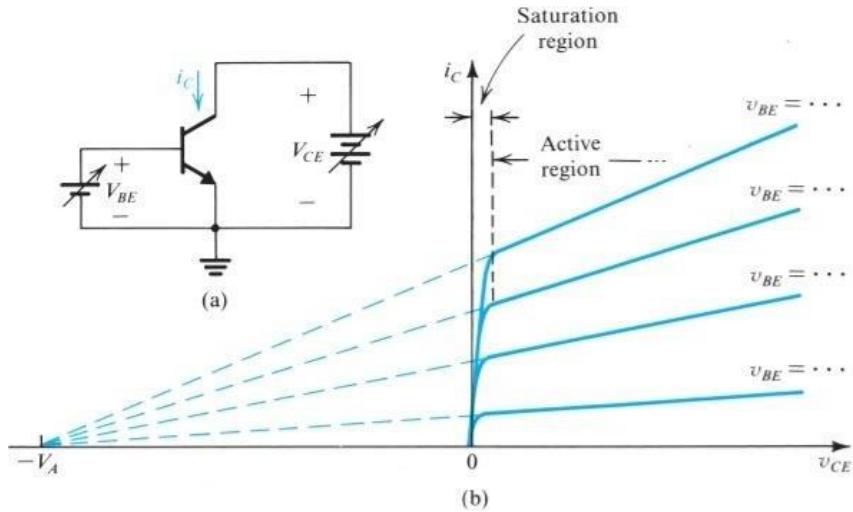


Figure 2.27 (a) Conceptual circuit for measuring the $i_C - v_{CE}$ Characteristics of the BJT. (b) The $i_C - v_{BE}$ characteristic of a practical BJT

clearly, consider the conceptual circuit shown in Fig. 2.27(a). The transistor is connected in the common-emitter configuration; that is, here the emitter serves as a common terminal between the input and output ports. The voltage V_{BE} can be set to any desired value by adjusting the dc source connected between base and emitter. At each value of V_{BE} , the corresponding $i_C - v_{CE}$ characteristic curve can be measured point by point by varying the dc source connected between collector and emitter and measuring the corresponding collector current. The result is the family of $i_C - v_{CE}$ characteristic curves shown in Fig. 2.27(b) and known as common-emitter characteristics.

At low values of v_{CE} (lower than about 0.3 V), as the collector voltage goes below that of the base by more than 0.4 V, the collector-base junction becomes forward biased and the transistor leaves the active mode and enters the saturation mode. Shortly, we shall look at the details of the $i_C - v_{CE}$ curves in the saturation region. At this time, however, we wish to examine the characteristic curves in the active region in detail. We observe that the characteristic curves, though still

straight lines, have finite slope. In fact, when extrapolated, the characteristic lines meet at a point on the negative v_{CE} axis, at $v_{CE} = -V_A$. The voltage V_A , a positive number, is a parameter for the particular BJT, with typical values in the range of 10 V to 100 V. As noted earlier, it is called the Early voltage, after J. M. Early, the engineering scientist who first studied this phenomenon.

At a given value of v_{BE} , increasing v_{CE} increases the reverse-bias voltage on the collector-base junction, and thus increases the width of the depletion region of this junction. This in turn results in a decrease in the effective base width W . Recalling that I_S is inversely proportional to W , we see that I_S will increase and that i_c increases proportionally. This is the Early effect. For obvious reasons, it is also known as the base-width modulation effect.

The linear dependence of i_c on v_{CE} can be explicitly accounted for by assuming that I_S remains constant and including the factor $(1+v_{CE}/V_A)$ in the equation for i_c as follows:

$$i_C = I_S e^{V_{BE}/VT} (1 + V_{CE}/V_A)$$

The nonzero slope of the i_C - V_{CE} straight lines indicates that the output resistance looking into the collector is not infinite. Rather, it is finite and defined by

$$[r_o \equiv \partial i_C / \partial V_{CE} \mid V_{BE} = \text{constant}]^{-1}$$

Using above Eq. we can show that

$$r_o = (V_A + V_{CE}) / I_C$$

where I_C and V_{CE} are the coordinates of the point at which the BJT is operating on the particular i_C - V_{CE} curve (i.e., the curve obtained for V_{BE} equal to constant value V_{BE} at which Eq. is evaluated). Alternatively, we can write

$$r_o = V_A / I'_C$$

where I'_C is the value of the collector current with the Early effect neglected; that is,

$$I'_C = I_S e^{V_{BE}/VT}$$

It is rarely necessary to include the dependence of i_C on V_{CE} in dc bias design and analysis that is performed by hand. Such an effect, however, can be easily included in the SPICE simulation of circuit operation, which is frequently used to "fine-tune" pencil-and-paper analysis or design.

The finite output resistance r_o can have a significant effect on the gain of transistor amplifiers. This is particularly the case in integrated-circuit amplifiers. Fortunately, there are many situations in which r_o can be included relatively easily in pencil-and-paper analysis. The output resistance r_o can be included in the circuit model of the transistor. This is illustrated in Fig. 2.28, where we show the two large-signal circuit models of a

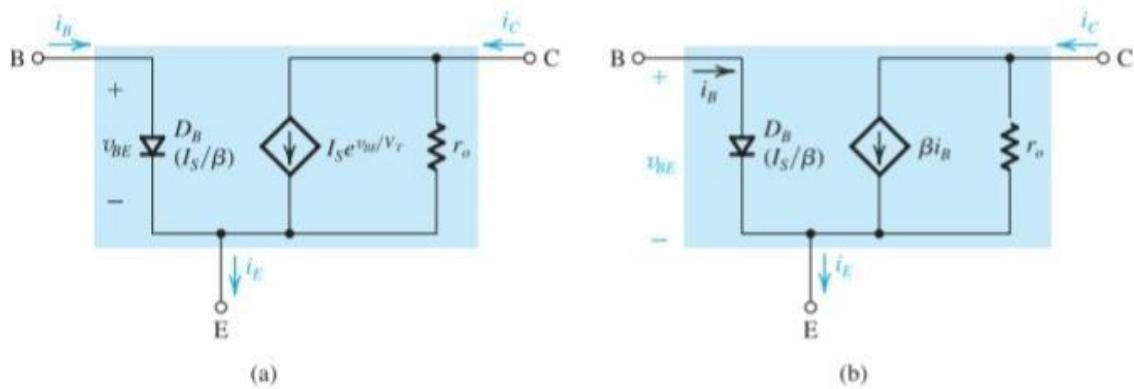


Figure 2.28 Large signal equivalent circuit models of an npn BJT operating in the active mode in the common emitter configuration with the output resistance r_o included

common-emitter npn transistor operating in the active mode, with the resistance r_o connected between the collector and the emitter terminals.

1. What is tunnel diode ?

The tunnel diode is a high conductive two terminal PN junction diode, doped heavily about 1000 times higher than a conventional PN junction diode.

2. Define Negative resistance of tunnel diode.

It is defined as the property of a tunnel diode, during its forward biased voltage increases current decreasing as a result of its dynamic resistance is negative ,hence it is called as negative resistance of the device.

3. Define Tunneling phenomenon. (or) How does tunnel diode works?

The width of the depletion region in a tunnel diode varies as the square root impurity concentration .i.e. if the concentration of impurity atom is greatly increased, the barrier width ‘W ‘ reduces. In this condition ,instead of crossing over the junction barrier, the electron penetrate through the barrier .this behavior of a diode is known as tunneling phenomenon.

4. Explain the advantages of tunnel diode?

- a. environmental immunity.

- b. Low cost
- c. Simplicity
- d. Low noise
- e. High speed
- f. Low power consumption

5. Explain the disadvantages of tunnel diode?

Only disadvantage of tunnel diode are its low output voltage swing and it is a two terminal device . hence there is no isolation between input and output . hence transistor is used along with a tunnel diode for frequencies below 1 GHz.

6. Explain the applications of tunnel diode?

- (i) . As a high speed switch .
- (ii). in pulse and digital circuits .
- (iii). in negative resistance and high frequency (micro wave) oscillator .
- (iv) in switch networks
- (v).in timing and computer logic circuitry .
- (vi). design of pulse generators and amplifiers

7. Explain the Advantages of photodiode?

- a. can be used as variable resistance device.
- b. Highly sensitive to the light.
- c. The speed of the operation is very high . the switching of current and hence the resistance value from high to low or otherwise is very fast .

8. Explain the Disadvantages of photodiode?

- (i) The dark current is temperature dependent .
- (ii) The overall photo diode characteristics are temperature dependent
Hence have poor temperature stability
- (iii) The current and change in current is in the range of TA which may not be sufficient to drive other circuits. Hence amplification is necessary

9. What is Pin diode ?

Pin diode is a high speed switching device, because its highly improved switching time

in comparison with a PN diode. In this diode high resistivity intrinsic layer is sandwiched between the heavily doped P and N regions thus it is named as PIN diode.

10. What is a varactor diode ?

The varactor diode is a semiconductor, voltage dependent variable capacitor diode. This special diode which is made for the application utilization of voltage variable properly hence it is called varactor diode or Varicaps (or) voltage cap. It is operated under reverse biased conditions so as to yield a variable junction capacitance.

11. What is the significance of varactor diode ?

The varactor diode is a semiconductor ,voltage dependent ,variable capacitors diode. Their mode of operation depends on the capacitance that exists at the PN junction when it is reverse biased.

12. Why germanium instead of silicon is used for construction of SCR?

For the construction of SCR germanium is preferred than silicon because ,more silicon per ampere current is required. Hence the current rating is increased ,it require more silicon.

13. What are the methods used to turn on SCR ?

- (i) Voltage triggering
- (ii) dV / dt triggering
- (iii) Gate triggering

14. What is SCR ?

A Silicon Controlled Rectifier (SCR)is a three terminal ,three junction semiconductor device .It is unidirectional device. It converts alternating current into direct current and control the amount of power fed to the load .

15. Write an two different characteristics of SCR ?

- 1. Forward characteristics
- 2. Reverse characteristics

16. Mention the application of SCR.

- (i) It can be used as a speed control element in DC and Ac motors.

- (ii) It can be used as an inverter.
- (iii) It can be used as a Converter.

17. Define breakdown voltage of SCR.

It can be defined as the minimum forward voltage at which the SCR starts conducting heavily.

18. Define latching current. It can be defined as the maximum anode current that an SCR capable of passing without destruction.

19. Define holding current of an SCR .

It can be defined as the minimum value of anode current required to keep the SCR in ON position.

20. What is DIAC ?

Diac is a two terminal ,bi-directional semiconductor switching depending upon the polarity of the voltage applied across its main terminals .In operation ,diac is equivalent to two 4 layer diodes connected in antiparallel.

21. List out the applications of DIAC.

- (i) It is used as a trigger device in TRIAC power control systems.
- (ii) It is used in lamp dimmer circuits
- (iii) It is used in heater control circuits
- (iv) It is used for speed control of universal motor.

22. A triac is considered as two SCRs connected in reverse parallel.Why?

The TRIAC is a bidirectional device.,i.e it conducts in both direction ,In order to achieve this characteristics two SCRs are connected in reverse and parallel.

23. Compare SCR with TRIAC.

SCR	TRIAC
<p>It is a unidirectional device</p> <p>It is triggered by a narrow positive pulse applied at the gate</p> <p>SCR are available only with large current</p>	<p>It is a bidirectional device</p> <p>It is triggered by a narrow pulse of the either polarity to the gate.</p> <p>Triac are available for both lower current and large current rating.</p>

rating.	
It has fast turn off	The turn off time is less than SCR
UJT is used for triggering	Diac is used for triggering
Applications: Phase control, Protection of power suppliers	Applications: Phase control, light dimmer

24. What is TRIAC ?

TRIAC is a three terminal bi-directional semiconductor switching device. It can conduct in both the directions for any desired period. In operation it is equivalent to two SCRs connected in antiparallel .

25. Give the applications of UJT.

- (i) It is used in timing circuits
- (ii) It is used in switching circuits.
- (iii) It is used in phase control circuits.
- (iv) It is used in saw –tooth generators.
- (v) It is used in pulse generation.

26. Define dark current of a photo diode .

When there is no light ,the reverse biased photodiode carries a current which is very small and is called as dark current.

27. What is photodiode ?

It is a light sensitive device used to convert light signal into electrical signal.

28. What is mean by solar cell ?

A solar cell is basically a PN junction diode which converts solar energy into electric energy.it is also called a solar energy converter.

29. What is Photo voltaic effect?

When the light is incident on the photodiode ,an internal voltage is generated,it causes

the current flow through internal circuit even though no external source is applied. This generated emf is proportional to the frequency and the intensity of the incident light. This phenomenon is called photo voltaic effect.

30. What is known as photo conductive effect?

This is the absorption of incident light by a semiconductor resulting in increase in conductivity.

31. What is an LCD?

LCD is a passive type display devices used for display of numeric and alphanumeric character in dotmatrix and seven segment display. The main advantage of LCD is the low power consumption because no light generation is required .

32. On what factor does the color of the light emitted by a LED depend ?

- (i) Energy gap of the material
- (ii) The colour of the emitted light depends on the type of the material used.

Essay Questions

1. List the applications of zener diode?

Mention the applications

2. Write the working principle of photo transistor?

Draw the circuit diagram

Explain working principle

3. What is intrinsic standoff ratio of UJT?

Express the equations and derive the UJT.

4. Discuss the operation and characteristics of the following:

a. SCR. b. UJT.

Explain about above operations.

5. Explain V-I characteristics of a tunnel diode and write its applications?

Draw the circuit diagrams,

Explain the V-I characteristics of Tunnel diode

MCQ'S

- 1. Zener diodes are also known as** (c)
 - a) Voltage regulators
 - b) Forward bias diode
 - c) Breakdown diode
 - d) None of the mentioned
- 2. Which of the following is true about the resistance of a Zener diode?** (d)
 - a) It has an incremental resistance
 - b) It has dynamic resistance
 - c) The value of the resistance is the inverse of the slope of the i-v characteristics of the Zener diode
 - d) All of the mentioned
- 3. Which of the following is true about the temperature coefficient or TC of the Zener diode?** (d)
 - a) For Zener voltage less than 5V, TC is negative
 - b) For Zener voltage around 5V, TC can be made zero
 - c) For higher values of Zener voltage, TC is positive
 - d) All of the mentioned
- 4. Which of the following can be used in series with a Zener diode so that combination has almost zero temperature coefficient?** (a)
 - a) Diode
 - b) Resistor
 - c) Transistor
 - d) MOSFET
- 5. In Zener diode, for currents greater than the knee current, the v-i curve is almost** (b)
 - a) Almost a straight line parallel to y-axis
 - b) Almost a straight line parallel to x-axis
 - c) Equally inclined to both the axes with a positive slope
 - d) Equally inclined to both the axes with a negative slope
- 6. Zener diodes can be effectively used in voltage regulator. However, they are these days being replaced by more efficient** (c)
 - a) Operational Amplifier
 - b) MOSFET
 - c) Integrated Circuits
 - d) None of the mentioned
- 7. A 9.1-V zener diode exhibits its nominal voltage at a test current of 28 mA. At this**

current the incremental resistance is specified as 5Ω . Find V_{Z0} of the Zener model (b).

- a) 8.96V
- b) 9.03V
- c) 9.17V
- d) 9.24V

8. A shunt regulator utilizing a zener diode with an incremental resistance of 5Ω is fed through an $82\text{-}\Omega$ resistor. If the raw supply changes by 1.0 V, what is the corresponding change in the regulated output voltage? (c)

- a) 72.7 mV
- b) 73.7 mV
- c) 74.7 mV
- d) 75.7 mV

9. A designer requires a shunt regulator of approximately 20 V. Two kinds of Zener diodes are available: 6.8-V devices with r_z of 10Ω and 5.1-V devices with r_z of 30Ω . For the two major choices possible, find the load regulation. In this calculation neglect the effect of the regulator resistance R . (d)

- a) -30mV/mA and 120mV/mA respectively
- b) 30mV/mA and 60mV/mA respectively
- c) -60mV/mA and +60mV/mA respectively
- d) -30mV/mA and -120mV/mA respectively

10. Partial specifications of a Zener diode is provided. $V_Z = 10.0 \text{ V}$, $V_{ZK} = 9.6 \text{ V}$, and $I_{ZT} = 50 \text{ mA}$. Assuming that the power rating of a breakdown diode is established at about twice the specified Zener current (I_{ZT}), what is the power rating of each of the diodes described above? (a)

- a) 1.04 W
- b) 0.104 W
- c) 10.4 mW
- d) 1.04 mW

UNIT – III
BI-POLAR JUNCTION
TRANSISTORS

BJT Circuits at DC:

We are now ready to consider the analysis of BJT circuits to which only dc voltages are applied. In the following examples we will use the simple model in which $|V_{BE}|$ of a conducting transistor is 0.7V and $|V_{CE}|$ of a saturated transistor is 0.2V, and we will neglect the Early effect. These models are shown in Table6.3. Better models can, of course, be used to obtain more accurate results. This, however, is usually achieved at the expense of speed of analysis; more importantly, the attendant complexity could impede the circuit designer's ability to gain insight regarding circuit behavior. Accurate results using elaborate models can be obtained using circuit simulation with SPICE. This is almost always done in the final stages of a design and certainly before circuit fabrication. Computer simulation, however, is not a substitute for quick pencil-and-paper circuit analysis, an essential ability that aspiring circuit designers must master. The following series of examples is a step in that direction.

As will be seen, in analyzing a circuit the first question that one must answer is: In which mode is the transistor operating? In some cases, the answer will be obvious. For instance, a quick check of the terminal voltages will indicate whether the transistor is cut off or conducting. If it is conducting, we have to determine whether it is operating in the active mode or in saturation. In some cases, however, this may not be obvious. Needless to say, as the reader gains practice and experience in transistor circuit analysis and design, the answer will be apparent in a much larger proportion of problems. The answer, however, can always be determined by utilizing the following procedure.

Assume that the transistor is operating in the active mode and, using the active-mode model in Table3.1, proceed to determine the various voltages and currents that correspond. Then check for consistency of the results with the assumption of active-mode operation; that is, is V_{CB} of an npn transistor greater than $-0.4V$ (or V_{CB} of a pnp transistor lower than $0.4V$)? If the answer is yes, then our task is complete. If the answer is no, assume saturation-mode operation and, using the saturation-mode model in Table3.1, proceed to determine currents and voltages and then check for consistency of the results with the assumption of saturation-mode operation. Here the test is usually to compute the ratio I_C/I_B and to verify that it is lower than the transistor β (i.e., $\beta_{\text{forced}} < \beta$). Since β for a given transistor type varies over a wide range,¹² one must use the lowest specified β for this test. Finally, note that the order of these two assumptions can be reversed.

A Note on Units: Except when otherwise specified, throughout this book we use a consistent set of units, namely, volts (V), millamps (mA), and kilohms (k).

Table 3.1:simplified models for the operation of BJT in Dc circuits:

		<i>npn</i>	<i>pnp</i>
Active EBJ: Forward Biased CBJ: Reverse Biased	$V_{BE} = 0.7 \text{ V}$		
Saturation EBJ: Forward Biased CBJ: Forward Biased	$V_{BE} = 0.7 \text{ V}$		

Applying BJT in amplifier design:

The Basis for Amplifier Operation The basis for the application of the transistor (a MOSFET or a BJT) in amplifier design is that when the device is operated in the active region, a voltage-controlled current source is realized. When a BJT is operated in the active region, the base-emitter voltage v_{BE} controls the collector current i_C according to the exponential relationship which, for an npn transistor, is expressed as

$$\dots\dots\dots(3.2)$$

Here, this first-order model of BJT operation indicates that the collector current i_C does not depend on the collector voltage v_{CE} because the collector-base junction is reverse biased, thus “isolating” the collector.

For the npn transistor in Fig.3.2, the CBJ reverse-bias condition is ensured by keeping $v_{CE} \geq 0.3 \text{ V}$. Since v_{BE} is usually in the vicinity of 0.7V, v_{BC} is thus kept.

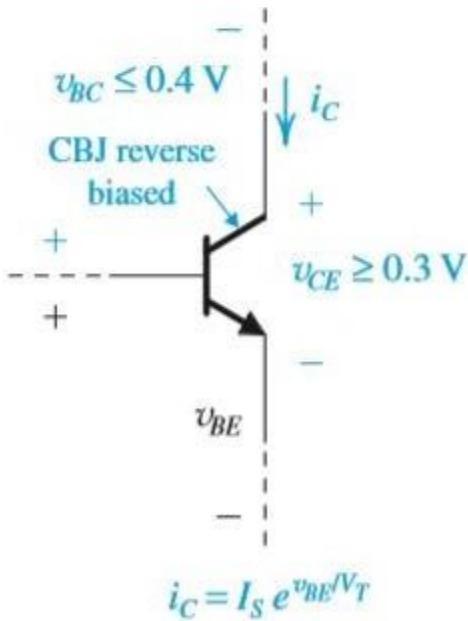


Fig:3.2 an npn transistor in the active mode

Note: $v_{BE} \approx 0.7$ V, and $v_{CE} \geq 0.3$ V results in $v_{BC} \leq 0.4$ V, which is sufficient to keep the CBJ from conducting.

Voltage amplifier:

The transistor is basically a transconductance amplifier: that is, an amplifier whose input signal is a voltage and whose output signal is a current. More commonly, however, one is interested in voltage amplifiers. A simple way to convert a transconductance amplifier to a voltage amplifier is to pass the output current through a resistor and take the voltage across the resistor as the output. Doing this for a BJT results in the common emitter amplifier circuit shown in Fig.3.3. Here V_{in} is the input voltage, R_C (known as a load resistance) converts the drain current I_C to a voltage ($I_C R_C$), and V_{in} (or) V_{cc} is the supply voltage that powers up the amplifier and, together with R_C , establishes operation in the active region, as will be shown shortly. In the amplifier circuit of Fig.3.3 the output voltage is taken between the collector and ground, rather than simply across R_C . This is done because of the need to maintain a common ground reference between the input and the output. The output voltage V_{ce} is given by

Thus it is an inverted version(note the minus sign)of $I_c R_c$ that is shifted by the constant value of the supply voltage V_{cc} .

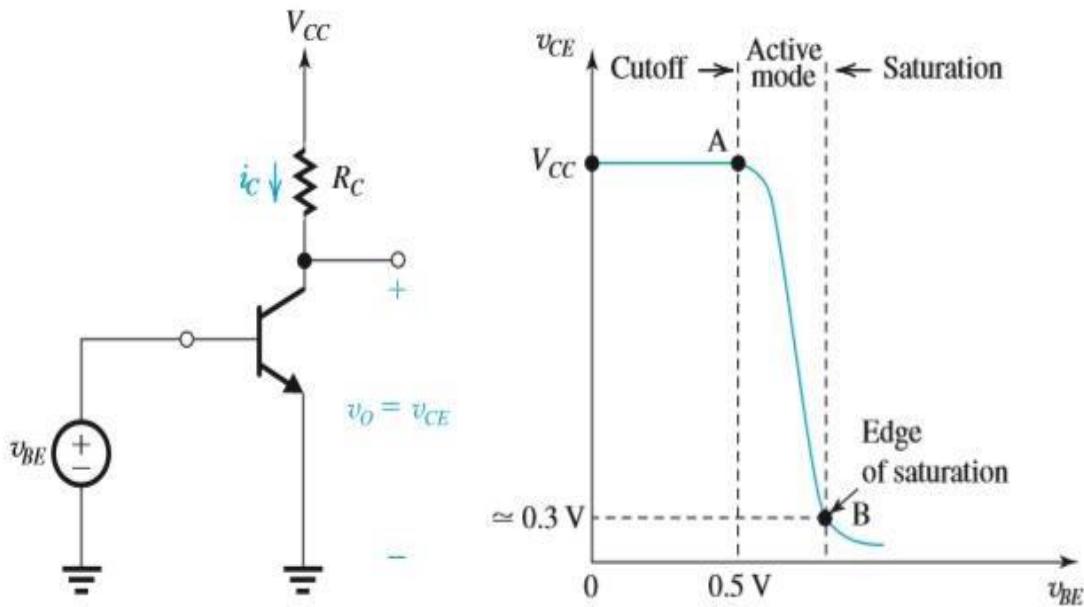


Fig:3.3 voltage amplifier using BJT & VTC

The Voltage-Transfer Characteristic (VTC) :

A useful tool that provides insight into the operation of an amplifier circuit is its voltage-transfer characteristic(VTC).This is simply a plot(or a clearly labeled sketch)of the output voltage versus the input voltage. This is the plot of v_{BE} versus v_{CE} shown in Fig.3.3

Observe that for $v_{BE} < V_t$, the transistor is cut off, $i_C = 0$ and, from Eq. $v_{CE} = V_{CC}$.As v_{GS} exceeds V_t , the transistor turns on and v_{CE} decreases. However, since initially v_{CE} is still high, the BJT will be operating in the active region. This continues as v_{BE} is increased until the value of v_{BE} is reached that results in v_{CE} becoming lower than v_{BE} by V_t volts [point B on the VTC in Fig.3.3]. For v_{BE} greater than that at point B, the transistor operates in the t saturation region and v_{BE} decreases more slowly. The VTC in Fig.3.3 indicates that the segment of greatest slope is that labeled AB,which corresponds to operation in the active region. When a BJT is operated as an amplifier, its operating point is confined to the segment AB at all times. An expression for the segment AB the output voltage v_{CE} is related to the input voltage v_{BE} by

$$v_{CE} = V_{CC} - R_C I_S e^{v_{BE}/V_T} \quad (3.4)$$

Here also, the input–output relationship is nonlinear. Nevertheless, linear (or almost-linear) amplification can be obtained by using the biasing technique.

Obtaining Linear Amplification by Biasing the Transistor:

Biasing enables us to obtain almost-linear amplification from the BJT. The technique is illustrated for the BJT case in Fig.3.4. A dc voltage V_{BE} is selected to obtain operation at a point Q on the segment AB of the VTC. How to select an appropriate location for the bias point Q will be discussed shortly. The resulting bias point Q will be characterized by dc voltages V_{BE} and V_{CE} , which are related by

$$V_{CE} = V_{CC} - R_C I_S e^{V_{BE}/V_T} \quad (3.5)$$

Point Q is known as the bias point or the dc operating point. Also, since at Q no signal component is present, it is also known as the quiescent point (which is the origin of the symbol Q).

A dc current I_C ,

$$I_C = I_S e^{V_{BE}/V_T} \quad (3.6)$$

Also, superimposing a small-signal v_{be} on the dc bias voltage V_{BE} results in

$$v_{BE}(t) = V_{BE} + v_{be}(t) \quad (3.7)$$

which can be substituted into Eq.3.5 to obtain the total instantaneous value of the output voltage $v_{CE}(t)$. Here again, almost-linear operation is obtained by keeping v_{be} small enough to restrict the excursion of the instantaneous operating point to a short, almost-linear segment of the VTC around the bias point Q. Similar comments also apply to the maximum allowable signal swing at the output.

The resulting $v_{CE}(t)$ can be obtained by substituting for $v_{BE}(t)$ into V_{CE} . Graphically, we can use the VTC to obtain $v_{CE}(t)$ point by point as illustrated in Fig.3.4. Here we show

Biasing the BJT to Obtain Linear Amplification

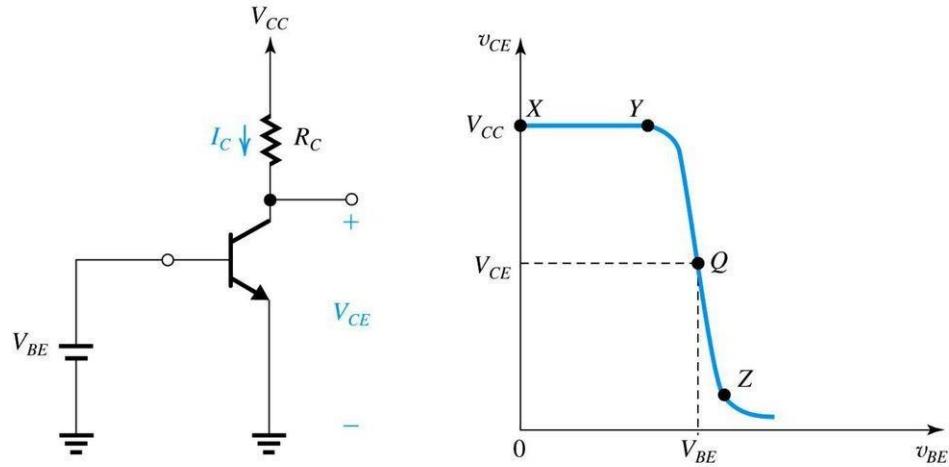


Figure 7.2 Biasing the BJT amplifier at a point Q located on the active⁸-mode segment of the VTC.

Fig:3.4 (a)

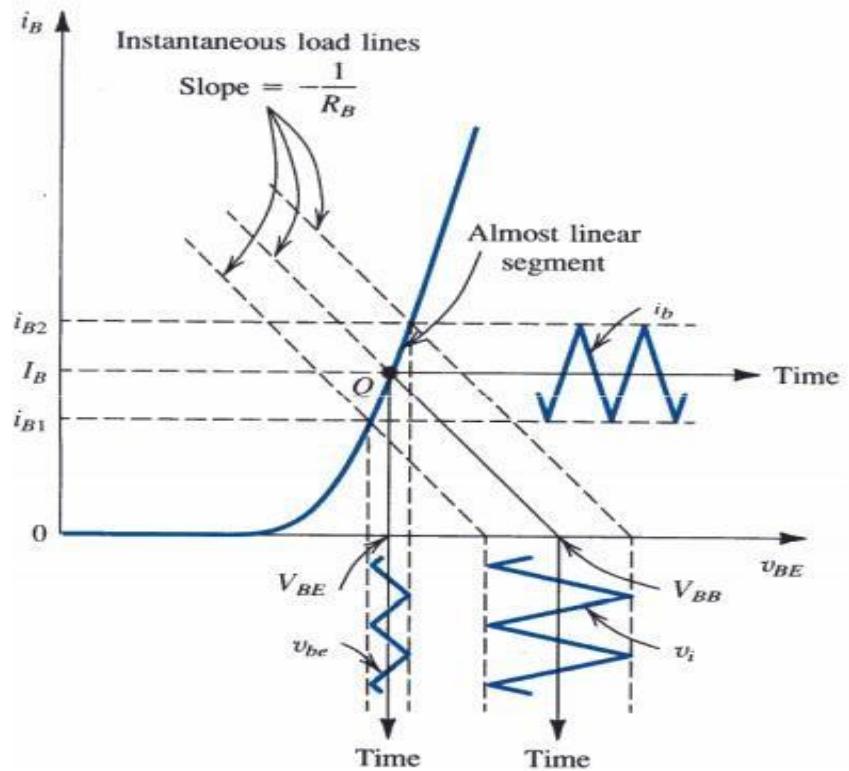


Fig:3.4(b) linear characteristics

The case of v_{BE} being a triangular wave of “small” amplitude. Specifically, the amplitude of v_{BE} is small enough to restrict the excursion of the instantaneous operating point to a short, almost-linear segment of the VTC around the bias point Q. The shorter the segment, the greater the linearity achieved, and the closer to an ideal triangular wave the signal component at the output, v_{CE} , will be. This is the essence of obtaining linear amplification from the nonlinear BJT.

Small signal voltage gain:

A similar development can be used to obtain the small-signal voltage gain of the BJT amplifier shown in Fig.3.5.

$$A_v = \frac{dv_{CE}}{dv_{BE}} \text{ when } v_{BE} = V_{BE} \quad (3.8)$$

Utilizing Eq.(3.4) together with Eq.(3.6), we obtain

$$A_v = -\left(\frac{I_C}{V_T}\right) R_C \quad (3.9)$$

We make the following observations on this expression for the voltage gain:

1. The gain is negative, which signifies that the amplifier is inverting; that is, there is a 180° phase shift between the input and the output. This inversion should have been anticipated from Eq.(3.4).
2. The gain is proportional to the collector bias current I_C and to the load resistance R_C . Additional insight into the voltage gain A_v can be obtained by expressing Eq.(3.9) as

$$A_v = -\left(\frac{I_C R_C}{V_T}\right) \quad (3.10)$$

Usually, $V_{ov}/2$ is larger than (V_T) , thus we can obtain higher voltage gain from the BJT amplifier. The gain A_v in Eq.(3.10) can be expressed alternately as

$$A_v = -\left(\frac{V_{CC} - V_{CE}}{V_T}\right) \quad (3.11)$$

from which we see that maximum gain is achieved when V_{CE} is at its minimum value of about 0.3V,

$$|A_{vmax}| = \left(\frac{V_{CC} - 0.3}{V_T}\right) \quad (3.12)$$

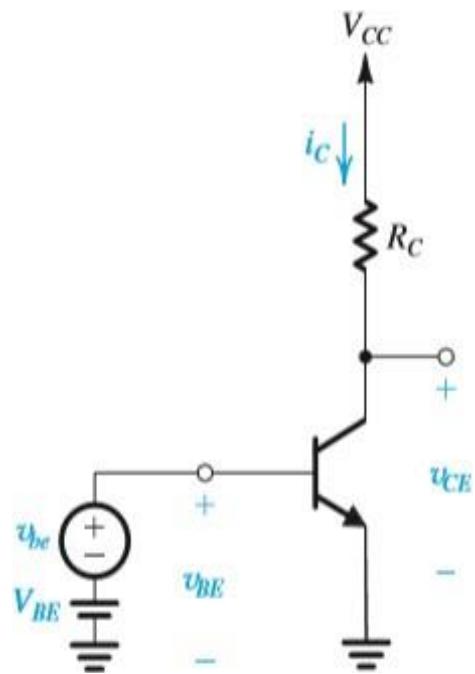


Fig:3.5 BJT amplifier biased at a point Q, with a small voltage signal v_{be} superimposed on the dc bias voltage V_{BE} . The resulting output signal v_{ce} appears superimposed on the dc collector voltage V_{CE} . The amplitude of v_{ce} is larger than that of v_{be} by the voltage gain A_v .

Here again, this is only a theoretical maximum, since biasing the BJT at the edge of saturation leaves no room for negative signals wing at the output. Equation(3.12)nevertheless provides an upper bound on the voltage gain achievable from the basic BJT amplifier. As an example, for $V_{CC} = 5$ V, the maximum gain is $188V/V$, considerably larger than in the MOSFET case. For modern low-voltage technologies, a V_{CC} of $1.3V$ provides a gain of $40V/V$,again much largerthan the MOSFETcase. The reader should not,however,jumptothe conclusion that the BJT is preferred to the MOSFET in the design of modern integrated-circuit amplifiers; in fact, the opposite is true.Finally,we conclude from Eq.(3.11)that to maximize $|Av|$ the transistor should be biased at the lowest possible V_{CE} consistent with the desired value of negative signal swing at the output.

Determining the VTC by Graphical Analysis:

Figure(3.6) shows a graphical method for determining the VTC of the BJT amplifier of Although graphical analysis of transistor circuits is rarely employed in practice,it is useful to us at this stage for gaining greater insight into circuit operation, especially in answering the question of where to locate the bias point Q. The graphical analysis is based on the observation

that for each value of I_B , the circuit will be operating at the point of intersection of the $i_C - v_{CE}$ graph corresponding to the particular

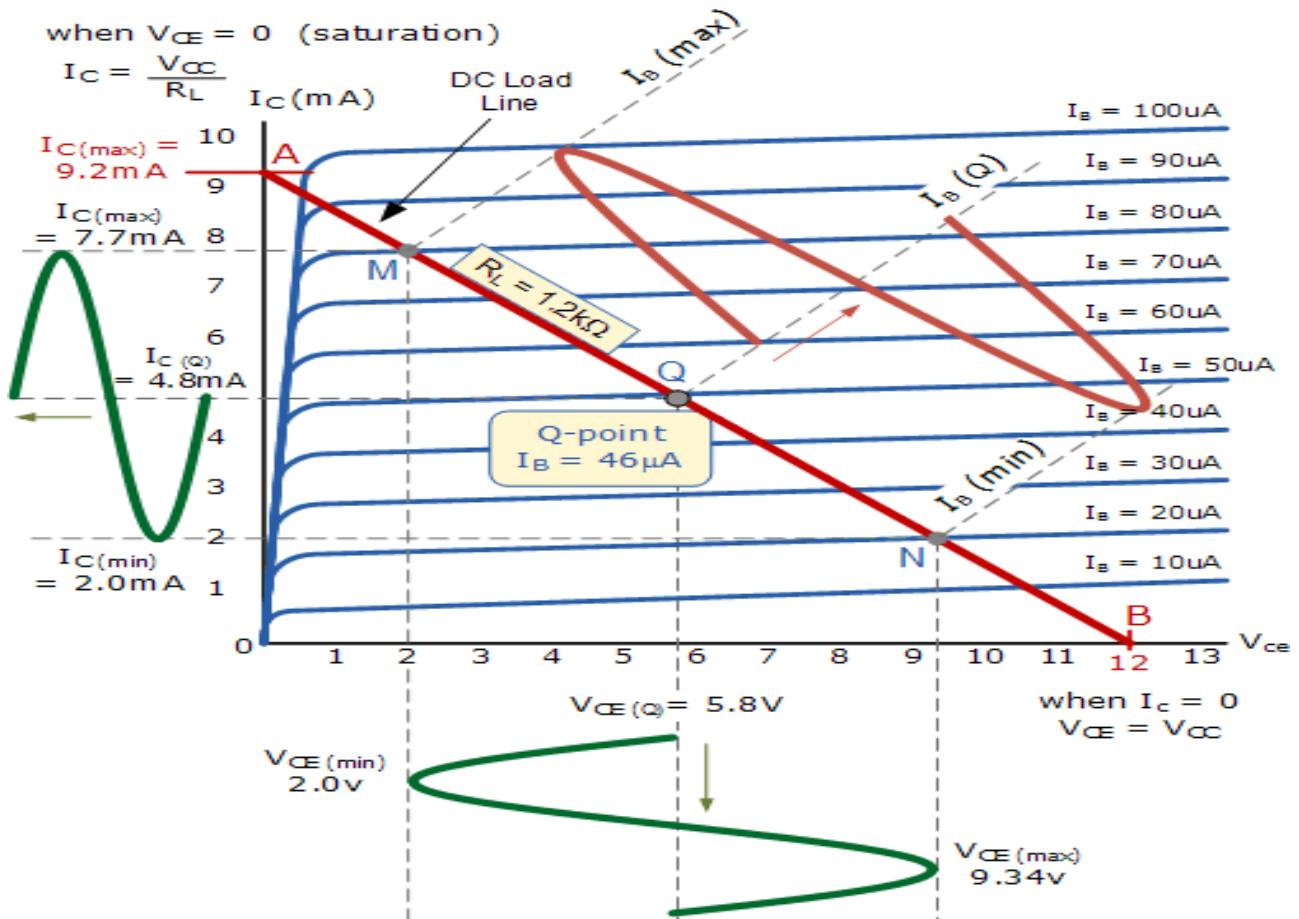


Fig:3.6: Graphical construction to determine the voltage-transfer characteristic of the amplifier in fig 3.5

The graphical analysis method above can be applied to determine the VTC of the BJT amplifier in Fig.3.3. Here point A, Fig.3.3, corresponds to the BJT just turning on ($v_{BE} \sim 0.5 \text{ V}$) and point B corresponds to the BJT leaving the active region and entering the saturation region. If the BJT is to be operated as a switch, the two modes of operation are cutoff (open switch) and saturation (closed switch). As discussed in Section 6.2, in saturation, the BJT has a small closure resistance $R_{CE\text{ sat}}$ as well as an offset voltage. More seriously, switching the BJT out of its saturation region can require a relatively long delay time to ensure the removal of the charge stored in the BJT base region. This phenomenon has made the BJT much less attractive in digital logic applications relative to the MOSFET.

Deciding on a Location for the Bias Point Q:

For the BJT amplifier, the bias point Q is determined by the value of V_{BE} and that of the load resistance R_C . Two important considerations in deciding on the location of Q are the required gain and the desired signal swing at the output. To illustrate, consider the VTC shown in Fig.3.6(a). Here the value of R_C is fixed and the only variable remaining is the value of V_{BE} . Since the slope increases as we move closer to point B, we obtain higher gain by locating Q as close to B as possible. However, the closer Q is to the boundary point B, the smaller the allowable magnitude of negative signal swing. Thus, as often happens in engineering design, we encounter a situation requiring a saturation. The answer here is relatively simple: For a given R_C , locate Q as close to the triode region (point B) as possible to obtain high gain but sufficiently distant to allow for the required negative signal swing.

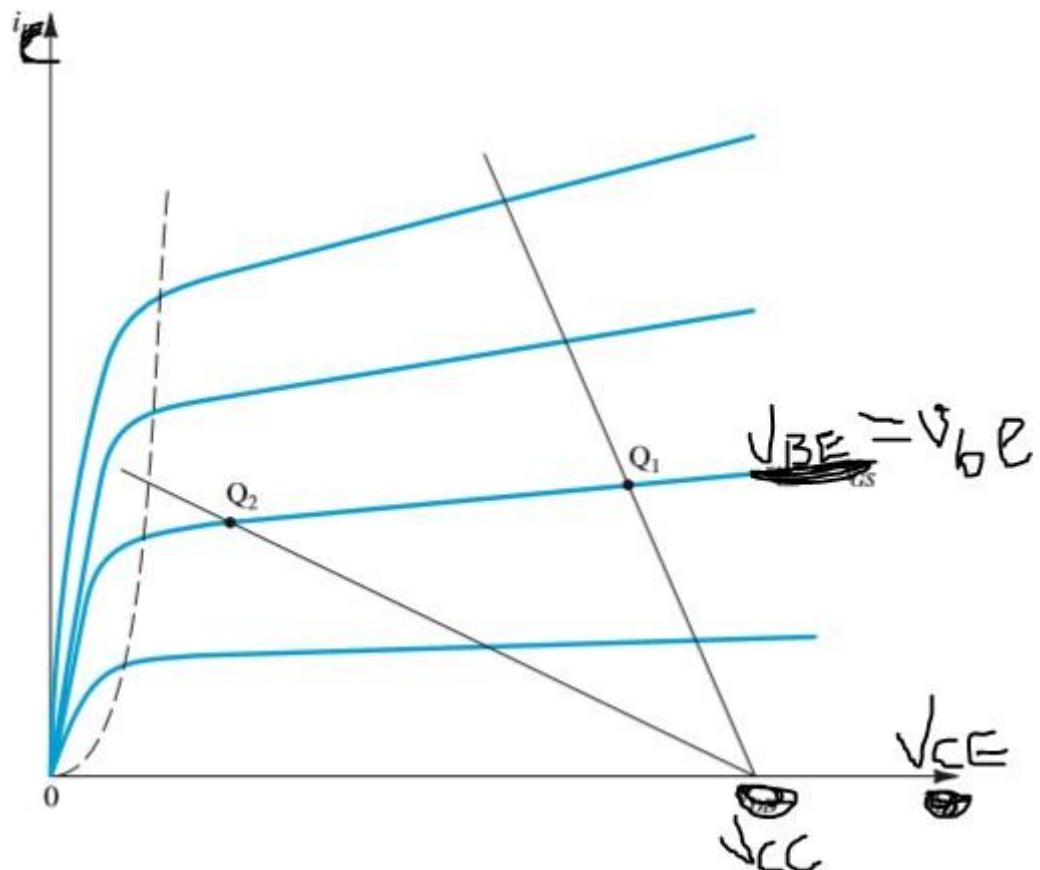


Fig:3.6(a) Two load lines and corresponding bias points. Bias point Q1 does not leave sufficient room for positive signal swing at the drain (too close to V_{CC}). Bias point Q2 is too close to the boundary of the triode region and might not allow for sufficient negative signal swing.

In deciding on a value for R_C , it is useful to refer to the i_C - V_{CE} plane. Figure 7.9 shows two load lines resulting in two extreme bias points: Point Q1 is too close to V_{CC} , resulting in a severe

constraint on the positive signal swing of v_{ce} . Exceeding the allowable positive maximum results in the positive peaks of the signal being clipped off, since the BJT will turn off for the part of each cycle near the positive peak. We speak of this situation by saying that the circuit does not have sufficient “headroom.” Similarly, point Q2 is too close to the boundary of the triode region, thus severely limiting the allowable negative signal swing of v_{ce} . Exceeding this limit would result in the transistor entering the triode region for part of each cycle near the negative peaks, resulting in a distorted output signal. In this situation we say that the circuit does not have sufficient “legroom.”

Small-Signal Operation and Models:

We next consider the small-signal operation of the BJT and develop small-signal equivalent-circuit models that represent its operation at a given bias point. The BJT draws a finite base current. As will be seen shortly, this phenomenon (finite β) manifests itself as a finite input resistance looking into the base of the BJT. Here the base-emitter junction is forward biased by a dc voltage V_{BE} . The reverse bias of the collector-base junction is established by connecting the collector to another power supply of voltage V_{CC} through a resistor R_C . The input signal to be amplified is represented by the voltage source v_{be} that is superimposed on V_{BE} .

The DC Bias Point : We consider first the dc bias conditions by setting the signal v_{be} to zero. The circuit reduces to that in Fig.3.7(b), and we can write the following relationships for the dc currents and voltages:

$$I_C = I_S e^{V_{BE}/V_T} \quad \dots \dots \dots (3.13)$$

$$I_E = I_C / \alpha \quad \dots \dots \dots (3.14)$$

$$I_B = I_C / \beta \quad \dots \dots \dots (3.15)$$

$$V_{CE} = V_{CC} - I_C R_C \quad \dots \dots \dots (3.16)$$

For active-mode operation, V_{CE} should be greater than $(V_{BE} - 0.4)$ by an amount that allows for the required negative signal swing at the collector.

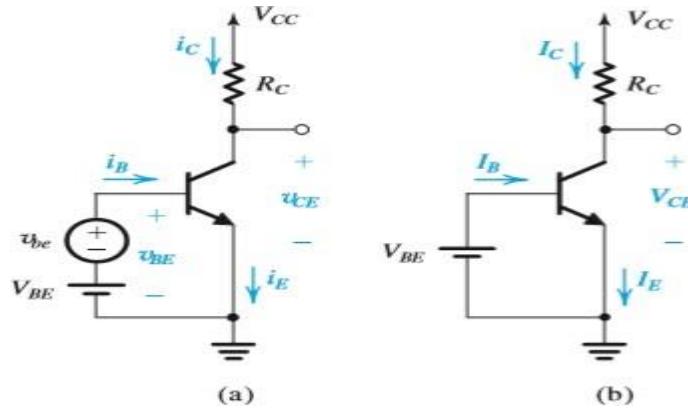


Fig:3.7 (a) Conceptual circuit to illustrate the operation of the transistor as an amplifier. (b) The circuit of (a) with the signal source v_{be} eliminated for dc (bias) analysis.

The Collector Current and the Transconductance

If a signal v_{be} is applied as shown in Fig.3.7(a), the total instantaneous base-emitter voltage v_{BE} becomes

$$v_{BE} = V_{BE} + v_{be} \quad \text{---(3.17)}$$

Correspondingly, the collector current becomes

$$\begin{aligned} i_C &= I_s e^{v_{BE}/V_T} \\ &= I_s e^{(V_{BE} + v_{be})/V_T} \\ &= I_s e^{V_{BE}/V_T} e^{v_{be}/V_T} \end{aligned} \quad \text{---(3.18)}$$

Use of Eq.(3.13) yields

$$i_C = I_C e^{v_{be}/V_T} \quad \text{---(3.19)}$$

Now, if $v_{be} \approx V_T$, we may approximate Eq.(3.19) as

$$i_C \approx I_C (1 + (v_{be}/V_T)) \quad \text{---(3.20)}$$

Here we have expanded the exponential in Eq.(3.19) in a series and retained only the first two terms. That is, we have assumed that

$$v_{be} \ll V_T \quad \text{---(3.21)}$$

so that we can neglect the higher-order terms in the exponential series expansion. The condition in Eq.(3.21) is the small-signal approximation for the BJT. The small-signal approximation for

the BJT is valid only for v_{be} less than 5mV or 10mV, at most. Under this approximation, the total collector current is given by Eq.(7.57) and can be rewritten

$$i_C = I_C + (I_C / V_T) v_{be} \quad \dots \dots \dots (3.22)$$

Thus the collector current is composed of the dc bias value I_C and a signal component i_c ,

$$i_c = (I_C / V_T) v_{be} \quad \dots \dots \dots (3.23)$$

This equation relates the signal current in the collector to the corresponding base-emitter signal voltage. It can be rewritten as

$$i_c = g_m v_{be} \quad \dots \dots \dots (3.24)$$

where g_m is the transconductance

from Eq.(3.23), it is given by

$$g_m = I_C / V_T \quad \dots \dots \dots (3.25)$$

We observe that the transconductance of the BJT is directly proportional to the collector bias current I_C . Thus to obtain a constant predictable value for g_m , we need a constant predictable I_C .

Also, we note that BJTs have relatively high transconductance :for instance, at $I_C = 1\text{mA}$, $g_m = 40\text{mA/V}$. g_m of a BJT depends only on the dc collector current at which it is biased to operate. A graphical interpretation for g_m is given in Fig.3.8, where it is shown that g_m is equal to the slope of the tangent to the i_C-v_{BE} characteristic curve at $i_C = I_C$ (i.e., at the bias point Q).

$$g_m = \partial i_C / \partial v_{BE} \quad \text{when } i_C = I_C \quad \dots \dots \dots (3.26)$$

The small-signal approximation implies keeping the signal amplitude sufficiently small that operation is restricted to an almost-linear segment of the i_C-v_{BE} exponential curve. Increasing the signal amplitude will result in the collector current having components non-linearly related to v_{be} .

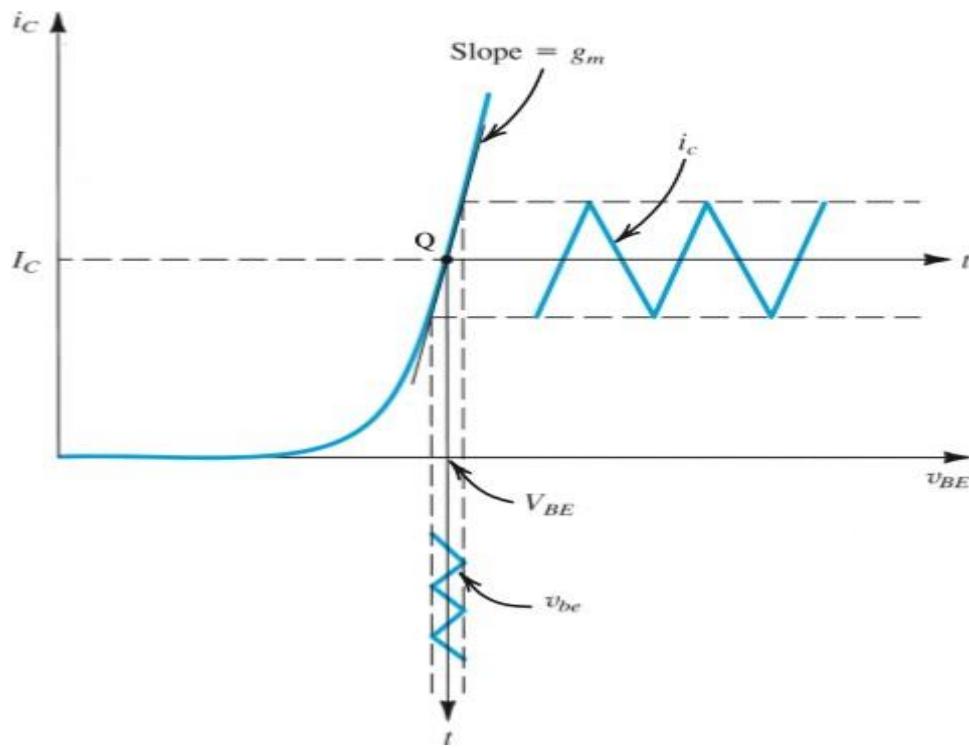


Fig:3.8 Linear operation of the transistor under the small-signal condition:

The Base Current and the Input Resistance at the Base:

To determine the resistance seen by v_{be} , we first evaluate the total base current i_B using Eq.(3.22), as follows:

$$i_B = i_C / \beta = I_C / \beta + (1 / \beta)(I_C / V_T)v_{be} \quad \dots \dots \dots (3.27)$$

Thus,

$$i_B = I_B + i_b \quad \dots \dots \dots (3.28)$$

where I_B is equal to I_C / β and the signal component i_b is given by

$$i_b = (1 / \beta)(I_C / V_T)v_{be} \quad \dots \dots \dots (3.29)$$

Substituting for I_C / V_T by g_m gives

$$i_b = (g_m / \beta)v_{be} \quad \dots \dots \dots (3.30)$$

The small-signal input resistance between base and emitter, looking into the base, is denoted by r_π and is defined as

$$r_\pi \equiv v_{be} / i_b \quad (3.31)$$

Using Eq.(3.26) gives

$$r_\pi = \beta / g_m \quad (3.32)$$

Thus r_π is directly dependent on β and is inversely proportional to the bias current I_C .

Substituting for g_m in Eq.(3.32) from Eq.(3.25) and replacing I_C/β by I_B gives an alternative expression for r_π ,

$$r_\pi = V_T / I_B \quad (3.33)$$

Here, we call that because the gate current of the MOSFET is zero (at dc and low frequencies) the input resistance at the gate is infinite; that is, in the MOSFET there is no counter part to r_π .⁵

The Emitter Current and the Input Resistance at the Emitter:

The total emitter current i_E can be determined using Eq.(0) as

$$i_E = i_C / \alpha = (I_C / \alpha) + (i_c / \alpha) \quad (3.34)$$

Thus,

$$i_E = I_E + i_e \quad (3.35)$$

where I_E is equal to I_C/α and the signal current i_e is given by

$$i_e = i_c / \alpha = (I_C / (\alpha VT)) v_{be} = (I_E / V_T) v_{be} \quad (3.36)$$

If we denote the small-signal resistance between base and emitter looking into the emitter by r_e , it can be defined as

$$r_e \equiv v_{be} / i_e \quad (3.37)$$

Using Eq.(3.36) we find that r_e , called the emitter resistance, is given by

$$r_e = V_T / I_E \quad (3.38)$$

Comparison with Eq.(3.25) reveals that

$$r_e = \alpha/g_m \approx 1/g_m \quad (3.39)$$

The relationship between r_π and r_e can be found by combining their respective definitions in Eqs.(3.31) and (3.37) as

$$v_{be} = i_b r_\pi = i_e r_e \quad (3.40)$$

Thus,

$$r_\pi = (i_e/i_b) r_e \quad (3.41)$$

which yields

$$r_\pi = (\beta + 1) r_e \quad (3.42)$$

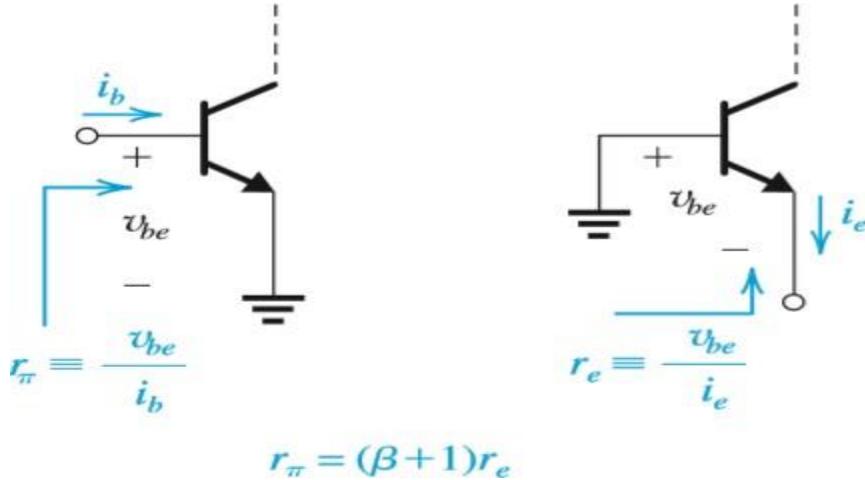


Fig:3.9 Illustrating the definition of r_π and r_e .

The Voltage Gain:

The total collector voltage v_{CE} is

$$\begin{aligned} v_{CE} &= V_{CC} - i_c R_C \\ &= V_{CC} - (I_C + i_c) R_C \quad (3.43) \\ &= (V_{CC} - I_C R_C) - i_c R_C \end{aligned}$$

$$= V_{CE} - i_c R_C$$

Thus, superimposed on the collector bias voltage V_{CE} we have signal voltage v_{ce} given by

$$\begin{aligned} v_{ce} &= -i_c R_C = -g_m v_{be} R_C \quad \dots \dots \dots (3.44) \\ &= (-g_m R_C) v_{be} \end{aligned}$$

from which we find the voltage gain A_v of this amplifier as

$$A_v \equiv v_{ce} / v_{be} = -g_m R_C \quad \dots \dots \dots (3.45)$$

Here again we note that because gm is directly proportional to the collector bias current, the gain will be as stable as the collector bias current is made. Substituting for gm from Eq.(3.25)

enables us to express the gain in the form

$$A_v = - (I_C R_C) / V_T \quad \dots \dots \dots (3.46)$$

Separating the Signal and the DC Quantities:

The analysis above indicates that every current and voltage in the amplifier circuit of Fig.3.7(a) is composed of two components: a dc component and a signal component. For instance, $v_{BE} = V_{BE} + v_{be}$, $I_C = I_C + i_c$, and so on. The dc components are determined from the dc circuit given in Fig.3.7(b) and from the relationships imposed by the transistor. On the other hand, a representation of the signal operation of the BJT can be obtained by eliminating the dc sources, as shown in Fig.3.10. Observe that since the voltage of an ideal dc supply does not change, the signal voltage across it will be zero. For this reason we have replaced V_{CC} and V_{BE} with short circuits. Had the circuit contained ideal dc current sources, these would have been replaced by open circuits. Note, however, that the circuit of Fig.3.10 is useful only insofar as it shows the various signal currents and voltages; it is not an actual amplifier circuit, since the dc bias circuit is not shown.

Note that this is a representation of the signal operation of the BJT and not an actual amplifier circuit.

a circuit should have three terminals—C, B, and E—and should yield the same terminal currents indicated in Fig.3.10. The resulting circuit is then equivalent to the transistor as far as small-signal operation is concerned, and thus it can be considered an equivalent small-signal circuit model.

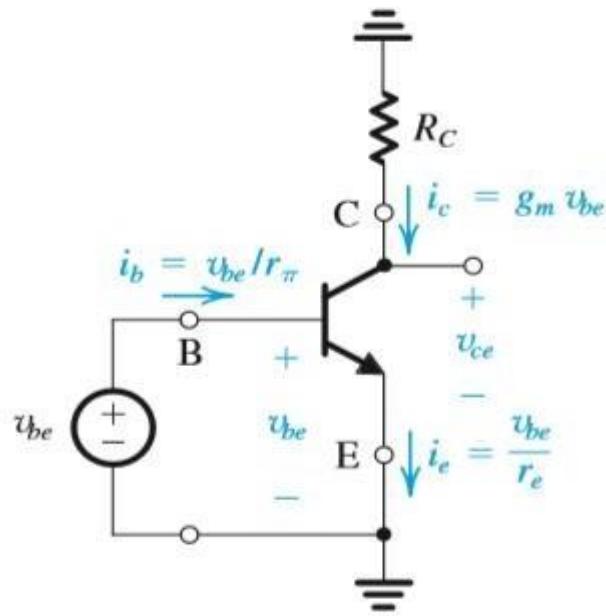


Fig:3.10 The amplifier circuit of Fig. 3.7(a) with the dc sources (V_{BE} and V_{CC}) eliminated (short-circuited). Thus only the signal components are present. .

The Hybrid- π Model:

An equivalent-circuit model for the BJT is shown in Fig.3.11(a). This model represents the BJT as a voltage-controlled current source and explicitly includes the input resistance looking into the base, r_π . The model obviously yield $i_c=g_m v_{be}$ and $i_b=v_{be}/r_\pi$. Not so obvious, however, is the fact that the model also yields the correct expression for i_e . This can be shown as follows: At the emitter node we have

$$\begin{aligned}
 i_e &= (v_{be} / r_\pi) + g_m v_{be} = v_{be} r_\pi (1 + g_m r_\pi) \\
 &= v_{be} / r_\pi (1 + \beta) = v_{be} / (r_\pi / (1 + \beta)) \\
 &= v_{be} / r_e
 \end{aligned}$$

A slightly different equivalent-circuit model can be obtained by expressing the current of the controlled source ($g_m v_{be}$) in terms of the base current i_b as follows:

$$g_m v_{be} = g_m (i_b r_\pi)$$

$$= (\mathbf{g}_m \mathbf{r}_\pi) \mathbf{i}_b = \beta \mathbf{i}_b$$

This results in the alternative equivalent-circuit model shown in Fig.3.11(b). Here the transistor is represented as a current-controlled current source, with the control current being i_b .

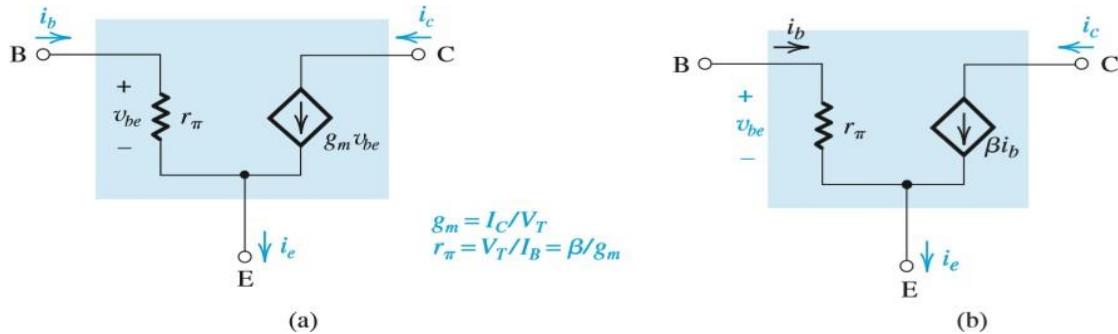


Fig:3.11 Two slightly different versions of the hybrid- π model for the small-signal operation of the BJT.

The equivalent circuit in (a) represents the BJT as a voltage-controlled current source (a transconductance amplifier), and that in (b) represents the BJT as a current-controlled current source (a current amplifier).

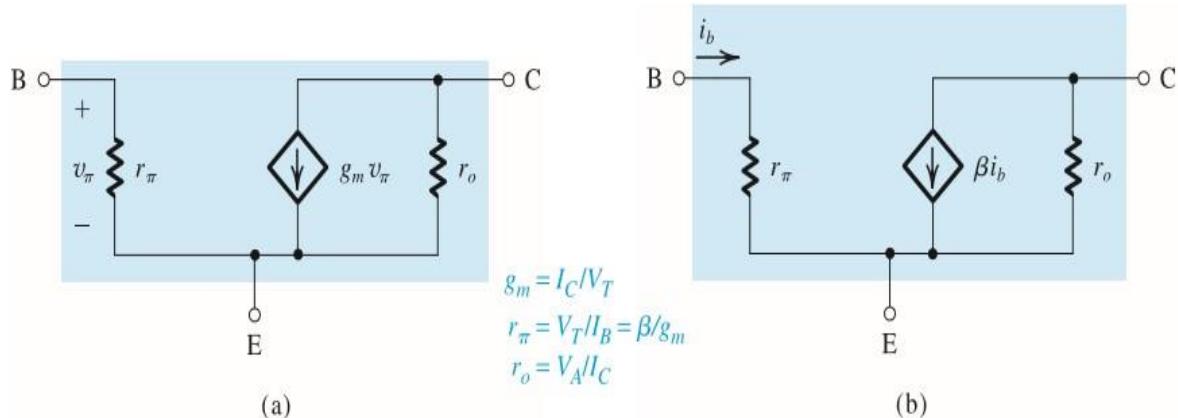


Fig:3.12 The hybrid- π small-signal model, in its two versions, with the resistance r_o included.

It is important to note that the small-signal equivalent circuits of Fig.3.12 model the operation of the BJT at a given bias point. This should be obvious from the fact that the model parameters g_m , r_π , and r_o depend on the value of the dc bias current I_C , as indicated in Fig.3.12. That is, these equivalent circuits model the incremental operation of the BJT around the bias point.

Including r_o in the BJT model causes the voltage gain of the conceptual amplifier of Fig.3.7(a) to become

$$v_{be} = -g_m(R_C \parallel r_o) \quad \text{----- (3.47)}$$

Thus, the magnitude of the gain is reduced somewhat.

The T Model:

Although the hybrid- π model can be used to carry out small-signal analysis of any transistor circuit, there are situations in which an alternative model, shown in Fig.3.13, is much more convenient. The model of Fig.3.13(a) represents the BJT as a voltage-controlled current source with the control voltage being v_{be} . Here, however, the resistance between base and emitter, looking into the emitter, is explicitly shown. From Fig.3.13(a) we see clearly that the model yields the correct expressions for i_c and i_e . It can also be shown to yield the correct expression for i_b .

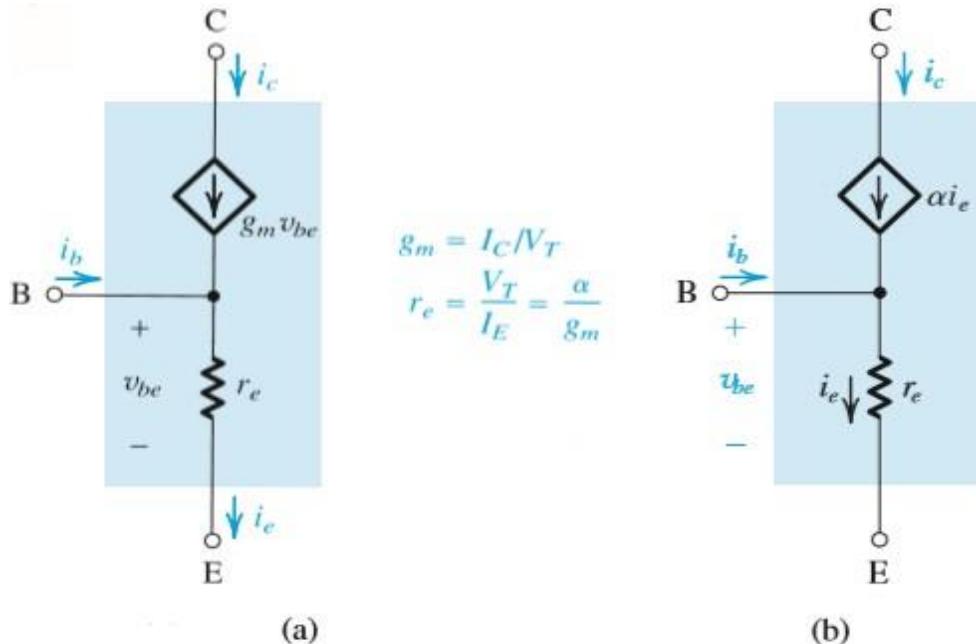


Fig:3.13 Two slightly different versions of what is known as the T model of the BJT. The circuit in (a) is a voltage-controlled current source representation and that in (b) is a current-controlled current source representation. These models explicitly show the emitter resistance r_e rather than the base resistance r_π featured in the hybrid- π model.

If in the model of Fig.3.13(a) the current of the controlled source is expressed in terms of the emitter current as

$$g_m v_{be} = g_m (i_e r_e)$$

$$= (g_m r_e) i_e = \alpha i_e$$

we obtain the alternative T model shown in Fig.3.13(b). Here the BJT is represented as a current-controlled current source but with the control signal being i_e .

Finally, the T models can be augmented by r_o to account for the dependence of i_c to v_{ce} (the Early effect) to obtain the equivalent circuits shown in Fig.3.14.

Small-Signal Models of the pnp Transistor:

Although the small-signal models in Figs.3.12 and 3.14 were developed for the case of the npn transistor, they apply equally well to the pnp transistor with no change in polarities.

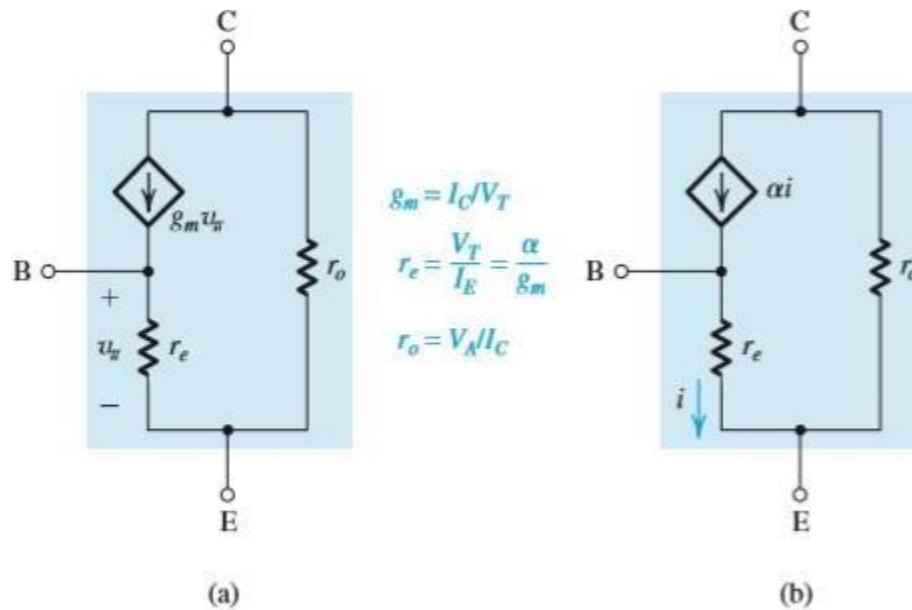


Fig:3.14 The T models of the BJT

Basic BJT amplifier configurations:

The Three Basic Configurations:

There are three basic configurations for connecting a BJT as an amplifier. Each of these configurations is obtained by connecting one of the device terminals to ground, thus creating a two-port network with the grounded terminal being common to the input and output ports. The resulting configurations are shown in Fig.3.15(d–f) for the BJT.

The BJT counterpart, the **grounded-emitter or common-emitter (CE)** amplifier in Fig.3.15(d). In the circuit of the emitter terminal is connected to ground, the input voltage signal v_i is applied between the base and ground, and the output voltage signal v_o is taken between the collector and ground, across the resistance R_C .

The **common-base (CB) or grounded-base amplifier** in Fig.3.15(e). Here the base is grounded, the input signal v_i is applied to the emitter, and the output signal v_o is taken at the collector across the resistance R_C . A CB amplifier in Example 7.7.

The **common-collector (CC) or grounded collector amplifier**. Here the collector terminal is grounded, the input signal v_i is applied between base and ground, and the output voltage v_o is taken between the emitter and ground, across a resistance R_L . For reasons that will become apparent shortly, this pair of configurations is more commonly called the **emitter follower**.

Our study of the three basic amplifier configurations of the BJT will reveal that each has distinctly different attributes, hence areas of application. As well, it will be shown that although each pair of configurations, (e.g., CE), has many common attributes, important differences remain.

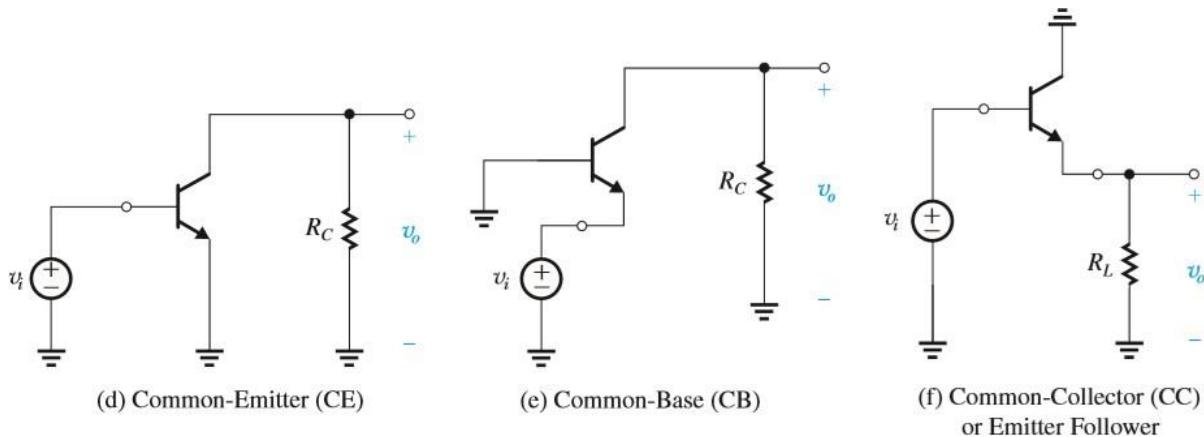


Fig:3.15 The basic configurations of transistor amplifiers. (d)–(f) for the BJT.

Our next step is to replace the transistor in each of the six circuits in Fig.3.15 by an appropriate equivalent-circuit model and analyze the resulting circuits to determine important characteristic parameters of the particular amplifier configuration. To simplify matters, we shall not include r_o in the initial analysis. At the end of the section we will offer a number of comments about when to include r_o in the analysis, and on the expected magnitude of its effect.

Characterizing Amplifiers:

Before we begin our study of the different transistor amplifier configurations, we consider how to characterize the performance of an amplifier as a circuit building block.

Figure 3.16(a) shows an amplifier fed with a signal source having an open-circuit voltage v_{sig} and an internal resistance R_{sig} . These can be the parameters of an actual signal source or, in a cascade amplifier, the Thevenin equivalent of the output circuit of another amplifier stage preceding the one under study. The amplifier is shown with a load resistance R_L connected to the output terminal. Here, R_L can be an actual load resistance or the input resistance of a succeeding amplifier stage in a cascade amplifier.

Figure 3.16(b) shows the amplifier circuit with the amplifier block replaced by its equivalent-circuit model. The input resistance R_{in} represents the loading effect of the amplifier input on the signal source. It is found from

$$R_{in} \equiv v_i / i_i$$

and together with the resistance R_{sig} forms a voltage divider that reduces v_{sig} to the value v_i that appears at the amplifier input,

$$v_i = (R_{in} / (R_{in} + R_{sig})) v_{sig} \quad \dots \quad (3.48)$$

Most of the amplifier circuits studied in this section are **unilateral**. That is, they do not contain internal feedback, and thus R_{in} will be independent of R_L . However, in general R_{in} may depend on the load resistance R_L . Indeed one of the six configurations studied in this section, the emitter follower, exhibits such dependence.

The second parameter in characterizing amplifier performance is the **open-circuit voltage gain** A_{vo} , defined as

$$A_{vo} \equiv v_o / v_i \quad \text{when, } R_L = \infty$$

The third and final parameter is the output resistance R_o . Observe from Fig.3.16(b) that R_o is the resistance seen looking back into the amplifier output terminal with v_i set to zero. Thus R_o can be determined, at least conceptually, as indicated in Fig.3.16(c) with

$$R_o = v_o / i_x$$

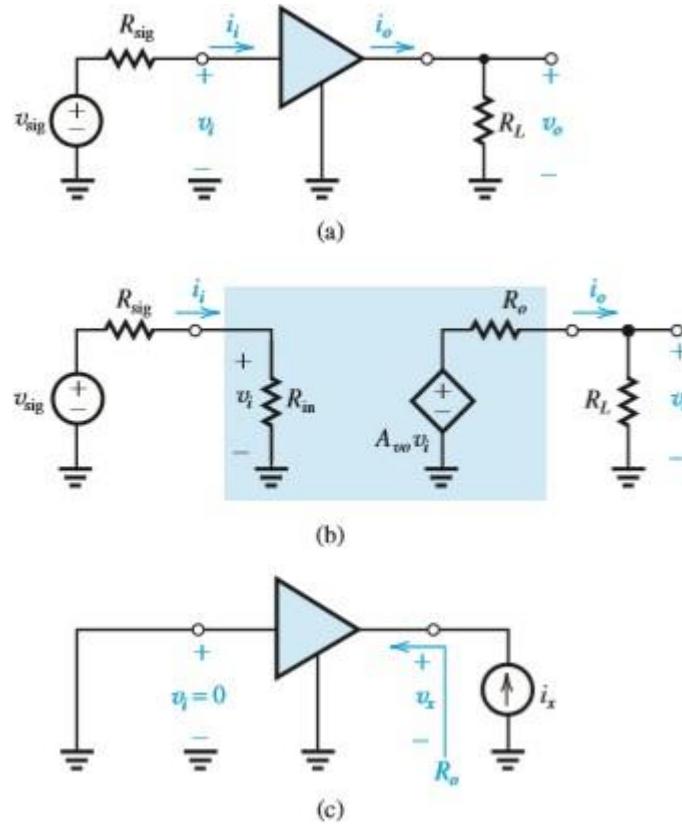


Fig:3.16 Characterization of the amplifier as a functional block:(a) An amplifier fed with a voltage signal v_{sig} having a source resistance R_{sig} , and feeding a load resistance R_L ; (b) equivalent-circuit representation of the circuit in (a); (c) determining the amplifier output resistance R_o

Because R_o is determined with $v_i = 0$, the value of R_o does not depend on R_{sig} .

The controlled source $A_{vo}v_i$ and the output resistance R_o represent the Thevenin equivalent of the amplifier output circuit, and the output voltage v_o can be found from

$$v_o = (R_L / (R_L + R_o)) A_{vo} v_i \dots \dots \dots \quad (3.49)$$

Thus the voltage gain of the amplifier proper, A_v , can be found as

$$A_v \equiv v_o / v_i = A_{vo} (R_L / (R_L + R_o)) \dots \dots \dots \quad (3.50)$$

and the overall voltage gain, G_v ,

$$G_v \equiv v_o / v_{sig}$$

can be determined by combining Eqs.(7.83) and (7.85):

$$G_v = (R_{in} / (R_{in} + R_{sig})) A_{vo} (R_L / (R_L + R_o)) \dots \quad (3.51)$$

The Common-Emitter (CE) Amplifiers:

Of the three basic transistor amplifier configurations, the common-emitter, for BJT, is the most widely used. Typically, in an amplifier formed by cascading a number of gain stages, the bulk of the voltage gain is obtained by using one or more common-emitter, for BJT stages in cascade.

Characteristic Parameters of the CE Amplifier:

Figure 3.17(a) shows a common-emitter amplifier. Its equivalent circuit, obtained by replacing the BJT with its hybrid- π model (without r_o), is shown in Fig.3.17(b). The latter circuit can be analyzed to obtain the characteristic parameters of the CE amplifier. Here we have the added complexity of a finite input resistance r_π . Tracing the signal through the amplifier from input to output, we can write by inspection

$$R_{in} = r_\pi$$

Then we write

$$v_i = (r_\pi / (r_\pi + R_{sig})) v_{sig} \dots \quad (3.52)$$

$$v_\pi = v_i$$

$$v_o = -g_m v_\pi R_C$$

Thus,

$$A_{vo} \equiv v_o / v_i = -g_m R_C \dots \quad (3.53)$$

$$R_o = R_C \dots \quad (3.54)$$

With a load resistance R_L connected across R_C ,

$$A_v = -g_m (R_C || R_L) \dots \quad (3.55)$$

and the overall voltage gain G_v can be found from

$$G_v \equiv v_o / v_{sig} = (v_i / v_{sig}) (v_o / v_i)$$

Thus

$$\mathbf{G}_v = -(\mathbf{r}_\pi) / (\mathbf{r}_\pi + \mathbf{R}_{sig}) \mathbf{g}_m(\mathbf{R}_C \| \mathbf{R}_L) \quad \dots \dots \dots \quad (3.56)$$

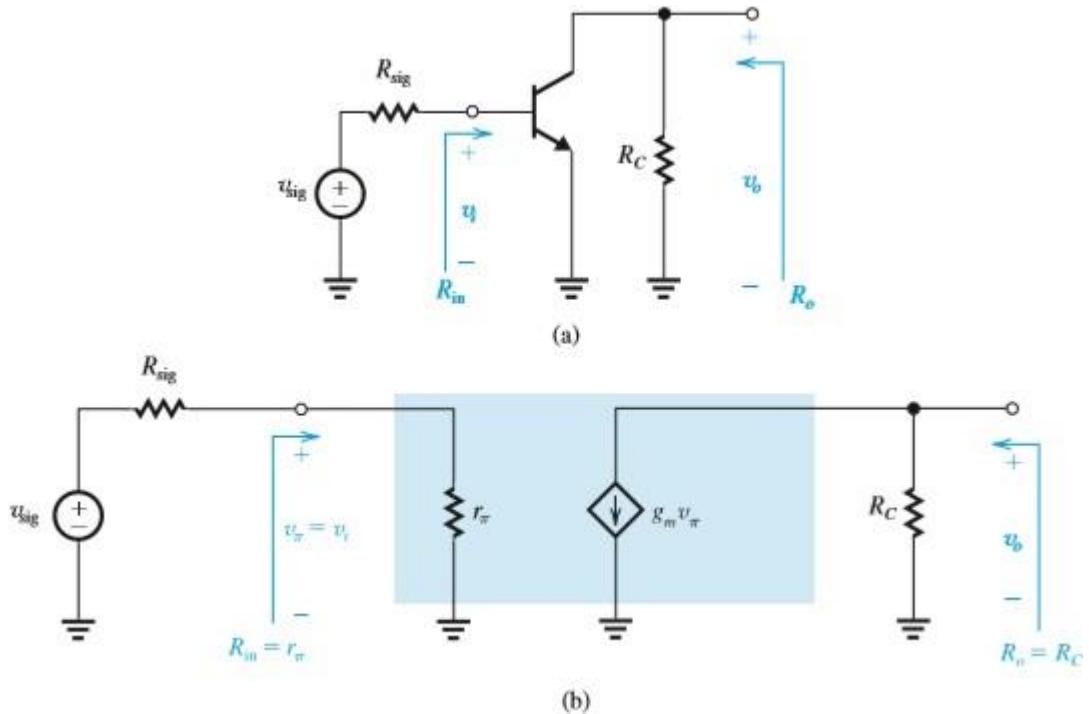


Fig:3.17 (a) Common-emitter amplifier fed with a signal v_{sig} from a generator with a resistance R_{sig} . The bias circuit is omitted.(b) The common-emitter amplifier circuit with the BJT replaced by its hybrid- π model.

It is important to note here the effect of the finite input resistance (r_π) in reducing the magnitude of the voltage gain by the voltage-divider ratio $r_\pi/(r_\pi + R_{sig})$. The extent of the gain reduction depends on the relative values of r_π and R_{sig} . However, there is a compensating effect in the CE amplifier: gm of the BJT is usually much higher.

Common-Emitter Amplifier with a Emitter Resistance:

It is often beneficial to insert a resistance R_E in the emitter lead of a common-emitter amplifier. Figure 3.18(a) shows a CE amplifier with a resistance R_E in its emitter lead. The corresponding small-signal equivalent circuit is shown in Fig. 3.18(b). Note that in the BJT case also, as a

general rule, the T model results in a simpler analysis and should be employed whenever there is a resistance in series with the emitter.

To determine the amplifier input resistance R_{in} , we note from Fig.3.18(b) that

$$R_{in} \equiv v_i / i_b$$

Where

$$i_b = (1 - \alpha)i_e = i_e / \beta + 1 \quad \dots \dots \quad (3.57)$$

and

$$i_e = v_i / (r_e + R_e) \quad \dots \dots \quad (3.58)$$

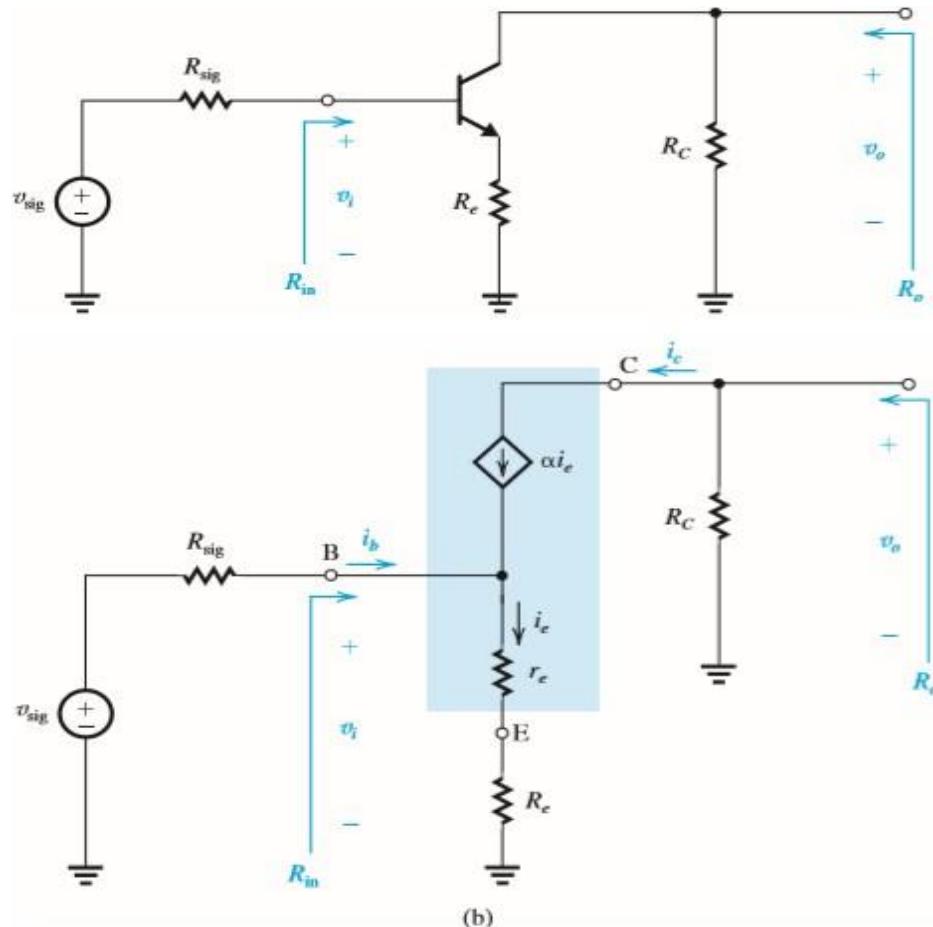


Fig:3.18 The CE amplifier with an emitter resistance R_e ; (a) circuit without bias details; (b) equivalent circuit with the BJT replaced with its T model.

Thus,

$$R_{in} = (\beta + 1)(r_e + R_e) \quad \dots \dots \dots (3.59)$$

This is a very important result. It states that the input resistance looking into the base is $(\beta + 1)$ times the total resistance in the emitter, and is known as the resistance-reflection rule. The factor $(\beta + 1)$ arises because the base current is $1/(\beta + 1)$ times the emitter current. The expression for R_{in} in Eq.(3.59) shows clearly that including a resistance R_e in the emitter can substantially increase R_{in} , a very desirable result. Indeed, the value of R_{in} is increased by the ratio.

$$\begin{aligned} & R_{in}(\text{with } R_e \text{ included}) / R_{in}(\text{without } R_e) \\ &= (\beta + 1)(r_e + R_e) / (\beta + 1)r_e \\ &= 1 + (R_e / r_e) \sim 1 + g_m R_e \quad \dots \dots \dots (3.60) \end{aligned}$$

Thus the circuit designer can use the value of R_e to control the value of R_{in} .

To determine the voltage gain A_{vo} , we see from Fig.3.18(b) that

$$v_o = -i_c R_C = -\alpha i_e R_C$$

Substituting for i_e from Eq.(3.58) gives

$$A_{vo} = -\alpha (R_C / (r_e + R_e)) \quad \dots \dots \dots (3.61)$$

This is a very useful result: It states that the gain from base to collector is α times the ratio of the total resistance in the collector to the total resistance in the emitter (in this case, $r_e + R_e$),

$$\begin{aligned} & \text{Voltage gain from base to collector} = -\alpha \left(\frac{\text{Total resistance in collector}}{\text{Total resistance in emitter}} \right) \quad \dots \dots \dots (3.62) \end{aligned}$$

This factor arises because $i_c = \alpha i_e$, unlike the MOSFET case where $i_d = i_s$. Usually, $\alpha \ll 1$ and can be dropped from Eq.(3.62). The open-circuit voltage gain in Eq.(3.61) can be expressed alternatively

$$A_{vo} = -(\alpha R_C / r_e (1 + R_e / r_e))$$

Thus,

$$A_{vo} = -(g_m R_C) / (1 + R_e / r_e) \quad \dots \dots \dots (3.63)$$

$$\sim - (g_m R_C) / (1 + g_m R_e)$$

Thus, including R_e reduces the voltage gain by the factor $(1+g_m R_e)$, which is the same factor by which R_{in} is increased. This points out an interesting trade-off between gain and input resistance, a trade-off that the designer can exercise through the choice of an appropriate value for R_e . The output resistance R_o can be found from the circuit in Fig.3.18(b) by inspection:

$$R_o = R_C$$

If a load resistance R_L is connected at the amplifier output, A_v can be found as

$$\begin{aligned} A_v &= A_{vo} [R_L / (R_L + R_o)] \\ &= -\alpha (R_C / (r_e + R_e)) (R_L / (R_L + R_C)) \\ &= -\alpha (R_C || R_L) / (r_e + R_e) \quad \text{----- (3.64)} \end{aligned}$$

which could have been written directly using Eq.(3.62). The overall voltage gain G_v can now be found:

$$G_v = (R_{in}) / (R_{in} + R_{sig}) \times -\alpha (R_C || R_L) / (r_e + R_e) \quad \text{----- (3.65)}$$

Substituting for R_{in} from Eq.(3.59) and replacing α with $\beta/(\beta+1)$ results in

$$G_v = -\beta [(R_C || R_L) / (R_{sig} + (\beta+1)(r_e + R_e))] \quad \text{----- (3.66)}$$

Careful examination of this expression reveals that the denominator comprises the total resistance in the base circuit [recall that $(\beta+1)(r_e + R_e)$ is the reflection of $(r_e + R_e)$ from the emitter side to the base side]. Thus the expression in Eq.(3.65) states that the voltage gain from base to collector is equal to β times the ratio of the total resistance in the collector to the total resistance in the base. The factor β appears because it is the ratio of the collector current to the base current. We observe that the overall voltage gain G_v is lower than the value without R_e , namely,

$$G_v = -\beta [(R_C || R_L) / (R_{sig} + (\beta+1)r_e)] \quad \text{----- (3.67)}$$

because of the additional term $(\beta+1)R_e$ in the denominator. The gain, however, is now less sensitive to the value of β , a desirable result because of the typical wide variability in the value of β . Another important consequence of including the resistance R_e in the emitter is that it enables the amplifier to handle larger input signals without incurring nonlinear distortion. This is because only a fraction of the input signal at the base, v_i , appears between the base and the emitter. Specifically, from the circuit in Fig.3.18(b), we see that

$$v_\pi / v_i = (r_e / (r_e + R_e)) \sim 1 / (1 + g_m R_e) \quad \text{----- (3.68)}$$

Thus, for the same v_π , the signal at the input terminal of the amplifier, v_i , can be greater than for the CE amplifier by the factor $(1+g_m R_e)$. To summarize, including a resistance R_e in the emitter of the CE amplifier results in the following characteristics:

1. The input resistance R_{in} is increased by the factor $(1+g_m R_e)$.
2. The voltage gain from base to collector, A_v , is reduced by the factor $(1+g_m R_e)$.
3. For the same nonlinear distortion, the input signal v_i can be increased by the factor $(1+g_m R_e)$.
4. The overall voltage gain is less dependent on the value of β .
5. The high-frequency response is significantly improved.

With the exception of gain reduction, these characteristics represent performance improvements. Indeed, the reduction in gain is the price paid for obtaining the other performance improvements. In many cases this is a good bargain; it is the underlying philosophy for the use of negative feedback. That the resistance R_e introduces negative feedback in the amplifier circuit .where we shall study negative feedback formally, we will find that the factor $(1+g_m R_e)$,which appears repeatedly,is the “amount of negative feedback”introduced by R_e . Finally,we note that the negative-feedback action of R_e gives it the name emitter degeneration resistance.

The Common-Base (CB) Amplifier:

The CB amplifier shown in Fig.3.19(a). Specifically, from the equivalent circuit in Fig.3.19(b) we can find

$$R_{in} = r_e = (\alpha / g_m) \approx (1/g_m) \quad \dots \dots \dots \quad (3.69)$$

$$A_{vo} = (\alpha / r_e) R_C = g_m R_C \quad \dots \dots \dots \quad (3.70)$$

$$R_o = R_C \quad \dots \dots \dots \quad (3.71)$$

and with a load resistance R_L connected to the output, the overall voltage gain is given by

$$G_v \equiv v_o / v_{sig} = \alpha [(R_C || R_L) / (R_{sig} + r_e)] \quad \dots \dots \dots \quad (3.72)$$

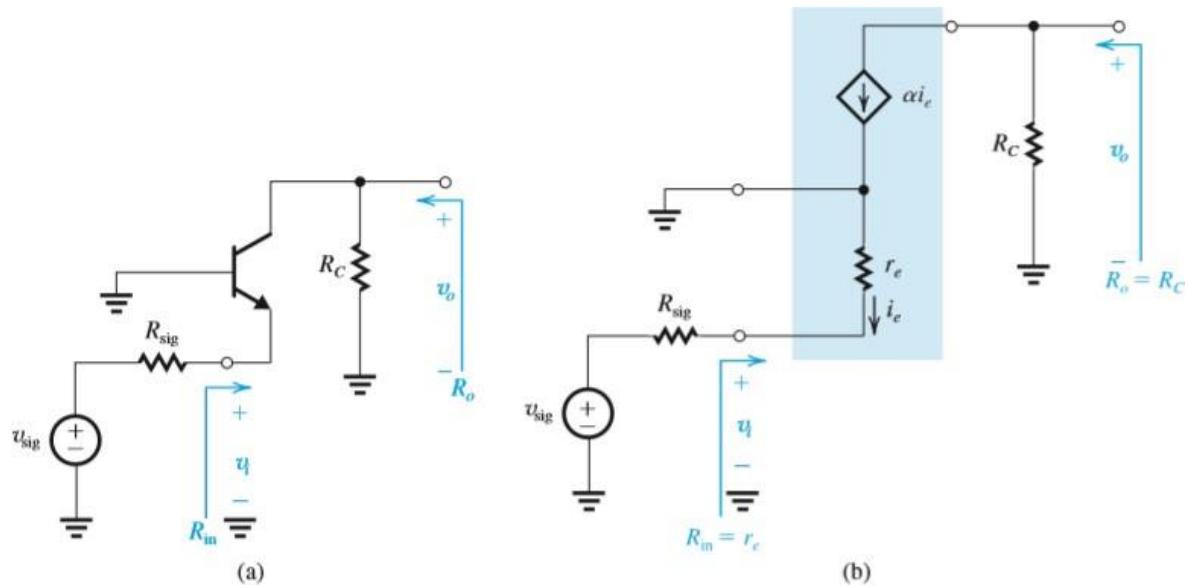


Fig:3.19 (a) CB amplifier with bias details omitted; (b) amplifier equivalent circuit with the BJT represented by its T model.

Since $\alpha \sim 1$, we see that as in the case of the CG amplifier, the overall voltage gain is simply the ratio of the total resistance in the collector to the total resistance in the emitter. We also note that the overall voltage gain is almost independent of the value of β (except through the small dependence of α on β), a desirable property. Observe that for R_{sig} of the same order as R_C and R_L , the gain will be very small. In summary, the CB amplifiers exhibit a very low input resistance ($1/gm$), an open-circuit voltage gain that is positive and equal in magnitude to that of the CE amplifier, and, like the CE amplifier, a relatively high output resistance. Because of its very low input resistance, the CB circuit alone is not attractive as a voltage amplifier except in specialized applications, such as the cable amplifier mentioned above. The CB amplifier has excellent high-frequency performance.

The Common-Collector or Emitter Followers:

The last of the basic transistor amplifier configurations is the common-collector amplifier, an important circuit that finds application in the design of both small-signal amplifiers and amplifiers that are required to handle large signals and deliver substantial amounts of signal power to a load. The common-collector amplifier is more commonly known as the emitter follower. The reason behind these names will become apparent shortly.

The Need for Voltage Buffers: Before embarking on the analysis of the source and the emitter followers, it is useful to look at one of their more common applications. Consider the situation depicted in Fig.3.20(a). A signal source delivering a signal of reasonable

strength (1V) with an internal resistance of 1M is to be connected to a 1-k load resistance. Connecting the source to the load directly as in Fig.3.20(b) would result in severe attenuation of the signal; the signal appearing across the load will be only $1/(1000+1)$ of the input signal, or about 1mV. An alternative course of action is suggested in Fig.3.20(c). Here we have interposed an amplifier between the source and the load. Our amplifier, however, is unlike the amplifiers we have been studying in this chapter thus far; it has a voltage gain of only unity. This is because our signal is already of sufficient strength and we do not need to increase its amplitude. Note, however, that our amplifier has a very high input resistance, thus almost all of v_{sig} (i.e., 1V) will appear at the input of the amplifier proper. Since the amplifier has a low output resistance (100), 90% of this signal (0.9V) will appear at the output, obviously a very significant improvement over the situation without the amplifier. As will be seen next, the source follower can easily implement the unity-gain buffer amplifier shown in Fig.3.20(c).

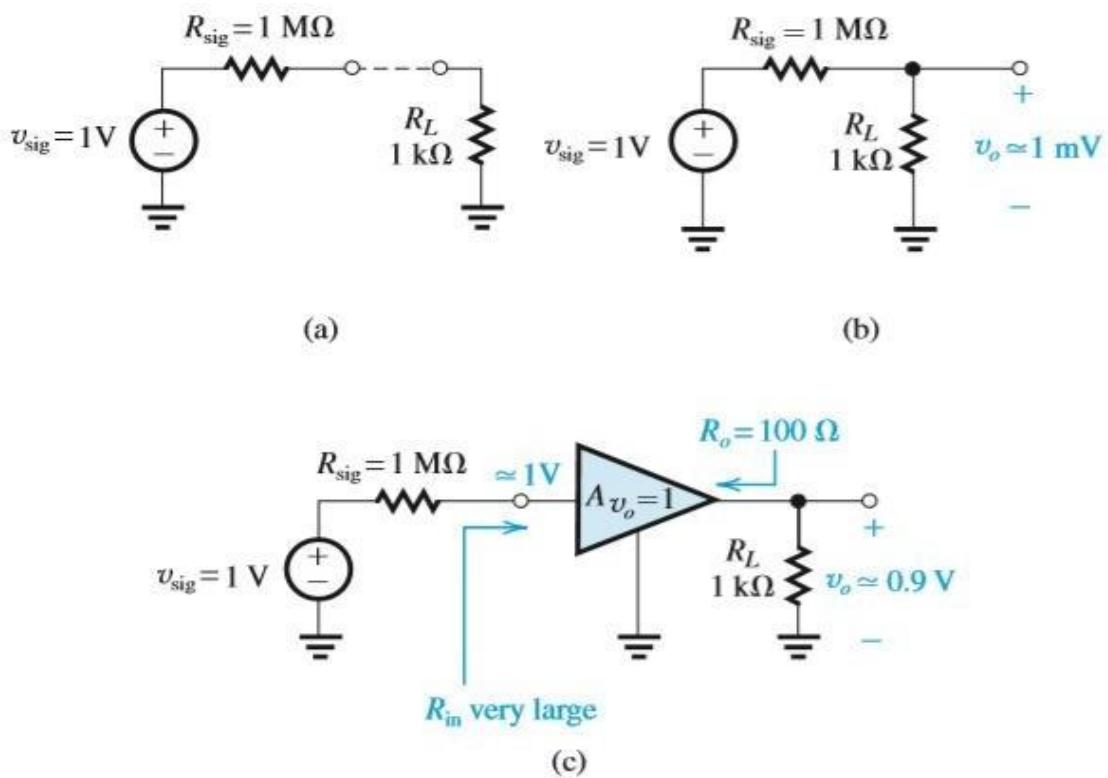


Fig:3.20 Illustrating the need for a unity-gain voltage buffer amplifier.

Characteristic Parameters of the Emitter Follower:

Although the emitter follower does not have an infinite input resistance, it is still widely used as a voltage buffer. In fact, it is a very versatile and popular circuit. We will therefore study it in

some detail. Figure 3.21(a) shows an emitter follower with the equivalent circuit shown in Fig.3.21(b). The input resistance R_{in} is found from

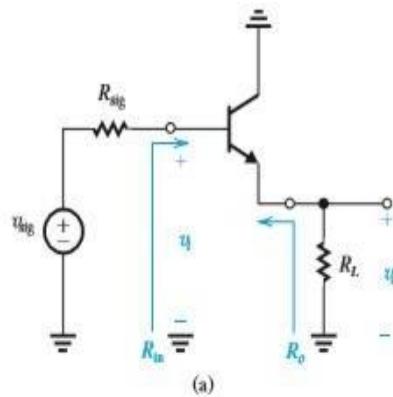
$$R_{in} = v_i / i_b$$

Substituting for $i_b = i_e / (\beta + 1)$ where i_e is given by

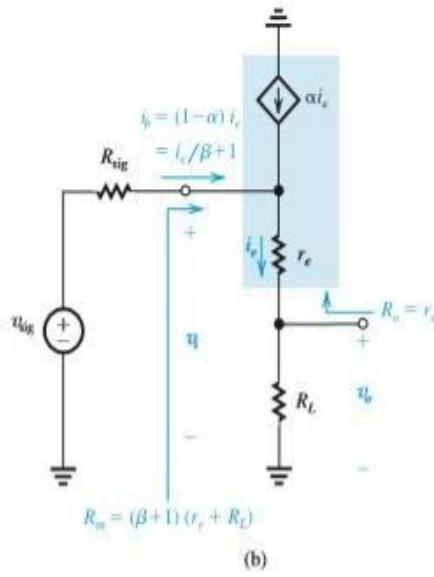
$$i_e = v_i / (r_e + R_L)$$

we obtain

$$R_{in} = (\beta + 1)(r_e + R_L) \dots\dots\dots (3.73)$$



(a)



(b)

Fig:3.21 (a) Common-collector amplifier or emitter follower with the bias circuit omitted.(b) Equivalent circuit obtained by replacing the BJT with its T model.

A result that we could have written directly, utilizing the resistance-reflection rule. Note that as expected the emitter follower takes the low load resistance and reflects it to the base side, where

the signal source is, after increasing its value by a factor $(\beta+1)$. It is this impedance transformation property of the emitter follower that makes it useful in connecting a low-resistance load to a high-resistance source, that is, to implement a buffer amplifier. The voltage gain A_v is given by

$$A_v = v_o / v_i = R_L / (R_L + r_e) \quad \dots \dots \dots (3.74)$$

Setting $R_L = \infty$ yields A_{vo} ,

$$A_{vo} = 1 \quad \dots \dots \dots (3.75)$$

Thus, as expected, the open-circuit voltage gain of the emitter follower proper is unity, which means that the signal voltage at the emitter follows that at the base, which is the origin of the name “emitter follower.” To determine R_o , refer to Fig. 3.21(b) and look back into the emitter (i.e., behind or excluding R_L) while setting $v_i = 0$ (i.e., grounding the base). You will see r_e of the BJT, thus

$$R_o = r_e \quad \dots \dots \dots (3.76)$$

This result together with $A_{vo}=1$ yields A_v in Eq. (3.74), thus confirming our earlier analysis. We next determine the overall voltage gain G_v , as follows:

$$\begin{aligned} v_i / v_{sig} &= R_{in} / [R_{in} + R_{sig}] \\ &= [(\beta+1)(r_e + R_L)] / [(\beta+1)(r_e + R_L) + R_{sig}] \\ G_v \equiv v_o / v_{sig} &= (v_i / v_{sig}) A_v \end{aligned}$$

Substituting for A_v from Eq. (7.130) results in

$$G_v = (\beta+1)R_L / [(\beta+1)R_L + (\beta+1)r_e + R_{sig}] \quad \dots \dots \dots (3.77)$$

This equation indicates that the overall gain, though lower than one, can be close to one if $(\beta+1)R_L$ is larger or comparable in value to R_{sig} . This again confirms the action of the emitter follower in delivering a large proportion of v_{sig} to a low-valued load resistance R_L even though R_{sig} can be much larger than R_L . The key point is that R_L is multiplied by $(\beta+1)$ before it is “presented to the source.” Figure 3.21(a) shows an equivalent circuit of the emitter follower obtained by simply reflecting r_e and R_L to the base side. The overall voltage gain $G_v \equiv v_o / v_{sig}$ can be determined directly and very simply from this circuit by using the voltage divider rule. The result is the expression for G_v already given in Eq. (3.77). Dividing all resistances in the circuit of Fig. 3.21(a) by $\beta+1$ does not change the voltage ratio v_o / v_{sig} . Thus we obtain another equivalent circuit, shown in Fig. 3.21(b), that can be used to determine $G_v \equiv v_o / v_{sig}$ of the emitter follower. A glance at this circuit reveals that it is simply the equivalent circuit obtained by reflecting v_{sig} and R_{sig} from the base side to the emitter side. In this reflection, v_{sig} does not change, but R_{sig} is

divided by $\beta+1$. Thus, we either reflect to the base side and obtain the circuit in Fig.7.3.21(a) or reflect to the emitter side and obtain the circuit in Fig.3.21(b). From the latter, G_v can be found as

$$G_v \equiv v_o / v_{sig} = R_L / (R_L + r_e + (R_{sig}/(\beta+1))) \quad \dots \quad (3.78)$$

Observe that this expression is the same as that in Eq.(3.77) except for dividing both the numerator and denominator by $\beta+1$.

The expression for G_v in Eq.(3.78) has an interesting interpretation :The emitter follower reduces R_{sig} by the factor $(\beta+1)$ before “presenting it to the load resistance R_L ”:an impedance transformation that has the same buffering effect.

At this point it is important to note that although the emitter follower does not provide voltage gain it has a current gain of $\beta+1$.

Thévenin Representation of the Emitter-Follower Output:

A more general representation of the emitter-follower output is shown in Fig.3.22(a). Here G_{vo} is the overall open-circuit voltage gain that can be obtained by setting $R_L=\infty$ in the circuit of Fig.3.22(b), as illustrated in Fig.3.22(b).The result is $G_{vo}=1$.The output resistance R_{out} is different from R_o . To determine R_{out} we set v_{sig} to zero (rather than setting v_i to zero). Again we can use the equivalent circuit in Fig.3.22(b) to do this, as illustrated in Fig.3.22(c). We see that

$$R_{out} = r_e + [R_{sig} / (\beta+1)]$$

Finally, we show in Fig.3.22(d) the emitter-follower circuit together with its R_{in} and R_{out} . Observe that R_{in} is determined by reflecting r_e and R_L to the base side (by multiplying their values by $\beta+1$).To determine R_{out} , grab hold of the emitter and walk (or just look!) backward while $v_{sig}=0$.You will see r_e in series with R_{sig} , which because it is in the base must be divided by $(\beta+1)$. We note that unlike the amplifier circuits we studied earlier, the emitter follower is not unilateral. This is manifested by the fact that R_{in} depends on R_L and R_{out} depends on R_{sig} .

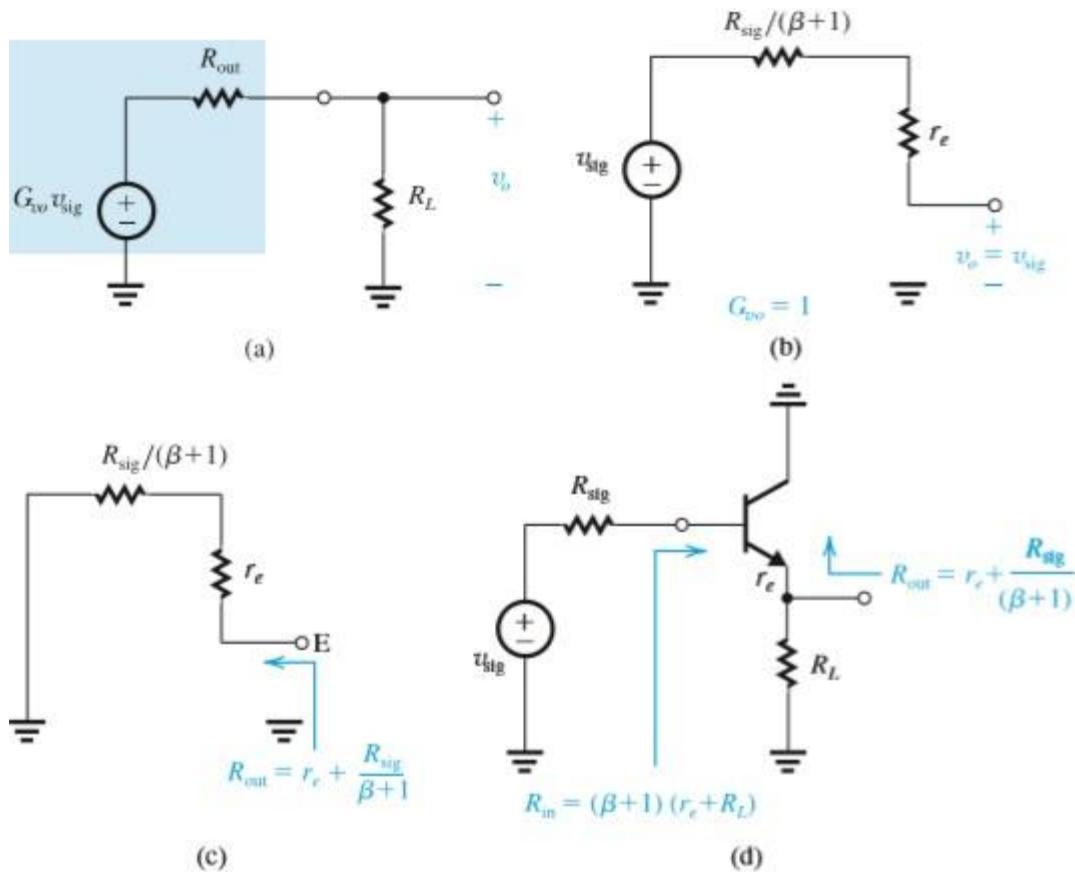


Fig:3.22 (a) Thévenin representation of the output of the emitter follower. (b) Obtaining G_{vo} from the equivalent circuit in Fig. 3.22(b). (c) Obtaining R_{out} from the equivalent circuit in Fig. 3.22(b) with v_{sig} set to zero. (d) The emitter follower with R_{in} and R_{out} determined simply by looking into the input and output terminals, respectively.

Tables and Comparisons:

The low input resistance of the CG and CB amplifiers makes them useful only in specific applications. As we shall see in Chapter 10, these two configurations exhibit a much better high-frequency response than that available from the CS and CE amplifiers. This makes them useful as high-frequency amplifiers, especially when combined with the CS or CE circuit.

The source follower (emitter follower) finds application as a voltage buffer for connecting a high-resistance source to a low-resistance load, and as the output stage in a multistage amplifier, where its purpose is to equip the amplifier with a low output resistance.

Table 7.5 Characteristics of BJT Amplifiers^{a,b}

	R_{in}	A_{vo}	R_o	A_v	G_v
Common emitter (Fig. 7.36)	$(\beta + 1)r_e$	$-g_m R_C$	R_C	$-g_m (R_C \parallel R_L)$ $-\alpha \frac{R_C \parallel R_L}{r_e}$	$-\beta \frac{R_C \parallel R_L}{R_{sig} + (\beta + 1)r_e}$
Common emitter with R_e (Fig. 7.38)	$(\beta + 1)(r_e + R_e)$	$-\frac{g_m R_C}{1 + g_m R_e}$	R_C	$\frac{-g_m (R_C \parallel R_L)}{1 + g_m R_e}$ $-\alpha \frac{R_C \parallel R_L}{r_e + R_e}$	$-\beta \frac{R_C \parallel R_L}{R_{sig} + (\beta + 1)(r_e + R_e)}$
Common base (Fig. 7.40)	r_e	$g_m R_C$	R_C	$g_m (R_C \parallel R_L)$ $\alpha \frac{R_C \parallel R_L}{r_e}$	$\alpha \frac{R_C \parallel R_L}{R_{sig} + r_e}$
Emitter follower (Fig. 7.43)	$(\beta + 1)(r_e + R_L)$	1	r_e	$\frac{R_L}{R_L + r_e}$	$\frac{R_L}{R_L + r_e + R_{sig}/(\beta + 1)}$ $G_{vo} = 1$ $R_{out} = r_e + \frac{R_{sig}}{\beta + 1}$

^a For the interpretation of R_m , A_{vo} , and R_o refer to Fig. 7.34.

^b Setting $\beta = \infty$ ($\alpha = 1$) and replacing r_e with $1/g_m$, R_C with R_D , and R_e with R_s results in the corresponding formulas for MOSFET amplifiers (Table 7.4).

When and How to Include the Output Resistance r_o So far we have been neglecting the output resistance r_o of the BJT. We have done this for two reasons:

1. To keep things simple and focus attention on the significant features of each of the basic configurations, and
2. Because our main interest in this chapter is discrete-circuit design, where the circuit resistances (e.g., R_C , R_D , and R_L) are usually much smaller than r_o .

Nevertheless, in some instances it is relatively easy to include r_o in the analysis. Specifically:

1. In the CS and CE amplifiers, it can be seen that r_o of the transistor appears in parallel with R_D and R_C , respectively, and can be simply included in the corresponding formulas in Tables 7.4 and 7.5 by replacing R_D with $(R_D \parallel r_o)$ and R_C with $(R_C \parallel r_o)$. The effect will be a reduction in the magnitude of gain, of perhaps 5% to 10%.
2. In the source and emitter followers, it can be seen that the transistor r_o appears in parallel with R_L and can be taken into account by replacing R_L in the corresponding formulas with $(R_L \parallel r_o)$. Thus, here too, the effect of taking r_o into account is a small reduction in gain. More significant, however, taking r_o into account reduces the open-circuit voltage gain A_{vo} from unity to

$$A_{vo} = r_o / (r_o + 1/g_m) \quad (3.79)$$

There are configurations in which taking r_o into account complicates the analysis considerably. These are the CE amplifiers with a emitter resistance, and the CB amplifier. Fortunately, for discrete implementation of these configurations, the effect of neglecting r_o is usually small (which can be verified by computer simulation).

Finally, a very important point: In the analysis and design of I_C amplifiers, r_o must always be taken into account. This is because, as will be seen in the next chapter, all the circuit resistances are of the same order of magnitude as r_o ; thus, neglecting r_o can result in completely erroneous results.

Biasing:

As discussed in Section 7.1, an essential step in the design of a transistor amplifier is the establishment of an appropriated cooperating point for the transistor. This is the step known as biasing or bias design. In this section, we study the biasing methods commonly employed in discrete-circuit amplifiers. Bias design aims to establish in the drain (collector) a dc current that is predictable and insensitive to variations in temperature and to the large variations in parameter values between devices of the same type. For instance, discrete BJTs belonging to the same manufacturer's part number can exhibit β values that range, say, from 50 to 150. Nevertheless, the bias design for an amplifier utilizing this particular transistor type may specify that the dc collector current shall always be within, say, $\pm 10\%$ of the nominal value of, say, 1mA.

A second consideration in bias design is locating the dc operating point in the active region of operation of the transistor so as to obtain high voltage gain while allowing for the required output signal swing without the transistor leaving the active region at any time (in order to avoid nonlinear distortion).

Although we shall consider the biasing of BJT amplifiers separately, the resulting circuits are very similar. Also, it will be seen that good bias designs incorporate a feedback mechanism that works to keep the dc bias point as constant as possible.

In order to keep matters simple and thus focus our attention on significant issues, we will neglect the Early effect; that is assume $\lambda=0$ or $V_A = \infty$. This is certainly allowed in initial designs of discrete circuits. Of course, the design can be fine-tuned at a later point with the assistance of a circuit-simulation program such as SPICE.

Biasing in BJT amplifiers circuits :

Before presenting the “good” biasing schemes, we should point out why two obvious arrangements are not good. First, attempting to bias the BJT by fixing the voltage V_{BE} by, for instance, using a voltage divider across the power supply V_{CC}, as shown in Fig.7.51(a), is not a viable approach: The very sharp exponential relationship iC-v_{BE} means that any small and inevitable differences in V_{BE} from the desired value will result in large differences in I_C and in V_{CE}. Second, biasing the BJT by establishing a constant current in the base, as shown in Fig.7.51(b), where $I_B = (V_{CC} - 0.7)/R_B$, is also not a recommended approach. Here the typically large variations in the value of β among units of the same device type will result in correspondingly large variations in I_C and hence in V_{CE}.

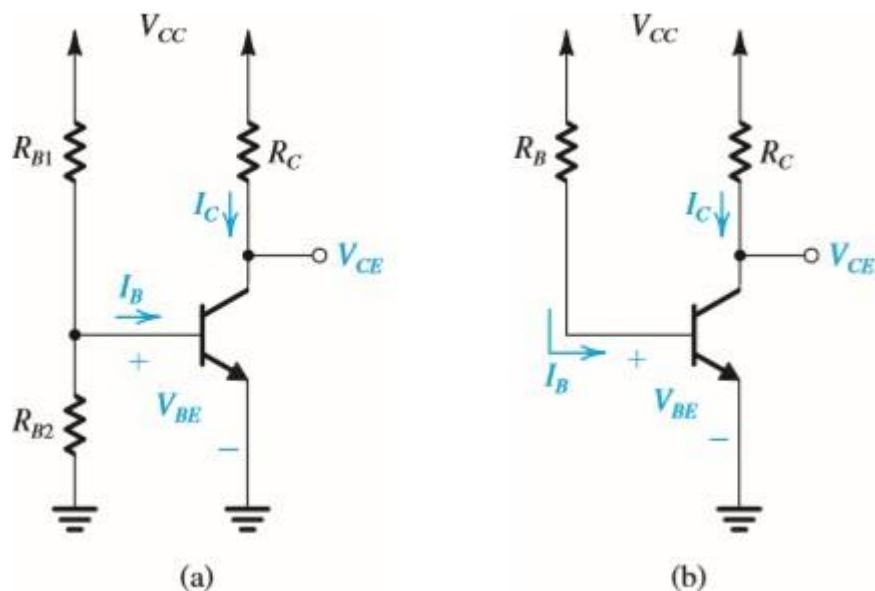


Fig:3.24 Two obvious schemes for biasing the BJT:(a) by fixing V_{BE};(b) by fixing I_B.Both result in wide variations in I_C and hence in V_{CE} and therefore are considered to be “bad.” Neither scheme is recommended.

Fixed bias :

The circuit shown is called as a “**fixed base bias circuit**”, because the transistors base current, I_B remains constant for given values of V_{cc}, and therefore the transistors operating point must also remain **fixed**.

The Classical Discrete-Circuit Bias Arrangement:

Figure3.25(a) shows the arrangement most commonly used for biasing a discrete-circuit transistor amplifier if only a single power supply is available. The technique consists of supplying the base of the transistor with a fraction of the supply voltage V_{CC} through the voltage

divider R_1 , R_2 . In addition, a resistor R_E is connected to the emitter. Here, however, the design must take into account the finite base current. Figure 3.25(b) shows the same circuit with the voltage-divider network replaced by its Thévenin equivalent,

$$V_{BB} = [R_2 / (R_1 + R_2)] \quad (3.80)$$

$$R_B = R_1 R_2 / (R_1 + R_2) \quad (3.81)$$

The current I_E can be determined by writing a Kirchhoff loop equation for the base-emitter-ground loop, labeled L, and substituting $I_B = I_E / (\beta + 1)$:

$$I_E = (V_{BB} - V_{BE}) / [R_E + R_B / (\beta + 1)] \quad (3.82)$$

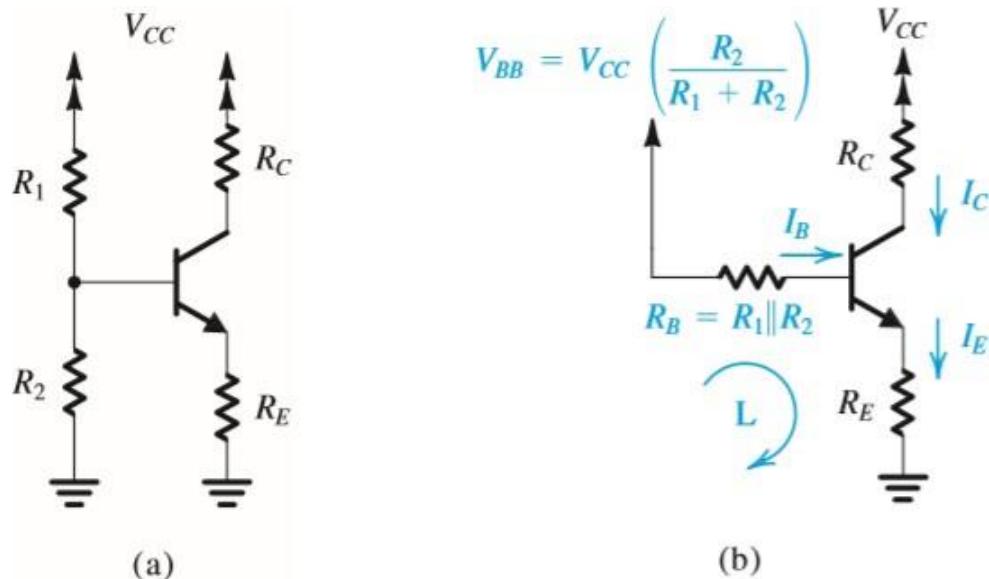


Fig:3.25 Classical biasing for BJTs using a single power supply: (a) circuit; (b) circuit with the voltage divider supplying the base replaced with its Thévenin equivalent.

To make I_E insensitive to temperature and β variation,⁸ we design the circuit to satisfy the following two constraints:

$$V_{BB} \gg V_{BE} \quad (3.83)$$

$$R_E \gg R_B / (\beta + 1) \quad (3.84)$$

Condition (3.83) ensures that small variations in V_{BE} ($\sim 0.7V$) will be swamped by the much larger V_{BB} . There is a limit, however, on how large V_{BB} can be: For a given value of the supply voltage V_{CC} , the higher the value we use for V_{BB} , the lower will be the sum of voltage across

RC and the collector–base junction (V_{CB}). On the other hand, we want the voltage across RC to be large in order to obtain high voltage gain and large signal swing (before transistor cutoff). We also want V_{CB} (or V_{CE}) to be large, to provide a large signal swing (before transistor saturation). Thus, as is the case in any design, we have a set of conflicting requirements, and the solution must be a trade-off. As a rule of thumb, one designs for V_{BB} about $1/3V_{CC}$, V_{CB} (or V_{CE}) about $1/3V_{CC}$, and $I_C R_C$ about $1/3V_{CC}$.

Condition (3.84) makes I_E insensitive to variation in β and could be satisfied by selecting R_B small. This in turn is achieved by using low values for R_1 and R_2 . Lower values for R_1 and R_2 , however, will mean a higher current drain from the power supply, and will result in a lowering of the input resistance of the amplifier (if the input signal is coupled to the base),⁹ which is the trade-off involved in this part of the design. It should be noted that condition (3.84) means that we want to make the base voltage independent of the value of β and determined solely by the voltage divider. This will obviously be satisfied if the current in the divider is made much larger than the base current. Typically one selects R_1 and R_2 such that their current is in the range of I_E to $0.1I_E$.

Further insight regarding the mechanism by which the bias arrangement of Fig.3.25(a) stabilizes the dc emitter (and hence collector)current is obtained by considering the feedback action provided by R_E . Consider that for some reason the emitter current increases. The voltage drop across R_E , and hence V_E , will increase correspondingly. Now, if the base voltage is determined primarily by the voltage divider R_1, R_2 , which is the case if R_B is small, it will remain constant, and the increase in V_E will result in a corresponding decrease in V_{BE} . This in turn reduces the collector (and emitter) current, a change opposite to that originally assumed. Thus R_E provides a negative feedback action that stabilizes the bias current. This should remind the reader of the resistance R_E that we included in the emitter lead of the CE amplifier in Section 7.3.4. Also, the feedback action of R_E in the circuit of Fig.3.25(a) is similar to the feedback action of R_S in the circuit of Fig.3.25(c).

A Two-Power-Supply Version of the Classical Bias Arrangement:

A somewhat simpler bias arrangement is possible if two power supplies are available ,as shown in Fig.3.27.

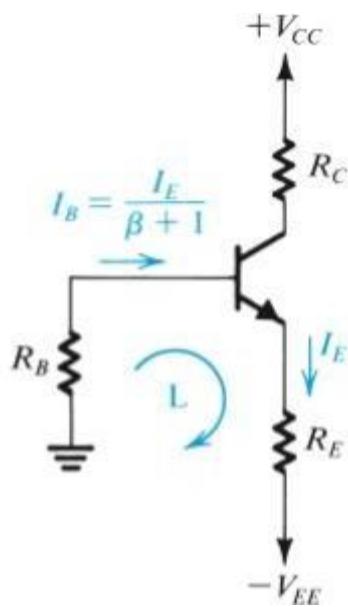


Fig:3.27 Biasing the BJT using two power supplies. Resistor R_B is needed only if the signal is to be capacitively coupled to the base. Otherwise, the base can be connected directly to ground, or to a grounded signal source, resulting in almost total β -independence of the bias current.

Writing a loop equation for the loop labeled L gives

$$I_E = (V_{EE} - V_{BE}) / [R_E + R_B / (\beta + 1)] \quad \text{-----(3.85)}$$

This equation is identical to Eq.(3.82) except for V_{EE} replacing V_{BB} . Thus the two constraints of Eqs.(3.83) and (3.84) apply here as well. Note that if the transistor is to be used with the base grounded (i.e., in the common-base configuration), then R_B can be eliminated altogether. On the other hand, if the input signal is to be coupled to the base, then R_B is needed. We shall study complete circuits of the various BJT amplifier configurations in Section.

Self bias:

A better method of biasing, known as *self-bias* is obtained by inserting the bias resistor directly between the base and collector, as shown in fig:3.28

Biasing Using a Collector-to-Base Feedback Resistor:

In the BJT case, Figure 3.28(a) shows such a simple but effective biasing arrangement that is suitable for common-emitter amplifiers. The circuit employs a resistor R_B connected between the collector and the base. Resistor R_B provides negative feedback, which helps to stabilize the bias point of the BJT.

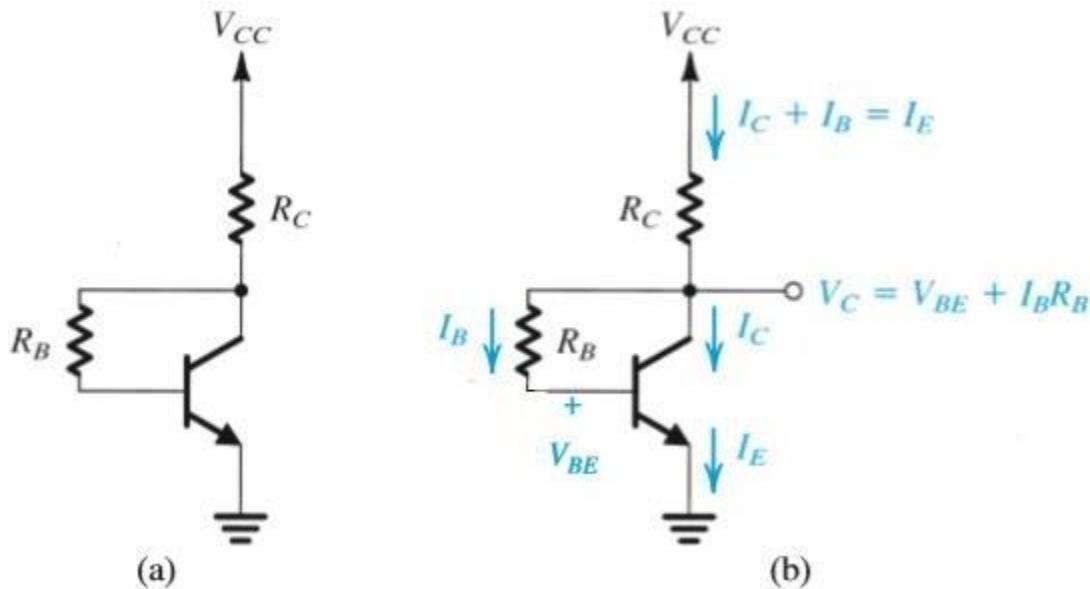


Fig:3.28 (a) A common-emitter transistor amplifier biased by a feedback resistor R_B .

(b) Analysis of the circuit in (a).

Analysis of the circuit is shown in Fig.3.28(b), from which we can write

$$\begin{aligned} V_{CC} &= I_E R_C + I_B R_B + V_{BE} \\ &= I_E R_C + (I_E / (\beta + 1)) R_B + V_{BE} \end{aligned}$$

Thus the emitter bias current is given by

$$I_E = (V_{CC} - V_{BE}) / [R_C + (R_B / (\beta + 1))] \quad \dots \dots \quad (3.86)$$

It is interesting to note that this equation is identical to Eq.(7.143), which governs the operation of the traditional bias circuit, except that V_{CC} replaces V_{BB} and R_C replaces R_E . It follows that to

obtain a value of I_E that is insensitive to variation of β , we select $R_B/(\beta+1) R_C$. Note, however, that the value of R_B determines the allowable negative signal swing at the collector since

$$V_{CB} = I_B R_B = I_E (R_B / (\beta + 1)) \quad \dots \dots \quad (3.87)$$

Biasing using a constant current source:

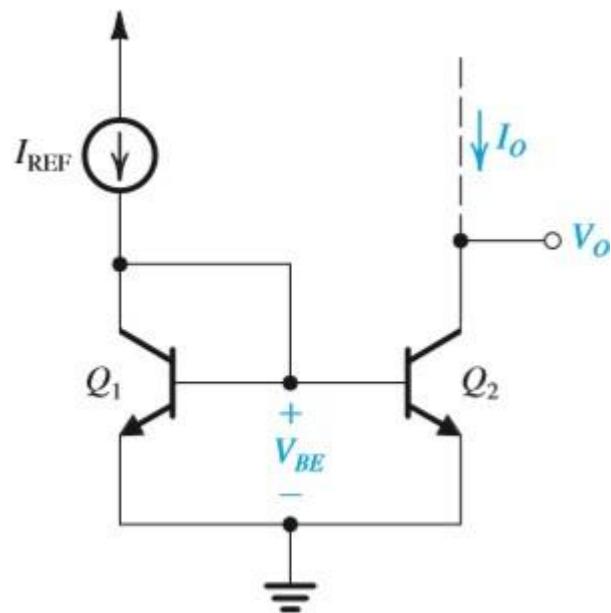


Fig:3.29 Basic BJT current source

The basic BJT current mirror is shown in Fig.3.29. However, there are two important differences: First, the nonzero base current of the BJT (or, equivalently, the finite β) causes an error in the current transfer ratio of the bipolar mirror. Second, the current transfer ratio is determined by the relative areas of the emitter–base junctions of Q1 and Q2.

Let us first consider the case of β sufficiently high that we can neglect the base currents. The reference current I_{REF} is passed through the diode-connected transistor Q1 and thus establishes a corresponding voltage V_{BE} , which in turn is applied between base and emitter of Q2. Now, if Q2 is matched to Q1 or, more specifically, if the EBJ area of Q2 is the same as that of Q1, and thus Q2 has the same scale current I_S as Q1, then the collector current of Q2 will be equal to that of Q1; that is,

$$I_O = I_{REF} \quad \dots \dots \quad (3.86)$$

For this to happen, however, Q2 must be operating in the active mode, which in turn is achieved as long as the collector voltage V_O is 0.3V or so higher than that of the emitter. To obtain a current transfer ratio other than unity, say m , we simply arrange that the area of the EBJ of Q2 is m times that of Q1. In this case,

$$I_O = m I_{REF} \quad \text{--- (3.87)}$$

In general, the current transfer ratio is given by

$$I_O / I_{REF} = I_{S2} / I_{S1} = (\text{Area of EBJ of Q2}) / (\text{Area of EBJ of Q1}) \quad \text{--- (3.88)}$$

Alternatively, if the area ratio m is an integer, one can think of Q2 as equivalent to m transistors, each matched to Q1 and connected in parallel.

Next we consider the effect of finite transistor β on the current transfer ratio. The analysis for the case in which the current transfer ratio is nominally unity—that is, for the case in which Q2 is matched to Q1—is illustrated in Fig. 8.8. The key point here is that since Q1 and Q2 are matched and have the same VBE, their collector currents will be equal. The rest of the analysis is straightforward. A node equation at the collector of Q1 yields

$$I_{REF} = I_C + 2I_C/\beta = I_C(1 + 2/\beta) \quad \text{--- (3.89)}$$

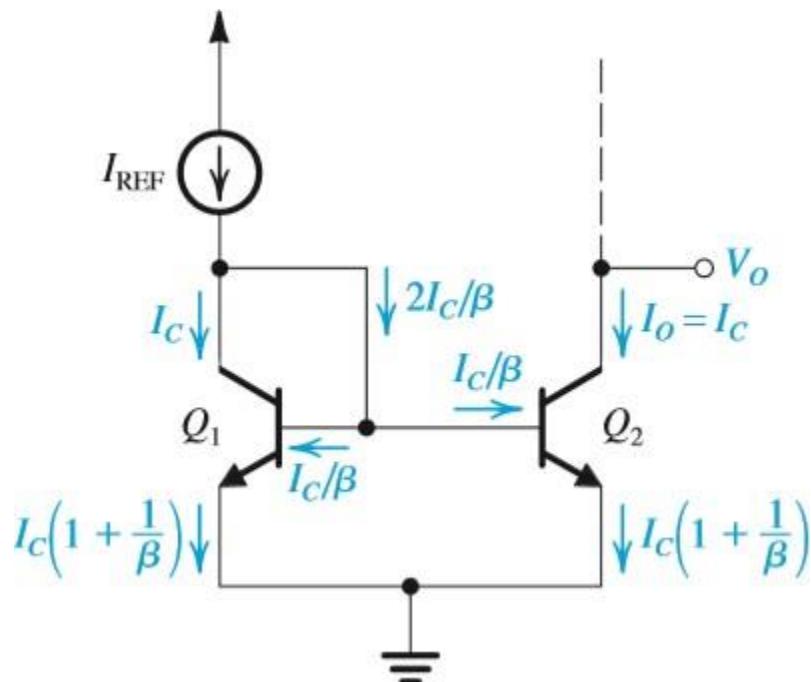


Fig:3.30 Analysis of the constant current source taking into account the finite β of the BJTs.

Finally, since $I_O = I_C$, the current transfer ratio can be found as

$$I_O / I_{REF} = I_C / I_C(1 + 2\beta) = 1 / (1 + 2\beta) \quad (3.90)$$

Note that as β approaches ∞ , I_O/I_{REF} approaches the nominal value of unity. For typical values of β , however, the error in the current transfer ratio can be significant. For instance, $\beta=100$ results in a 2% error in the current transfer ratio. Furthermore, the error due to the finite β increases as the nominal current transfer ratio is increased. The reader is encouraged to show that for a mirror with a nominal current transfer ratio m —that is, one in which $I_{S2}=mI_{S1}$ —the actual current transfer ratio is given by

$$I_O / I_{REF} = m / (1 + ((m+1)/\beta)) \quad (3.91)$$

In common with the MOS current mirror, the BJT mirror has a finite output resistance R_o

$$R_o \equiv \Delta V_O / \Delta I_O = r_{o2} = V_{A2} / I_O \quad (3.92)$$

where V_{A2} and r_{o2} are the Early voltage and the output resistance, respectively, of Q2. Thus, even if we neglect the error due to finite β , the output current I_O will be at its nominal value only when Q2 has the same V_{CE} as Q1, namely, at $V_O = V_{BE}$. As V_O is increased, I_O will correspondingly increase. Taking both the finite β and the finite R_o into account, we can express the output current of a BJT mirror with a nominal current transfer ratio m as

$$I_O = I_{REF} [m / 1 + (m + 1/\beta)] [1 + (V_O - V_{BE}) / V_{A2}] \quad (3.93)$$

where we note that the error term due to the Early effect is expressed in a form that shows that it reduces to zero for $V_O = V_{BE}$.

A Simple Current Source:

The basic BJT current mirror can be used to implement a simple current source, as shown in Fig.8.9. Here the reference current is

$$I_{REF} = (V_{CC} - V_{BE}) / R \quad (3.94)$$

where V_{BE} is the base-emitter voltage corresponding to the desired value of I_{REF} . The output current I_O is given by

$$I_O = (I_{REF} / 1 + (2/\beta)) (1 + (V_O - V_{BE}) / V_A) \quad (3.95)$$

The output resistance of this current source is r_o of Q2,

$$R_o = r_{o2} \approx (V_A / I_O) \approx (V_A / I_{REF}) \quad (3.96)$$

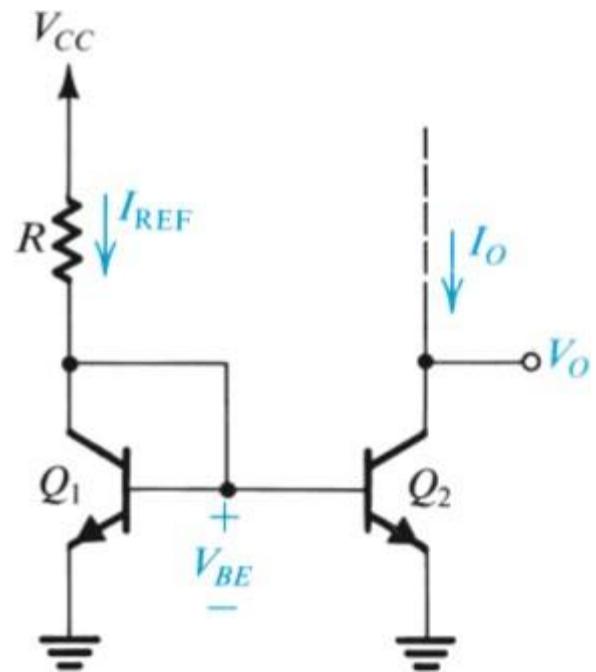


Fig:3.31 A simple BJT current source.

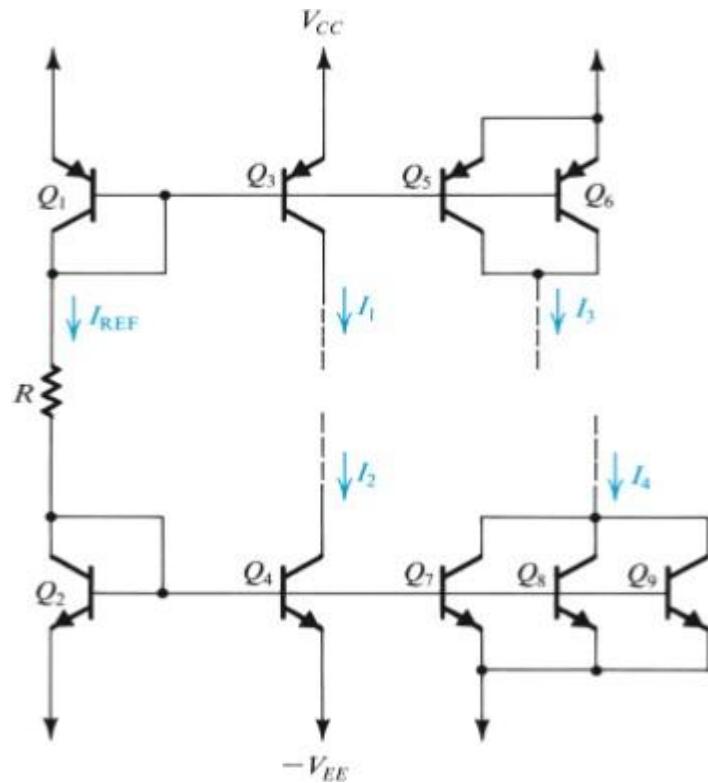


Fig:3.32 Generation of a number of constant currents of various magnitudes.

Current Steering:

To generate bias currents for different amplifier stages in an IC, the current-steering approach described for MOS circuits can be applied in the bipolar case. As an example, consider the circuit shown in Fig.3.32. The dc reference current I_{REF} is generated in the branch that consists of the diode-connected transistor Q1, resistor R, and the diode-connected transistor Q2:

$$I_{REF} = (V_{CC} + V_{EE} - V_{EB1} - V_{BE2}) / R \quad (3.97)$$

Now ,for simplicity ,assume that all the transistors have high β and thus that the base currents are negligibly small. We will also neglect the Early effect. The diode-connected transistor Q1 forms a current mirror with Q3; thus Q3 will supply a constant current I_1 equal to I_{REF} . Transistor Q3 can supply this current to any load as long as the voltage that develops at the collector does not exceed ($V_{CC} - 0.3V$); otherwise Q3 would enter the saturation region.

To generate a dc current twice the value of I_{REF} ,two transistors ,Q5 and Q6,each of which Is matched to Q1,are connected in parallel ,and the combination forms a mirror with Q1.Thus

$$I_3 = 2I_{REF}.$$

Note that the parallel combination of Q5 and Q6 is equivalent to a transistor with an EBJ area double that of Q1, which is precisely what is done when this circuit is fabricated in IC form.

Transistor Q4 forms a mirror with Q2; thus Q4 provides a constant current I_2 equal to I_{REF} . Note that while Q3 sources its current to parts of the circuit whose voltage should not exceed ($V_{CC} - 0.3V$),Q4 sinks its current from parts of the circuit whose voltage should not decrease below ($-V_{EE} + 0.3V$). Finally, to generate a current three times I_{REF} , three transistors, Q7, Q8, and Q9, each of which is matched to Q2, are connected in parallel, and the combination is placed in a mirror configuration with Q2. Again, in an IC implementation, Q7, Q8, and Q9 would be replaced with a transistor having a junction area three times that of Q2.

A Bipolar Mirror with Base-Current Compensation:

Figure 3.33 shows a bipolar current mirror with a current transfer ratio that is much less dependent on β than that of the simple current mirror. The reduced dependence on β is achieved by including transistor Q3, the emitter of which supplies the base currents of Q1 and Q2. The sum of the base currents is then divided by $(\beta_3 + 1)$, resulting in a much smaller error current that has to be supplied by I_{REF} . Detailed analysis is shown on the circuit diagram; it is based on the assumption that Q1 and Q2 are matched and thus have equal collector currents, I_C . A node equation at the node labeled x gives

$$I_{REF} = I_C [1 + (2/\beta(\beta+1))]$$

Since,

$$I_O = I_C$$

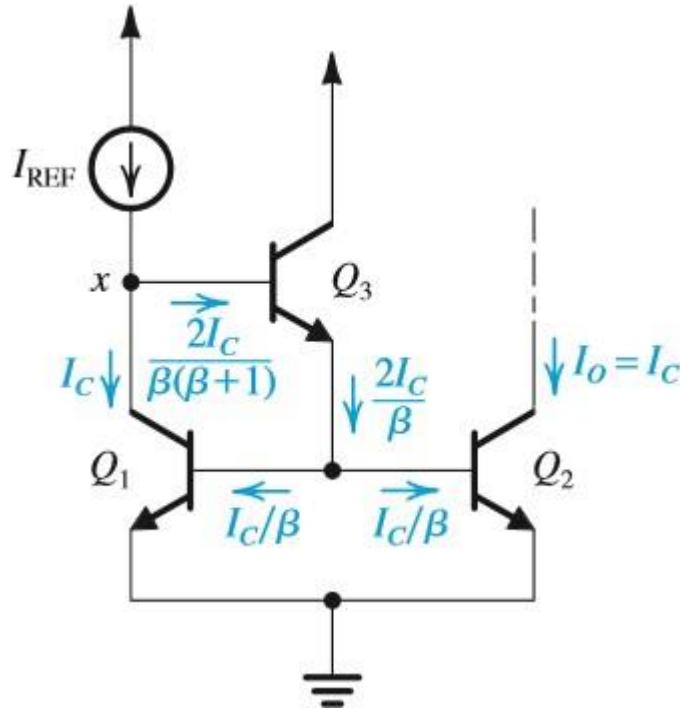


Fig:3.33 A current mirror with base-current compensation.

the current transfer ratio of the mirror will be

$$I_O / I_{REF} = 1 / (1 + 2/(\beta^2 + \beta))$$

$$\approx 1 / (1 + 2/\beta^2) \text{ ----- (3.98)}$$

which means that the error due to finite β has been reduced from $2/\beta$ in the simple mirror to $2/\beta^2$, a tremendous improvement. Unfortunately, however, the output resistance remains approximately equal to that of the simple mirror, namely r_o . Finally, note that if a reference current I_{REF} is not available, we simply connect node x to the power supply, V_{CC} , through a resistance R . The result is a reference current given by

$$I_{REF} = (V_{CC} - V_{BE1} - V_{BE3}) / R \text{ ----- (3.99)}$$

A Common-Emitter Amplifier:

A Common-Emitter Amplifier The common-emitter (CE) amplifier is the most widely used of all BJT amplifier configurations. Figure 7.56(a) shows a CE amplifier utilizing the classical biasing arrangement of Fig.7.48(c), the design of which was considered in Section7.4. The CE circuit in Fig.7.54(a) is the BJT counterpart of the CS amplifier of Fig.7.55(a). It utilizes coupling capacitors CC1 and CC2 and bypass capacitor CE. Here we assume that these capacitors, while blocking dc, behave as perfect short circuits at all signal frequencies of interest. To determine the characteristic parameters of the CE amplifier, we replace the BJT with its hybrid- π model, replace VCC with a short circuit to ground, and replace the coupling and bypass capacitor with short circuits. The resulting small-signal equivalent circuit of the CE amplifier is shown in Fig.7.56(b). The analysis is straightforward and is given in the

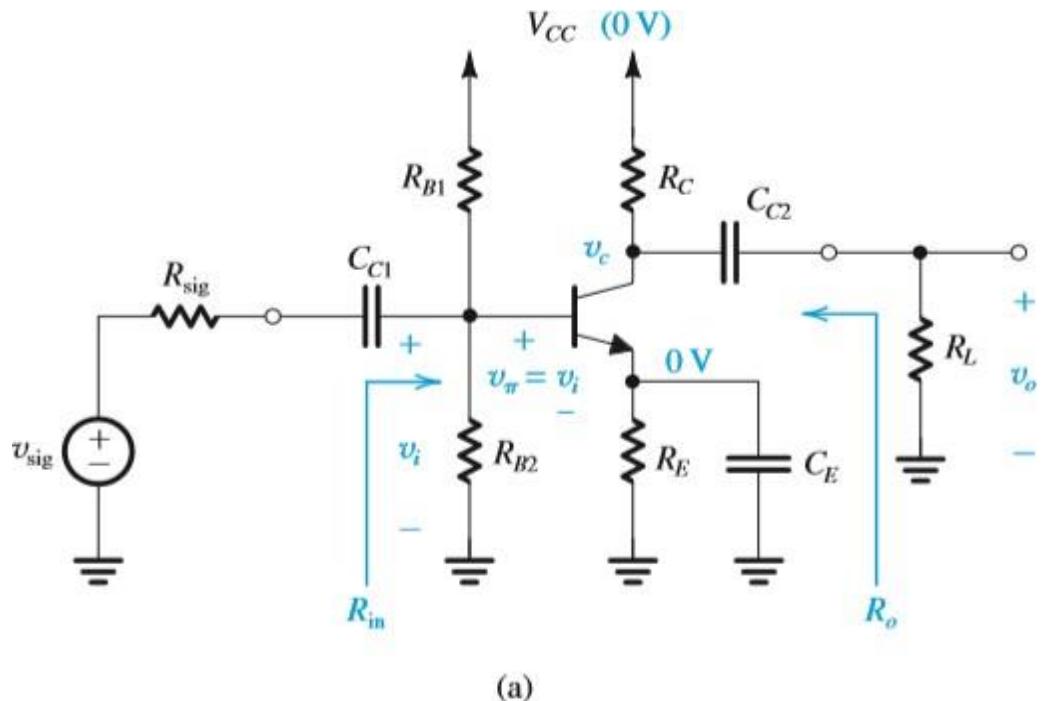
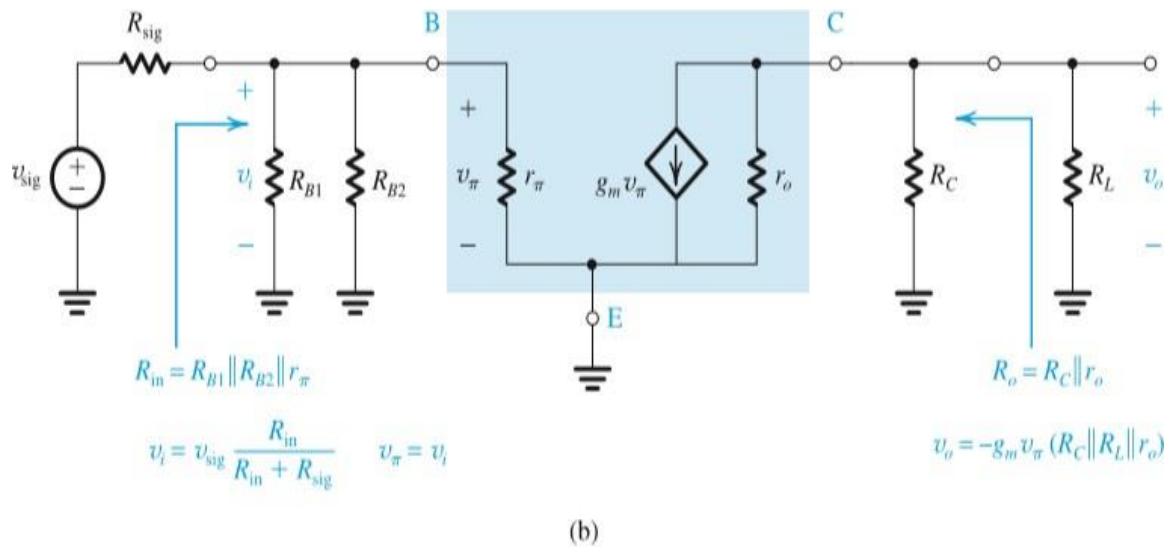


Fig:3.34 (a) A common-emitter amplifier using the classical biasing arrangement of Fig. 7.52(a).



(b) Equivalent circuit and analysis.

figure, thus

$$R_{in} = R_{B1} || R_{B2} || r_\pi \quad \dots \dots \dots \quad (3.100)$$

Which indicates that to keep R_{in} relatively high, R_{B1} and R_{B2} should be selected large (typically in the range of tens or hundreds of kilohms). This requirement conflicts with the need to keep R_{B1} and R_{B2} low so as to minimize the dependence of the dc current I_C on the transistor β . We discussed this design trade-off in Section 7.4. The voltage gain G_v is given by

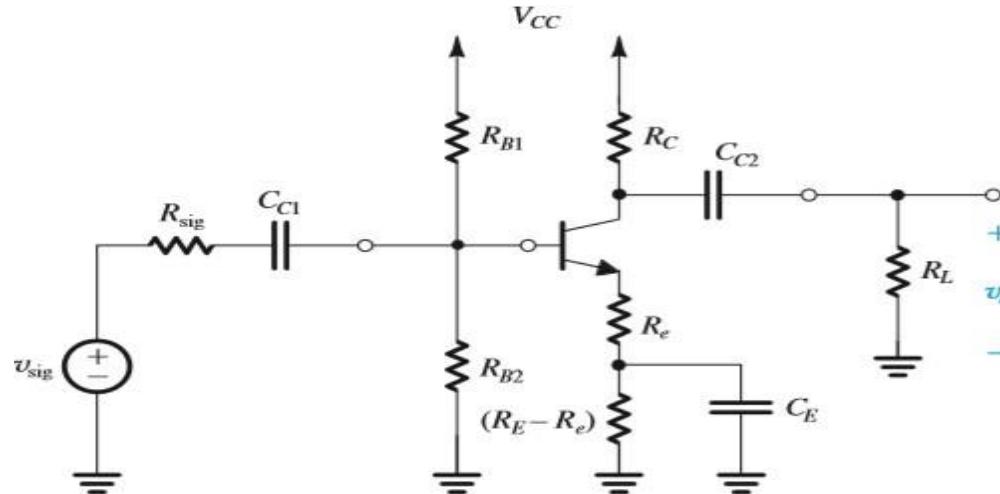
$$G_v = -(R_{in} / (R_{in} + R_{sig})) (g_m (R_C || R_L || r_o)) \quad \dots \dots \dots \quad (3.101)$$

Note that we have taken r_o into account because it is easy to do so. However, as already mentioned, the effect of this parameter on discrete-circuit amplifier performance is usually small.

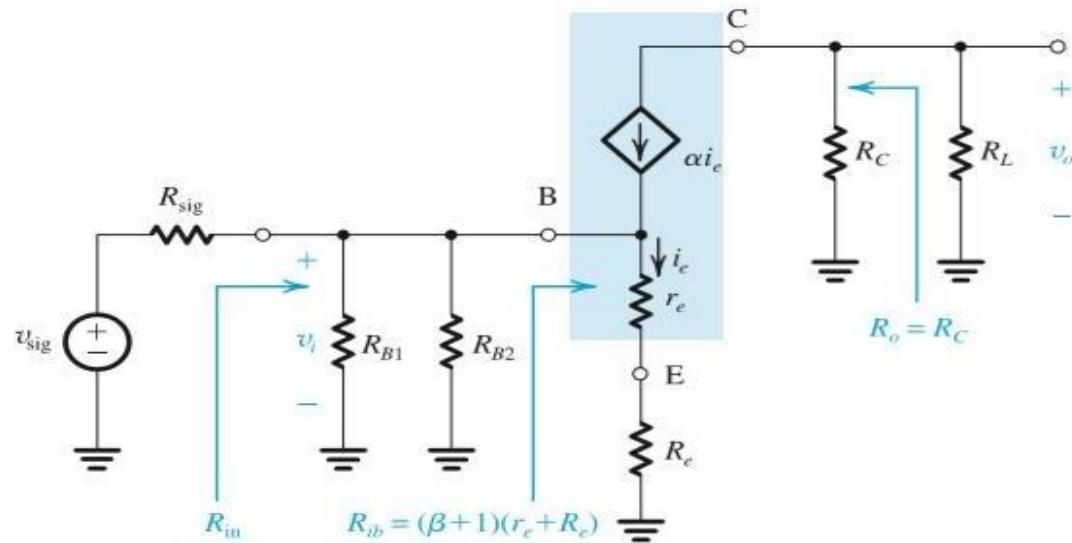
A Common-Emitter Amplifier with an Emitter Resistance R_E :

It is beneficial to include a small resistance in the transistor emitter lead. This can be implemented in the circuit of Fig.3.35(a) by splitting the emitter bias resistance R_E into two components: an unbypassed resistance R_E , and a resistance $(R_E - R_E)$ across which the bypass capacitor C_E is connected. The resulting circuit is shown in Fig.3.35(a) and its small-signal model is shown in Fig.3.35(b). In the latter we utilize the T model of the BJT because it results in much simpler analysis (recall that this is always the case when a resistance is connected in series with the emitter). Also note that we have not included r_o , for doing so would complicate the

analysis significantly. This burden would not be justified given that r_o has little effect on the performance of discrete-circuit amplifiers.



(a)



(b)

Fig:3.35 (a) A common-emitter amplifier with an unbiased emitter resistance R_e . (b) The amplifier small-signal model and analysis.

$$G_v = - \frac{R_o}{R_{in} + R_{sig}} \times \alpha \frac{(R_C \parallel R_L)}{r_e + R_e}$$

Analysis of the circuit in Fig.3.35(b) is straightforward and is shown in the figure. Thus,

$$\begin{aligned} R_{in} &= R_{B1} \parallel R_{B2} \parallel (\beta + 1)(r_e + R_e) \\ &= R_{B1} \parallel R_{B2} \parallel [r_\pi + (\beta + 1)R_e] \end{aligned} \quad (3.102)$$

from which we note that including R_e increases R_{in} because it increases the input resistance looking into the base by adding a component $(\beta + 1)R_e$ to r_π . The overall voltage gain G_v is

$$\begin{aligned} G_v &= -[R_{in} / (R_{in} + R_{sig})] \times \alpha \left(\frac{\text{Total resistance in collector}}{\text{Total resistance in emitter}} \right) \\ &= -\alpha \left[\frac{R_{in}}{(R_{in} + R_{sig})} \right] \left[\frac{(R_C \parallel R_L)}{(r_e + R_e)} \right] \end{aligned} \quad (3.103)$$

Breakdown mechanisms in BJTs:

The breakdown mechanisms of BJTs are similar to that of p-n junctions. Since the base-collector junction is reversed biased, it is this junction where breakdown typically occurs. Just like for a p-n junction the breakdown mechanism can be due to either avalanche multiplication as well as tunneling. However, the collector doping in power devices tends to be low-doped either to ensure a large enough breakdown voltage – also called blocking voltage – or to provide a high Early voltage. The collector doping in microwave BJTs is typically higher than that of power devices, yet based on the trade-off between having a short transit time through the base-collector depletion region and having a low base-collector capacitance. As a result, one finds that the collector doping density rarely exceeds 10^{18} cm^{-3} and tunneling does not occur.

Instead, breakdown is dominated by avalanche multiplication. The large electric field in the base-collector depletion region causes carrier multiplication due to impact ionization. Just like in a p-n diode, this breakdown is not destructive. However, the high voltage and rapidly increasing current does cause large heat dissipation in the device, which can cause permanent damage to the semiconductor or the contacts.

The breakdown voltage of a BJT also depends on the chosen circuit configuration: In a common base mode (i.e. operation where the base is grounded and forms the common electrode between the emitter-base input and collector-base output of the device) the breakdown resembles that of a p-n diode. In a common emitter mode (i.e. operation where the emitter is grounded and forms the common electrode between the base-emitter input and the collector-emitter output of the device) the transistor action further influences the $I-V$ characteristics and breakdown voltage. Base width modulation was described in section 5.4.1 to result in an increase in the collector current with increased collector-emitter voltage. In the extreme case of punchthrough where the base is completely depleted, an even larger increase is observed be it nowhere as abrupt as in the case of avalanche breakdown. Avalanche breakdown of the base-collector junction is further influenced by transistor action in common-emitter mode of operation, since the holes generated by impact ionization are pulled back into the base region which results in an additional base current. This additional base current causes an even larger additional flow of electrons through the base and into the

collector due to the current gain of the BJT. This larger flow of electrons in the base-collector junction causes an even larger generation of electron-hole pairs.

To further analyze this effect quantitatively we first write the total collector current, I_C , in response to an applied base current, I_B :

$$I_C = \mathcal{A}(I_B + (M-1)I_C)$$

Where the term $(M - 1) I_C$ was added to the base current to include the holes generated due to impact ionization. This equation can be rearranged to yield:

$$I_C = \frac{\beta I_B}{1 + \beta - \beta M}$$

The collector current will therefore approach infinity as the denominator approaches zero. From this equation and combining with equation 4.5.6 one finds that the common emitter breakdown voltage equals:

$$VB_{CEO} = \frac{VB_{CBO}}{(\beta + 1)^{1/m}}$$

The common emitter breakdown voltage as characterized by the open base breakdown voltage, VB_{CEO} , is therefore significantly less than the open emitter breakdown voltage, VB_{CBO} .

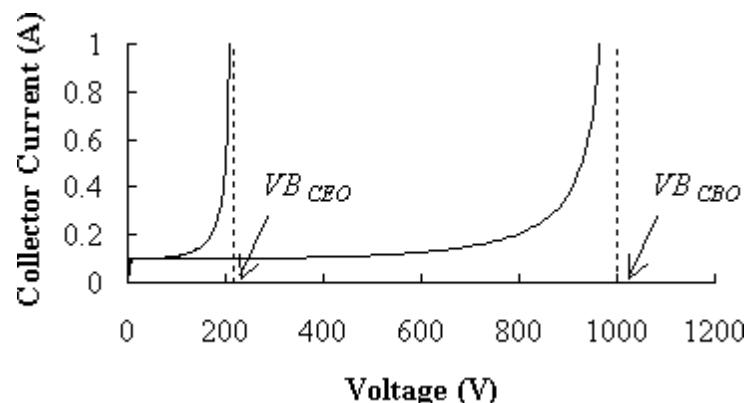


Fig:3.36 Comparison of BJT breakdown in common emitter mode (left curve) versus breakdown in common base mode (right curve) for a BJT with $VB_{CBO} = 1000V$ and $b = 100$.

Temperature dependent effects in BJT:

The temperature dependence of bipolar transistors depends on a multitude of parameters affecting the bipolar transistor characteristics in different ways.

First we will discuss the temperature dependence of the current gain. Since the current gain depends on both the emitter efficiency and base transport factor, we will discuss these separately.

The emitter efficiency depends on the ratio of the carrier density, diffusion constant and width of the emitter and base. As a result, it is not expected to be very temperature dependent. The carrier densities are linked to the doping densities. Barring incomplete ionization, which can be very temperature dependent, the carrier densities are independent of temperature as long as the intrinsic carrier density does not exceed the doping density in either region. The width is very unlikely to be temperature dependent and therefore also the ratio of the emitter and base width. The ratio of the mobility is expected to be somewhat temperature dependent due to the different temperature dependence of the mobility in n-type and p-type material.

The base transport is more likely to be temperature dependent since it depends on the product of the diffusion constant and carrier lifetime. The diffusion constant in turn equals the product of the thermal voltage and the minority carrier mobility in the base. The recombination lifetime depends on the thermal velocity. The result is therefore moderately dependent on temperature. Typically the base transport reduces with temperature, primarily because the mobility and recombination lifetime are reduced with increasing temperature. Occasionally the transport factor initially increases with temperature, but then reduces again

2 Marks Questions:

1. What is a transistor? What are the types?

Transistor consists of two junctions formed by sandwiching either p type or n type semiconductor between a pair of opposite types.

Two types - 1. NPN transistor

2. PNP transistor

2. Define BJT?

A bipolar junction transistor is a three terminal semiconductor device in which the operation depends on the interaction of both majority and minority carriers and hence the name bipolar.

3. Why is the transistor called a current controlled device?

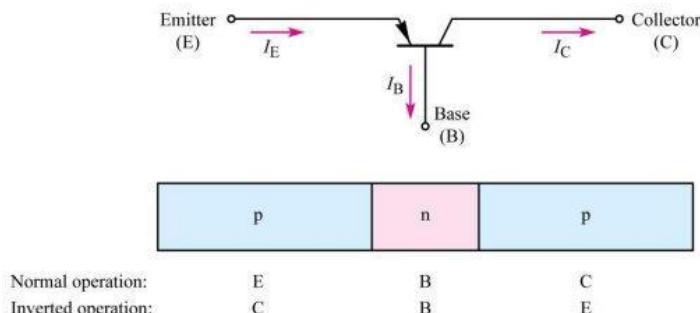
A transistor is called a current controlled device. This is because; collector current is controlled by base current. The changes in collector current are proportional to the corresponding changes in base current.

4. Explain about the characteristics of a transistor?

In common emitter configuration, input characteristics is the plot obtained by tracing the variation of input current I_B with the input voltage V_{BE} . Similarly, the variation of output current I_C with the Collector to emitter voltage V_{CE} is known as output characteristics.

5. Draw the Symbol of BJT?

PNP transistor:



6. Why we use h-parameters to describe a transistor?

Any linear circuit can be analyzed by four parameters (input resistance, reverse voltage gain, forward current gain and output admittance) of mixed dimensions. Since the dimensions of the parameters have mixed units they are referred as h-parameters. The h-parameters are determined by both open circuit and short circuit terminations.

7. Differentiate FET and BJT (any two)?

FET	BJT
1. Unipolar device (that is current conduction by only one type of either electron or hole).	Bipolar device (current conduction by both electron and hole).
2. High input impedance due to reverse bias.	2. Low input impedance due to forward bias.
3. Gain is characterized by trans conductance	Gain is characterized by voltage gain.
4. Low noise level	High noise level.

8. In a bipolar transistor which region is wider and which region is thinner? Why?

The middle region of bipolar junction transistor is called as the base of the transistor. Input signal of small amplitude is applied to the base. This region is thin and lightly doped. The magnified output signal is obtained at the collector. This region is thick and heavily doped.

9. Define the delay time and rise time in the switching characteristics of transistor?

In the transistor switching characteristics the delay time is the time that elapses the application of the input pulse and current to rise to 10 percent of its maximum value. The time required for IC to reach 90% of its maximum level from 10% level is called the rise time.

10. When a transistor is used as a switch, in which region of output characteristics it is operated?

When a transistor is used as a switch it is operated alternately in the cut off region and saturation region of the output characteristics.

11. When does a transistor act as a switch?

A transistor should be operated in saturation and cut off regions to use it as a switch. While operating in saturation region, transistor carry heavy current hence considered as ON state. In cut-off, it carries no current and it is equivalent to open switch.

12. Why do the output characteristics of a CB transistor have a slight upward slope?

The emitter and collector are forward biased under the saturation region. Hence a small change in collector voltage causes a significant change in collector current. Therefore the slight upward slope is found in output characteristics.

13. Define current gain in CE configuration?

The current gain (β) of common emitter configuration is defined as the ratio of change in collector current to change in base current when collector emitter voltage is kept constant.

I_C / I_B β is also referred as h_{fe} .

14. Define FET?

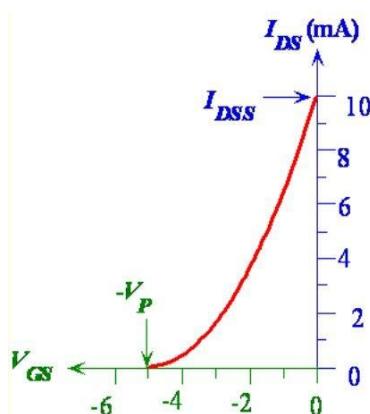
A field effect transistor (FET) is a three terminal semiconductor device which can be used as an amplifier or switch. The three terminals are Drain (D), Source (S), and Gate (G).

15. Define channel?

It is a bar like structure which determines the type of FET. Different types of N channel are FET and P channel FET.

16. Draw the transfer characteristic for n-channel depletion type MOSFET?

Transfer characteristic:



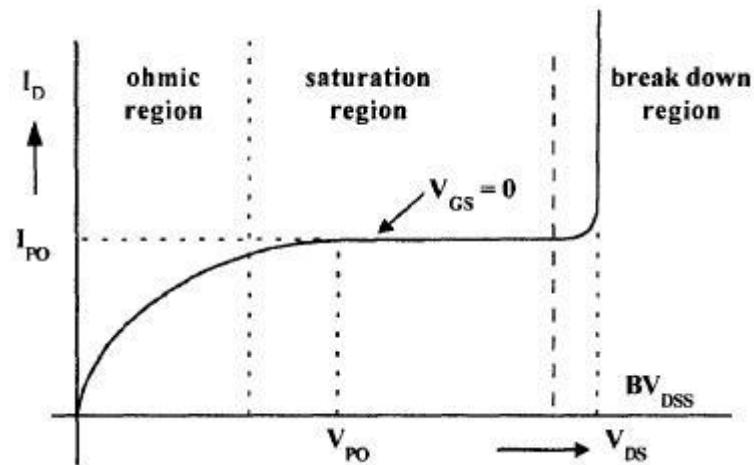
17. What do you understand by pinch off voltage and out of voltage?

As the reverse bias is further increased, the effective width of the channel decreases, the depletion region or the space charge region widens, reaching further into the channel and restricting the passage of electrons from the source to drain. Finally at a certain gate to source voltage $V_{GS} = V_P$.

18. Why FET is called as “voltage operated device”?

In FET the output current, I_D is controlled by the voltage applied between gate and source (V_{GS}). Therefore FET is said to be voltage controlled device.

19. Sketch the ohmic region in drain characteristics of JFET? Drain characteristics:



20. Define Amplification factor in JFET?

It is defined as the ratio of change in drain-source voltage V_{DS} to the change in gate-source voltage V_{GS} at constant drain current I_D . It is also called mutual conductance

$$\mu = \frac{\Delta V_{DS}}{\Delta V_{GS}} \left| \begin{array}{l} I_D \\ \text{constant} \end{array} \right.$$

Amplification factor = Drain resistance X Trans conductance.

$$\mu = R_o \times g_m$$

21. Which MOSFET is called as Normally ON MOSFET and NORMALLY OFF MOSFET? Why?

This is just one type of MOSFET, called '**normally -off**' because it is only the application of a positive gate voltage above the critical voltage which allows it to pass current between source and drain..

Another type of MOSFET is the '**normally-on**', which has a conductive channel of less heavily doped *n*-type material between the source and drain electrodes.

22. What are the advantages of FET over BJT?

In FET input resistance is high compared to BJT
Construction is smaller than BJT

Less sensitive to changes in applied voltage

Thermal stability is more and Thermal noise is much lower Thermal runaway does not exist in JFET

23. What are the important features of FET?

The parameters of FET are temperature dependent. In FET, as temperature increases drain resistance also increases, reducing the drain current. Thus unlike BJT, thermal runaway does not occur with FET. Thus we can say FET is more temperature stable.

FET has very high input impedance. Hence FET is preferred in amplifiers. It is less noisy. Requires less space.

It exhibits no offset voltage at zero drain current.

24 Explain the biasing of JFET?

Input is always reverse biased and output is forward biased. (Note: In transistor input is forward biased and output is reverse biased).

25. Define Drain resistance.

It is the ratio of change in Drain – source voltage (ΔV_{DS}) to the change in Drain current (ΔI_D) at constant gate source voltage (V_{GS}).

.

Essay Questions:

- 1. What is rectifier?**
Draw the diagram & explain the concept.

- 2. Define regulation and efficiency of a rectifie?.**
Explain the above concept.

- 3. List out different filters used in association with rectifiers?**
Explain the different filters with rectifiers.

- 4. With the help of a neat circuit diagram, input and output waveforms, describe the operation of Half-wave rectifier?**
Draw the block diagram
Explain the operation of Half-wave rectifier

- 5. Derive the expressions for ripple factor and maximum efficiency of HWR?**
Draw the circuit diagrams
Express the equations for ripple factor.

- 6. Design a Full-wave center-tap rectifier with capacitor filter and explain its operation?.**
Draw the circuit diagram
Explain the operation for filters.

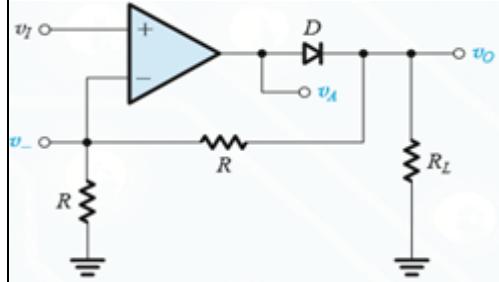
MCQ's:

- 1. Which of the following isn't a type of rectifier? (d)**
- a) Precision Half-wave Rectifier
 - b) Bridge Rectifier
 - c) Peak Rectifier
 - d) None of the mentioned
- 2. For a half wave or full wave rectifier the Peak Inverse Voltage of the rectifier is always (b)**
- a) Greater than the input voltage
 - b) Smaller than the input voltage
 - c) Equal to the input voltage
 - d) Greater than the input voltage for full wave rectifier and smaller for the half wave rectifier
- 3. Bridge rectifier is an alternative for (a)**
- a) Full wave rectifier
 - b) Peak rectifier
 - c) Half wave rectifier
 - d) None of the mentioned
- 4. Which of the following is true for a bridge rectifier? (d)**
- a) The peak inverse voltage or PIV for the bridge rectifier is lower when compared to an identical center tapped rectifier
 - b) The output voltage for the center tapped rectifier is lower than the identical bridge rectifier
 - c) A transistor of higher number of coil is required for center tapped rectifier than the identical bridge rectifier
 - d) All of the mentioned
- 5. The diode rectifier works well enough if the supply voltage is much than greater than 0.7V. For smaller voltage (of few hundreds of millivolt) input which of the following can be used? (a)**
- a) Superdiode
 - b) Peak rectifier
 - c) Precision rectifier
 - d) None of the mentioned
- 6. Consider a peak rectifier fed by a 60-Hz sinusoid having a peak value $V_p = 100$ V. Let the load resistance $R = 10 \text{ k}\Omega$. Calculate the fraction of the cycle during which the diode is conducting (c)**

- a) 1.06 %
- b) 2.12 %
- c) 3.18%
- d) 4.24%

7.The op amp in the precision rectifier circuit is ideal with output saturation levels of ± 12 V. Assume that when conducting the diode exhibits a constant voltage drop of 0.7 V. Find V_- when V_I is -1V.

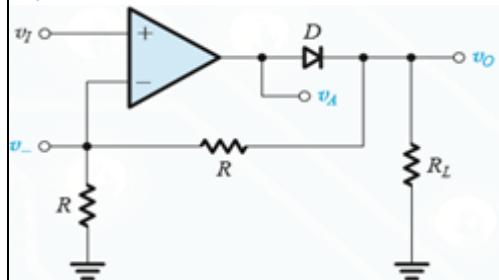
(c)



- a) 0V
- b) 0.7V
- c) 1V
- d) 1.7V

8.The op amp in the precision rectifier circuit is ideal with output saturation levels of ± 12 V. Assume that when conducting the diode exhibits a constant voltage drop of 0.7 V. Find V_0 when V_I is 2V.

(a)



- a) 0V
- b) 0.7V
- c) 1V
- d) 1.7V

UNIT – IV

**MOSFETS (MOS FIELD EFFECT
TRANSISTORS)**

UNIT-4

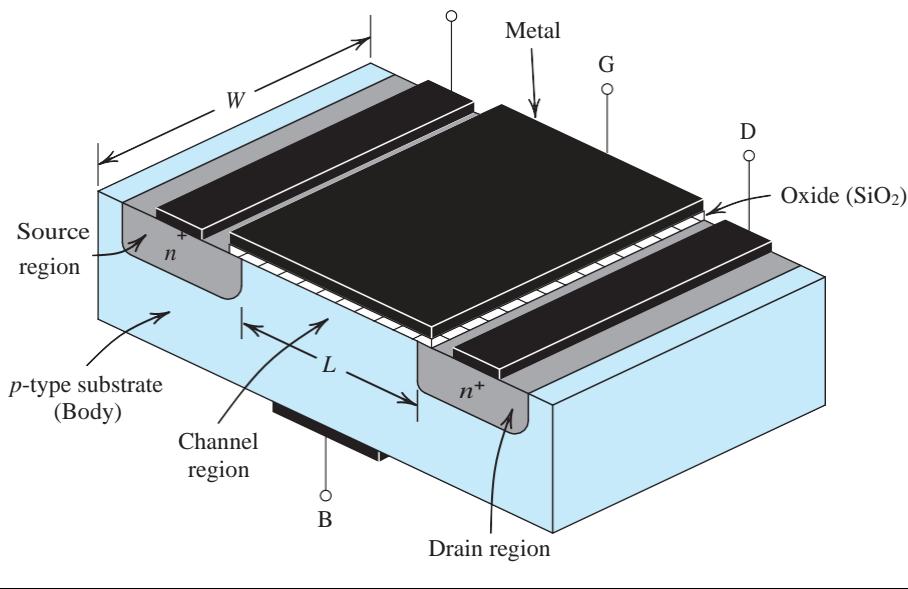
MOSFETs(MOS Field Effect Transistors)

4.1 Device Structure and Physical Operation

The enhancement-type MOSFET is the most widely used field-effect transistor. Except for the last section, this chapter is devoted to the study of the enhancement-type MOSFET. We begin in this section by learning about its structure and physical operation. This will lead to the current–voltage characteristics of the device, studied in the next section.

4.1.1 Device Structure

Figure 4.1 shows the physical structure of the *n*-channel enhancement-type MOSFET. The meaning of the names “enhancement” and “*n*-channel” will become apparent shortly. The transistor is fabricated on a *p*-type substrate, which is a single-crystal silicon wafer that provides physical support for the device (and for the entire circuit in the case of an integrated circuit). Two heavily doped *n*-type regions, indicated in the figure as the *n*⁺ **source**¹ and the *n*⁺ **drain** regions, are created in the substrate. A thin layer of silicon dioxide (SiO_2) of thickness t_{ox} (typically 1 nm to 10 nm), which is an excellent electrical insulator, is grown on the surface of the substrate, covering the area between the source and drain regions. Metal is deposited on top of the oxide layer to form the **gate electrode** of the device.s



(a)

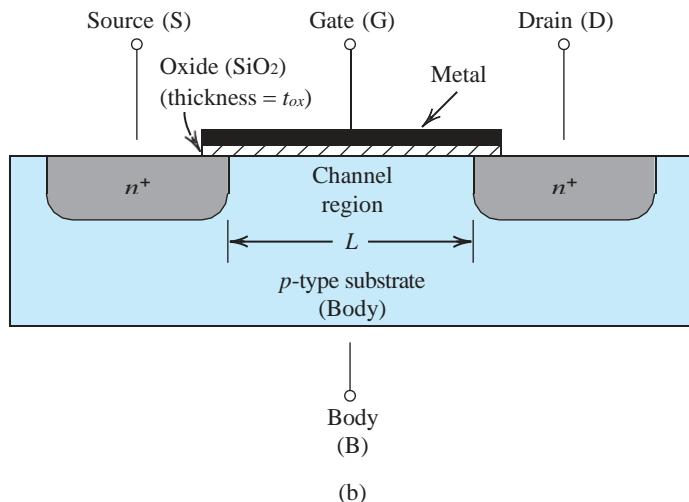


Figure 4.1 Physical structure of the enhancement-type NMOS transistor: **(a)** perspective view; **(b)** cross section. Typically $L = 0.03 \mu\text{m}$ to $1 \mu\text{m}$, $W = 0.05 \mu\text{m}$ to $100 \mu\text{m}$, and the thickness of the oxide layer (t_{ox}) is in the range of 1 to 10 nm.

Metal contacts are also made to the source region, the drain region, and the substrate, also known as the **body**.³ Thus four terminals are brought out: the gate terminal (G), the source terminal (S), the drain terminal (D), and the substrate or body terminal (B).

At this point it should be clear that the name of the device (metal-oxide-semiconductor FET) is derived from its physical structure. The name, however, has become a general one and is used also for FETs that do not use metal for the gate electrode. In fact, most modern MOSFETs are fabricated using a process known as silicon-gate technology, in which a certain type of silicon, called polysilicon, is used to form the gate electrode (see Appendix A). Our description of MOSFET operation and characteristics applies irrespective of the type of gate electrode.

Another name for the MOSFET is the **insulated-gate FET** or **IGFET**. This name also arises from the physical structure of the device, emphasizing the fact that the gate electrode is electrically insulated from the device body (by the oxide layer). It is this insulation that causes the current in the gate terminal to be extremely small (of the order of 10^{-15} A).

Observe that the substrate forms *pn* junctions with the source and drain regions. In normal operation these *pn* junctions are kept reverse biased at all times. Since, as we shall see shortly, the drain will always be at a positive voltage relative to the source, the two *pn* junctions can be effectively cut off by simply connecting the substrate terminal to the source terminal. We shall assume this to be the case in the following description of MOSFET operation. Thus, here, the substrate will be considered as having no effect on device operation, and the MOSFET will be treated as a three-terminal device, with the terminals being the gate (G), the source (S), and the drain (D). It will be shown that a voltage applied to the gate controls current flow between source and drain. This current will flow in the longitudinal direction from drain to source in the region labeled “channel region.” Note that this region has a length L and a width W , two

important parameters of the MOSFET. Typically, L is in the range of 0.03 μm to 1 μm , and W is in the range of 0.05 μm to 100 μm . Finally, note that the MOSFET is a symmetrical device; thus its source and drain can be interchanged with no change in device characteristics.

4.1.1 Operation with Zero Gate Voltage

With zero voltage applied to the gate, two back-to-back diodes exist in series between drain and source. One diode is formed by the *pn* junction between the n^+ drain region and the *p*-type substrate, and the other diode is formed by the *pn* junction between the *p*-type substrate and the n^+ source region. These back-to-back diodes prevent current conduction from drain to source when a voltage v_{DS} is applied. In fact, the path between drain and source has a very high resistance (of the order of 10^{12} Δ).

4.1.1 Creating a Channel for Current Flow

Consider next the situation depicted in Fig. 5.2. Here we have grounded the source and the drain and applied a positive voltage to the gate. Since the source is grounded, the gate voltage appears in effect between gate and source and thus is denoted v_{GS} . The positive voltage on the gate causes, in the first instance, the free holes (which are positively charged) to be repelled from the region of the substrate under the gate (the channel region). These holes are pushed downward into the substrate, leaving behind a carrier-depletion region. The depletion region is populated by the bound negative charge associated with the acceptor atoms. These charges are “uncovered” because the neutralizing holes have been pushed downward into the substrate.

As well, the positive gate voltage attracts electrons from the n^+ source and drain regions (where they are in abundance) into the channel region. When a sufficient number of electrons accumulate near the surface of the substrate under the gate, an n region is in effect created, connecting the source and drain regions, as indicated in Fig. 5.2. Now if a voltage is applied between drain and source, current flows through this induced n region, carried by the mobile electrons.

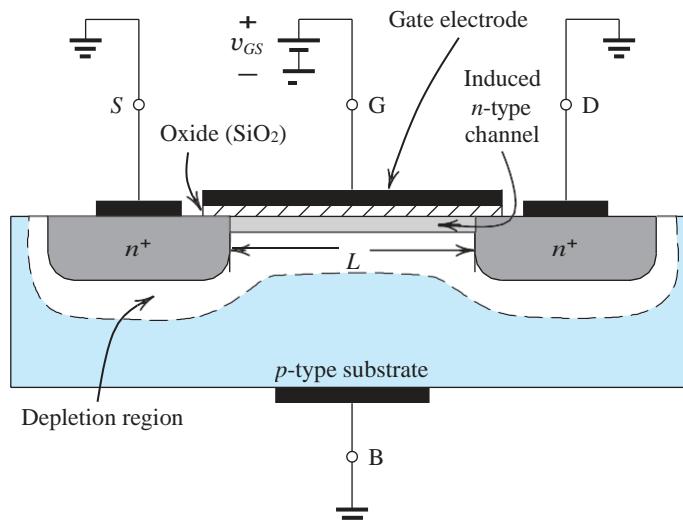


Figure 4.2 The enhancement-type NMOS transistor with a positive voltage applied to the gate. An n channel is induced at the top of the substrate beneath the gate.

The *induced n* region thus forms a **channel** for current flow from drain to source and is aptly called so. Correspondingly, the MOSFET of Fig. 5.2 is called an **n-channel MOSFET** or, alternatively, an **NMOS transistor**. Note that an *n*-channel MOSFET is formed in a *p*-type substrate: The channel is created by *inverting* the substrate surface from *p* type to *n* type. Hence the induced channel is also called an **inversion layer**.

The value of v_{GS} at which a sufficient number of mobile electrons accumulate in the channel region to form a conducting channel is called the **threshold voltage** and is denoted V_t .

⁴ Obviously, V for an *n*-channel FET is positive. The value of V is controlled during device fabrication and typically lies in the range of 0.3 V to 1.0 V.

The gate and the channel region of the MOSFET form a parallel-plate capacitor, with the oxide layer acting as the capacitor dielectric. The positive gate voltage causes positive charge to accumulate on the top plate of the capacitor (the gate electrode). The corresponding negative charge on the bottom plate is formed by the electrons in the induced channel. An electric field thus develops in the vertical direction. It is this field that controls the amount of charge in the channel, and thus it determines the channel conductivity and, in turn, the current that will flow through the channel when a voltage v_{DS} is applied. This is the origin of the name “field-effect transistor” (FET).

The voltage across this parallel-plate capacitor, that is, the voltage across the oxide, must exceed $\overline{V_t}$ for a channel to form. When $v_{DS} = 0$, as in Fig. 5.2, the voltage at every point along the channel is zero, and the voltage across the oxide (i.e., between the gate and the points along the channel) is uniform and equal to v_{GS} . The excess of v_{GS} over $\overline{V_t}$ is termed the **effective voltage** or the **overdrive voltage** and is the quantity that determines the charge in the channel. In this book, we shall denote $(v_{GS} - \overline{V_t})$ by v_{OV} ,

$$v_{GS} - \overline{V_t} \equiv v_{OV} \quad (4.1)$$

We can express the magnitude of the electron charge in the channel by

$$|Q| = C_{ox}(WL)v_{ov} \quad (4.2)$$

where C_{ox} , called the **oxide capacitance**, is the capacitance of the parallel-plate capacitor per unit gate area (in units of F/m²), W is the width of the channel, and L is the length of the channel. The oxide capacitance

where ϵ_{ox} is the permittivity of the silicon dioxide,

$$\epsilon_{ox} = 3.9\epsilon_0 = 3.9 \times 8.854 \times 10^{-12} = 3.45 \times 10^{-11} \text{ F/m}$$

The oxide thickness t_{ox} is determined by the process technology used to fabricate the MOSFET. As an example, for a process with $t_{ox} = 4 \text{ nm}$,

$$C_{ox} = \frac{3.45 \times 10^{-11}}{4 \times 10^{-9}} = 8.6 \times 10^{-3} \text{ F/m}^2$$

$$C = C_{ox}WL = 8.6 \times 0.18 \times 0.72 = 1.1 \text{ fF}$$

Finally, note from Eq. (4.2) that as v_{OV} is increased, the magnitude of the channel charge increases proportionately. Sometimes this is depicted as an increase in the depth of the channel; that is, the larger the overdrive voltage, the deeper the channel.

4.1.1 Applying a Small v_{DS}

Having induced a channel, we now apply a positive voltage v_{DS} between drain and source, as shown in Fig. 4.3. We first consider the case where v_{DS} is small (i.e., 50 mV or so). The voltage v_{DS} causes a current i_D to flow through the induced n channel. Current is carried by free electrons traveling from source to drain (hence the names source and drain). By convention, the direction of current flow is opposite to that of the flow of negative charge. Thus the current in the channel, i_D , will be from drain to source, as indicated in Fig. 4.3.

We now wish to calculate the value of i_D . Toward that end, we first note that because v_{DS} is small, we can continue to assume that the voltage between the gate and various points along the channel remains approximately constant and equal to the value at the source end, v_{GS} . Thus, the effective voltage between the gate and the various points along the channel remains equal to v_{OV} , and the channel charge Q is still given by Eq. (4.2). Of particular interest

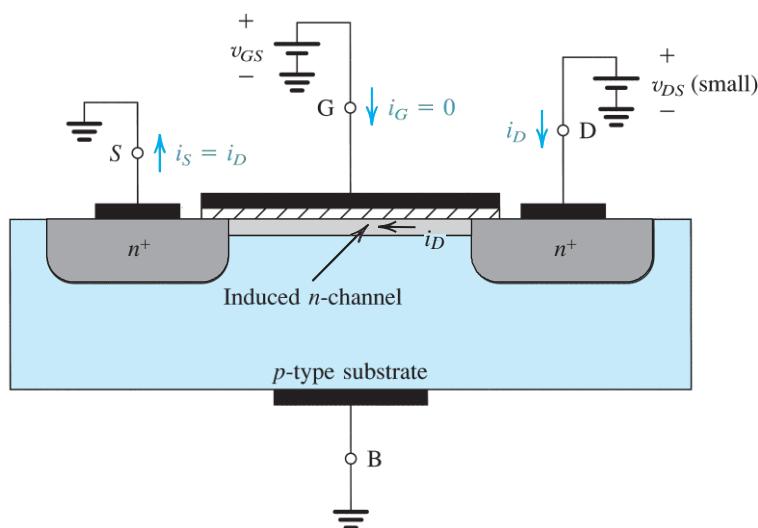


Figure 4.3 An NMOS transistor with $v_{GS} > V_t$ and with a small v_{DS} applied. The device acts as a resistance whose value is determined by v_{GS} . Specifically, the channel conductance is proportional to $v_{GS} - V_t$, and thus i_D is proportional to $(v_{GS} - V_t)v_{DS}$. Note that the depletion region is not shown (for simplicity).

the current i_D is the charge per unit length channel in calculating, which can be found from Eq. (4.2) as

$$\frac{|O|}{\text{unit channel length}} = C_{ox} W v_{OV} \quad (4.4)$$

The voltage v_{DS} establishes an electric field E across the length of the channel,

$$|E| = \frac{v_{DS}}{L} \quad (4.5)$$

This electric field in turn causes the channel electrons to drift toward the drain with a velocity given by

$$\text{Electron drift velocity} = \mu_n |E| = \mu_n \frac{v_{DS}}{L} \quad (4.6)$$

where μ_n is the mobility of the electrons at the surface of the channel. It is a physical parameter whose value depends on the fabrication process technology. The value of i_D can now be found by multiplying the charge per unit channel length (Eq. 4.4) by the electron drift velocity (Eq. 4.6),

$$i_D = (\mu_n C_{ox}) \frac{W}{L} v_{OV} v_{DS} \quad (4.7)$$

Thus, for small v_{DS} , the channel behaves as a linear resistance whose value is controlled by the overdrive voltage v_{OV} , which in turn is determined by v_{GS} :

$$i_D = (\mu_n C_{ox}) \frac{W}{L} (v_{GS} - V_t) v_{DS} \quad (4.8)$$

The conductance g_{DS} of the channel can be found from Eq. (4.7) or (4.8) as

$$g_{DS} = (\mu_n C_{ox}) \frac{W}{L} v_{OV} \quad (4.9)$$

$$g_{DS} = (\mu_n C_{ox}) \frac{W}{L} (v_{GS} - V_t) \quad (4.10)$$

that the conductance is determined by the product of three factors: ($\mu_n C_{ox}$), (W/L), and v_{ov} (or equivalently, $v_{GS} V_t$). To gain insight into MOSFET operation, we consider each of the three factors in turn.

The first factor, ($\mu_n C_{ox}$), is determined by the process technology used to fabricate the MOSFET. It is the product of the electron mobility, μ_n , and the oxide capacitance, C_{ox} . It makes physical sense for the channel conductance to be proportional to each of μ_n and C_{ox} (why?) and hence to their product, which is termed the **process transconductance** parameter⁵ and given the symbol k_n^r , where the subscript n denotes n channel,

$$k_n^r = \mu_n C_{ox} \quad (4.11)$$

It can be shown that with μ_n having the dimensions of meters squared per volt-second ($m^2/V \cdot s$) and C_{ox} having the dimensions of farads per meter squared (F/m^2), the dimensions of k^r are amperes per volt squared (A/V^2).

The second factor in the expression for the conductance g_{DS} in Eqs. (4.9) and (4.10) is the transistor **aspect ratio** (W/L). That the channel conductance is proportional to the channel width W and inversely proportional to the channel length L should make perfect physical sense. The (W/L) ratio is obviously a dimensionless quantity that is determined by the device designer. Indeed, the values of W and L can be selected by the device designer to give the device the i_v characteristics desired. For a given fabrication process, however, there is a minimum channel length, L_{min} . In fact, the minimum channel length that is possible with a given fabrication process is used to characterize the process and is being continually reduced as technology advances. For instance, in 2014 the state-of-the-art in commercially available MOS technology was a 32-nm process, meaning that for this process the minimum channel length possible was 32 nm. Finally, we should note that the oxide thickness t_{ox} scales down with L_{min} . Thus, for a 0.13-μm technology, t_{ox} is 2.7 nm, but for the currently popular 65-nm technology, t_{ox} is about 2.2 nm.

The product of the process transconductance parameter k_n^r and the transistor aspect ratio(W/L) is the **MOSFET transconductance parameter** k_n ,

$$k_n = k_n^r (W/L) \quad (4.12a)$$

$$k_n = (\mu_n C_{ox})(W/L) \quad (4.12b)$$

The third term in the expression of the channel conductance g_{DS} is the overdrive voltage v_{OV} . This is hardly surprising, since v_{OV} directly determines the magnitude of electron charge in the channel. As will be seen, v_{OV} and V_t is a very important circuit-design parameter. In this book, we will use v interchangeably.

We conclude this subsection by noting that with v_{DS} kept small, the MOSFET behaves as a linear resistance r_{DS} whose value is controlled by the gate voltage v_{GS} ,

$$r_{DS} = \frac{1}{g_{DS}} = \frac{1}{(\mu_n C_{ox})(W/L)v_{OV}} \quad (4.13a)$$

$$r_{DS} = \frac{1}{(\mu_n C_{ox})(W/L)(v_{GS} - V_t)} \quad (4.13b)$$

The operation of the MOSFET as a voltage-controlled resistance is further illustrated in Fig. 4.4, which is a sketch of i_D versus v_{DS} for various values of v_{GS} . Observe that the resistance is infinite for $v_{GS} < V_t$ and decreases as v_{GS} is increased above V_t . It is interesting to note that although v_{GS} is used as the parameter for the set of graphs in Fig. 5.4, the graphs in fact depend only on v_{OV} (and, of course, k_n).

The description above indicates that for the MOSFET to conduct, a channel has to be induced. Then, increasing v_{GS} above the threshold voltage V_t enhances the channel, hence the names **enhancement-mode operation** and **enhancement-type MOSFET**. Finally, we note that the current that leaves the source terminal (i_S) is equal to the current that enters the drain terminal (i_D), and the gate current $i_G = 0$.

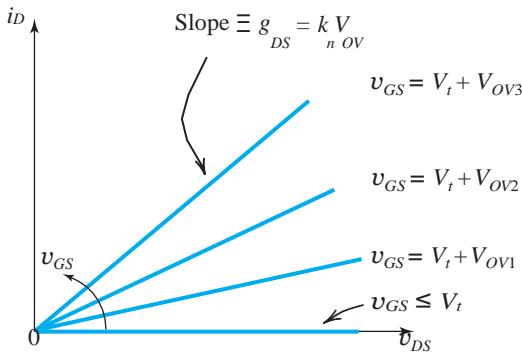


Figure 5.4 The i_D - v_{DS} characteristics of the MOSFET in Fig. 5.3 when the voltage applied between drain and source, v_{DS} , is kept small. The device operates as a linear resistance whose value is controlled by v_{GS} .

4.1.1 Operation as v_{DS} Is Increased

We next consider the situation as v_{DS} is increased. For this purpose, let v_{GS} be held constant at a value greater than V_t ; that is, let the MOSFET be operated at a constant overdrive voltage V_{OV} . Refer to Fig. 5.5, and note that v_{DS} appears as a voltage drop across the length of the channel. That is, as we travel along the channel from source to drain, the voltage (measured relative to the source) increases from zero to v_{DS} . Thus the voltage between the gate and points along the channel decreases from $v_{GS} - V_t - V_{OV}$ at the source end to $v_{GS} - v_{DS} - V_t - V_{OV} - v_{DS}$ at the drain end. Since the channel depth depends on this voltage, and specifically on the amount by which this voltage exceeds V_t , we find that the channel is no longer of uniform depth; rather, the channel will take the tapered shape shown in Fig. 5.5, being deepest at the source end (where the depth is proportional to V_{OV}) and shallowest at the drain end (where the depth is proportional to $V_{OV} v_{DS}$). This point is further illustrated in Fig. 4.6.

As v_{DS} is increased, the channel becomes more tapered and its resistance increases correspondingly. Thus, the i_D - v_{DS} curve does not continue as a straight line but bends as shown in Fig. 4.7. The equation describing this portion of the i_D - v_{DS} curve can be easily derived by utilizing the information in Fig. 4.6. Specifically, note that the charge in the tapered channel is proportional to the channel cross-sectional area shown in Fig. 4.6(b). This area in turn can

be easily seen as proportional to $\frac{1}{2}[V_{ov} + (V_{ov} - v_{DS})]$ or $V_{ov} - \frac{1}{2}v_{DS}$. Thus, the relationship between i_D and v_{DS} can be found by replacing V_{ov} in Eq. (5.7) by $V_{ov} - \frac{1}{2}v_{DS}$,

$$i_D = k_n \frac{W}{L} \left(V_{ov} - \frac{1}{2}v_{DS} \right)^2$$

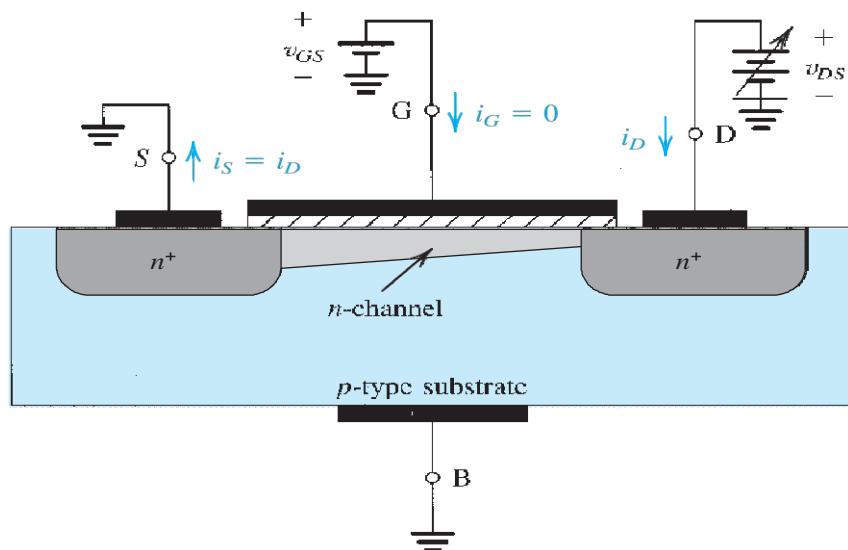
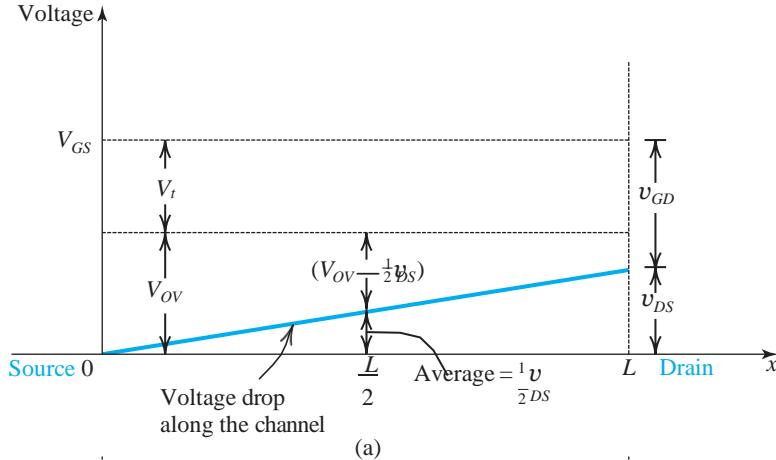
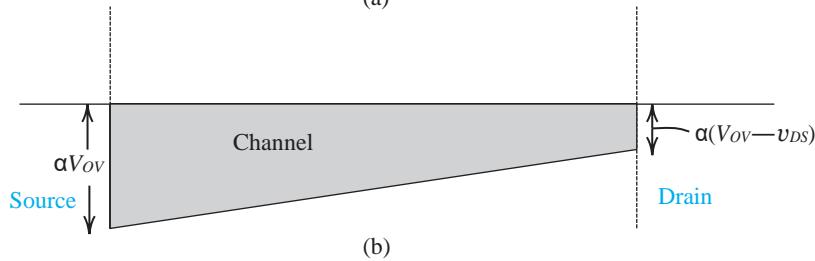


Figure 4.5 Operation of the enhancement NMOS transistor as v_{DS} is increased. The induced channel acquires a tapered shape, and its resistance increases as v_{DS} is increased. Here, v_{GS} is kept constant at a value $> V_t$; $v_{GS} = V_t + V_{ov}$.



(a)



(b)

Figure 4.6. (a) For a MOSFET with a value of $\bar{V}_t + V_{ov}$ at the midpoint of v_{ds} , causes the voltage drop along the channel to vary linearly, with an average value of $\frac{1}{2}V_{ov}$ at the midpoint. Since $V_{GD} < V_t$, the channel still exists at the drain end. (b) The channel shape corresponding to the situation in (a). While the depth of the channel at the source end is still proportional to V_{ov} , that at the drain end is proportional to $(V_{ov} - v_{DS})$.

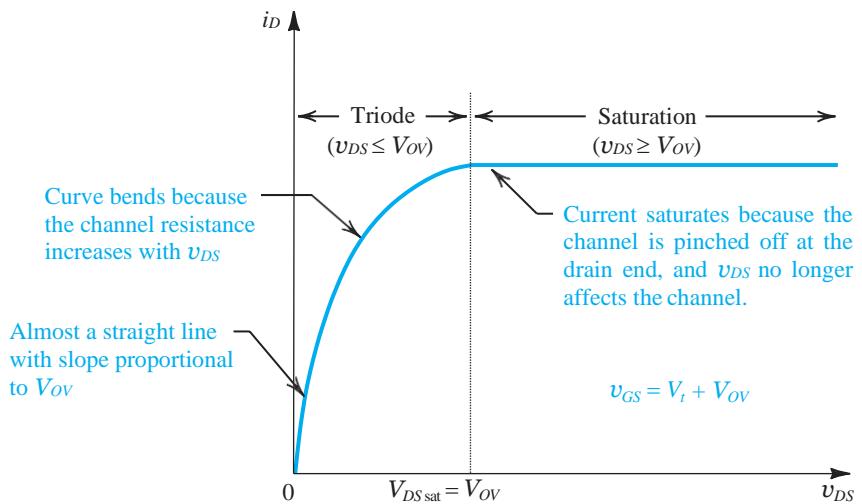


Figure 5.7 The drain current i_D versus the drain-to-source voltage v_{DS} for an enhancement-type NMOS transistor operated with $v_{GS} = V_t + V_{OV}$.

alternate form

$$i_D = k_n^r \frac{W}{L} \left(\frac{V_{OV}}{v_{DS}} - \frac{1}{2} \frac{v_{DS}^2}{V_{DS}} \right) \quad (5.15)$$

Furthermore, for an arbitrary value of V_{OV} , we can replace V_{OV} by $(v_{GS} - V_t)$ and rewrite Eq. (5.15) as

$$i_D = k_n^r \frac{W}{L} \left(v_{GS} - V_t - \frac{1}{2} \frac{v_{DS}^2}{V_{DS}} \right) \quad (5.16)$$

4.1.1 Operation for $v_{DS} > V_{OV}$: Channel Pinch-Off and Current Saturation

The above description of operation assumed that even though the channel became tapered, it still had a finite (nonzero) depth at the drain end. This in turn is achieved by keeping v_{DS} sufficiently small that the voltage between the gate and the drain, v_{GD} , exceeds V_t . This is indeed the situation shown in Fig. 5.6(a). Note that for this situation to obtain, v_{DS} must not exceed V_{OV} , for as $v_{DS} > V_{OV}$, $v_{GD} = V_t$, and the channel depth at the drain end reduces to zero.

Figure 5.8 shows v_{DS} reaching V_{OV} and v_{GD} correspondingly reaching V_t . The zero depth of the channel at the drain end gives rise to the term **channel pinch-off**. Increasing v_{DS} beyond this value (i.e., $v_{DS} > V_{OV}$) has no effect on the channel shape and charge, and the current through the channel remains constant at the value reached for $v_{DS} = V_{OV}$. The drain current thus **saturates** at the value found by substituting $v_{DS} = V_{OV}$ in Eq. (5.14),

$$i_D = \frac{1}{2} k_n^r \frac{W}{L} V_{OV}^2 \quad (4.17)$$

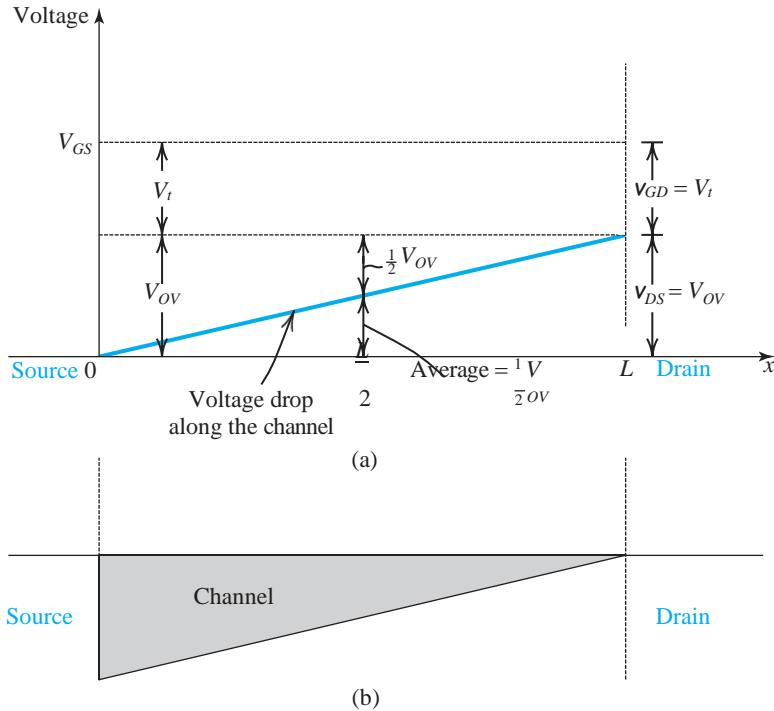


Figure 5.8 Operation of MOSFET with $v_{GS} = V_t + V_{OV}$, as v_{DS} is increased to V_{OV} . At the drain end, v_{GD} decreases to V_t and the channel depth at the drain end reduces to zero (pinch-off). At this point, the MOSFET enters the saturation mode of operation. Further increasing v_{DS} (beyond $V_{DSsat} = V_{OV}$) has no effect on the channel shape and i_D remains constant.

The MOSFET is then said to have entered the **saturation region** (or, equivalently, the saturation mode of operation). The voltage v_{DS} at which saturation occurs is denoted V_{DSsat} ,

$$V_{DSsat} = V_{OV} = V_{GS} - V_t \quad (4.18)$$

It should be noted that channel pinch-off does *not* mean channel blockage: Current continues to flow through the pinched-off channel, and the electrons that reach the drain end of the channel are accelerated through the depletion region that exists there (not shown in Fig. 5.5) and into the drain terminal. Any increase in v_{DS} above V_{DSsat} appears as a voltage drop across the depletion region. Thus, both the current through the channel and the voltage drop across it remain constant in saturation.

The saturation portion of the i_D v_{DS} curve is, as expected, a horizontal straight line, as indicated in Fig. 5.7. Also indicated in Fig. 5.7 is the name of the region of operation obtained with a continuous (non-pinched-off) channel, the **triode region**. This name is a carryover from the days of vacuum-tube devices, whose operation a FET resembles.

Finally, we note that the expression for i_D in saturation can be generalized by replacing the constant overdrive voltage V_{OV} by a variable one, v_{OV} :

$$i_D = \frac{1}{2} k' \frac{W}{L} v_{OV}^2 \quad (4.19)$$

Also, v_{OV} can be replaced by $(v_{GS} - V_t)$ to obtain the alternate expression for saturation-mode i_D ,

$$i_D = \frac{1}{2} k_n^r \frac{W}{L} (v_{GS} - V_t)^2 \quad (4.20)$$

Example 5.1

Consider a process technology for which $L_{min} = 0.4 \mu\text{m}$, $t_{ox} = 8 \text{ nm}$, $\mu_n = 450 \text{ cm}^2/\text{V} \cdot \text{s}$, and $V = 0.7 \text{ V}$.

- (a) Find C_{ox} and k_n^r .
- (b) For a MOSFET with $W/L = 8 \mu\text{m}/0.8 \mu\text{m}$, calculate the values of V_{OV} , V_{GS} , and V_{DSmin} needed to operate the transistor in the saturation region with a dc current $I_D = 100 \mu\text{A}$.
- (c) For the device in (b), find the values of V_{OV} and V_{GS} required to cause the device to operate as a 1000- Ω resistor for very small v_{DS} .

Solution

(a)

$$\begin{aligned} C_{ox} &= \frac{e_{ox}}{t_{ox}} = \frac{3.45 \times 10^{-11}}{8 \times 10^{-9}} = 4.32 \times 10^{-3} \text{ F/m}^2 \\ &= 4.32 \text{ fF}/\mu\text{m} \\ k_n^r &= \mu_n C_{ox} = 450 \left(\text{cm}^2/\text{V} \cdot \text{s} \right) \times 4.32 \left(\text{fF}/\mu\text{m} \right)^2 \\ &= 450 \times 10^8 \left(\mu\text{m}^2/\text{V} \cdot \text{s} \right) \times 4.32 \times 10^{-15} \left(\text{F}/\mu\text{m}^2 \right) \\ &= 194 \times 10^{-6} \left(\text{F}/\text{V} \cdot \text{s} \right) \\ &= 194 \mu\text{A/V} \end{aligned}$$

(b) For operation in the saturation region,

$$i_D = \frac{1}{2} k_n^r \frac{W}{L} V_{OV}^2$$

Thus,

$$100 = \frac{1}{2} \times 194 \times \frac{8}{0.8} V_{OV}^2$$

which results in

$$V_{OV} = 0.32 \text{ V}$$

Thus,

$$V_{GS} = V_t + V_{OV} = 1.02 \text{ V}$$

and

$$V_{DSmin} = V_{OV} = 0.32 \text{ V}$$

(c) For the MOSFET in the triode region with v_{DS} very small,

$$r_{DS} = \frac{1}{k_n \frac{W}{L} V_{ov}}$$

Thus

$$1000 = \frac{1}{194 \times 10^{-6} \times 10 \times V_{ov}}$$

which yields

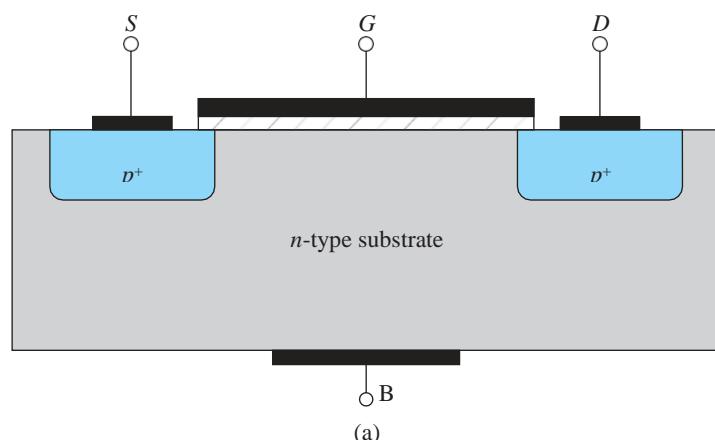
$$V_{ov} = 0.52 \text{ V}$$

Thus,

$$V_{GS} = 1.22 \text{ V}$$

4.1.1 The *p*-Channel MOSFET

Figure 5.9(a) shows a cross-sectional view of a *p*-channel enhancement-type MOSFET. The structure is similar to that of the NMOS device except that here the substrate is *n* type and the source and the drain regions are *p*⁺ type; that is, all semiconductor regions are reversed in polarity relative to their counterparts in the NMOS case. The PMOS and NMOS transistors are said to be *complementary* devices.



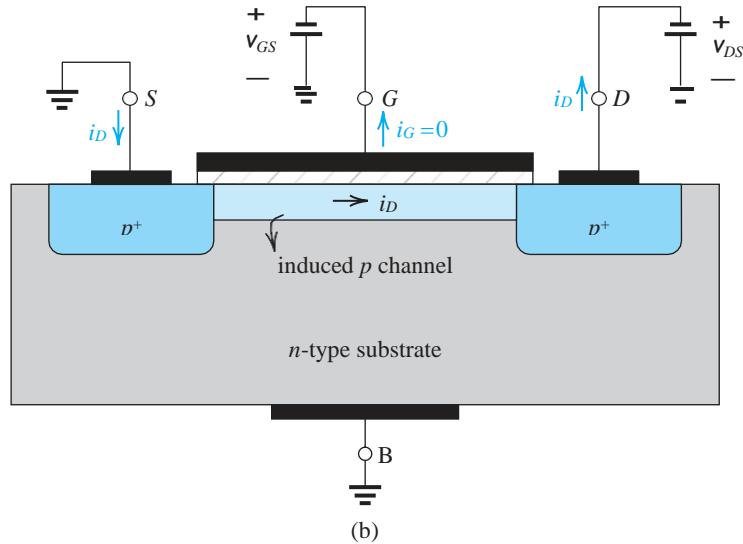


Figure 4.9 (a) Physical structure of the PMOS transistor. Note that it is similar to the NMOS transistor shown in Fig. 5.1(b) except that all semiconductor regions are reversed in polarity. (b) A negative voltage v_{GS} of magnitude greater than V_{tp} induces a p channel, and a negative v_{DS} causes a current i_D to flow from source to drain.

To induce a channel for current flow between source and drain, a negative voltage is applied to the gate, that is, between gate and source, as indicated in Fig. 4.9(b). By increasing the magnitude of the negative v_{GS} beyond the magnitude of the threshold voltage V_{tp} , which by convention is negative, a p channel is established as shown in Fig. 4.9(b). This condition can be described as

$$v_{GS} \leq V_{tp}$$

or, to avoid dealing with negative signs,

$$|v_{GS}| \geq |V_{tp}|$$

Now, to cause a current i_D to flow in the p channel, a negative voltage v_{DS} is applied to the drain.⁷ The current i is carried by holes and flows through the channel from source to drain. As we have done for the NMOS transistor, we define the process trans conductance parameter for the PMOS device as

$$k_p^r = \mu_p C_{p\text{ ox}}$$

where μ_p is the mobility of the holes in the induced p channel. Typically, $\mu_p = 0.25 \mu_n$ to $0.5 \mu_n$ and is process-technology dependent. The transistor transconductance parameter k_p is obtained by multiplying k_p^r by the aspect ratio W/L ,

$$k_p = k_p^r (W/L)$$

The remainder of the description of the physical operation of the p -channel MOSFET follows that for the NMOS device, except of course for the sign reversals of all voltages. We will present the complete current–voltage characteristics of both NMOS and PMOS transistors in the next section.

PMOS technology originally dominated MOS integrated-circuit manufacturing, and the original microprocessors utilized PMOS transistors. As the technological difficulties of fabricating NMOS transistors were solved, NMOS completely supplanted PMOS. The main reason for this change is that electron mobility μ_n is higher by a factor of 2 to 4 than the hole mobility μ_p , resulting in NMOS transistors having greater gains and speeds of operation than PMOS devices. Subsequently, a technology was developed that permits the fabrication of both NMOS and PMOS transistors on the same chip. Appropriately called **complementary MOS**, or **CMOS**, this technology is currently the dominant electronics technology.

4.1.1 Complementary MOS or CMOS

As the name implies, complementary MOS technology employs MOS transistors of both polarities. Although CMOS circuits are somewhat more difficult to fabricate than NMOS, the availability of complementary devices makes possible many powerful circuit configurations. Indeed, at the present time CMOS is the most widely used of all the IC technologies. This statement applies to both analog and digital circuits. CMOS technology has virtually replaced designs based on NMOS transistors alone. Furthermore, by 2014 CMOS technology had taken over many applications that just a few years

earlier were possible only with bipolar devices. Throughout this book, we will study many CMOS circuit techniques.

Figure 5.10 shows a cross section of a CMOS chip illustrating how the PMOS and NMOS transistors are fabricated. Observe that while the NMOS transistor is implemented directly in the *p*-type substrate, the PMOS transistor is fabricated in a specially created *n* region, known as an ***n* well**. The two devices are isolated from each other by a thick region of oxide that functions as an insulator. Not shown on the diagram are the connections made to the *p*-type body and to the *n* well. The latter connection serves as the body terminal for the PMOS transistor.

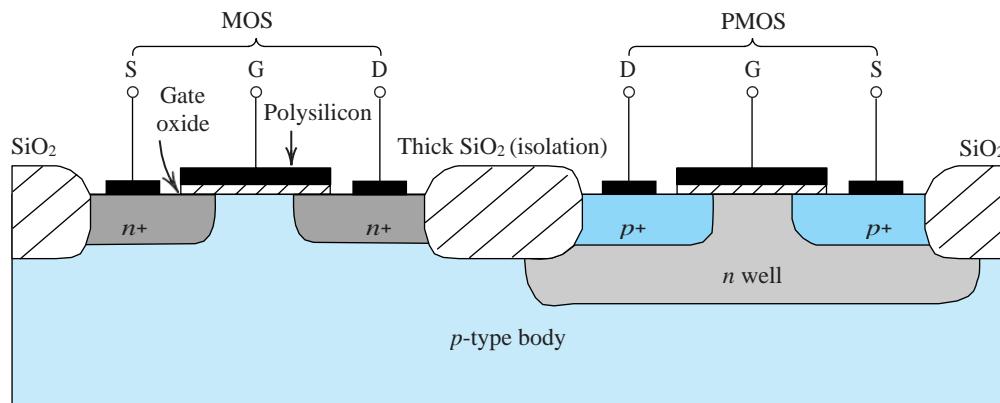


Figure 4.10 Cross section of a CMOS integrated circuit. Note that the PMOS transistor is formed in a separate *n*-type region, known as an *n* well. Another arrangement is also possible in which an *n*-type substrate (body) is used and the *n* device is formed in a *p* well. Not shown are the connections made to the *p*-type body and to the *n* well; the latter functions as the body terminal for the *p*-channel device.

4.1.1 Operating the MOS Transistor in the Subthreshold Region

The above description of the *n*-channel MOSFET operation implies that for $v_{GS} < V_t$, no current flows and the device is cut off. This is not entirely true, for it has been found that for values of v_{GS} smaller than but close to V_t , a small drain current flows. In this **subthreshold region** of operation, the drain current is exponentially related to v_{GS} , much like the $i_C - v_{BE}$ relationship of a BJT, as will be shown in the next chapter.

Although in most applications the MOS transistor is operated with $v_{GS} > V_t$, there are special, but a growing number of, applications that make use of subthreshold operation. In Chapter 14, we will briefly consider subthreshold operation.

4.2 Current-Voltage Characteristics

Building on the physical foundation established in the previous section for the operation of the enhancement MOS transistor, in this section we present its complete current–voltage characteristics. These characteristics can be measured at dc or at low frequencies and thus are called static characteristics. The dynamic effects that limit the operation of the MOSFET at high frequencies and high switching speeds will be discussed in Chapter 10.

4.2.1 Circuit Symbol

Figure 5.11(a) shows the circuit symbol for the *n*-channel enhancement-type MOSFET. Observe that the spacing between the two vertical lines that represent the gate and the channel indicates the fact that the gate electrode is insulated from the body of the device. The polarity of the *p*-type substrate (body) and the *n* channel is indicated by the arrowhead on the line representing the body (B). This arrowhead also indicates the polarity of the transistor, namely, that it is an *n*-channel device.

Although the MOSFET is a symmetrical device, it is often useful in circuit design to designate one terminal as the source and the other as the drain (without having to write S and

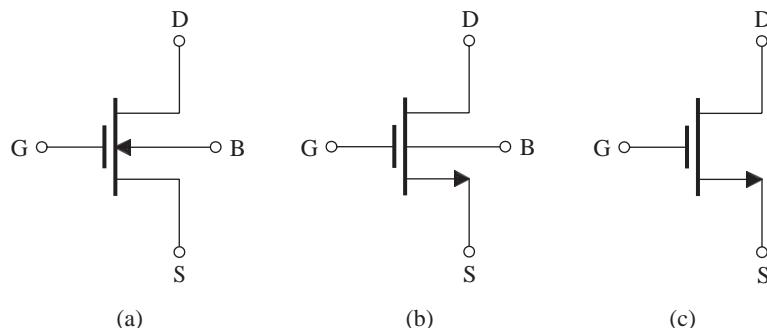


Figure 4.11 (a) Circuit symbol for the *n*-channel enhancement-type MOSFET. (b) Modified circuit symbol with an arrowhead on the source terminal to distinguish it from the drain and to indicate device polarity (i.e., *n* channel). (c) Simplified circuit symbol to be used when the source is connected to the body or when the effect of the body on device operation is unimportant.

D beside the terminals). This objective is achieved in the modified circuit symbol shown in Fig. 4.11(b). Here an arrowhead is placed on the source terminal, thus distinguishing it from the drain terminal. The arrowhead points in the normal direction of current flow and thus indicates the polarity of the device (i.e., n channel). Observe that in the modified symbol, there is no need to show the arrowhead on the body line. Although the circuit symbol of Fig. 4.11(b) clearly distinguishes the source from the drain, in practice it is the polarity of the voltage impressed across the device that determines source and drain; *the drain is always positive relative to the source in an n-channel FET.*

In applications where the source is connected to the body of the device, a further simplification of the circuit symbol is possible, as indicated in Fig. 4.11(c). This symbol is also used in applications when the effect of the body on circuit operation is not important, as will be seen later.

4.2.2 The $i_D - v_{DS}$ Characteristics

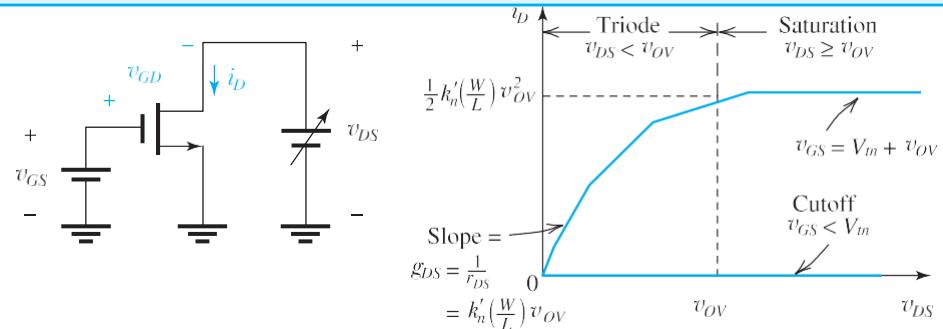
Table 5.1 provides a compilation of the conditions and the formulas for the operation of the NMOS transistor in each of the three possible regions: the cutoff region, the triode region, and the saturation region. The first two are useful if the MOSFET is to be utilized as a switch. On the other hand, if the MOSFET is to be used to design an amplifier, it must be operated in the saturation region. The rationale for these choices will be addressed in Chapter 7.

At the top of Table 4.1 we show a circuit consisting of an NMOS transistor and two dc supplies providing v_{GS} and v_{DS} . This conceptual circuit can be used to measure the $i_D - v_{DS}$ characteristic curves of the NMOS transistor. Each curve is measured by setting v_{GS} to a desired constant value, varying v_{DS} , and measuring the corresponding i_D . Two of these characteristic curves are shown in the accompanying diagram: one for $v_{GS} < V_{tn}$ and the other for $v_{GS} > V_{tn}$ (v_{ov}). (Note that we now use V_{tn} to denote the threshold voltage of the NMOS transistor, to distinguish it from that of the PMOS transistor, denoted V_{tp} .)

As Table 4.1 shows, the boundary between the triode region and the saturation region is determined by whether v_{DS} is less or greater than the overdrive voltage v_{ov} at which the transistor is

operating. An equivalent way to check for the region of operation is to examine the relative values of the drain and gate voltages. To operate in the triode region, the gate voltage must exceed the drain voltage by at least V_{tn} volts, which ensures that the channel remains continuous (not pinched off). On the other hand, to operate in saturation, the channel must be

Table 5.1 Regions of Operation of the Enhancement NMOS Transistor



- $v_{GS} < V_m$: no channel; transistor in cutoff; $i_D = 0$
- $v_{GS} = V_m + v_{OV}$: a channel is induced; transistor operates in the triode region or the saturation region depending on whether the channel is continuous or pinched off at the drain end;

↓
Triode Region

↓
Saturation Region

Continuous channel, obtained by:

$$v_{GD} > V_m$$

or equivalently:

$$v_{DS} < v_{OV}$$

Then,

$$i_D = k_n' \left(\frac{W}{L}\right) \left[(v_{GS} - V_m)v_{DS} - \frac{1}{2}v_{DS}^2 \right]$$

or equivalently,

$$i_D = k_n' \left(\frac{W}{L}\right) \left(v_{OV} - \frac{1}{2}v_{DS} \right) v_{DS}$$

Pinched-off channel, obtained by:

$$v_{GD} \leq V_m$$

or equivalently:

$$v_{DS} \geq v_{OV}$$

Then

$$i_D = \frac{1}{2} k_n' \left(\frac{W}{L}\right) (v_{GS} - V_m)^2$$

or equivalently,

$$i_D = \frac{1}{2} k_n' \left(\frac{W}{L}\right) v_{OV}^2$$

pinched off at the drain end; pinch-off is achieved here by keeping v_D higher than $v_G - V_m$, that is, not allowing v_D to fall below v_G by more than V_m volts. The graphical construction of Fig. 4.12 should serve to remind the reader of these conditions.

A set of i_D v_{DS} characteristics for the NMOS transistor is shown in Fig. 4.13. Observe that each graph is obtained by setting v_{GS} above V_m by a specific value of overdrive voltage, denoted V_{OV1} , V_{OV2} , V_{OV3} , and V_{OV4} . This in turn is the value of v_{DS} at which the corresponding graph.

Finally, observe that the boundary between the triode and the saturation regions, that is, the locus of the saturation points, is a parabolic curve described by

$$i_D = \frac{1}{2} k_n' \frac{W}{L} v_{DS}^2$$

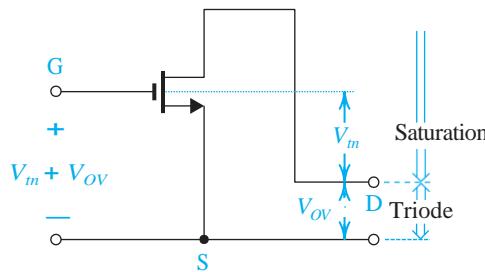


Figure 5.12 The relative levels of the terminal voltages of the enhancement NMOS transistor for operation in the triode region and in the saturation region.

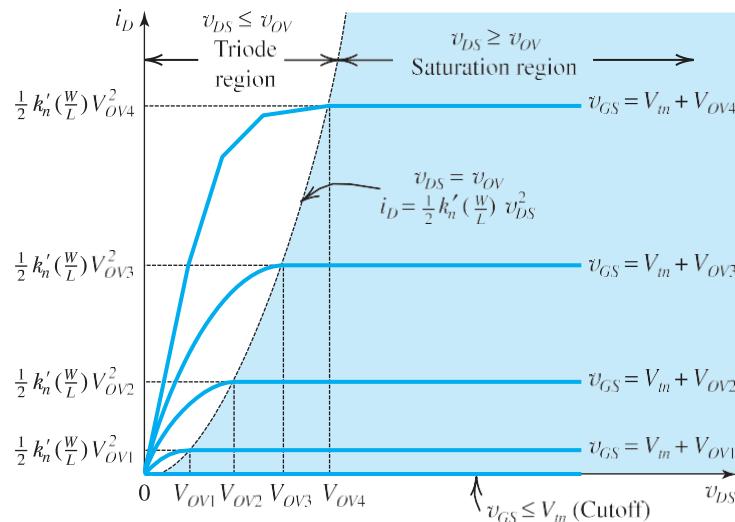


Figure 5.13 The $i_D - v_{DS}$ characteristics for an enhancement-type NMOS transistor.

4.2.2 The $i_D - v_{GS}$ Characteristic

When the MOSFET is used to design an amplifier, it is operated in the saturation region. As Fig. 5.13 indicates, in saturation the drain current is constant determined by v_{GS} (or v_{OV}) and is independent of v_{DS} . That is, the MOSFET operates as a constant-current source where the value of the current is determined by v_{GS} . In effect, then, the MOSFET operates as a voltage-controlled current source with the control relationship described by

$$i_D = \frac{1}{2} k^r \frac{W}{L} (v_{GS} - V_m)^2 \quad (5.21)$$

or in terms of v_{OV} ,

$$i_D = \frac{1}{2} k^r \frac{W}{L} v_{OV}^2 \quad (5.22)$$

This is the relationship that underlies the application of the MOSFET as an amplifier. That it is nonlinear should be of concern to those interested in designing linear amplifiers. Nevertheless, in Chapter 7, we will see how one can obtain linear amplification from this nonlinear control or transfer characteristic.

Figure 5.14 shows the $i_D - v_{GS}$ characteristic of an NMOS transistor operating in saturation. Note that if we are interested in a plot of i_D versus v_{OV} , we simply shift the origin to the point $v_{GS} = V_{tn}$.

The view of the MOSFET in the saturation region as a voltage-controlled current source is illustrated by the equivalent-circuit representation shown in Fig. 5.15. For reasons that will become apparent shortly, the circuit in Fig. 5.15 is known as a **large-signal equivalent circuit**. Note that the current source is ideal, with an infinite output resistance representing the independence, in saturation, of i_D from v_{DS} . This, of course, has been assumed in the idealized model of device operation utilized thus far. We are about to rectify an important shortcoming of this model. First, however, we present an example.

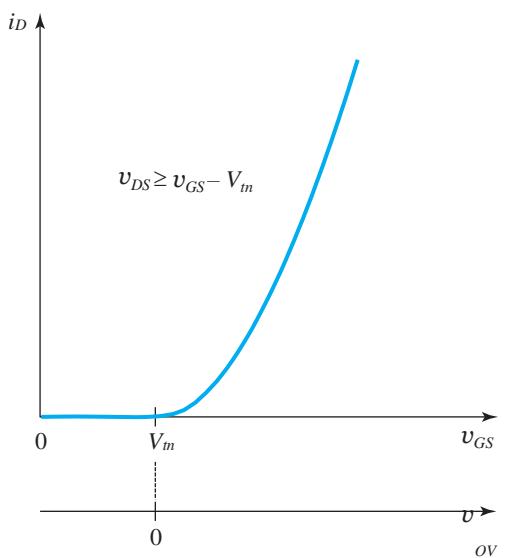


Figure 5.14 The $i_D - v_{GS}$ characteristic of an NMOS transistor operating in the saturation region. The $i_D - v_{OV}$ characteristic can be obtained by simply relabeling the horizontal axis, that is, shifting the origin to the point $v_{GS} = V_m$.

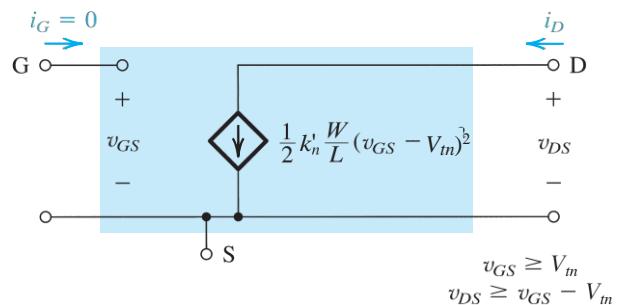


Figure 5.15 Large-signal, equivalent-circuit model of an n-channel MOSFET operating in the saturation region.

Example 5.2

Consider an NMOS transistor fabricated in a 0.18- μm process with $L = 0.18 \mu\text{m}$ and $W = 2 \mu\text{m}$. The process technology is specified to have $C_{ox} = 8.6 \text{ fF}/\mu\text{m}^2$, $\mu_n = 450 \text{ cm}^2/\text{V}\cdot\text{s}$, and $V_m = 0.5 \text{ V}$.

- Find V_{GS} and V_{DS} that result in the MOSFET operating at the edge of saturation with $I_D = 100 \mu\text{A}$.
- If V_{GS} is kept constant, find V_{DS} that results in $I_D = 50 \mu\text{A}$.
- To investigate the use of the MOSFET as a linear amplifier, let it be operating in saturation with $V_{DS} = 0.3 \text{ V}$. Find the change in i_D resulting from v_{GS} changing from 0.7 V by +0.01 V and by -0.01 V.

Solution

First we determine the process transconductance parameter k_n^r ,

$$\begin{aligned} k_n^r &= \mu_n C_{ox} \\ &= 450 \times 10^{-4} \times 8.6 \times 10^{-15} \times 10^{12} \text{ A/V}^2 \\ &= 387 \mu\text{A/V} \end{aligned}$$

and the transistor transconductance parameter k_n ,

$$\begin{aligned} k_n &= k_n^r \frac{W}{L} \\ 387 &\frac{2}{0.18} = 4.3 \text{ mA/V}^2 \end{aligned}$$

- With the transistor operating in saturation,

$$I_D = \frac{1}{2} k_n V_{OV}^2$$

Thus,

$$100 = \frac{1}{2} \times 4.3 \times 10^3 \times V_{OV}^2$$

which results in

$$V_{OV} = 0.22 \text{ V}$$

Thus,

$$V_{GS} = V_m + V_{OV} = 0.5 + 0.22 = 0.72 \text{ V}$$

and since operation is at the edge of saturation,

$$V_{DS} = V_{OV} = 0.22 \text{ V}$$

Example 5.2 continued

- (b) With V_{GS} kept constant at 0.72 V and I_D reduced from the value obtained at the edge of saturation, the MOSFET will now be operating in the triode region, thus

$$I_D = k \frac{V_{GS} - V_{OV}}{2} V_{DS}^2$$
$$50 = 4.3 \times 10^3 \frac{0.72 - 0.22}{2} V_{DS}^2$$

which can be rearranged to the form

$$\frac{V_{DS}^2}{2} - 0.44V_{DS} + 0.023 = 0$$

This quadratic equation has two solutions

$$V_{DS} = 0.06 \text{ V} \quad \text{and} \quad V_{DS} = 0.39 \text{ V}$$

The second answer is greater than V_{OV} and thus is physically meaningless, since we know that the transistor is operating in the triode region. Thus we have

$$V_{DS} = 0.06 \text{ V}$$

- (c) For $v_{GS} = 0.7 \text{ V}$, $v_{OV} = 0.2 \text{ V}$, and since $V_{DS} = 0.3 \text{ V}$, the transistor is operating in saturation and

$$I_D = \frac{1}{2} k \frac{V_{GS}^2}{V_{OV}}$$
$$= \frac{1}{2} \times 4300 \times 0.04$$
$$= 86 \mu\text{A}$$

Now for $v_{GS} = 0.710 \text{ V}$, $v_{OV} = 0.21 \text{ V}$ and

$$I_D = \frac{1}{2} \times 4300 \times 0.21^2 = 94.8 \mu\text{A}$$

and for $v_{GS} = 0.690 \text{ V}$, $v_{OV} = 0.19 \text{ V}$, and

$$I_D = \frac{1}{2} \times 4300 \times 0.19^2 = 77.6 \mu\text{A}$$

Thus, with $\Delta v_{GS} = +0.01 \text{ V}$, $\Delta I_D = 8.8 \mu\text{A}$; and for $\Delta v_{GS} = -0.01 \text{ V}$, $\Delta I_D = -8.4 \mu\text{A}$.

We conclude that the two changes are almost equal, an indication of almost-linear operation when the changes in v_{GS} are kept small. This is just a preview of the “small-signal operation” of the MOSFET studied in Chapter 7.

4.2.2 Finite Output Resistance in Saturation

Equation (5.21) and the corresponding large-signal equivalent circuit in Fig. 5.15, as well as the graphs in Fig. 5.13, indicate that in saturation, i_D is independent of v_{DS} . Thus, a change Δv_{DS} in the drain-to-source voltage causes a zero change in i_D , which implies that the incremental resistance looking into the drain of a saturated MOSFET is infinite. This, however, is an idealization based on the premise that once the channel is pinched off at the drain end, further increases in v_{DS} have no effect on the channel's shape. But, in practice, increasing v_{DS} beyond v_{OV} does affect the channel somewhat. Specifically, as v_{DS} is increased, the channel pinch-off point is moved slightly away from the drain, toward the source. This is illustrated in Fig. 5.16, from which we note that the voltage across the channel remains constant at v_{OV} , and the additional voltage applied to the drain appears as a voltage drop across the narrow depletion region between the end of the channel and the drain region. This voltage accelerates the electrons that reach the drain end of the channel and sweeps them across the depletion region into the drain. Note, however, that (with depletion-layer widening) the channel length is in effect reduced, from L to $L - \Delta L$, a phenomenon known as **channel-length modulation**. Now, since i_D is inversely proportional to the channel length (Eq. 5.21), i_D increases with v_{DS} .

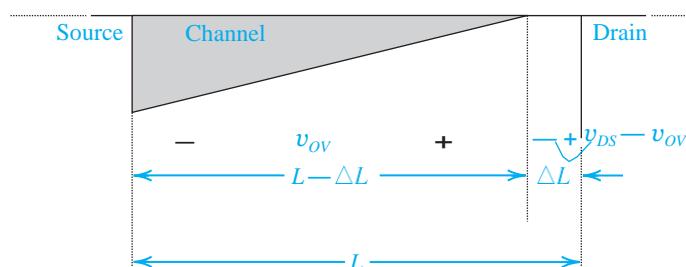


Figure 5.16 Increasing v_{DS} beyond v_{DSsat} causes the channel pinch-off point to move slightly away from the drain, thus reducing the effective channel length (by ΔL).

This effect can be accounted for in the expression for i_D by including a factor $1 + \lambda(v_{DS} - v_{OV})$ or, for simplicity, $(1 + \lambda v_{DS})$,

$$i_D = \frac{1}{2} k_n^r \frac{W}{L} (v_{GS} - V_{m0})^2 (1 + \lambda v_{DS}) \quad (4.23)$$

Here λ is a device parameter having the units of reciprocal volts V^{-1} . The value of λ depends both on the process technology used to fabricate the device and on the channel length L that the circuit designer selects. Specifically, the value of λ is much larger for newer submicron technologies than for older technologies. This makes intuitive sense: Newer technologies have very short channels, and are thus much more greatly impacted by the channel-length modulation effect. Also, for a given process technology, λ is inversely proportional to L .

A typical set of i_D-v_{DS} characteristics showing the effect of channel-length modulation is displayed in Fig. 5.17. The observed linear dependence of i_D on v_{DS} in the saturation region is represented in Eq. (4.23) by the factor $(1 + \lambda v_{DS})$. From Fig. 4.17 we observe that when the straight-line i_D-v_{DS} characteristics are extrapolated, they intercept the v_{DS} axis at the point, $v_{DS} = -V_A$, where V_A is a positive voltage. Equation (5.23), however, indicates that $i_D = 0$ at $v_{DS} = -1/\lambda$. It follows that

$$V_A = \frac{1}{\lambda}$$

and thus V_A is a device parameter with the dimensions of V. For a given process, V_A is proportional to the channel length L that the designer selects for a MOSFET. We can isolate the dependence of V_A on L by expressing it as

$$V_A = V_A^r L$$

where V_A^r is entirely process-technology dependent, with the dimensions of volts per micron. Typically, V_A^r falls in the range of 5 V/ μm to 50 V/ μm . The voltage V_A is usually referred to as the Early voltage, after J. M. Early, who discovered a similar phenomenon for the BJT (Chapter 6).

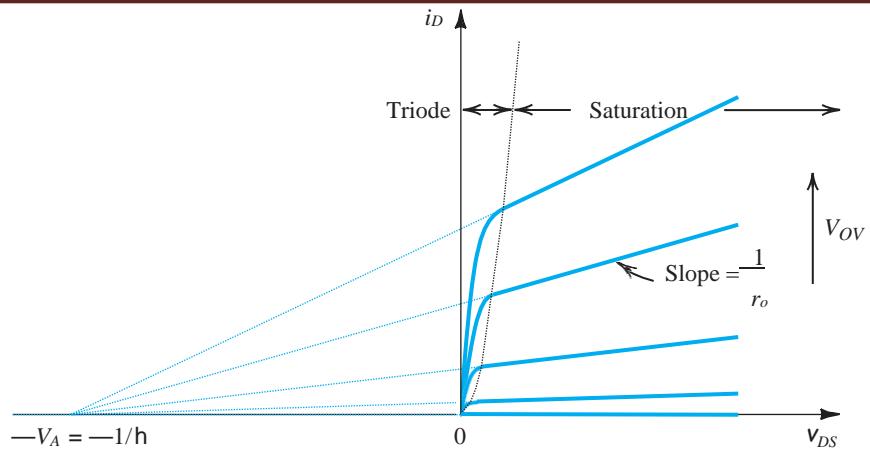


Figure 4.17 Effect of v_{DS} on i_D in the saturation region. The MOSFET parameter V_A depends on the process technology and, for a given process, is proportional to the channel length L .

Equation (4.23) indicates that when channel-length modulation is taken into account, the saturation values of i_D depend on v_{DS} . Thus, for a given v_{GS} , a change Δv_{DS} yields a corresponding change Δi_D in the drain current i_D . It follows that the output resistance of the current source representing i_D in saturation is no longer infinite. Defining the output

resistance r_o
as

$$r_o \equiv \frac{\partial i_D}{\partial v_{DS}} \Big|_{v_{GS} \text{ constant}} \quad (4.24)$$

and using Eq. (4.23) results in

$$r_o = \frac{k_n^r W}{2L} \left(V_{GS} - V_m \right)^{-1} \quad (4.25)$$

which can be written as

$$r_o = \frac{1}{A_L} \quad (4.26)$$

or, equivalently,

$$r_o = \frac{V_A}{I_D^r} \quad (4.27)$$

where I_D^r is the drain current *without* channel-length modulation taken into account; that is,

$$I_D^r = \frac{1}{2} k_n^r \frac{W}{L} (V_{GS} - V_m)^2 \quad (4.27^r)$$

Thus the output resistance is inversely proportional to the drain current.⁹ Finally, we show in Fig. 4.18 the large-signal, equivalent-circuit model incorporating r_o .

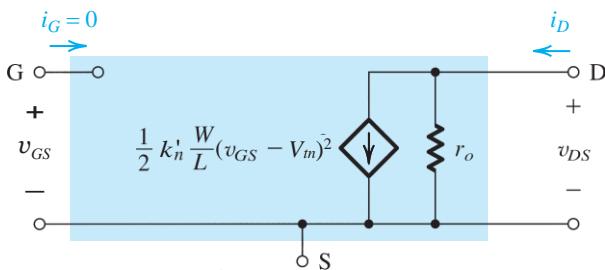


Figure 4.18 Large-signal, equivalent-circuit model of the *n*-channel MOSFET in saturation, incorporating the output resistance r_o . The output resistance models the linear dependence of i_D on v_{DS} and is given by Eq. (5.27).

4.2.2 Characteristics of the *p*-Channel MOSFET

The circuit symbol for the *p*-channel enhancement-type MOSFET is shown in Fig. 4.19(a). Figure 4.19(b) shows a modified circuit symbol in which an arrowhead pointing in the normal direction of current flow is included on the source terminal. For the case where the source is connected to the substrate, the simplified symbol of Fig. 4.19(c) is usually used.

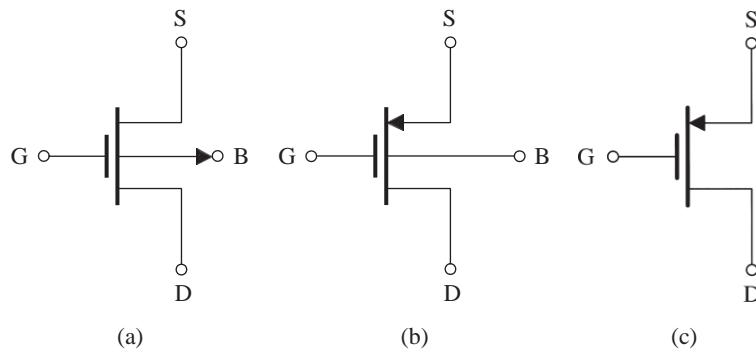
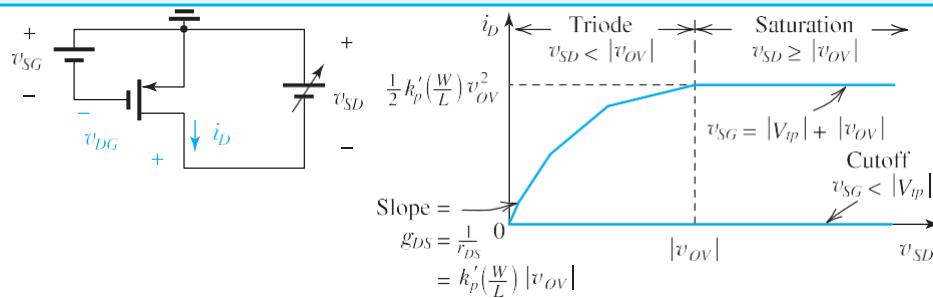


Figure 4.19 (a) Circuit symbol for the *p*-channel enhancement-type MOSFET. (b) Modified symbol with an arrowhead on the source lead. (c) Simplified circuit symbol for the case where the source is connected to the body.

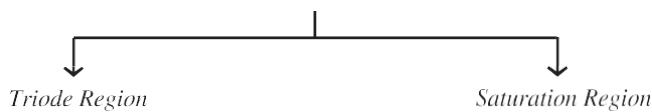
The regions of operation of the PMOS transistor and the corresponding conditions and expression for i_D are shown in Table 4.2. Observe that the equations are written in a way that emphasizes physical intuition and avoids the confusion of negative signs. Thus while V_{tp} is by convention negative, we use V_{tp} , and the voltages v_{SG} and v_{SD} are positive. Also, in all of our circuit diagrams we will always draw *p*-channel devices with their sources on top so that current flows from top to bottom. Finally, we note that PMOS devices also suffer from the channel-length modulation effect. This can be taken into account by including a factor $(1 + |\lambda|v_{SD})$ in the saturation-region expression for i_D as follows

$$i_D = \frac{1}{2} k_p^r \frac{W}{L} (v_{SG} - V_{tp})^2 (1 + |\lambda| v_{SD}) \quad (4.28)$$

Table 5.2 Regions of Operation of the Enhancement PMOS Transistor



- $v_{SG} < |V_{tp}|$: no channel; transistor in cutoff; $i_D = 0$
- $v_{SG} = |V_{tp}| + |v_{OV}|$: a channel is induced; transistor operates in the triode region or in the saturation region depending on whether the channel is continuous or pinched off at the drain end;



Continuous channel, obtained by:

$$v_{DG} > |V_{tp}|$$

or equivalently

$$v_{SD} < |v_{OV}|$$

$$\text{Then } k'_p \left(\frac{W}{L}\right) \left[(v_{SG} - |V_{tp}|) v_{SD} - \frac{1}{2} v_{SD}^2 \right]$$

or equivalently

$$i_D = k'_p \left(\frac{W}{L}\right) \left(|v_{OV}| - \frac{1}{2} v_{SD} \right) v_{SD}$$

Pinched-off channel, obtained by:

$$v_{DG} \leq |V_{tp}|$$

or equivalently

$$v_{SD} \geq |v_{OV}|$$

Then

$$\text{or equivalently } i_D = \frac{1}{2} k'_p \left(\frac{W}{L}\right) (v_{SG} - |V_{tp}|)^2$$

$$i_D = \frac{1}{2} k'_p \left(\frac{W}{L}\right) v_{OV}^2$$

or equivalently

$$i_D = \frac{1}{2} k_p \left(\frac{W}{L}\right) v_{SG} - \cdot V_{tp} \cdot^2 \frac{1 + \frac{v_{SD}}{|V_A|}}{|V_A|} \quad (4.29)$$

where λ and V_A (the Early voltage for the PMOS transistor) are by convention negative quantities, hence we use $|\lambda|$ and $|V_A|$.

Finally, we should note that for a given CMOS fabrication process λ_n and λ_p are generally not equal, and similarly for V_{An} and $\cdot V_{Ap} \cdot$.

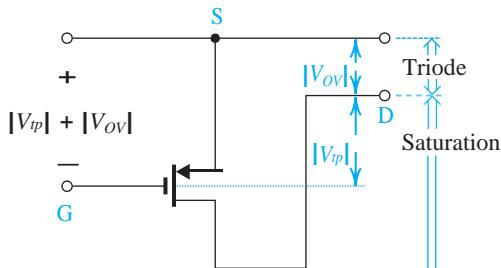


Figure 4.20 The relative levels of the terminal voltages of the enhancement-type PMOS transistor for operation in the triode region and in the saturation region.

EXERCISE

5.7 The PMOS transistor shown in Fig. E5.7 has $V_{tp} = -1$ V, $k^r = 60 \mu\text{A/V}_2$, and $W/L = 10$.

- Find the range of V_G for which the transistor conducts.
- In terms of V_G , find the range of V_D for which the transistor operates in the triode region.
- In terms of V_G , find the range of V_D for which the transistor operates in saturation.
- Neglecting channel-length modulation (i.e., assuming $\lambda = 0$), find the values of V_{ov} and V_G and the corresponding range of V_D to operate the transistor in the saturation mode with $I_D = 75 \mu\text{A}$.
- If $\lambda = -0.02 \text{ V}^{-1}$, find the value of r_o corresponding to the overdrive voltage determined in (d).
- For $\lambda = -0.02 \text{ V}^{-1}$ and for the value of V_{ov} determined in (d), find I_D at $V_D = +3$ V and at $V_D = 0$ V.

$V_D = 0$ V; hence, calculate the value of the apparent output resistance in saturation. Compare to the value found in (e).

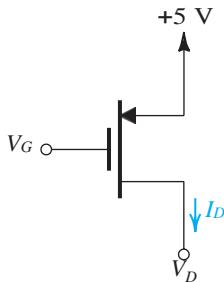


Figure E5.7

Ans. (a) $V_G \leq +4$ V; (b) $V_D \geq V_G + 1$; (c) $V_D \leq V_G + 1$; (d) 0.5 V, 3.5 V, ≤ 4.5 V; (e) 0.67 M Δ ; (f) 78 μA , 82.5 μA , 0.67 M Δ (same)

4.3 MOSFET Circuits at DC

Having studied the current–voltage characteristics of MOSFETs, we now consider circuits in which only dc voltages and currents are of concern. Specifically, we shall present a series of design and analysis examples of MOSFET circuits at dc. The objective is to instill in the reader a familiarity with the device and the ability to perform MOSFET circuit analysis both rapidly and effectively.

In the following examples, to keep matters simple and thus focus attention on the essence of MOSFET circuit operation, we will generally neglect channel-length modulation; that is, we will assume $\lambda = 0$. We will find it convenient to work in terms of the overdrive voltage;

$$V_{OV} = V_{GS} - V_{tn} \text{ for NMOS and } |V_{OV}| = V_{SG} - V_{tp} \text{ for PMOS.}$$

Example 5.3

Design the circuit of Fig. 5.21: that is, determine the values of R_D and R_S so that the transistor operates at $I_D = 0.4$ mA and $V_D = +0.5$ V. The NMOS transistor has $V_t = 0.7$ V, $\mu_n C_{ox} = 100 \mu\text{A/V}^2$, $L = 1 \mu\text{m}$, and $W = 32 \mu\text{m}$. Neglect the channel-length modulation effect (i.e., assume that $\lambda = 0$).

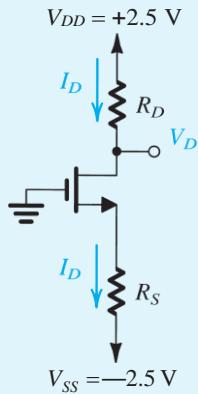


Figure 5.21 Circuit for Example 5.3.

Solution

To establish a dc voltage of +0.5 V at the drain, we must select R_D as follows:

$$R_D = \frac{V_{DD} - V_D}{I_D} = \frac{2.5 - 0.5}{0.4} = 5 \text{ k}\Omega$$

To determine the value required for R_S , we need to know the voltage at the source, which can be easily found if we know V_{GS} . This in turn can be determined from V_{OV} . Toward that end, we note that since $V_D = 0.5$ V is greater than V_G , the NMOS transistor is operating in the saturation region, and we can use the saturation-region expression of i_D to determine the required value of V_{OV} ,

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_G - V_{OV})^2$$

Then substituting $I_D = 0.4$ mA = 400 μA , $\mu_n C_{ox} = 100 \mu\text{A/V}^2$ and $W/L = 32/1$ gives

$$400 = \frac{1}{2} \times 100 \times \frac{32}{1} (V_G - V_{OV})^2$$

Example 4.3 *continued*

which results in

$$V_{OV} = 0.5 \text{ V}$$

Thus,

$$V_{GS} = V_t + V_{OV} = 0.7 + 0.5 = 1.2 \text{ V}$$

Referring to Fig. 4.21, we note that the gate is at ground potential. Thus, the source must be at -1.2 V , and the required value of R_S can be determined from

$$\begin{aligned} R_S &= \frac{V_S - V_{SS}}{I_D} \\ &= \frac{-1.2 - (-2.5)}{0.4} = 3.25 \text{ k}\Delta \end{aligned}$$

Example 4.4

Figure 5.22 shows an NMOS transistor with its drain and gate terminals connected together. Find the $i-v$ relationship of the resulting two-terminal device in terms of the MOSFET parameters $k_n = k_f W/L$ and V_m . Neglect channel-length modulation (i.e., $\lambda = 0$). Note that this two-terminal device is known as a **diode-connected transistor**.

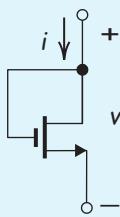


Figure 4.22

Solution

Since $v_D = v_G$ implies operation in the saturation mode,

$$i_D = k \frac{1}{2} n^r \frac{W}{L} (v_{GS} - V_m)^2$$

Now, $i = i_D$ and $v = v_{GS}$, thus

$$i = \frac{1}{2} k n^r \frac{W}{L} (v - V_m)^2$$

Replacing $k n^r \frac{W}{L}$ by k_n results in

$$i = \frac{1}{2} k_n v - V_m^2$$

Example 4.5

Design the circuit in Fig. 5.23 to establish a drain voltage of 0.1 V. What is the effective resistance drain and source at this operating point? Let $V_t = 1$ V and $k' W/L = 1 \text{ mA/V}^2$

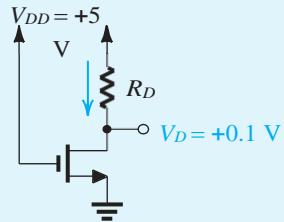


Figure 4.23 Circuit for Example

Solution

Since the drain voltage is lower than the gate voltage by 4.9 V and $V_m = 1$ V, the MOSFET is in the triode region. Thus the current I_D is given

$$I_D = k' \frac{W}{L} (V_{GS} - V_m)^2$$
$$I_D = 1 \times (5 - 1) \times 0.1 - \frac{1}{2} \times$$
$$= 0.395$$

The required value for R_D can be found as follows:

$$R_D = \frac{V_{D\bar{S}} - V_D}{I_D}$$
$$= \frac{5 - 0.1}{0.395} = 12.4 \text{ k}\Delta$$

In a practical discrete-circuit design problem, one selects the closest standard value available for, say, 5% resistors—in this case, 12 k Δ ; see Appendix J. Since the transistor is operating in the triode region with a small V_{DS} , the effective drain-to-source resistance can be determined as follows:

$$r_{DS} = \frac{V_{DS}}{I_D}$$
$$= \frac{0.1}{0.395} = 253 \Delta$$

Alternatively, we can determine r_{DS} by using the formula

$$r_{DS} = \frac{1}{k_n b_V}$$

to obtain

$$r_{DS} = \frac{1}{1 \times (5 - 1)} = 0.25 \text{ k}\Delta = 250 \Delta$$

which is close to the value found above.

Example 4.6

Analyze the circuit shown in Fig. 5.24(a) to determine the voltages at all nodes and the currents through all branches. Let $V_m = 1$ V and $k (W/L) = 1 \text{ mA/V}^2$. Neglect the channel-length modulation effect (i.e., assume $\lambda = 0$).

Example 4.6 *continued*

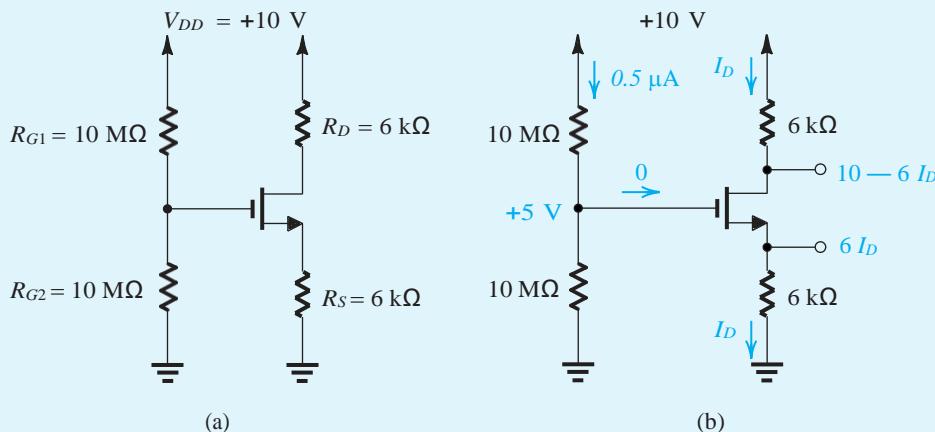


Figure 4.24 (a) Circuit for Example 5.6. (b) The circuit with some of the analysis details shown.

Solution

Since the gate current is zero, the voltage at the gate is simply determined by the voltage divider formed by the two 10-M Δ resistors,

$$V = V \frac{R_{G2}}{\frac{R}{DD_{C2}} + R_{G1}} = 10 \times \frac{10}{10 + 10} = +5V$$

With this positive voltage at the gate, the NMOS transistor will be turned on. We do not know, however, whether the transistor will be operating in the saturation region or in the triode region. We shall assume saturation-region operation, solve the problem, and then check the validity of our assumption. Obviously, if our assumption turns out not to be valid, we will have to solve the problem again for triode-region operation.

Refer to Fig. 5.24(b). Since the voltage at the gate is 5 V and the voltage at the source is I_D (mA) \times 6 (Δ) = $6I_D$ (V), we have

$$V_{GS} = 5 - 6I_D$$

Thus I_D is given by

$$I_D = \frac{1}{2} k_r \frac{W}{L} V_{GS} - V_m$$

$$= 2 \times 1 \times 5 - 6 I_D - 1_2$$

which results in the following quadratic equation in I_D :

$$18I_D^2 - 25I_D + 8 = 0$$

This equation yields two values for I_D : 0.89 mA and 0.5 mA. The first value results in a source voltage of $6 \times 0.89 = 5.34$ V, which is greater than the gate voltage and does not make physical sense as it would imply that the NMOS transistor is cut off. Thus,

$$I_D = 0.5 \text{ mA}$$

$$V_S = 0.5 \times 6 = +3\text{V}$$

$$V_{GS} = 5 - 3 = 2\text{V}$$

$$V_D = 10 - 6 \times 0.5 = +7\text{V}$$

Since $V_D > V_G - V_m$, the transistor is operating in saturation, as initially assumed.

Example 4.7

Design the circuit of Fig. 4.25 so that the transistor operates in saturation with $I_D = 0.5$ mA and $V_D = +3$ V. Let the PMOS transistor have $V_{T_P} = -1$ V and $k' (W/L) = 1$ mA/V.² Assume $\lambda = 0$. What is the largest value that R_D can have while maintaining saturation-region operation?

Example 4.7 continued

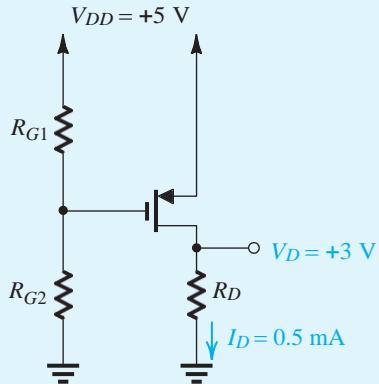


Figure 4.25 Circuit for Example 5.7.

Solution

Since the MOSFET is to be in saturation, we can write

$$I_D = \frac{1}{2} k_p \frac{W}{L} \cdot V_{ov}^2$$

Substituting $I_D = 0.5 \text{ mA}$ and $k_p W/L = 1 \text{ mA/V}^2$, we obtain

$$V_{ov} = 1 \text{ V}$$

and

$$V_{SG} = V_{tp} + V_{ov} = 1 + 1 = 2 \text{ V}$$

Since the source is at +5 V, the gate voltage must be set to +3 V. This can be achieved by the appropriate selection of the values of R_{G1} and R_{G2} . A possible selection is $R_{G1} = 2 \text{ M}\Omega$ and $R_{G2} = 3 \text{ M}\Omega$.

The value of R_D can be found from

$$R_D = \frac{V_D}{I_D} = \frac{3}{0.5} = 6 \text{ k}\Omega$$

Saturation-mode operation will be maintained up to the point that V_D exceeds V_G by V_{tp} ; that is, until

$$V_{D_{max}} = 3 + 1 = 4 \text{ V}$$

This value of drain voltage is obtained with R_D given by

$$R_D = \frac{4}{0.5} = 8 \text{ k}\Omega$$

Example 4.8

The NMOS and PMOS transistors in the circuit of Fig. 5.26(a) are matched, with $k_n^r W_n/L_n = k_p^r W_p/L_p = 1 \text{ mA/V}^2$ and $V_{m^-} = V_{l_p} = 1.5 \text{ V}$. Assuming $\lambda = 0$ for both devices, find the drain currents i_{DN} and i_{DP} , as well as the voltage v_o , for $v_I = 0 \text{ V}$, $+2.5 \text{ V}$, and -2.5 V .

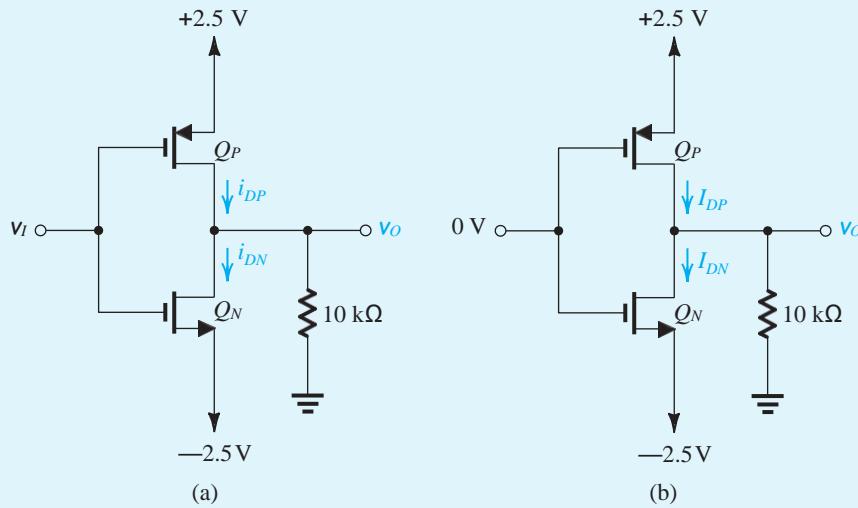


Figure 4.26 Circuits for Example 4.8.

Example 4.8 continued

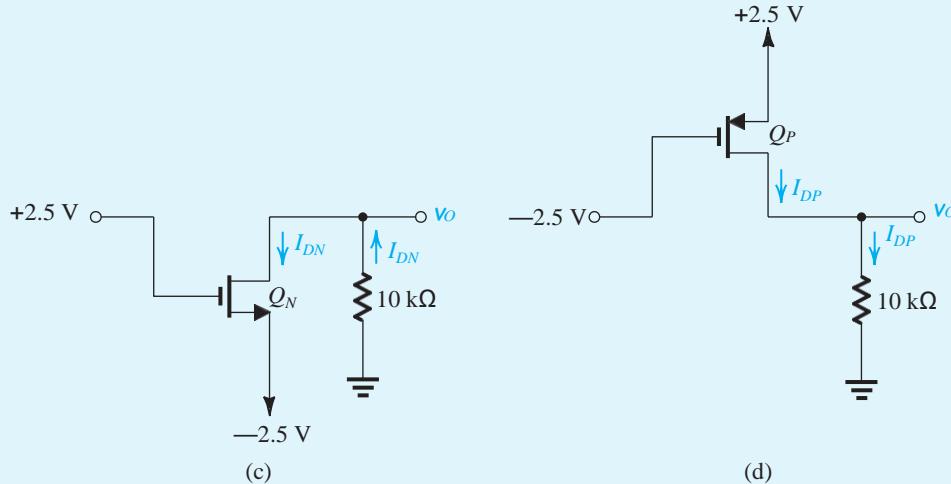


Figure 4.26 continued

Solution

Figure 4.26(b) shows the circuit for the case $v_I = 0$ V. We note that since Q and Q' are perfectly matched and are operating at equal values of $|V_{GS}| = 2.5$ V, the circuit is symmetrical, which dictates that $v_O = 0$ V. Thus both Q_N and Q_P are operating with $V_{DG} = 0$ and, hence, in saturation. The drain currents can now be found from

$$I_{DP} = I_{DN} = \frac{1}{2} \times 1 \times (2.5 - 1)^2 = 1.125 \text{ mA}$$

Next, we consider the circuit with $v_I = +2.5$ V. Transistor Q_P will have a V_{SG} of zero and thus will be cut off, reducing the circuit to that shown in Fig. 5.26(c). We note that v_O will be negative, and thus v_{GD} will be greater than V_m , causing Q_N to operate in the triode region. For simplicity we shall assume that v_{DS} is small and thus use

$$V_{DS} = \frac{1}{2} [V_D - V_{GS}] = \frac{1}{2} [v_O - (-2.5)]$$

From the circuit diagram shown in Fig. 5.26(c), we can also write

$$I_{DN} (\text{mA}) = \frac{0 - v_O}{10(\text{k}\Delta)}$$

These two equations can be solved simultaneously to yield

$$I_{DN} = 0.244 \text{ mA} \quad v_O = -2.44 \text{ V}$$

Note that $V_{DS} = -2.44 - (-2.5) = 0.06 \text{ V}$, which is small as assumed.

Finally, the situation for the case $v_I = -2.5 \text{ V}$ [Fig. 5.26(d)] will be the exact complement of the case $v_I = +2.5 \text{ V}$: Transistor Q_N will be off. Thus $I_{DN} = 0$, Q_P will be operating in the triode region with $I_{DP} = 0.244 \text{ mA}$ and $v_O = +2.44 \text{ V}$.

EXERCISE

- 4.15** The NMOS and PMOS transistors in the circuit of Fig. E5.15 are matched with $k^r W/L = n \text{ mA/V}^2$ and $V_m = \frac{1}{2} V_p = 1 \text{ V}$. Assuming $\lambda = 0$ for both devices, find the drain currents i_{DN} and i_{DP} and the voltage v_O for $v_I = 0 \text{ V}$, $+2.5 \text{ V}$, and -2.5 V .

Ans. $v_I = 0 \text{ V}$: 0 mA, 0 mA, 0 V; $v_I = +2.5 \text{ V}$: 0.104 mA, 0 mA, 1.04 V; $v_I = -2.5 \text{ V}$: 0 mA, 0.104 mA, -1.04 V

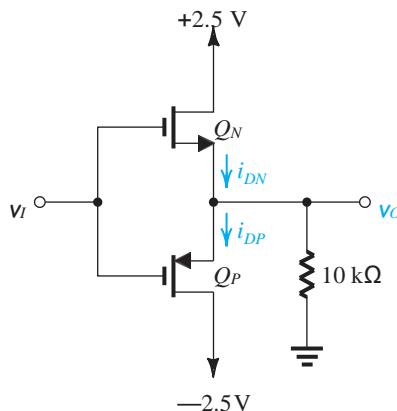


Figure E4.15

Concluding Remark If a MOSFET is conducting but its mode of operation (saturation or triode) is not known, we assume operation in the saturation region, solve the problem, and check whether the conditions for saturation-mode operation are satisfied. If not, then the MOSFET is operating in the triode region and the analysis is done accordingly.

GORDON MOORE— HIS LAW:

A half-century ago, Gordon Moore, who would go on to become a cofounder first of Fairchild Semiconductor and then of Intel, presented a startling idea in the issue of *Electronics Magazine* dated April 19, 1965. Moore, who had a doctorate in chemistry, had projected the potential growth of the integrated-circuit industry based on five points spanning a seven-year period from 1958 to 1965. The conclusion he reached—that the number of transistors per chip had been increasing and would continue to increase by a factor of 2 every two years or so—was destined to propel progress in integrated circuits over the succeeding decades into the twenty-first century. Doubling of the number of transistors was predicted on the basis of another prediction: the continuing shrinkage of transistor dimensions. In early recognition of the importance of this prediction, Carver Mead, a pioneer in very large scale integration (VLSI), soon began to refer to this prediction as “Moore’s law.” (See Chapter 15, Section 15.1, for the implications of Moore’s law).

4.4 The Body Effect and Other Topics

In this section we briefly consider a number of important though secondary issues.

4.4.1 The Role of the Substrate—The Body Effect

In many applications the source terminal is connected to the substrate (or body) terminal B, which results in the *pn* junction between the substrate and the induced channel (review Fig. 5.5) having a constant zero (cutoff) bias. In such a case the substrate does not play any role in circuit operation and its existence can be ignored altogether.

In integrated circuits, however, the substrate is usually common to many MOS transistors. In order to maintain the cutoff condition for all the substrate-to-channel junctions, the substrate is usually connected to the most negative power supply in an NMOS circuit (the most positive in a PMOS circuit). The resulting reverse-bias voltage between source and body (V_{SB} in an *n*-channel device) will have an effect on device operation. To appreciate this fact, consider an NMOS transistor and let its substrate be made negative relative to the source. The reverse-bias voltage will widen the depletion region (refer to Fig. 4.2). This in turn reduces the channel depth. To return the channel to its former state, v_{GS} has to be increased.

The effect of V_{SB} on the channel can be most conveniently represented as a change in the threshold voltage V_t . Specifically, it has been shown that increasing the reverse substrate bias voltage V_{SB} results in an increase in V_t according to the relationship

$$V_t = V_{t0} + \gamma \sqrt{\frac{2\varphi_f + V_{SB}}{2\varphi_f}} \quad (4.30)$$

where V_{t0} is the threshold voltage for $V_{SB} = 0$; φ_f is a physical parameter with $(2\varphi_f)$ typically 0.6 V; γ is a fabrication-process parameter given by

$$\gamma = \frac{2qN_Ae}{C_{ox}} \quad (4.31)$$

where q is the magnitude of the electron charge (1.6×10^{-19} C), N_A is the doping concentration of the p -type substrate, and e is the permittivity of silicon ($11.7\epsilon_0 = 11.7 \times 8.854 \times 10^{-14} = 1.04$ F/cm). The parameter γ has the dimension of $\frac{V}{V}$ and is typically 0.4 V. Finally, note that Eq. (5.30) applies equally well for p -channel devices with V_{SB} replaced by

the reverse bias of the substrate, V_{BS} (or, alternatively, replace $|V_{SB}|$ by V_{SB}) and note that γ is negative. Also, in evaluating γ , N_A must be replaced with N_D , the doping concentration of the n well in which the PMOS is formed. For p -channel devices, $2\varphi_f$ is typically 0.75 V, and γ is typically $-0.5 \text{ V}^{1/2}$.

Equation (4.30) indicates that an incremental change in V_{SB} gives rise to an incremental change in V_t , which in turn results in an incremental change in i_D even though v_{GS} might have been kept constant. It follows that the body voltage controls i_D ; thus the body acts as another gate for the MOSFET, a phenomenon known as the **body effect**. Here we note that the parameter γ is known as the **body-parameter**

4.4.1 Temperature Effects

Both V_t and k^r are temperature sensitive. The magnitude of V_t decreases by about 2 mV for every 1°C rise in temperature. This decrease in V_t gives rise to a corresponding increase in drain current as temperature is increased. However, because k^r decreases with temperature and its effect is a dominant one, the overall observed effect of a temperature increase is a *decrease* in drain current. This very interesting result is put to use in applying the MOSFET in power circuits (Chapter 12).

4.4.1 Breakdown and Input Protection

As the voltage on the drain is increased, a value is reached at which the *pn* junction between the drain region and substrate suffers avalanche breakdown (see Section 3.5.3). This breakdown usually occurs at voltages of 20 V to 150 V and results in a somewhat rapid increase in current (known as a **weak avalanche**).

Another breakdown effect that occurs at lower voltages (about 20 V) in modern devices is called **punch-through**. It occurs in devices with relatively short channels when the drain voltage is increased to the point that the depletion region surrounding the drain region extends through the channel to the source. The drain current then increases rapidly. Normally, punch-through does not result in permanent damage to the device.

Yet another kind of breakdown occurs when the gate-to-source voltage exceeds about 30 V. This is the breakdown of the gate oxide and results in permanent damage to the device. Although 30 V may seem high, it must be remembered that the MOSFET has a very high input resistance and a very small input capacitance, and thus small amounts of static charge accumulating on the gate capacitor can cause its breakdown voltage to be exceeded.

To prevent the accumulation of static charge on the gate capacitor of a MOSFET, gate-protection devices are usually included at the input terminals of MOS integrated circuits. The protection mechanism invariably makes use of clamping diodes.

4.4.1 Velocity Saturation

At high longitudinal electric fields, the drift velocity of charge carriers in the channel reaches an upper limit (approximately 10^7 cm/s for electrons and holes in silicon). This effect, whichin modern very-short-channel devices can occur for v_{DS} lower than 1 V, is called velocity saturation. It can be shown that when velocity saturation occurs, the current i_D will no longer be related to v_{GS} by the square-law relationship. Rather, i_D becomes linearly dependent on

v_{GS} and the transconductance g_m becomes constant and independent of v_{GS} . In Chapter 15, we shall consider velocity saturation in our study of deep-submicron (i.e., $L < 0.25 \mu\text{m}$) CMOS digital circuits.

4.4.1 The Depletion-Type MOSFET

We conclude this section with a brief discussion of another type of MOSFET, the depletion-type MOSFET. Its structure is similar to that of the enhancement-type MOSFET with one important difference: The depletion MOSFET has a physically implanted channel. Thus an n -channel depletion-type MOSFET has an n -type silicon region connecting the n^+ source and the n^+ drain regions at the top of the p -type substrate. Thus if a voltage v_{DS} is applied between drain and source, a current i_D flows for $v_{GS} > 0$. In other words, there is no need to induce a channel, unlike the case of the enhancement MOSFET.

The channel depth and hence its conductivity can be controlled by v_{GS} in exactly the samemanner as in the enhancement-type device. Applying a positive v_{GS} enhances the channel by attracting more electrons into it. Here, however, we also can apply a negative v_{GS} , whichcauses electrons to be repelled from the channel, and thus the channel becomes shallower andits conductivity decreases. The negative v_{GS} is said to **deplete** the channel of its charge carriers,

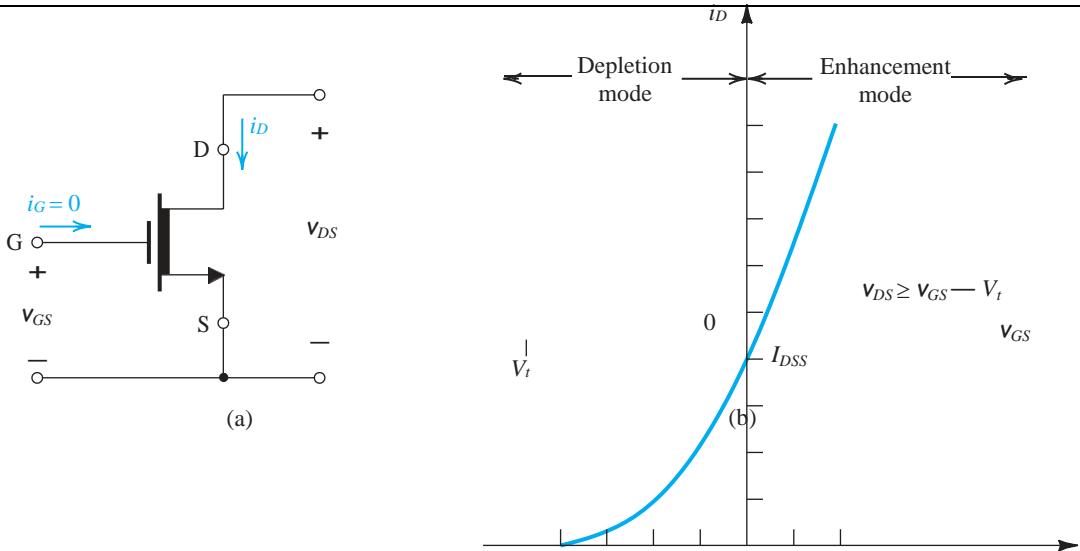


Figure 4.27 The circuit symbol (a) and the i_D-v_{GS} characteristic in saturation (b) for an n-channel depletion-type MOSFET.

and this mode of operation (negative v_{GS}) is called **depletion mode**. As the magnitude of v_{GS} is increased in the negative direction, a value is reached at which the channel is completely depleted of charge carriers and i_D is reduced to zero even though v_{DS} may be still applied. This negative value of v_{GS} is the threshold voltage of the n-channel depletion-type MOSFET.

The description above suggests (correctly) that a depletion-type MOSFET can be operated in the enhancement mode by applying a positive v_{GS} and in the depletion mode by applying a negative v_{GS} . This is illustrated in Fig. 4.27, which shows both the circuit symbol for the depletion NMOS transistor (Fig. 4.27a) and its i_D-v_{GS} characteristic. Observe that here the threshold voltage V_t is negative. The i_D-v_{DS} characteristics (not shown) are similar to those for the enhancement-type MOSFET except for the negative V_t . Finally, note that the device symbol denotes the existing channel via the shaded area next to the vertical line.

Depletion-type MOSFETs can be fabricated on the same IC chip as enhancement-type devices, resulting in circuits with improved characteristics, as will be shown in a later chapter. The depletion-type MOSFET, however, is a specialty device and is not commonly used.

4.5 The Voltage-Transfer Characteristic (VTC)

A useful tool that provides insight into the operation of an amplifier circuit is its voltage-transfer characteristic (VTC). This is simply a plot (or a clearly labeled sketch) of the output voltage versus the input voltage. For the MOS amplifier in Fig. 4.2(a), this is the plot of v_{DS} versus v_{GS} shown in Fig. 4.2(b).

Observe that for $v_{GS} < V_t$, the transistor is cut off, $i_D = 0$ and, from Eq. (4.3), $v_{DS} = V_{DD}$. As v_{GS} exceeds V_t , the transistor turns on and v_{DS} decreases. However, since initially v_{DS} is still high, the MOSFET will be operating in saturation or the active region. This continues as v_{GS} is increased until the value of v_{GS} is reached that results in v_{DS} becoming lower than v_{GS} by V_t volts [point B on the VTC in Fig. 4.2(b)]. For v_{GS} greater than that at point B, the transistor operates in the triode region and v_{DS} decreases more slowly.

The VTC in Fig. 4.2(b) indicates that the segment of greatest slope (hence potentially the largest amplifier gain) is that labeled AB, which corresponds to operation in the active region. When a MOSFET is operated as an amplifier, its operating point is confined to the segment AB at all times. An expression for the segment AB can be obtained by substituting for i_D in Eq. (4.3) by its active-region value from Eq. (4.1), thus

$$v_{DS} = V_{DD} - \frac{1}{2} k_n R_D (v_{GS} - V_t) \quad (4.5)$$

This is obviously a nonlinear relationship. Nevertheless, linear (or almost-linear) amplification can be obtained by using the technique of biasing the MOSFET. Before considering biasing, however, it is useful to determine the coordinates of point B, which is at the boundary between the saturation and the triode regions of operation. These can be obtained by substituting in

Eq. (4.5), $v_{GS} = V_{GS,B}$ and $v_{DS} = V_{DS,B} = V_{GS,B} - V_t$. The result is

$$V_{GS,B} = V_t + \frac{2k_n R_D V_{DD} + 1 - 1}{k_n R_D} \quad (4.5)$$

Point B can be alternatively characterized by the overdrive voltage

$$V_{OV_B} \equiv V_{GS_B} - V_t = \frac{\sqrt{2k_n R_D V_{DD} + 1} - 1}{k_n R_D} \quad (4.7)$$

and

$$V_{DS_B} = V_{OV_B} \quad (4.8)$$

An exactly similar development applies to the BJT case. This is illustrated in Fig. 4.2(c) and (d). In this case, over the active-region or amplifier segment AB, the output voltage v_{CE} is related to the input voltage v_{BE} by

$$v_{CE} = V_{CC} - R_E I e^{v_{BE}/V_T} \quad (4.9)$$

Here also, the input–output relationship is nonlinear. Nevertheless, linear (or almost-linear) amplification can be obtained by using the biasing technique discussed next.

4.5.1 Obtaining Linear Amplification by Biasing the Transistor

Biasing enables us to obtain almost-linear amplification from the MOSFET and the BJT. The technique is illustrated for the MOSFET case in Fig. 4.3(a). A dc voltage V_{GS} is selected to obtain operation at a point Q on the segment AB of the VTC. How to select an appropriate location for the bias point Q will be discussed shortly. For the time being, observe that the coordinates of Q are the dc voltages V_{GS} and V_{DS} , which are related by

$$V_{DS} = V_{DD} - k_n R_D (V_{GS} - V_t) \quad (7.10)$$

Point Q is known as the **bias point** or the **dc operating point**. Also, since at Q no signal component is present, it is also known as the **quiescent point** (which is the origin of the symbol Q).

Next, the signal to be amplified, v_{gs} , a function of time t , is superimposed on the bias voltage V_{GS} , as shown in Fig. 4.4(a). Thus the total instantaneous value of v_{GS} becomes

$$v_{GS}(t) = V_{GS} + v_{gs}(t)$$

The resulting $v_{DS}(t)$ can be obtained by substituting for $v_{GS}(t)$ into Eq. (4.5). Graphically, we can use the VTC to obtain $v_{DS}(t)$ point by point, as illustrated in Fig. 4.4(b). Here we show

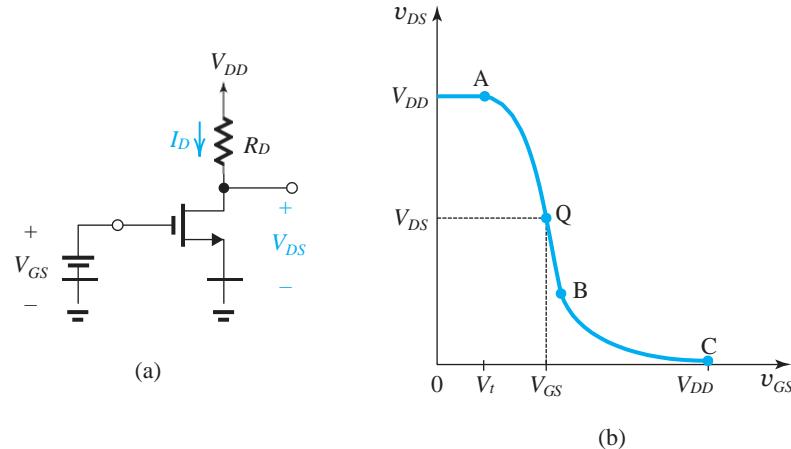


Figure 4.3 Biasing the MOSFET amplifier at a point Q located on the segment AB of the VTC.

the case of v_{gs} being a triangular wave of “small” amplitude. Specifically, the amplitude of v_{gs} is small enough to restrict the excursion of the instantaneous operating point to a short, almost-linear segment of the VTC around the bias point Q. The shorter the segment, the greater the linearity achieved, and the closer to an ideal triangular wave the signal component at the output, v_{ds} , will be. This is the essence of obtaining linear amplification from the nonlinear MOSFET.

Before leaving Fig. 4.4(b) we wish to draw the reader’s attention to the consequence of increasing the amplitude of the signal v_{gs} . As the instantaneous operating point will no longer be confined to the almost-linear segment of the VTC, the output signal v_{ds} will deviate from its ideal triangular shape; that is, it will exhibit nonlinear distortion. Worse yet, if the input signal amplitude becomes sufficiently large, the instantaneous operating point may leave the segment AB altogether. If this happens at the negative peaks of v_{gs} , the transistor will cut off for a portion of the cycle and the positive peaks of v_{ds} will be “clipped off.” If it occurs at the positive peaks of v_{gs} , the transistor will enter the triode region for a portion of the cycle, and the negative peaks of v_{ds} will become flattened. It follows that the selection of the location of the bias point Q can have a profound effect on the maximum allowable amplitude of v_{ds} , referred to as the *allowable signal swing at the output*. We will have more to say later on this important point.

An exactly parallel development can be applied to the BJT amplifier. In fact, all we need to do is replace the NMOS transistor in Figs. 4.3 and 4.4 with an *npn* transistor and change the voltage and current symbols to their BJT counterparts. The resulting bias point Q will be characterized by dc voltages V_{BE} and V_{CE} , which are related by

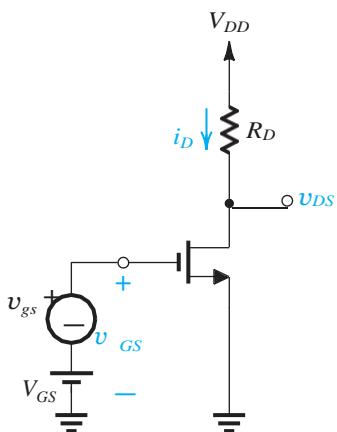
$$V_{CE} = V_{CC} - R_C I_S e^{V_{BE}/V_T} \quad (4.11)$$

and a dc current I_C ,

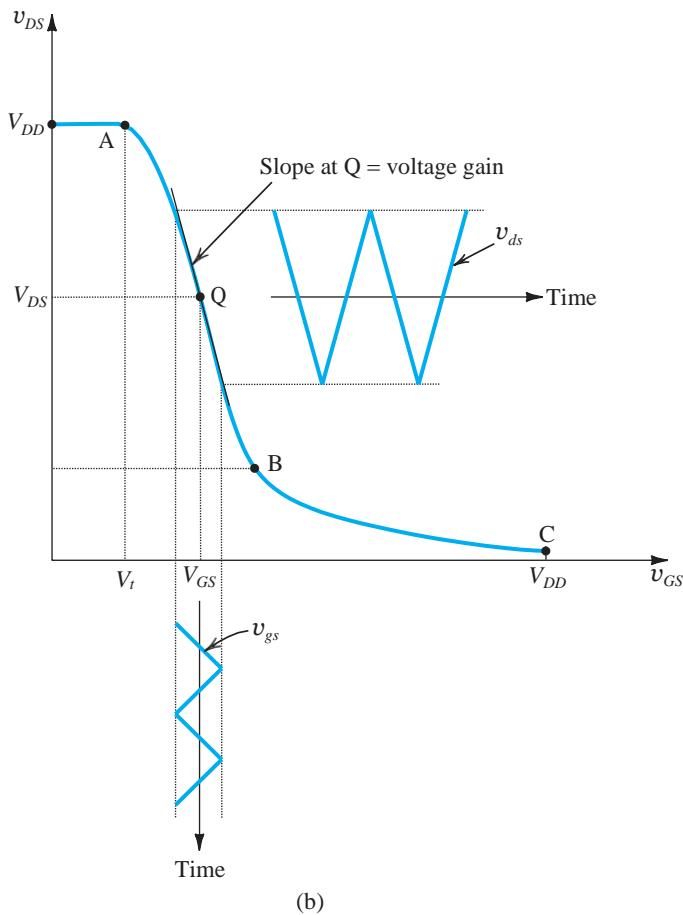
$$I_C = I_S e^{V_{BE}/V_T} \quad (4.12)$$

Also, superimposing a small-signal v_{be} on the dc bias voltage V_{BE} results in

$$v_{BE}(t) = V_{BE} + v_{be}(t)$$



(a)



(b)

Figure 4.4 The MOSFET amplifier with a small time-varying signal $v_{gs}(t)$ superimposed on the dc bias voltage V_{GS} . The MOSFET operates on a short almost-linear segment of the VTC around the bias point Q and provides an output voltage $v_{ds} = A_v v_{gs}$.

which can be substituted into Eq. (7.9) to obtain the total instantaneous value of the output voltage $v_{CE}(t)$. Here again, almost-linear operation is obtained by keeping v_{be} small enough to restrict the excursion of the instantaneous operating point to a short, almost-linear segment of the VTC around the bias point Q. Similar comments also apply to the maximum allowable signal swing at the output.

4.5.2 The Small-Signal Voltage Gain

The MOSFET Case Consider the MOSFET amplifier in Fig. 4.4(a). If the input signal v_{gs} is kept small, the corresponding signal at the output v_{ds} will be nearly proportional to v_{gs} with the constant of proportionality being the slope of the almost-linear segment of the VTC around Q. This is the voltage gain of the amplifier, and its value can be determined by evaluating the slope of the tangent to the VTC at the bias point Q,

$$A = \frac{dv_{DS}}{dv_{GS}} \Big|_{v_{GS}=V_{GS}} \quad (4.13)$$

Utilizing Eq. (7.5) we obtain

$$A_v = -k_n(V_{GS} - V_t)R_D \quad (4.14)$$

which can be expressed in terms of the overdrive voltage at the bias point, V_{OV} , as

$$A_v = -k_n V_{OV} R_D \quad (4.15)$$

We make the following observations on this expression for the voltage gain.

1. The gain is negative, which signifies that the amplifier is inverting; that is, there is a 180° phase shift between the input and the output. This inversion is obvious in Fig. 4.4(b) and should have been anticipated from Eq. (4.5).
2. The gain is proportional to the load resistance R_D , to the transistor transconductance parameter k_n , and to the overdrive voltage V_{OV} . This all makes intuitive sense.

Another simple and insightful expression for the voltage gain A_v can be derived by recalling that the dc current in the drain at the bias point is related to V_{OV} by

$$I_D = \frac{1}{2} k_n V_{OV}^2$$

This equation can be combined with Eq. (4.15) to obtain

$$A_v = -\frac{I_D R_D}{V_{OV}/2} \quad (4.16)$$

That is, the gain is simply the ratio of the dc voltage drop across the load resistance R_D to $V_{OV}/2$. It can be expressed in the alternative form

$$A_v = -\frac{V_{DD} - V_{DS}}{V_{OV}/2} \quad (7.17)$$

Since the maximum slope of the VTC in Fig. 4.4(b) occurs at point B, the maximum gain magnitude $|A_{vmax}|$ is obtained by biasing the transistor at point B,

$$|A_{vmax}| = \frac{V_{DD} - V_{DS_B}}{V_{OV_B}/2}$$

Example 4.1

Consider the amplifier circuit shown in Fig. 4.4(a). The transistor is specified to have $V_t = 0.4$ V, $k_n^t = 0.4$ mA/V², $W/L = 10$, and $\lambda = 0$. Also, let $V_{DD} = 1.8$ V, $R_D = 17.5$ k Ω , and $V_{GS} = 0.6$ V.

- (a) For $v_{gs} = 0$ (and hence $v_{ds} = 0$), find V_{OV} , I_D , V_{DS} , and A_v .
- (b) What is the maximum symmetrical signal swing allowed at the drain? Hence, find the maximum allowable amplitude of a sinusoidal v_{gs} .

Solution

- (a) With $V_{GS} = 0.6$ V, $V_{OV} = 0.6 - 0.4 = 0.2$ V. Thus,

$$\begin{aligned} I_D &= \frac{1}{2} k_n^t \frac{W}{L} \frac{V_{OV}^2}{V_t^2} \\ &= \frac{1}{2} \times 0.4 \times 10 \times 0.2 = 0.08 \text{ mA} \end{aligned}$$

$$\begin{aligned} V_{DS} &= V_{DD} - R_D I_D \\ &= 1.8 - 17.5 \times 0.08 = 0.4 \text{ V} \end{aligned}$$

Since V_{DS} is greater than V_{OV} , the transistor is indeed operating in saturation. The voltage gain can be found from Eq. (7.15),

$$\begin{aligned} A_v &= -k_n V_{OV} R_D \\ &= -0.4 \times 10 \times 0.2 \times 17.5 \\ &= -14 \text{ V/V} \end{aligned}$$

An identical result can be found using Eq. (7.17).

- (b) Since $V_{OV} = 0.2$ V and $V_{DS} = 0.4$ V, we see that the maximum allowable negative signal swing at the drain is 0.2 V. In the positive direction, a swing of +0.2 V would not cause the transistor to

Example 4.1 continued

cut off (since the resulting v_{DS} would be still lower than V_{DD}) and thus is allowed. Thus the maximum symmetrical signal swing allowable at the drain is ± 0.2 V. The corresponding amplitude of v_{gs} can be found from

$$v_{gs} = \frac{\hat{v}_{ds}}{|A_v|} = \frac{0.2 \text{ V}}{14} = 14.2 \text{ mV}$$

Since $\hat{v}_{gs} < V_{OV}$, the operation will be reasonably linear (more on this in later sections).

Greater insight into the issue of allowable signal swing can be obtained by examining the signal waveforms shown in Fig. 4.5. Note that for the MOSFET to remain in saturation at the negative peak of v_{ds} , we must ensure that

$$v_{DS\min} \geq v_{GS\max} - V_t$$

that is,

$$0.4 - |A_v| \hat{v}_{gs} \geq 0.6 + \hat{v}_{gs} - 0.4$$

which results in

$$\hat{v}_{gs} \leq \frac{0.2}{|A_v| + 1} = 13.3 \text{ mV}$$

This result differs slightly from the one obtained earlier.

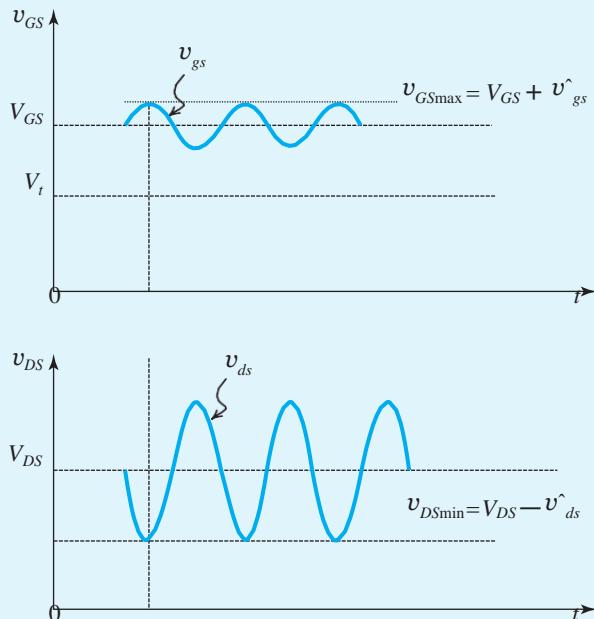


Figure 4.5 Signal waveforms at gate and drain for the amplifier in Example 7.1. Note that to ensure operation in the saturation region at all times, $v_{DS\min} \geq v_{GS\max} - V_t$.

The BJT Case A similar development can be used to obtain the small-signal voltage gain of the BJT amplifier shown in Fig. 4.6,

$$A = \frac{dv_{CE}}{dv_{BE}} \Big|_{v_{BE} = V_{BE}} \quad (4.19)$$

Utilizing Eq. (4.9) together with Eq. (4.12), we obtain

$$A = -\frac{I_C}{V_T} R_C \quad (4.20)$$

We make the following observations on this expression for the voltage gain:

The gain is negative, which signifies that the amplifier is inverting; that is, there is a 180° phase shift between the input and the output. This inversion should have been anticipated from Eq. (4.9).

The gain is proportional to the collector bias current I_C and to the load resistance R_C . Additional insight into the voltage gain A_v can be obtained by expressing Eq. (4.20) as

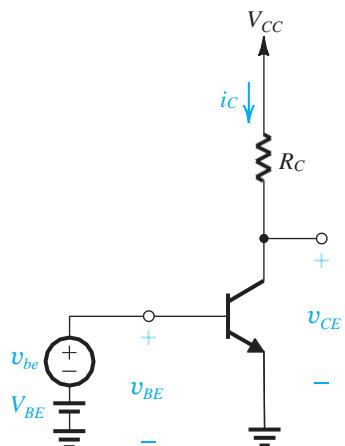


Figure 4.6 BJT amplifier biased at a point Q, with a small voltage signal v_{be} superimposed on the dc bias voltage V_{BE} . The resulting output signal v_{ce} appears superimposed on the dc collector voltage V_{CE} . The amplitude of v_{ce} is larger than that of v_{be} by the voltage gain A_v .

That is, the gain is the ratio of the dc voltage drop across the load resistance R_C to the physical constant V_T (recall that the thermal voltage V_T 25 mV at room temperature). This relationship is similar in form to that for the MOSFET (Eq. 4.16) except that here the denominator is a physical constant (V_T) rather than a design parameter ($V_{ov}/2$). Usually, $V_{ov}/2$ is larger than (V_T), thus we can obtain higher voltage gain from the BJT amplifier than from the MOSFET amplifier. This should not be surprising, as the exponential i_C-v_{BE} relationship is much steeper than the square-law relationship i_D-v_{GS} .

The gain A_v in Eq. (4.21) can be expressed alternately as

$$A_v = -\frac{V_{CC} - V_{CE}}{V_T} \quad (7.22)$$

from which we see that maximum gain is achieved when V_{CE} is at its minimum value of about 0.3 V,

$$|A_{v_{max}}| = \frac{V_{CC} - 0.3}{V_T} \quad (4.23)$$

Here again, this is only a theoretical maximum, since biasing the BJT at the edge of saturation leaves no room for negative signal swing at the output. Equation (4.23) nevertheless provides an upper bound on the voltage gain achievable from the basic BJT amplifier. As an example, for V_{CC} 5 V, the maximum gain is 188 V/V, considerably larger than in the MOSFET case. For modern low-voltage technologies, a V_{CC} of 1.3 V provides a gain of 40 V/V, again much larger than the MOSFET case. The reader should not, however, jump to the conclusion that the BJT is preferred to the MOSFET in the design of modern integrated-circuit amplifiers; in fact, the opposite is true, as we shall see in Chapter 8 and beyond.

Finally, we conclude from Eq. (4.22) that to maximize A_v the transistor should be biased at the lowest possible V_{CE} consistent with the desired value of negative signal swing at the output.

Example 4.2

Consider an amplifier circuit using a BJT having $I_s = 10^{-15}$ A, a collector resistance $R_c = 6.8 \text{ k}\Omega$, and a power supply $V_{CC} = 10$ V.

- (a) Determine the value of the bias voltage V_{BE} required to operate the transistor at $V_{CE} = 3.2$ V. What is the corresponding value of I_C ?
- (b) Find the voltage gain A_v at this bias point. If an input sine-wave signal of 5-mV peak amplitude is superimposed on V_{BE} , find the amplitude of the output sine-wave signal (assume linear operation).
- (c) Find the positive increment in v_{BE} (above V_{BE}) that drives the transistor to the edge of saturation, where $v_{CE} = 0.3$ V.
- (d) Find the negative increment in v_{BE} that drives the transistor to within 1% of cutoff (i.e., to $v_{CE} = 0.99V_{CC}$).

Solution

(a)

$$\begin{aligned} I_c &= \frac{V_{CC} - V_{CE}}{R_C} \\ &= \frac{10 - 3.2}{6.8} = 1 \text{ mA} \end{aligned}$$

The value of V_{BE} can be determined from

$$1 \times 10^{-3} = 10^{-15} e^{V_{BE}/VT}$$

which results in

$$V_{BE} = 690.8 \text{ mV}$$

(b)

$$\begin{aligned} A_v &= -\frac{V_{CC} - V_{CE}}{V_T} \\ &= \frac{10 - 3.2}{0.025} = -272 \text{ V/V} \\ v_{ce}^{\hat{v}} &= 272 \times 0.005 = 1.36 \text{ V} \end{aligned}$$

(c) For $v_{CE} = 0.3 \text{ V}$,

$$i_c = \frac{10 - 0.3}{6.8} = 1.617 \text{ mA}$$

To increase i_c from 1 mA to 1.617 mA, v_{BE} must be increased by

$$\begin{aligned} \Delta v_{BE} &= V_T \ln \frac{1.617}{1} \\ &= 12 \text{ mV} \end{aligned}$$

(d) For $v_{CE} = 0.99V_{CC} = 9.9 \text{ V}$,

$$i_c = \frac{10 - 9.9}{6.8} = 0.0147 \text{ mA}$$

To decrease i_c from 1 mA to 0.0147 mA, v_{BE} must change by

$$\begin{aligned} \Delta v_{BE} &= V_T \ln \frac{0.0147}{1} \\ &= -105.5 \text{ mV} \end{aligned}$$

4.5.3 Determining the VTC by Graphical Analysis

Figure 7.7 shows a graphical method for determining the VTC of the amplifier of Fig. 7.4(a). Although graphical analysis of transistor circuits is rarely employed in practice, it is useful to us at this stage for gaining greater insight into circuit operation, especially in answering the question of where to locate the bias point Q.

The graphical analysis is based on the observation that for each value of v_{GS} , the circuit will be operating at the point of intersection of the $i_D - v_{DS}$ graph corresponding to the particular

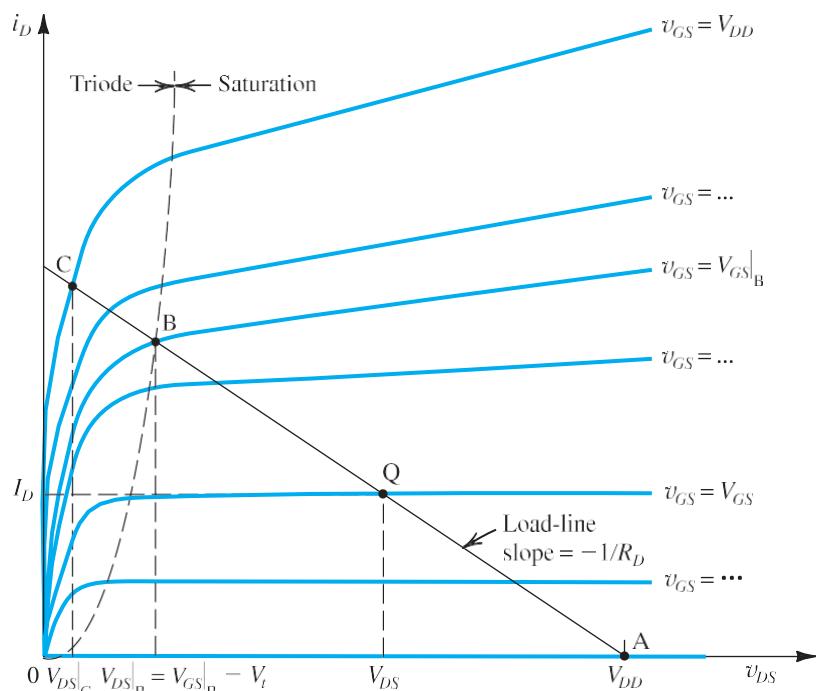


Figure 4.7 Graphical construction to determine the voltage-transfer characteristic of the amplifier in Fig. 7.4(a).

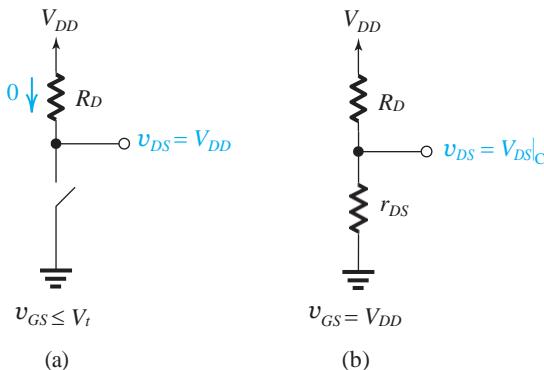


Figure 4.8 Operation of the MOSFET in Fig. 7.4(a) as a switch: (a) open, corresponding to point A in Fig. 4.7; (b) closed, corresponding to point C in Fig. 7.7. The closure resistance is approximately equal to r_{DS} because V_{DS} is usually very small.

value of v_{GS} and the straight line representing Eq. (4.3), which can be rewritten in the form

$$i_D = \frac{V_{DD}}{R_D} - \frac{1}{R_D} v_{DS} \quad (4.24)$$

The straight line representing this relationship is superimposed on the i_D v_{DS} characteristics in Fig. 4.7. It intersects the horizontal axis at $v_{DS} = V_{DD}$ and has a slope of $1/R_D$. Since this straight line represents in effect the load resistance R_D , it is called the **load line**. The VTC is then determined point by point. Note that we have labeled four important points: point A at which $v_{GS} = V_t$, point Q at which the MOSFET can be biased for amplifier operation ($v_{GS} = V_{GS}$ and $v_{DS} = V_{DS}$), point B at which the MOSFET leaves saturation and enters the triode region, and point C, which is deep into the triode region and for which $v_{GS} = V_{DD}$. If the MOSFET is to be used as a switch, then operating points A and C are applicable: At A the transistor is off (open switch), and at C the transistor operates as a low-valued resistor r_{DS} and has a small voltage drop (closed switch). The incremental resistance at point C is also known as the **closure resistance**. The operation of the MOSFET as a switch is illustrated in Fig. 4.8. A detailed study of the application of the MOSFET as a switch is undertaken in Chapter 14, dealing with CMOS digital logic circuits.

The graphical analysis method above can be applied to determine the VTC of the BJT amplifier in Fig. 4.2(c). Here point A, Fig. 4.2(d), corresponds to the BJT just turning on ($v_{BE} = 0.5$ V) and point B corresponds to the BJT leaving the active region and entering the saturation region. If the BJT is to be operated as a switch, the two modes of operation are cutoff (open switch) and saturation (closed switch). As discussed in Section 6.2, in saturation, the BJT has a small closure resistance R_{CEsat} as well as an offset voltage. More seriously, switching the BJT out of its saturation region can require a relatively long delay time to ensure the removal of the charge.

4.5.4 Deciding on a Location for the Bias Point Q

For the MOSFET amplifier, the bias point Q is determined by the value of V_{GS} and that of the load resistance R_D . Two important considerations in deciding on the location of Q

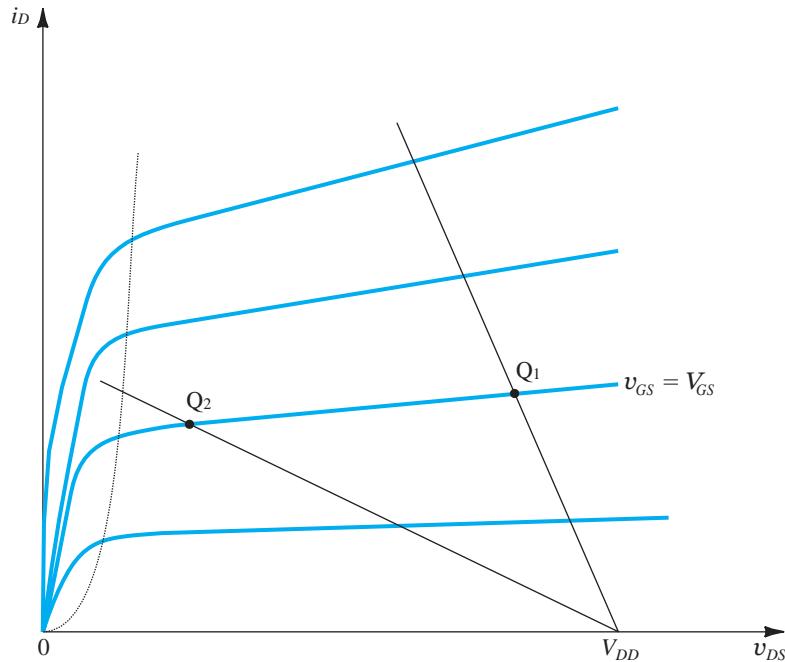


Figure 4.9 Two load lines and corresponding bias points. Bias point Q_1 does not leave sufficient room for positive signal swing at the drain (too close to V_{DD}). Bias point Q_2 is too close to the boundary of the triode region and might not allow for sufficient negative signal swing.

are the required gain and the desired signal swing at the output. To illustrate, consider the VTC shown in Fig. 4.4(b). Here the value of R_D is fixed and the only variable remaining is the value of V_{GS} . Since the slope increases as we move closer to point B, we obtain higher gain by locating Q as close to B as possible. However, the closer Q is to the boundary point B, the smaller the allowable magnitude of negative signal swing. Thus, as often happens in engineering design, we encounter a situation requiring a trade-off. The answer here is relatively simple: For a given R_D , locate Q as close to the triode region (point B) as possible to obtain high gain but sufficiently distant to allow for the required negative signal swing.

In deciding on a value for R_D , it is useful to refer to the i_D v_{DS} plane. Figure 4.9 shows two load lines resulting in two extreme bias points: Point Q_1 is too close to V_{DD} , resulting in a severe constraint on the positive signal swing of v_{ds} . Exceeding the allowable positive maximum results in the positive peaks of the signal being clipped off, since the MOSFET will turn off for the part of each cycle near the positive peak. We speak of this situation by saying that the circuit does not have sufficient “headroom.” Similarly, point Q_2 is too close to the boundary of the triode region, thus

severely limiting the allowable negative signal swing of v_{ds} . Exceeding this limit would result in the transistor entering the triode region for part of each cycle near the negative peaks, resulting in a distorted output signal. In this situation we say that the circuit does not have sufficient “legroom.” We will have more to say on bias design in Section 4.4.

Finally, we note that exactly similar considerations apply to the case of the BJT amplifier.

2 Marks Questions:

1. Mention the methods used for biasing circuits in FET?

Self-bias.

Voltage divider bias.

2. Explain the term MOSFET?

In the insulated gate FET, conductivity is controlled by the potential on the insulated metal plate lying on the top of the channel the insulated gate field effect transistor is often called metallic oxide semiconductor FET.

3. Mention the three regions that are present in the drain source characteristics of JFET?

Saturation region
Break down region
Ohmic region

4. List the characteristics of JFET.

Drain characteristics

Transfer characteristics.

5. What is operating point?

The Q point or quiescent point or operating point where DC load line intersects proper base current curve. The coordinates of Q point decides the zero values of IC and VCE in a common emitter transistor.

6. Define current amplification factor, β ?

It is the ratio of injected carrier current reaching at collector base junction to injected carrier current at emitter base junction.

$$\beta = I_C / I_B$$

7. What do you understand by thermal runaway?

The excess heat produced at the collector base junction may even burn and destroy the transistor. The self destruction of an unbiased transistor is known as thermal runaway. To avoid thermal run away the operating point of the circuit is to be stabilized.

8. What is meant by biasing a transistor?

Process of maintaining proper flow of zero signal collector current and collector emitter voltage during the passage of signal. Biasing keeps emitter base junction forward biased and collector base junction reverse biased during the passage of signal.

9. What are the various methods used for transistor biasing? Which one is popular?

(or) Which is the most commonly used biasing technique and why?

1. two battery biasing
2. fixed biasing
3. collector to base biasing
4. collector to base biasing with emitter feedback
5. Voltage divider or Self biasing

Voltage divider bias is wide popular because it offers excellent stabilization to the circuit.

10. What are the limitations of h-parameters?

Obtaining the exact value of h-parameters for a particular transistor is quite difficult. Highly suitable only for small ac signals.

11. What is the basic difference between bias compensation and stabilization?

Stabilization is the process of making operating point independent of temperature variations or changes in transistor parameters using dc biasing circuits. In the case of compensation technique, in order to stabilize the Q point, we use temperature sensitive devices like diodes, thermistors, transistors instead of DC biasing circuits.

12. List the 3 sources of instability of collector current?

1. Individual variations
2. Temperature dependence of collector current
3. Thermal runaway

13. Define Transconductance?

It is the ratio of change in drain current (ΔI_D) to the change in Gate to Source Voltage (ΔV_{GS}) at constant Drain to Source voltage (V_{DS})

$$g_m = \Delta I_D / \Delta V_{GS} \quad (\text{at constant } V_{DS})$$

14. Write the advantages of JFET?

Input impedance of JFET is very high.

This allows high degree of Isolation between the Input and Output circuit.

Current carriers are not crossing the junction hence noise is reduced drastically

15. List the JFET parameters?

Drain resistance (r_d)

Trans conductance(g_m)

Amplification factor (μ)

16. Explain the depletion mode of operation in MOSFET?

When the gate is at negative bias, the thickness of the depletion layer further increases owing to the further increase of the induced positive charge. Thus the drain current decreases, as the gate is made more negative. This is called depletion mode of operation.

17. Explain the term Drain in FET?

The drain is the terminal through which the current leaves the bar. Convention current entering the bar is designated as I_D .

18. Define the term Gate in FET?

The gate consists of either P⁺ or N⁺ impurity regions, heavily doped and diffused to the bar. This region is always reverse biased and in fact, controls the drain current I_D .

19. Write the relative disadvantages of an FET over that of a BJT?

1. The gain bandwidth product in case of a FET is low as compared with a BJT.
2. The category, called MOSFET, is extremely sensitive to handling therefore additional precautions have to be considered while handling.

20. Give the drain current equation of JFET.

$$I_D = I_{DSS} \frac{(1 - V_{GS})^2}{V_p}$$

I_D = drain current

I_{DSS} = saturation drain current

V_{GS} = gate source voltage

V_P = pinch-off voltage

21. Why MOSFET is called IGFET?

MOSFET is constructed with gate terminal insulated from the channel. So it is also called as insulated gate FET or IGFET.

22. Comparison between JFET and MOSFET

JFET	MOSFET
Gate is not insulated from channel	Gate is insulated from channel by a thin layer of SiO ₂
There are two types – N-channel and P-Channel	Four types - P-channel enhancement, P-channel depletion, N-channel enhancement, N-channel Depletion
Cannot be operated in depletion and enhancement modes	Can be operated in depletion and enhancement Modes
There is a continuous channel	There is a continuous channel only in depletion type, but not in enhancement type

23. Define biasing?

Applying external DC voltage to any electronic devices is called biasing. Depending upon the polarity of the DC voltage externally applied to it, the biasing is classified as forward and reverse biasing.

24. Define forward biasing?

If an external DC voltage is connected in such a way that the p region terminal is connected to the positive of the dc voltage and the n region is connected to the negative of the dc voltage, the biasing condition is called forward biasing.

25. Define reverse biasing?

If an external dc voltage is connected in such a way that the p region terminal of a p-n junction is connected to the negative of the battery and the n region terminal of a p-n junction is connected to the positive terminal of the battery, the biasing condition is called reverse biasing.

Essay Questions:

1. (a)Define operating point?

(b)Define alpha and beta DC amplification factors of BJT?

Draw the diagrams,

Explain the operation with expressions.

2. Compare CE, CB and CC amplifier configurations?

Compare the amplifiers mentioned above.

3. Write the relation between stability factors?

Draw the diagrams,

Explain the concept as above.

4. When the emitter resistor is bypassed by the capacitor, how is the gain of the amplifier affected?

Draw the diagrams,

Express the operation of gain of amplifier affected.

5. What is meant by transistor biasing? Why it is needed? Explain?

Define the biasing,

Draw the diagrams,

Explain the biasing

6. Illustrate the Input and output characteristics of all three configurations of a BJT. Also give the important equations related to those configurations?

Draw the diagram,

Express the equations of configurations.

MCQ's:

- 1. The AC current gain in a common base configuration is** (a)
a) $-\Delta I_C/\Delta I_E$ b) $\Delta I_C/\Delta I_E$ c) $\Delta I_E/\Delta I_C$ d) $-\Delta I_E/\Delta I_C$
- 2. The negative sign in the formula of amplification factor indicates** (a)
a) that I_E flows into transistor while I_C flows out it
b) that I_C flows into transistor while I_E flows out it
c) that I_B flows into transistor while I_C flows out it
d) that I_C flows into transistor while I_B flows out it
- 3. The relation between α and β is** (b)
a) $\beta=\alpha/(1-\alpha)$ b) $\alpha=\beta/(1+\beta)$ c) $\beta=\alpha/(1+\alpha)$ d) $\alpha=\beta/(1-\beta)$
- 4. The base current amplification factor β is given by** _____ (a)
a) I_C/I_B b) I_B/I_C c) I_E/I_B d) I_B/I_E
- 5. In CE configuration, if the voltage drop across $5\text{k}\Omega$ resistor connected in the collector circuit is 5V. Find the value of I_B when $\beta=50$.** (d)
a) 0.01mA b) 0.25mA c) 0.03mA d) 0.02mA
- 6. When the signal is applied, the ratio of change of collector current to the ratio of change of base current is called** _____ (d)
a) dc current gain
b) base current amplification factor
c) emitter current amplification factor
d) ac current gain
- 7. The range of β is** _____ (a)
a) 20 to 500 b) 50 to 300 c) 30 to 400 d) 10 to 20
- 8. The CC configuration has an input resistance** _____ (b)
a) $500\text{k}\Omega$ b) $750\text{k}\Omega$ c) $600\text{k}\Omega$ d) $400\text{k}\Omega$
- 9. The application of a CC configured transistor is** _____ (b)
a) voltage multiplier
b) level shifter
c) rectification
d) impedance matching
- 10. When is the transistor said to be saturated** (a)
a) when V_{CE} is very low
b) when V_{CE} is very high
c) when V_{BE} is very low
d) when V_{BE} is very high

UNIT – V

MOSFET - SMALL SIGNAL

OPERATION MODELS

UNIT-5

5.1 MOSFET Small-Signal Operation and Models

In our study of the operation of the MOSFET and BJT amplifiers in Section 5.1 we learned that linear amplification can be obtained by biasing the transistor to operate in the active region and by keeping the input signal small. In this section, we explore the small-signal operation in greater detail.

Consider the conceptual amplifier circuit shown in Fig. 5.10. Here the MOS transistor is biased by applying a dc voltage³ V_{GS} and the input signal to be v_{gs} is superimposed on the dc bias voltage V_{GS} . The output voltage is taken at the drain.

5.1.1 The DC Bias Point The dc bias current I_D can be found by setting the signal v_{gs} to zero; thus,

$$I_D = \frac{1}{2} k_n (V_{GS} - V_t)^{\frac{1}{2}} = \frac{k_n V_{OV}^2}{2} \quad (5.1)$$

where we have neglected channel-length modulation (i.e., we have assumed $\lambda = 0$). Here V_{OV} is the overdrive voltage at which the MOSFET is biased to operate. The dc voltage at the drain, V_{DS} , will be

$$V_{DS} = V_{DD} - R_D I_D \quad (5.2)$$

To ensure saturation-region operation, we must have

$$V_{DS} > V_{OV}$$

Furthermore, since the total voltage at the drain will have a signal component superimposed on V_{DS} , V_{DS} has to be sufficiently greater than V_{OV} to allow for the required negative signal swing.

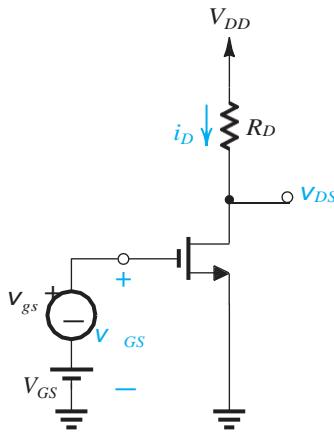


Figure 5.1 Conceptual circuit utilized to study the operation of the MOSFET as a small-signal ampli

The Signal Current in the Drain Terminal Next, consider the situation with the input signal v_{gs} applied. The total instantaneous gate-to-source voltage will be

$$v_{GS} = V_{GS} + v_{gs} \quad (5.3)$$

resulting in a total instantaneous drain current i_D ,

$$\begin{aligned} i_D &= \frac{1}{2} k_n \left(V_{GS} + v_{gs} - V_t \right)^2 \\ &= \frac{1}{2} k_n (V_{GS} - V_t)^2 + k_n (V_{GS} - V_t) v_{gs} + \frac{1}{2} k_n v_{gs}^2 \end{aligned} \quad (5.4)$$

The first term on the right-hand side of Eq. (5.3) can be recognized as the dc bias current I_D (Eq. 5.2). The second term represents a current component that is directly proportional to the input signal v_{gs} . The third term is a current component that is proportional to the square of the input signal. This last component is undesirable because it represents *nonlinear distortion*. To reduce the nonlinear distortion introduced by the MOSFET, the input signal should be kept small so tha

$$\frac{1}{2} k_n v_{gs}^2 \quad k_n (V_{GS} - V_t) v_{gs} \quad (5.5)$$

resulting in

$$v_{gs} = 2(V_{GS} - V_t) \quad (5.5)$$

or, equivalently,

$$v_{gs} = 2V_{ov} \quad (5.6)$$

If this **small-signal condition** is satisfied, we may neglect the last term in Eq. (5.4) and express i_D as

$$i_D = I_D + i_d \quad (5.7)$$

where

$$i_d = k_n(V_{GS} - V_t)v_{gs}$$

The parameter that relates i_d and v_{gs} is the MOSFET **transconductance** g_m ,

$$g_m \equiv \frac{i_d}{v_{gs}} = k_n(V_{GS} - V_t) \quad (5.8)$$

or in terms of the overdrive voltage V_{OV} ,

$$g_m = k_n V_{OV} \quad (5.9)$$

Figure 5.1 presents a graphical interpretation of the small-signal operation of the MOSFET amplifier. Note that g_m is equal to the slope of the i_D-v_{GS} characteristic at the bias point,

$$g_m \equiv \left. \frac{\partial i_D}{\partial v_{GS}} \right|_{V_{GS}=V_{GS}} \quad (5.10)$$

This is the formal definition of g_m , which can be shown to yield the expressions given in Eqs. (5.32) and (5.33).

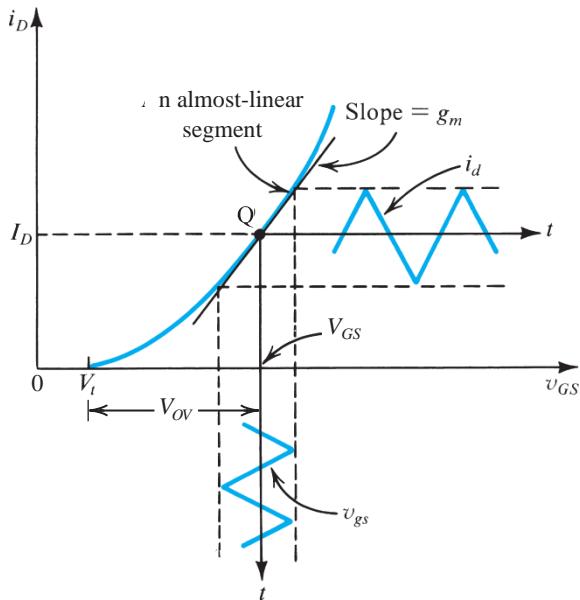


Figure 5.11 Small-signal operation of the MOSFET amplifier.

The Voltage Gain Returning to the circuit of Fig. 5.10, we can express the total instantaneous drain voltage v_{DS} as follows:

$$v_{DS} = V_{DD} - R_D i_D$$

Under the small-signal condition, we have

$$v_{DS} = V_{DD} - R_D (I_D + i_d)$$

which can be rewritten as

$$v_{DS} = V_{DS} - R_D i_d$$

Thus the signal component of the drain voltage is

$$v_{ds} = -i_d R_D = -g_m v_{gs} R_D \quad (5.35)$$

which indicates that the voltage gain is given by

The minus sign in Eq. (5.36) indicates that the output signal v_{ds} is 180° out of phase with respect to the input signal v_{gs} . This is illustrated in Fig. 5.12, which shows v_{GS} and v_{DS} . The input signal is assumed to have a triangular waveform with an amplitude much smaller than $2(V_{GS} - V_t)$, the small-signal condition in Eq. (5.29), to ensure linear operation. For operation in the saturation (active) region at all times, the minimum value of v_{DS} should not fall below the corresponding value of v_{GS} by more than V_t . Also, the maximum value of v_{DS} should be

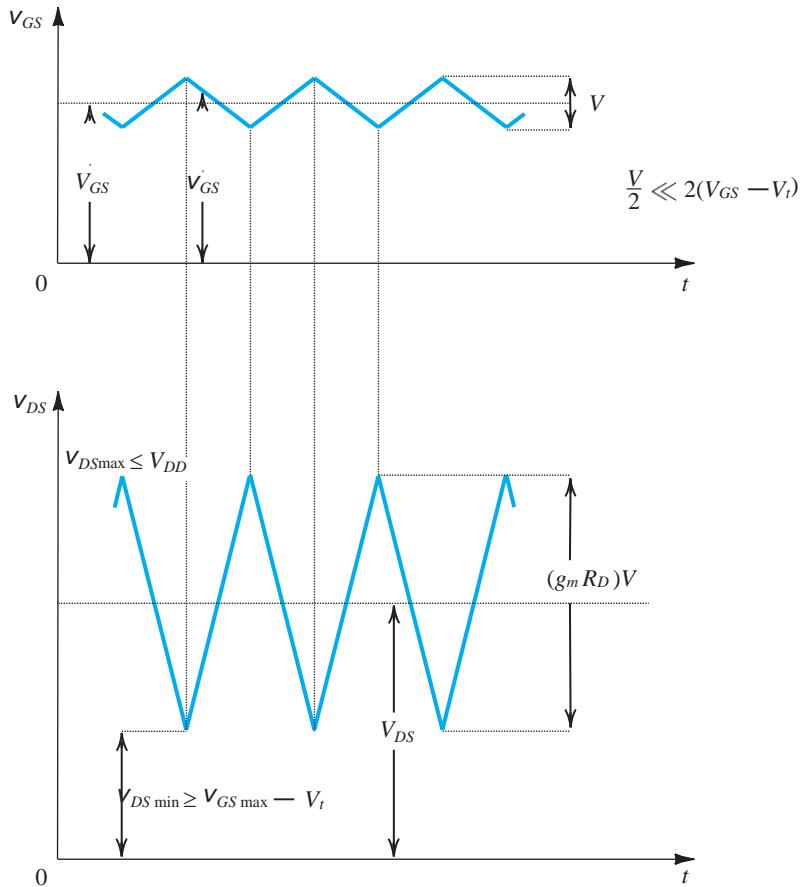


Figure 7.12 Total instantaneous voltages v_{GS} and v_{DS} for the circuit in Fig. 7.10.

smaller than V_{DD} ; otherwise the FET will enter the cutoff region and the peaks of the output signal waveform will be clipped off.

Finally, we note that by substituting for g_m from Eq. (7.33) the voltage-gain expression in Eq. (5.36) becomes identical to that derived in Section 5.1—namely, Eq. (5.15).

5.1.2 Separating the DC Analysis and the Signal Analysis From the preceding analysis, we see that under the small-signal approximation, signal quantities are superimposed on dc quantities. For instance, the total drain current i_D equals the dc current I_D plus the signal current i_d , the total drain voltage v_{DS} $=$ v_{ds} , and so on. It follows that the analysis and design can be greatly simplified by separating dc or bias calculations from small-signal calculations. That is, once a stable dc operating point has been established and all dc quantities calculated, we may then perform signal

analysis ignoring dc quantities.

5.1.3 Small-Signal Equivalent-Circuit Models From a signal point of view, the FET behaves as a voltage-controlled current source. It accepts a signal v_{gs} between gate and source and provides a current $g_m v_{gs}$ at the drain terminal. The input resistance of this controlled source is very high—ideally, infinite. The output resistance—that is, the resistance looking into the

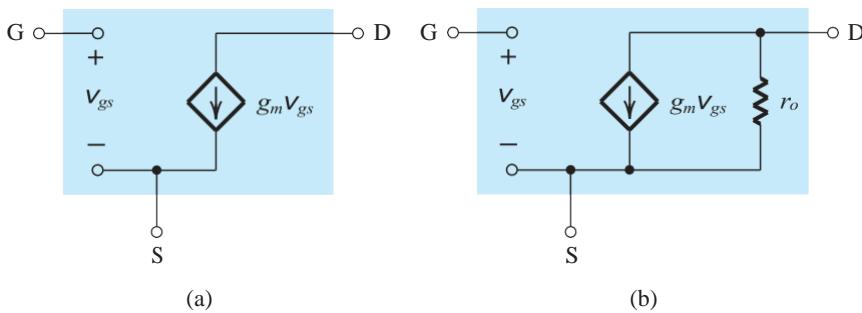


Figure 5.13 Small-signal models for the MOSFET: (a) neglecting the dependence of i_D on v_{DS} in the active region (the channel-length modulation effect) and (b) including the effect of channel-length modulation, modeled by output resistance $r_o = V_A/I_D$. These models apply equally well for both NMOS and PMOS transistors.

drain—also is high, and we have assumed it to be infinite thus far. Putting all of this together, we arrive at the circuit in Fig. 5.13(a), which represents the small-signal operation of the MOSFET and is thus a **small-signal model** or a **small-signal equivalent circuit**.

In the analysis of a MOSFET amplifier circuit, the transistor can be replaced by the equivalent-circuit model shown in Fig. 5.13(a). The rest of the circuit remains unchanged except that *ideal constant dc voltage sources are replaced by short circuits*. This is a result of the fact that the voltage across an ideal constant dc voltage source does not change, and thus there will always be a zero voltage signal across a constant dc voltage source. A dual statement applies for constant dc current sources; namely, the signal current of an ideal constant dc current source will always be zero, and thus *an ideal constant dc current source can be replaced by an open circuit* in the small-signal equivalent circuit of the amplifier. The circuit resulting can then be used to perform any required signal analysis, such as calculating voltage gain.

The most serious shortcoming of the small-signal model of Fig. 5.13(a) is that it assumes the drain current in saturation to be independent of the drain voltage. From our study of the MOSFET

characteristics in saturation, we know that the drain current does in fact depend on v_{DS} in a linear manner. Such dependence was modeled by a finite resistance r_o between drain and source, whose value was given by Eq. (5.27) in Section 5.2.4, which we repeat here (with the prime on I_D dropped) as

$$r_o = \frac{|V_A|}{I_D} \quad (5.37)$$

where V_A $1/\lambda$ is a MOSFET parameter that either is specified or can be measured. It should be recalled that for a given process technology, V_A is proportional to the MOSFET channel length. The current I_D is the value of the dc drain current without the channel-length modulation taken into account; that is,

$$I_D = \frac{1}{2} k_n V_{ov}^2 \quad (5.38)$$

Typically, r_o is in the range of $10 \text{ k}\Omega$ to $1000 \text{ k}\Omega$. It follows that the accuracy of the small-signal model can be improved by including r_o in parallel with the controlled source, as shown in Fig. 5.13(b).

It is important to note that the small-signal model parameters g_m and r_o depend on the dc bias point of the MOSFET.

Returning to the amplifier of Fig. 5.10, we find that replacing the MOSFET with the small-signal model of Fig. 5.13(b) results in the voltage-gain expression

Thus, the finite output resistance r_o results in a reduction in the magnitude of the voltage gain. Although the analysis above is performed on an NMOS transistor, the results, and the equivalent-circuit models of Fig. 7.13, apply equally well to PMOS devices, except for using

$|V_{GS}|$, $|V_t|$, $|V_{ov}|$, and $|V_A|$ and replacing k_n with k_p

The Trans conductance g_m We shall now take a closer look at the MOSFET transconductance given by Eq. (5.32), which we rewrite with $k_n = k^\pm(W/L)$ as follows:

$$g_m = k_n^\pm(W/L)(V_{GS} - V_t) = k_n^\pm(W/L)V_{OV} \quad (5.40)$$

This relationship indicates that g_m is proportional to the process transconductance parameter $k_n^\pm = \mu_n C_{ox}$ and to the W/L ratio of the MOS transistor; hence to obtain relatively large transconductance the device must be short and wide. We also observe that for a given device the transconductance is proportional to the overdrive voltage, $V_{OV} = V_{GS} - V_t$, the amount by which the bias voltage V_{GS} exceeds the threshold voltage V_t . Note, however, that increasing g_m by biasing the device at a larger V_{GS} has the disadvantage of reducing the allowable voltage signal swing at the drain.

Another useful expression for g_m can be obtained by substituting for V_{OV} in Eq. (7.40) by $2I_D/(k_n^\pm(W/L))$ [from Eq. (5.25)]:

$$g_m = \frac{\sqrt{2k_n^\pm}}{W/L} I_D \quad (5.41)$$

This expression shows two things:

1. For a given MOSFET, g_m is proportional to the square root of the dc bias current.
2. At a given bias current, g_m is proportional to W/L .

In contrast, as we shall see shortly, the transconductance of the bipolar junction transistor (BJT) is proportional to the bias current and is independent of the physical size and geometry of the device.

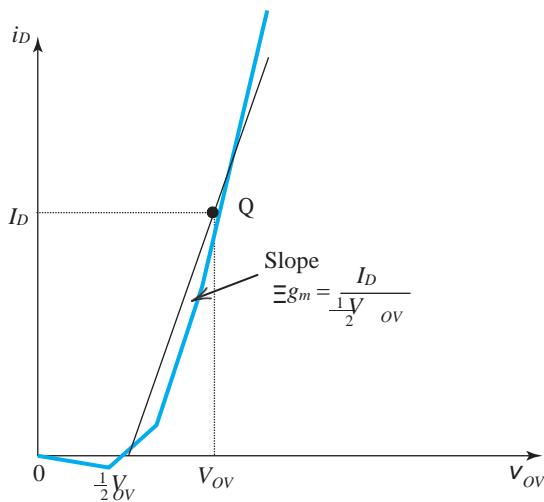


Figure 7.14 The slope of the tangent at the bias point Q intersects the v_{ov} axis at $\frac{1}{2} V_{ov}$. Thus, $g_m = I_D / (\frac{1}{2} V_{ov})$.

A convenient graphical construction that clearly illustrates this relationship is shown in Fig. 5.14.

In summary, there are three different relationships for determining g_m —Eqs. (5.40), (5.41), and (5.42)—and there are three design parameters— (W/L) , V_{ov} , and I_D , any two of which can be chosen independently. That is, the designer may choose to operate the MOSFET with a certain overdrive voltage V_{ov} and at a particular current I_D ; the required W/L ratio can then be found and the resulting g_m determined.

Example 5.3

Figure 7.15(a) shows a discrete MOSFET amplifier utilizing a drain-to-gate resistance R_G for biasing purposes. Such a biasing arrangement will be studied in Section 7.4. The input signal v_i is coupled to the gate via a large capacitor, and the output signal at the drain is coupled to the load resistance R_L via another large capacitor. We wish to analyze this amplifier circuit to determine its small-signal voltage gain, its input resistance, and the largest allowable input signal. The transistor has $V_t = 1.5$ V, $k_n^t(W/L) = 0.25$ mA/V², and $V_A = 50$ V. Assume the coupling capacitors to be sufficiently large so as to act as short circuits at the signal frequencies of interest.

⁴This assumes that the circuit designer is also designing the device, as is typically the case in IC design. On the other hand, a circuit designer working with a discrete-circuit MOSFET obviously does not have the freedom to change its W/L ratio. Thus, in this case there are only two design parameters— V_{ov} and I_D , and only one can be specified by the designer.

Example 5.3 continued

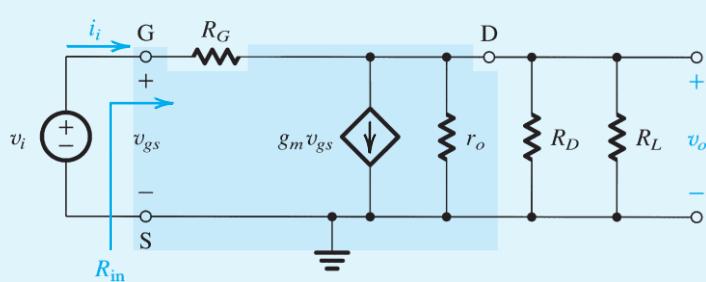
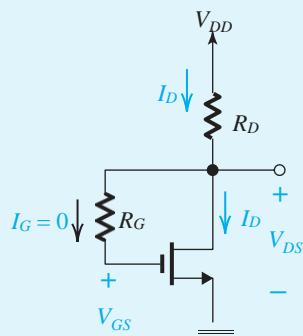
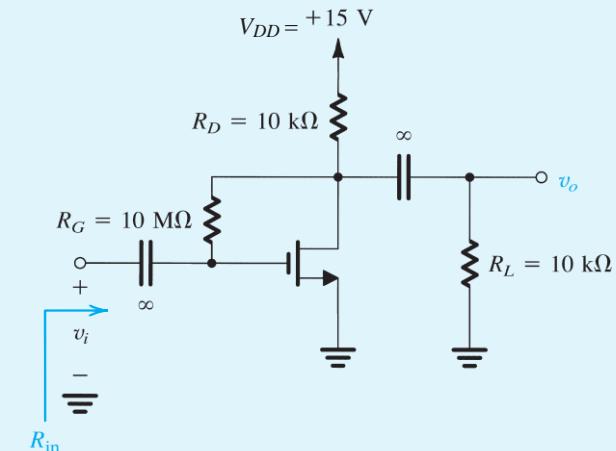
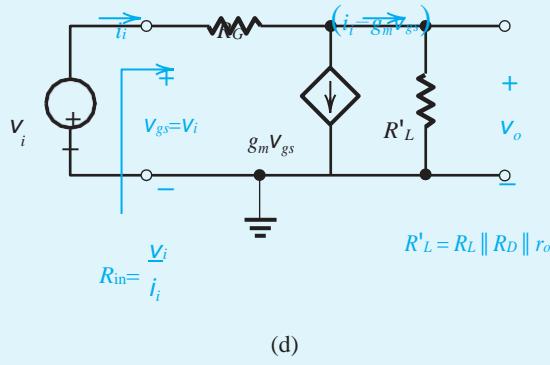


Figure 5.15 Example 7.3: (a) amplifier circuit; (b) circuit for determining the dc operating point; (c) the amplifier small-signal equivalent circuit; (d) a simplified version of the circuit in (c).



(d)

Figure 5.15 continued

Solution

We first determine the dc operating point. For this purpose, we eliminate the input signal v_i , and open-circuit the two coupling capacitors (since they block dc currents). The result is the circuit shown in Fig. 5.14(b). We note that since $I_G = 0$, the dc voltage drop across R_G will be zero, and

$$V_{GS} = V_{DS} = V_{DD} - R_D I_D \quad (5.43)$$

With $V_{DS} = V_{GS}$, the NMOS transistor will be operating in saturation. Thus,

$$I_D = \frac{1}{2} k_n V_{GS} - V_t^2 \quad (5.44)$$

where, for simplicity, we have neglected the effect of channel-length modulation on the dc operating point. Substituting $V_{DD} = 15$ V, $R_D = 10$ k Ω , $k_n = 0.25$ mA/V², and $V_t = 1.5$ V in Eqs. (5.43) and (5.44), and substituting for V_{GS} from Eq. (5.43) into Eq. (5.44) results in a quadratic equation in I_D . Solving the latter and discarding the root that is not physically meaningful yields the solution

$$I_D = 1.06 \text{ mA}$$

which corresponds to

$$V_{GS} = V_{DS} = 4.4 \text{ V}$$

and

$$V_{OV} = 4.4 - 1.5 = 2.9 \text{ V}$$

Next we proceed with the small-signal analysis of the amplifier. Toward that end we replace the MOSFET with its small-signal model to obtain the small-signal equivalent circuit of the amplifier, shown in Fig. 5.15(c). Observe that we have replaced the coupling capacitors with short circuits. The dc voltage supply V_{DD} has also been replaced with a short circuit to ground.

Example 5.3 continued

The values of the transistor small-signal parameters g_m and r_o can be determined by using the dc bias quantities found above, as follows:

$$\begin{aligned} g_m &= k_n V_{OV} \\ &= 0.25 \times 2.9 = 0.725 \text{ mA/V} \\ r_o &= \frac{V_A}{I_D} = \frac{50}{1.06} = 47 \text{ k}\Delta \end{aligned}$$

Next we use the equivalent circuit of Fig. 5.15(c) to determine the input resistance $R_{in} \equiv v_i/i_i$ and the voltage gain $A_v = v_o/v_i$. Toward that end we simplify the circuit by combining the three parallel resistances r_o , R_D , and R_L in a single resistance R_L^*

$$\begin{aligned} R_L^* &= R_L || R_D || r_o \\ &= 10 || 10 || 47 = 4.52 \text{ k}\Delta \end{aligned}$$

as shown in Fig. 5.15(d). For the latter circuit we can write the two equations

$$v_o = i_i - g_m v_{gs} R_L^* \quad (5.45)$$

and

$$i_i = \frac{v_{gs} - v_o}{R_G} \quad (5.46)$$

Substituting for i_i from Eq. (5.46) into Eq. (5.45) results in the following expression for the voltage gain $A_v \equiv v_o/v_i = v_o/v_{gs}$:

$$A_v = -g_m R_L^* \frac{1 - 1/g_m R_G}{1 + R_L^*/R_G}$$

Since R_G is very large, $g_m R_G \ll 1$ and $R_L^*/R_G \ll 1$ (the reader can easily verify this), and the gain expression can be approximated as

$$A_v \approx -g_m R_L^* \quad (5.47)$$

Substituting $g_m = 0.725 \text{ mA/V}$ and $R_L^* = 4.52 \text{ k}\Delta$ yields

$$A_v = -3.3 \text{ V/V}$$

To obtain the input resistance, we substitute in Eq. (7.46) for $v_o = A_v v_{gs} = -g_m R_L^* v_{gs}$, then use $R_{in} \equiv v_i/i_i = v_{gs}/i_i$ to obtain

$$R_{in} = \frac{R_G}{1 + g_m R_L^*} \quad (5.48)$$

This is an interesting relationship: The input resistance decreases as the gain $g_m R_i$ is increased. The value of R_{in} can now be determined; it is

$$R_{in} = \frac{10 \text{ M}\Delta}{1 + 3.3} = 2.33 \text{ M}\Delta$$

which is still very large.

The largest allowable input signal \hat{v}_i is constrained by the need to keep the transistor in saturation at all times; that is,

$$v_{DS} \geq v_{GS} - V_t$$

Enforcing this condition with equality at the point v_{GS} is maximum and v_{DS} is minimum, we write

$$v_{DS\ min} = v_{GS\ max} - V_t$$

$$V_{DS} - A_v \cdot \hat{v}_i = V_{GS} + \hat{v}_i - V_t$$

Since $V_{DS} = V_{GS}$, we obtain

$$\hat{v}_i = \frac{V_t}{A_v + 1}$$

This is a general relationship that applies to this circuit irrespective of the component values. Observe that it simply states that the maximum signal swing is determined by the fact that the bias arrangement makes $V_D = V_G$ and thus, to keep the MOSFET out of the triode region, the signal between D and G is constrained to be equal to V_t . For our particular design,

$$\hat{v}_i = \frac{1.5}{3.3 + 1} = 0.35 \text{ V}$$

Since $V_{OV} = 2.9 \text{ V}$, a v_i of 0.35 is indeed much smaller than $2V_{OV} = 5.8 \text{ V}$; thus the assumption of linear operation is justified.

A modification of this circuit that increases the allowable signal swing is investigated in Problem 7.103.

§ The T Equivalent-Circuit Model Through a simple circuit transformation it is possible to develop an alternative equivalent-circuit model for the MOSFET. The development of such a model, known as the T model, is illustrated in Fig. 5.16. Figure 5.16(a) shows the equivalent circuit studied above without r_o . In Fig. 5.16(b) we have added a second $g_m v_{gs}$ current source in series with the original controlled source. This addition obviously does not change the terminal currents and is thus allowed. The newly created circuit node, labeled X, is joined to the gate terminal G in Fig. 5.16(c). Observe that the gate current does not change—that is, it remains equal to zero—and thus this connection does not alter the terminal characteristics. We now note that we have a controlled current source $g_m v_{gs}$ connected across its control voltage v_{gs} . We can replace this controlled source by a resistance as long as this resistance draws an equalcurrent as the source. (See the source-absorption theorem in Appendix D.) Thus the value ofthe resistance is $v_{gs}/g_m v_{gs} = 1/g_m$. This replacement is shown in Fig. 5.16(d), which depicts

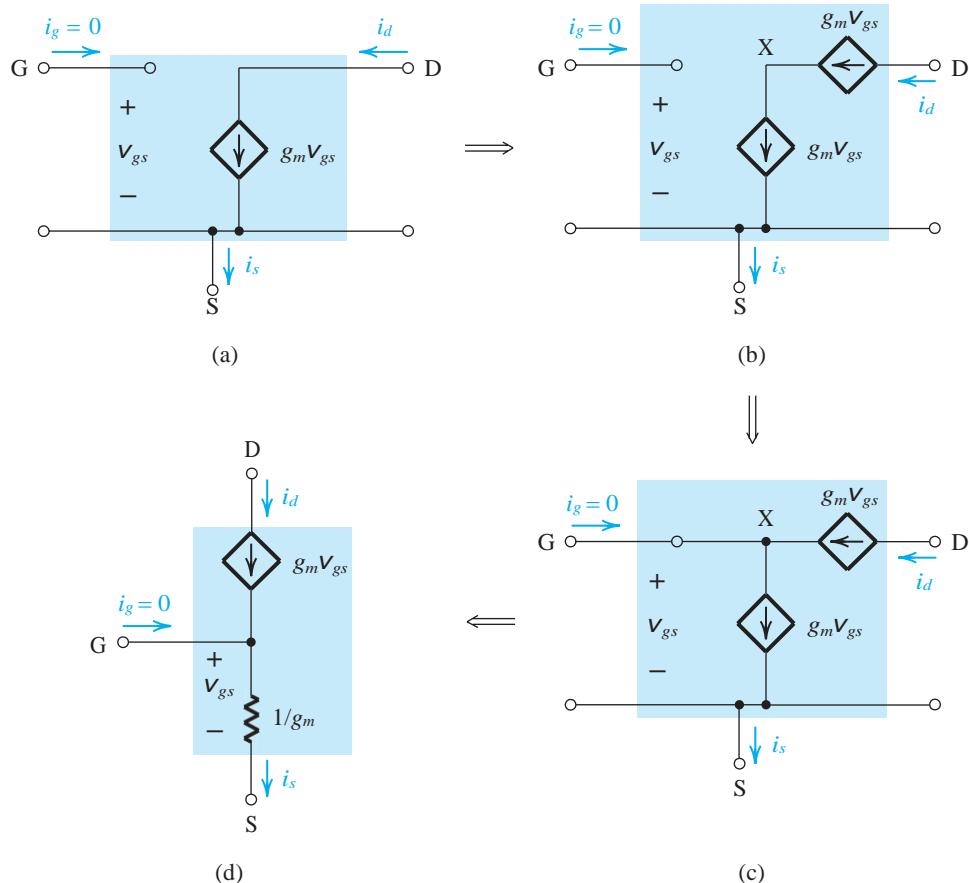


Figure 5.16 Development of the T equivalent-circuit model for the MOSFET. For simplicity, r_o has been omitted; however, it may be added between D and S in the T model of (d).

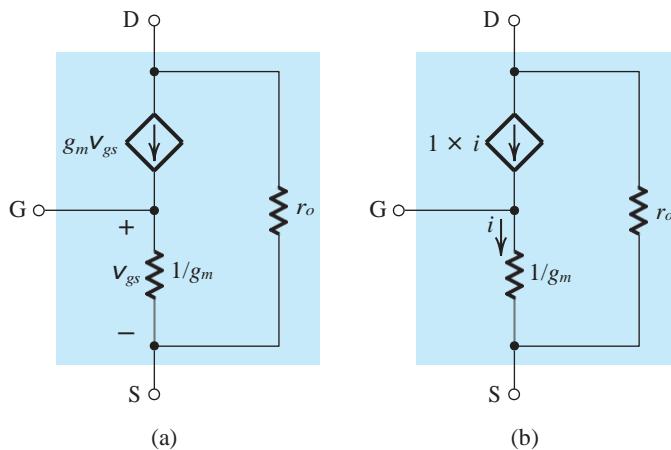


Figure 5.17 (a) The T model of the MOSFET augmented with the drain-to-source resistance r_o .
(b) An alternative representation of the T model.

the alternative model. Observe that i_g is still zero, $i_d = g_m v_{gs}$, and $i_s = v_{gs}/(1/g_m) = g_m v_{gs}$, all the same as in the original model in Fig. 5.16(a).

The model of Fig. 5.16(d) shows that the resistance between gate and source looking into the source is $1/g_m$. This observation and the T model prove useful in many applications. Note that the resistance between gate and source, looking into the gate, is infinite.

In developing the T model we did not include r_o . If desired, this can be done by incorporating in the circuit of Fig. 5.16(d) a resistance r_o between drain and source, as shown in Fig. 5.17(a). An alternative representation of the T model, in which the voltage-controlled current source is replaced with a current-controlled current source, is shown in Fig. 5.17(b).

Finally, we should note that in order to distinguish the model of Fig. 5.13(b) from the equivalent T model, the former is sometimes referred to as the **hybrid- π model**, a carryover from the bipolar transistor literature. The origin of this name will be explained shortly.

Example 5.4 continued

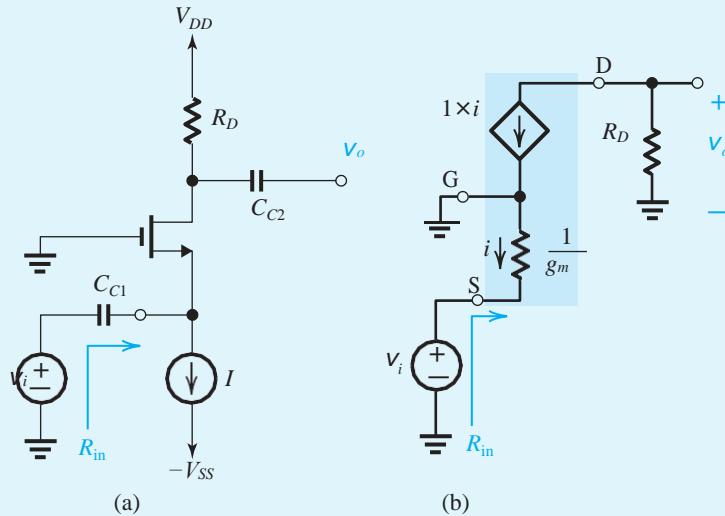


Figure 5.18 (a) Amplifier circuit for Example 7.4. (b) Small-signal equivalent circuit of the amplifier in (a).

Solution

Replacing the MOSFET with its T equivalent-circuit model results in the amplifier equivalent circuit shown in Fig. 7.18(b). Observe that the dc current source I is replaced with an open circuit and the dc voltage source V_{DD} is replaced by a short circuit. The large coupling capacitors have been replaced by short circuits. From the equivalent-circuit model we determine

$$R_{in} = \frac{v_i}{-i} = 1/g_m$$

and

$$v_o = -i R_D = \frac{v_i}{1/g_m} R_D = g_m R_D v_i$$

Thus,

$$A_v \equiv \frac{v_o}{v_i} = g_m R_D$$

We note that this amplifier, known as the common-gate amplifier because the gate at ground potential is common to both the input and output ports, has a low input resistance $1/g_m$ and a noninverting gain. We shall study this amplifier type in Section 5.3.5.

Modeling the Body Effect As mentioned earlier (see Section 5.4), the body effect occurs in a MOSFET when the source is not tied to the substrate (which is always connected to the most negative power supply in the integrated circuit for *n*-channel devices and to the most positive for *p*-channel devices). Thus the substrate (body) will be at signal ground, but since the source is not, a signal voltage v_{bs} develops between the body (B) and the source (S). The substrate then acts as a “second gate” or a **backgate** for the MOSFET. Thus the signal v_{bs} gives rise to a drain-current component, which we shall write as $g_{mb}v_{bs}$, where g_{mb} is the **bodytransconductance**, defined as

$$g_{mb} \equiv \frac{\partial i_D}{\partial v_{BS}} \Big|_{\substack{v_{GS} = \text{constant} \\ v_{DS} = \text{constant}}} \quad (7.49)$$

Recalling that i_D depends on v_{BS} through the dependence of V_t on V_{BS} , we can show that

$$g_{mb} = \chi g_m \quad (5.50)$$

where

$$\chi \equiv \frac{\partial V_t}{\partial V_{SB}} = \frac{V}{2\varphi_f + V_{SB}} \quad (5.51)$$

Typically the value of χ lies in the range 0.1 to 0.3.

Figure 5.19 shows the MOSFET model augmented to include the controlled source $g_{mb}v_{bs}$ that models the body effect. Ideally, this is the model to be used whenever the source is not connected to the substrate. It has been found, however, that except in some very particular

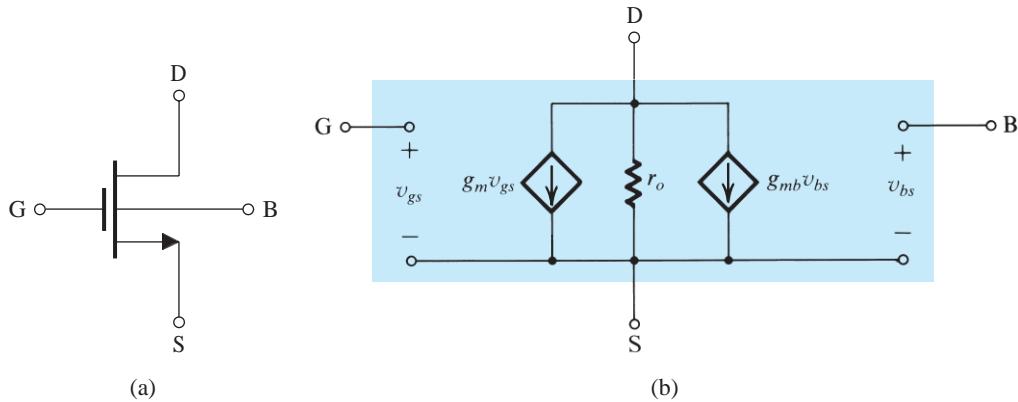


Figure 5.19 Small-signal, equivalent-circuit model of a MOSFET in which the source is not connected to the body.

situations, the body effect can generally be ignored in the initial, pencil-and-paper design of MOSFET amplifiers.

Finally, although the analysis above was performed on an NMOS transistor, the results and the equivalent circuit of Fig. 5.19 apply equally well to PMOS transistors, except for using $|V_{GS}|$, $|V_t|$, $|V_{ov}|$, $|V_A|$, $|V_{SB}|$, $|\gamma|$, and $|\lambda|$ and replacing k_n^{\pm} with k_p^{\pm} in the appropriate formula.

5.2 Basic MOSFET Amplifier Configurations

It is useful at this point to take stock of where we are and where we are going in our study of transistor amplifiers. In Section 5.1 we examined the underlying principle for the application of the MOSFET, and of the BJT, as an amplifier. There we found that almost-linear amplification can be obtained by dc biasing the transistor at an appropriate point in its active region of operation, and by keeping the input signal (v_{gs} or v_{be}) small. We then developed, in Section 5.2, circuit models that represent the small-signal operation of each of the two transistor types (Tables 5.2 and 5.3), thus providing a systematic procedure (Table 5.1) for the analysis of transistor amplifiers.

We are now ready to consider the various possible configurations of MOSFET amplifiers, and we will do that in the present section. To focus our attention on the salient features of the various configurations, we shall present them in their most simple, or “stripped-down,” version. Thus, we will not show the dc biasing arrangements, leaving the study of bias design to the next section. Finally, in Section 5.5 we will bring everything together and present practical *discrete-*

circuit amplifiers, namely, amplifier circuits that can be constructed using discrete components.

5.2.1 The Three Basic Configurations

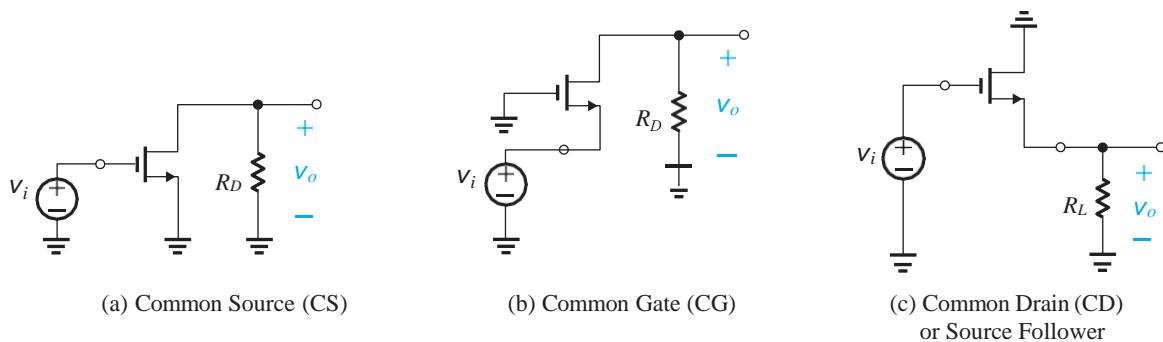
There are three basic configurations for connecting a MOSFET as an amplifier. Each of these configurations is obtained by connecting one of the device terminals to ground, thus creating a two-port network with the grounded terminal being *common* to the input and output ports. The resulting configurations are shown in Fig. 5.33(a–c) for the MOSFET.

In the circuit of Fig. 5.33(a) the source terminal is connected to ground, the input voltage signal v_i is applied between the gate and ground, and the output voltage signal v_o is taken between the drain and ground, across the resistance R_D . This configuration, therefore, is called the **grounded-source or common-source (CS)** amplifier. It is by far the most popular MOS amplifier configuration, and we utilized it in Sections 5.1 and 5.2 to study MOS amplifier operation.

The **common-gate (CG)** or grounded-gate amplifier is shown in Fig. 5.33(b). Here the gate is grounded, the input signal v_i is applied to the source, and the output signal v_o is taken at the drain across the resistance R_D . We encountered a CG amplifier in Example 5.4.

Finally, Fig. 5.33(c) shows the **common drain (CD)** or grounded-drain amplifier. Here the drain terminal is grounded, the input signal v_i is applied between gate and ground, and the output voltage v_o is taken between the source and ground, across a resistance R_L . For reasons that will become apparent shortly, this pair of configurations is more commonly called the **source follower**.

Our study of the three basic amplifier configurations of the MOSFET and of the BJT will reveal that each has distinctly different attributes, hence areas of application. As well, it will be shown that although each pair of configurations, (e.g., CS and CE), has many common attributes, important differences remain.



Our next step is to replace the transistor in each of the six circuits in Fig. 5.33 by an appropriate equivalent-circuit model (from Tables 5.2 and 5.3) and analyze the resulting circuits to determine important characteristic parameters of the particular amplifier configuration. To simplify matters, we shall not include r_o in the initial analysis.

5.2.2 Characterizing Amplifiers

Before we begin our study of the different transistor amplifier configurations, we consider how to characterize the performance of an amplifier as a circuit building block. An introduction to this topic was presented in Section 1.5.

Figure 5.34(a) shows an amplifier fed with a signal source having an open-circuit voltage v_{sig} and an internal resistance R_{sig} . These can be the parameters of an actual signal source or, in a cascade amplifier, the Thevenin equivalent of the output circuit of another amplifier stage preceding the one under study. The amplifier is shown with a load resistance R_L connected to the output terminal. Here, R_L can be an actual load resistance or the input resistance of a succeeding amplifier stage in a cascade amplifier.

Figure 5.34(b) shows the amplifier circuit with the amplifier block replaced by its equivalent-circuit model. The input resistance R_{in} represents the loading effect of the amplifier

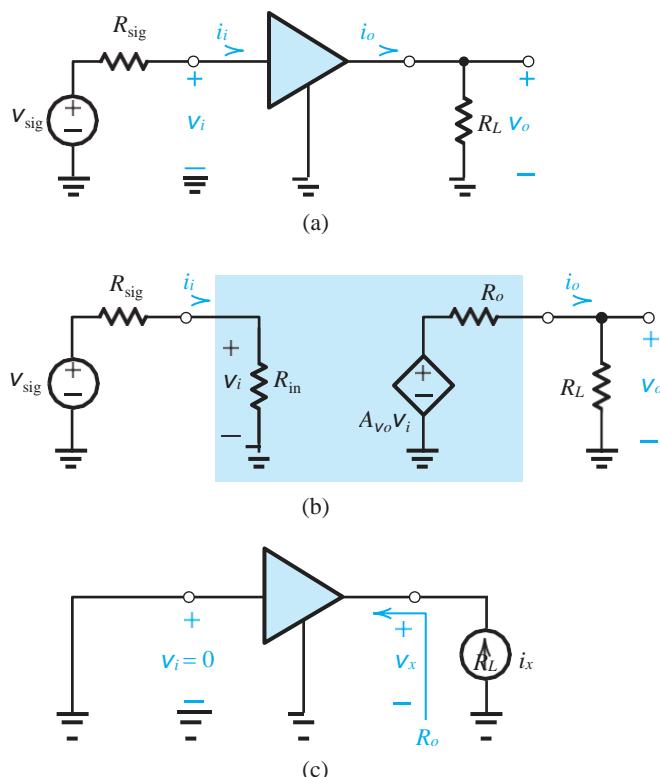


Figure 5.34 Characterization of the amplifier as a functional block: (a) An amplifier fed with a voltage signal v_{sig} having a source resistance R_{sig} , and feeding a load resistance R_L ; (b) equivalent-circuit representation of the circuit in (a); (c) determining the amplifier output resistance R_o .

input on the signal source. It is found from

$$R_{\text{in}} \equiv \frac{\underline{V}_i}{i_i}$$

and together with the resistance R_{sig} forms a voltage divider that reduces $\underline{V}_{\text{sig}}$ to the value \underline{v}_i that appears at the amplifier input,

$$\underline{v}_i = \frac{R_{\text{in}}}{R_{\text{in}} + R_{\text{sig}}} \underline{V}_{\text{sig}} \quad (7.83)$$

Most of the amplifier circuits studied in this section are **unilateral**. That is, they do not contain internal feedback, and thus R_{in} will be independent of R_L . However, in general R_{in} may depend on the load resistance R_L . Indeed one of the six configurations studied in this section, the emitter follower, exhibits such dependence.

The second parameter in characterizing amplifier performance is the **open-circuit voltagegain** A_{v_o} , defined as

$$A_{v_o} \equiv \frac{\underline{V}_o}{\underline{v}_i} \Big|_{R_L=\infty}$$

The third and final parameter is the output resistance R_o . Observe from Fig. 5.34(b) that R_o is the resistance seen looking back into the amplifier output terminal with \underline{v}_i set to zero. Thus R_o can be determined, at least conceptually, as indicated in Fig. 5.34(c) with

$$R_o = \frac{\underline{V}_x}{i_x}$$

Because R_o is determined with $\underline{v}_i = 0$, the value of R_o does not depend on R_{sig} .

The controlled source $A_{v_o}\underline{v}_i$ and the output resistance R_o represent the Thevenin equivalent of the amplifier output circuit, and the output voltage \underline{V}_o can be found from

5.2.3 The Common-Source (CS) Amplifier

Of the three basic transistor amplifier configurations, the common-source is the most widely used. Typically, in an amplifier formed by cascading a number of gain stages, the bulk of the voltage gain is obtained by using one or more common-source stages in cascade.

Characteristic Parameters of the CS Amplifier Figure 5.35(a) shows a common-source amplifier (with the biasing arrangement omitted) fed with a signal source v_{sig} having a source resistance R_{sig} . We wish to analyze this circuit to determine R_{in} , A_{v_o} , and R_o . For this purpose, we assume that R_D is part of the amplifier; thus if a load resistance R_L is connected to the amplifier output, R_L appears in parallel with R_D . In such a case, we wish to determine A_v and G_v as well.

Replacing the MOSFET with its hybrid- π model (without r_o), we obtain the CS amplifier equivalent circuit in Fig. 5.35(b) for which, tracing the signal from input to output, we can write by inspection

$$R_{\text{in}} = \infty \quad (5.87)$$

$$v_i = v_{\text{sig}}$$

$$v_{gs} = v_i$$

$$v_o = -g_m v_{gs} R_D$$

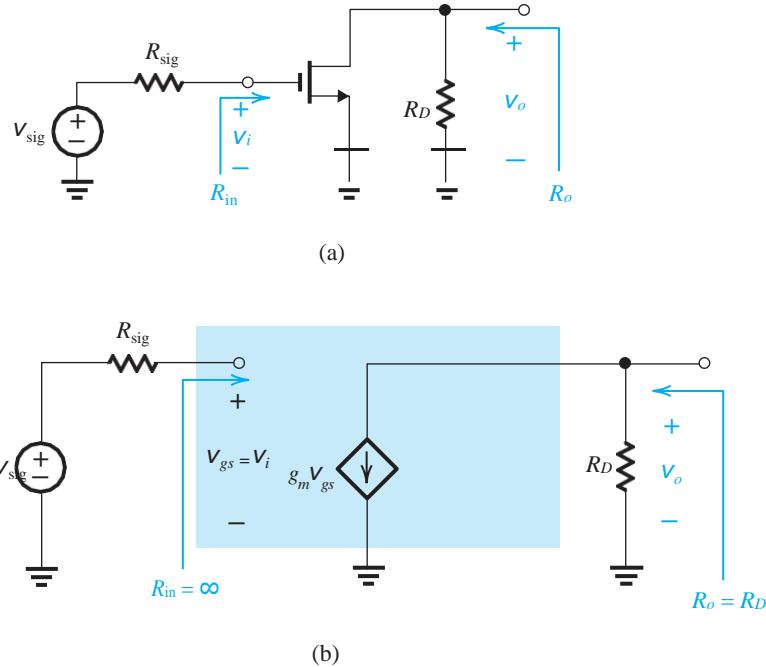


Figure 5.35 (a) Common-source amplifier fed with a signal V_{sig} from a generator with a resistance R_{sig} . The bias circuit is omitted. (b) The common-source amplifier with the MOSFET replaced with its hybrid- π model.

$$A_{vo} \equiv \frac{V_o}{V_i} = -g_m R_D \quad (5.88)$$

$$R_o = R_D \quad (5.89)$$

If a load resistance R_L is connected across R_D , the voltage gain A_v can be obtained from

$$A_v = A_{vo} \frac{R_L}{R_{vo L} + R_o} \quad (5.90)$$

where A_{v_o} is given by Eq. (7.88) and R_o by Eq. (7.89), or alternatively by simply adding R_L in parallel with R_D in Eq. (7.88), thus

$$A_v = -g_m(R_D \| R_L) \quad (5.91)$$

The reader can easily show that the expression obtained from Eq. (5.90) is identical to that in Eq. (5.91). Finally, since $R_{in} = \infty$ and thus $v_i = v_{sig}$, the overall voltage gain G_v is equal to A_v ,

Then we
write

$$\begin{aligned} v_i &= \frac{r_\pi}{r_\pi + R_{sig}} v \\ v_\pi &= v_i \\ v_o &= -g_m v_\pi R_C \end{aligned} \quad (5.93)$$

Thus,

$$A_{v_o} \equiv \frac{v_o}{v_i} = -g_m R_C \quad (5.94)$$

$$R_o = R_C \quad (5.95)$$

With a load resistance R_L connected across R_C ,

$$A_v = -g_m(R_C \| R_L) \quad (5.96)$$

and the overall voltage gain G_v can be found from

$$G \equiv \frac{v_o}{v} = \frac{v_i v_o}{v_{sig} v_i}$$

Thus,

$$G = \frac{r_\pi}{r_\pi + R_{sig}} g_m (R_C \| R_L) \quad (5.97)$$

It is important to note here the effect of the finite input resistance (r_π) in reducing the magnitude of the voltage gain by the voltage-divider ratio $r_\pi/(r_\pi + R_{sig})$. The extent of the gain reduction depends on the relative values of r_π and R_{sig} . However, there is a compensating effect in the CE amplifier: g_m of the BJT is usually much higher than the corresponding value of the MOSFET.

The amplifier characteristic parameters can now be found as

$$\begin{aligned}
 R_{in} &= r_\pi = 2.5 \text{ k}\Delta \\
 A_{vo} &= -g_m R_C \\
 &= -40 \text{ mA/V} \times 5 \text{ k}\Delta \\
 &= -200 \text{ V/V} \\
 R_o &= R_C = 5 \text{ k}\Delta
 \end{aligned}$$

With a load resistance $R_L = 5 \text{ k}\Delta$ connected at the output, we can find A_v by either of the following two approaches:

$$\begin{aligned}
 A_v &= A \frac{R_L}{R_{vo} + R_o} \\
 &= -200 \times \frac{5}{5+5} = -100 \text{ V/V}
 \end{aligned}$$

or

$$\begin{aligned}
 A_v &= -g_m (R_C \parallel R_L) \\
 &= -40(5|5) = -100 \text{ V/V}
 \end{aligned}$$

The overall voltage gain G_v can now be determined as

$$\begin{aligned}
 G_v &= \frac{R_{in}}{R_{in} + R_{sig}} A_v \\
 &= \frac{2.5}{2.5 + 5} \times -100 = -33.3 \text{ V/V}
 \end{aligned}$$

If the maximum amplitude of v_π is to be 5 mV, the corresponding value of \hat{v}_{sig} will be

$$\hat{v}_{sig} = \frac{R_{in} + R_{sig}}{R_{in}} v_\pi = \frac{2.5 + 5}{2.5} \times 5 = 15 \text{ mV}$$

and the amplitude of the signal at the output will be

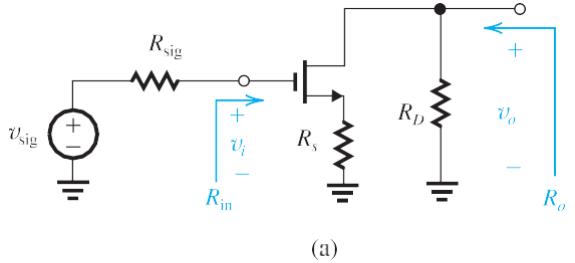
$$\hat{v}_o = G_v \hat{v}_{sig} = 33.3 \times 0.015 = 0.5 \text{ V}$$

Final Remarks

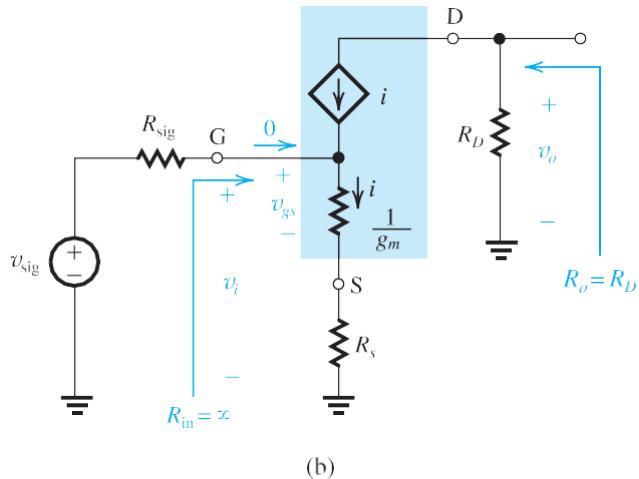
1. The CS amplifier are the most useful of all transistor amplifier configurations. They exhibit a moderate to high input resistance (infinite for the CS), a moderate to high output resistance, and reasonably high voltage gain.
2. The input resistance of the CS amplifier, $R_{in} = \beta/g_m$, is inversely proportional to the dc bias current I_C . To increase R_{in} one is tempted to lower the bias current I_C ; however, this also lowers g_m and hence the voltage gain. This is a significant design trade-off. If a much higher input resistance is desired, then a modification of the CS configuration (to be discussed in Section 5.3.4) can be applied, or an emitter-follower stage can be inserted between the signal source and the CE amplifier (see Section 5.3.6).
3. Reducing R_D or R_C to lower the output resistance of the CS amplifier, respectively, is usually not a viable proposition because the voltage gain is also reduced. Alternatively, if a very low output resistance (in the ohms or tens-of-ohms range) is needed, a source-follower or an emitter-follower stage can be utilized between the output of the CS amplifier and the load resistance (see Section 5.3.6).
4. Although the CS and configurations are the workhorses of transistor amplifiers, both suffer from a limitation on their high-frequency response. As will be shown in Chapter 10, combining the CS amplifier with a CG amplifier can extend the bandwidth considerably. The CG amplifier is studied in Section 5.3.5.

5.2.4 The Common-Source Amplifier with a Source Resistance

It is often beneficial to insert a resistance R_s in the source lead of a common-source amplifier. Figure 7.37(a) shows a CS amplifier with a resistance R_s in its source lead. The corresponding small-signal equivalent circuit is shown



(a)



(b)

Figure 5.37 The CS amplifier with a source resistance R_s : (a) circuit without bias details; (b) equivalent circuit with the MOSFET represented by its T model.

in Fig. 5.37(b), where we have utilized the T model for the MOSFET. The T model is used in preference to the hybrid- π model because it makes the analysis in this case considerably simpler. In general, *whenever a resistance is connected in the source lead, the T model is preferred*. The source resistance then simply appears in series with the model resistance $1/g_m$ and can be added to it.

From Fig. 5.37(b) we see that as expected, the input resistance R_{in} is infinite and thus $v_i = v_{sig}$. Unlike the CS amplifier, however, here only a fraction of v_i appears between gate and source as v_{gs} . The voltage divider composed of $1/g_m$ and R_s , which appears across the amplifier input, can be used to determine v_{gs} , as follows:

$$v_{gs} = v_i \frac{1/g_m}{1/g_m + R_s} = \frac{v_i}{1 + g_m R_s} \quad (5.98)$$

Thus we can use the value of R_s to control the magnitude of the signal v_{gs} and thereby ensure that v_{gs} does not become too large and cause unacceptably high nonlinear distortion. This is the first benefit

of including resistor R_s . Other benefits will be encountered in later sections and chapters. For instance, it will be shown in Chapter 10 that R_s causes the useful bandwidth of the amplifier to be extended. The mechanism by which R_s causes such improvements in amplifier performance is *negative feedback*. To see how R_s introduces negative feedback, refer to Fig. 5.37(a): If with v_{sig} and hence v_i kept constant, the drain current increases for some

reason, the source current also will increase, resulting in an increased voltage drop across R_s . Thus the source voltage rises, and the gate-to-source voltage decreases. The latter effect causes the drain current to decrease, counteracting the initially assumed change, an indication of the presence of negative feedback. In Chapter 11 we shall study negative feedback formally. There we will learn that the improvements that negative feedback provides are obtained at the expense of a reduction in gain. We will now show this to be the case in the circuit of Fig. 5.37.

The output voltage v_o is obtained by multiplying the controlled-source current i by R_D ,

$$v_o = -iR_D$$

The current i in the source lead can be found by dividing v_i by the total resistance in the source,

$$i = \frac{v_i}{1/g_m + R_s} = \frac{g_m v_i}{1 + g_m R_s} \quad (5.99)$$

Thus, the voltage gain A_{vo} can be found as

$$v_o A \equiv \frac{v_o}{v_i} = -\frac{g_m R_D}{1 + g_m R_s} \quad (5.100)$$

which can also be expressed as

$$A_{vo} = -\frac{R_D}{1/g_m + R_s} \quad (5.101)$$

Equation (5.100) indicates that including the resistance R_s reduces the voltage gain by the factor $(1 + g_m R_s)$. This is the price paid for the improvements that accrue as a result of R_s . It is interesting to note that in Chapter 11, we will find that the factor $(1 + g_m R_s)$ is the “amount of negative feedback” introduced by R_s . It is also the same factor by which linearity, bandwidth, and other performance parameters improve. Because of the negative-feedback action of R_s it is known as a **source-degeneration resistance**.

There is another useful interpretation of the expression for the drain current in Eq. (7.99): The quantity between brackets on the right-hand side can be thought of as the “effective transconductance with R_s included.” Thus, including R_s reduces the transconductance by the factor $(1 + g_m R_s)$. This, of course, is simply the result of the fact that only a fraction $1/(1 + g_m R_s)$ of v_i appears as v_{gs} (see Eq. 5.98).

The alternative gain expression in Eq. (5.101) has a powerful and insightful interpretation: The voltage gain between gate and drain is equal to the ratio of the total resistance in the drain (R_D) to the total resistance in the source $(1/g_m + R_s)$,

$$\text{Voltage gain from gate to drain} = -\frac{\text{Total resistance in drain}}{\text{Total resistance in source}} \quad (5.102)$$

This is a general expression. For instance, setting $R_s = 0$ in Eq. (5.101) yields A_{vo} of the CS amplifier.

Finally, we consider the situation of a load resistance R_L connected at the output. We can obtain the gain A_v using the open-circuit voltage gain A_{vo} together with the output resistance R_o , which can be found by inspection to be

$$R_o = R_D$$

Alternatively, A_v can be obtained by simply replacing R_D in Eq. (7.101) or (7.100) by $(R_D \parallel R_L)$; thus,

$$A_v = -\frac{g_m(R_D \parallel R_L)}{1 + g_m R_s} \quad (7.103)$$

or

$$A_v = -\frac{R_D \parallel R_L}{1 + g_m R_s} \quad (7.104)$$

Example 5.9

For the CE amplifier specified in Example 5.8, what value of R_e is needed to raise R_{in} to a value four times that of R_{sig} ? With R_e included, find A_{vo} , R_o , A_v , and G_v . Also, if v^{\wedge}_n is limited to 5 mV, what are the corresponding values of v^{\wedge}_{sig} and v^{\wedge}_o ?

Solution

To obtain $R_{in} = 4R_{sig} = 4 \times 5 = 20 \text{ k}\Delta$, the required R_e is found from

$$20 = (\beta + 1) r_e + R_e$$

With $\beta = 100$,

$$r_e + R_e = 200 \Delta$$

Thus,

$$\begin{aligned} R_e &= 200 - 25 = 175 \Delta \\ A_{vo} &= -\alpha \frac{R_C}{r_e + R_e} \\ &= -\frac{25}{25+175} = -25 \text{ V/V} \\ R_o &= R_C = 5 \text{ k}\Delta \text{ (unchanged)} \\ A &= A_v \frac{R_L}{R_{in}} = -25 \times \frac{5}{5} = -12.5 \text{ V/V} \\ G &= \frac{A_v}{A} = \frac{-12.5}{-12.5} = 10 \text{ V/V} \end{aligned}$$

For $v^{\wedge}_n = 5 \text{ mV}$,

$$\begin{aligned} v^{\wedge}_n &= v^{\wedge}_n \frac{r_e + R_e}{r_e} \\ &= 5 \left(1 + \frac{175}{25} \right) = 40 \text{ mV} \\ v^{\wedge}_{sig} &= v^{\wedge}_n \frac{R_{in} + R_{sig}}{R} \\ &= 40 \left(1 + \frac{5}{20} \right) = 50 \text{ mV} \\ v^{\wedge}_o &= v^{\wedge}_{sig} \times G_v \\ &= 50 \times 10 = 500 \text{ mV} = 0.5 \text{ V} \end{aligned}$$

Thus, while G_v has decreased to about a third of its original value, the amplifier is able to produce as large an output signal as before for the same nonlinear distortion.

5.2.5 The Common-Gate (CG) Amplifier

Figure 5.39(a) shows a common-gate amplifier with the biasing circuit omitted. The amplifier is fed with a signal source characterized by v_{sig} and R_{sig} . Since R_{sig} appears in series with the source, it is more convenient to represent the transistor with the T model than with the π model. Doing this, we obtain the amplifier equivalent circuit shown in Fig. 5.39(b).

From inspection of the equivalent circuit of Fig. 5.39(b), we see that the input resistance

$$R_{\text{in}} = \frac{1}{g_m} \quad (5.116)$$

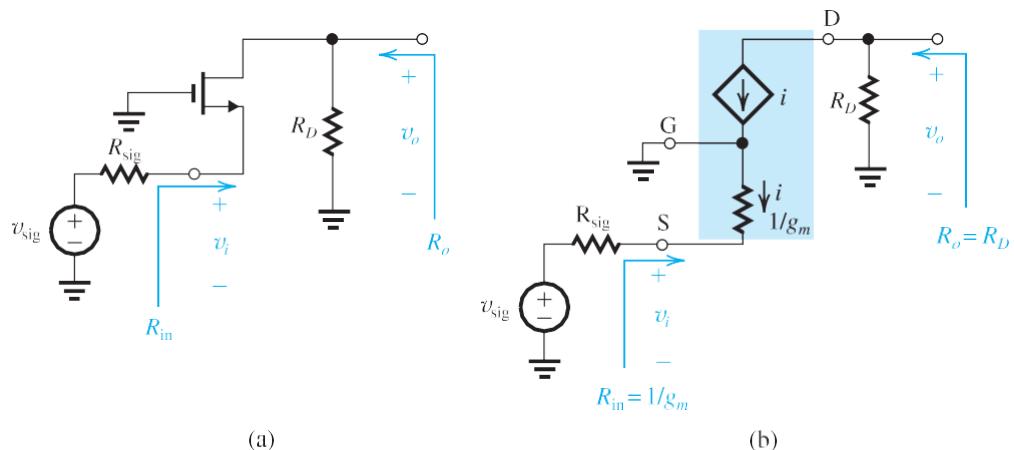


Figure 5.39 (a) Common-gate (CG) amplifier with bias arrangement omitted. (b) Equivalent circuit of the CG amplifier with the MOSFET replaced with its T model.

This should have been expected, since we are looking into the source and the gate is grounded.

Typically $1/g_m$ is a few hundred ohms; thus the CG amplifier has a low input resistance.

To determine the voltage gain A_{v_o} , we write at the drain node

$$v_o = -iR_D$$

and substitute for the source current i from

$$i = \frac{v_i}{1/g_m}$$

to obtain

$$A_{v_o} \equiv \frac{v_o}{v_i} = g_m R_D \quad (5.117)$$

which except for the positive sign is identical to the expression for A_{V_o} of the CS amplifier. The output resistance of the CG circuit can be found by inspection of the circuit in Fig. 5.39(b) as

$$R_o = R_D \quad (5.118)$$

which is the same as in the case of the CS amplifier.

Although the gain of the CG amplifier proper has the same magnitude as that of the CS amplifier, this is usually not the case as far as the overall voltage gain is concerned. The low input resistance of the CG amplifier can cause the input signal to be severely attenuated. Specifically,

$$\frac{v_i}{v_{\text{sig}}} = \frac{R_{\text{in}}}{R_{\text{sig}} + R_{\text{in}}} = \frac{1/g_m}{1/g_m + R_{\text{sig}}} \quad (5.119)$$

from which we see that except for situations in which R_{sig} is on the order of $1/g_m$, the signal transmission factor v_i/v_{sig} can be very small and the overall voltage gain G_v can be correspondingly small. Specifically, with a resistance R_L connected at the output

$$G_v = \frac{1/g_m}{R_{\text{sig}} + 1/g_m} [g_m (R_D \| R_L)]$$

Thus,

$$G_v = \frac{(R_D \| R_L)}{R_{\text{sig}} + 1/g_m} \quad (5.120)$$

Observe that *the overall voltage gain is simply the ratio of the total resistance in the drain circuit to the total resistance in the source circuit*. If R_{sig} is of the same order as R_D and R_L , G_v will be very small.

Because of its low input resistance, the CG amplifier alone has very limited application. One such application is to amplify high-frequency signals that come from sources with relatively low resistances. These include cables, where it is usually necessary for the input resistance of the amplifier to match the characteristic resistance of the cable. As will be shown in Chapter 10, the CG amplifier has excellent high-frequency response. Thus it can be combined with the CS amplifier in a very beneficial way that takes advantage of the best features of each of the two configurations. A very significant circuit of this kind will be studied in Chapter 8.

Since $\alpha \approx 1$, we see that as in the case of the CG amplifier, the overall voltage gain is simply the ratio of the total resistance in the collector to the total resistance in the emitter. We also note that the overall voltage gain is almost independent of the value of β (except through the small dependence of α on β), a desirable property. Observe that for R_{sig} of the same order as R_C and R_L , the gain will be very small.

In summary, the CB and CG amplifiers exhibit a very low input resistance ($1/g_m$), an open-circuit voltage gain that is positive and equal in magnitude to that of the CG amplifier ($g_m R_C$ or $g_m R_D$), and, like the CS amplifier, a relatively high output resistance (R_C or R_D). Because of its very low input resistance, the CG circuit *alone* is not attractive as a voltage amplifier except in specialized applications, such as the cable amplifier mentioned above. The CG amplifier has excellent high-frequency performance, which as we shall see in Chapters 8 and 10, makes it useful in combination with other circuits in the implementation of high-frequency amplifiers.

5.2.5.1 The Source Followers

The last of the basic transistor amplifier configurations is the common-drain amplifier, an important circuit that finds application in the design of both small-signal amplifiers and amplifiers that are required to handle large signals and deliver substantial amounts of signal power to a load. This latter variety will be studied in Chapter 12. The common-drain amplifier is more commonly known as the *source follower*, and the common-collector amplifier is more commonly known as the *emitter follower*. The reason behind these names will become apparent shortly.

The Need for Voltage Buffers Before embarking on the analysis of the source and the emitter followers, it is useful to look at one of their more common applications. Consider the situation depicted in Fig. 5.41(a). A signal source delivering a signal of reasonable strength (1 V) with an internal resistance of 1 M Ω is to be connected to a 1-k Ω load resistance. Connecting the source to the load directly as in Fig. 5.41(b) would result in severe attenuation

5.2.5 Frequency Response of the Common-Gate and Cascade Amplifiers

Although common-source and common-emitter amplifiers provide substantial gain at midband frequencies, their gain falls off in the high-frequency band at a relatively low frequency. This is primarily due to the large input capacitance C_{in} , whose value is significantly increased by the Miller component. The latter is large because of the Miller multiplication effect, which the bridging capacitance C_{gd} (or C_μ) experiences. It follows that the key to obtaining wideband operation, that is, high f_H , is to use circuit configurations that do not suffer from the Miller effect. One such configuration is the common-gate circuit.

5.2.5 High-Frequency Response of the CG Amplifier

Figure 5.26(a) shows the CG amplifier with the MOSFET internal capacitances C_{gs} and C_{gd} pulled out of the model and indicated. For generality, a capacitance C_L is included. For an amplifier utilizing this particular transistor type may specify that the dc collector current shall always be within, say, 10% of the nominal value of, say, 1 mA. A similar statement can be made about the desired insensitivity of the dc drain current to the wide variations encountered in V_t of discrete MOSFETs.

A second consideration in bias design is locating the dc operating point in the active region of operation of the transistor so as to obtain high voltage gain while allowing for the required output signal swing without the transistor leaving the active region at any time (in order to avoid nonlinear distortion). We discussed this point in Section 5.1.7.

In biasing of MOSFET, it will be seen that good bias designs incorporate a feedback mechanism that works to keep the dc bias point as constant as possible.

In order to keep matters simple and thus focus our attention on significant issues, we will neglect the Early effect; that is assume $\lambda = 0$ or $V_A = \infty$. This is certainly allowed in initial designs of discrete circuits. Of course, the design can be fine-tuned at a later point with the assistance of a circuit-simulation program such as SPICE.

5.3 Biasing in MOSFET Amplifier Circuits

5.3.1 Biasing by Fixing V_{GS}

The most straightforward approach to biasing a MOSFET is to fix its gate-to-source voltage V_{GS} to the value required to provide the desired I_D . This voltage value can be derived from the power-supply voltage V_{DD} through the use of an appropriate voltage divider, as shown in Fig. 7.47(a). Alternatively, it can be derived from another suitable reference voltage that might be available in the system. Independent of how the voltage V_{GS} may be generated, this is *not* a good approach to biasing a MOSFET. To understand the reason for this statement, recall that²

$$I_D = \frac{1}{2} n \mu_{ox} \frac{W}{L} (V_{GS} - V_t)$$

and note that the values of the threshold voltage V_t , the oxide-capacitance C_{ox} , and (to a lesser extent) the transistor aspect ratio W/L vary widely among devices of supposedly the same size and type.

This is certainly the case for discrete devices, in which large spreads in the values of these parameters occur among devices of the same manufacturer's part number. The spread can also be large in integrated circuits, especially among devices fabricated on different wafers and certainly between different batches of wafers. Furthermore, both V_t and μ_n depend on temperature, with the result that if we fix the value of V_{GS} , the drain current I_D becomes very much temperature dependent.

To emphasize the point that biasing by fixing V_{GS} is not a good technique, we show in Fig. 5.47 two i_D-v_{GS} characteristic curves representing extreme values in a batch of MOSFETs of the same type. Observe that for the fixed value of V_{GS} , the resultant spread in the values of the drain current can be substantial.

5.3.1 Biasing by Fixing V_G and Connecting a Resistance in the Source

An excellent biasing technique for discrete MOSFET circuits consists of fixing the dc voltage at the gate, V_G , and connecting a resistance in the source lead, as shown in Fig. 7.48(a). For this circuit

²That is indeed what we were doing in Section 7.1. However, the amplifier circuits studied there were

conceptual ones, not actual practical circuits. Our purpose in this section is to study the latter.

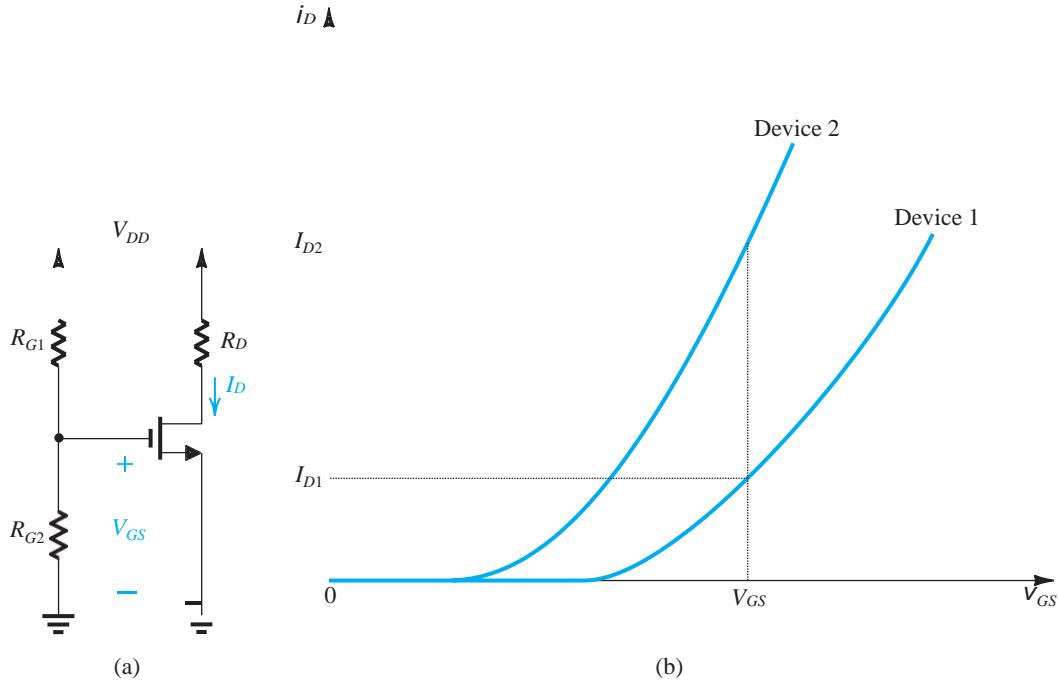


Figure 5.47 (a) Biasing the MOSFET with a constant V_{GS} generated from V_{DD} using a voltage divider (R_{G1}, R_{G2}); (b) the use of fixed bias (constant V_{GS}) can result in a large variability in the value of I_D . Devices 1 and 2 represent extremes among units of the same type.

we can write

$$V_G = V_{GS} + R_S I_D \quad (5.137)$$

Now, if V_G is much greater than V_{GS} , I_D will be mostly determined by the values of V_G and R_S . However, even if V_G is not much larger than V_{GS} , resistor R_S provides *negative feedback*, which acts to stabilize the value of the bias current I_D . To see how this comes about, consider what happens when I_D increases for whatever reason. Equation (5.137) indicates that since V_G is constant, V_{GS} will have to decrease. This in turn results in a decrease in I_D , a change that is opposite to that initially assumed. Thus the action of R_S works to keep I_D as constant as possible.

Figure 5.48(b) provides a graphical illustration of the effectiveness of this biasing scheme. Here too we show the i_D-v_{GS} characteristics for two devices that represent the extremes of a batch of MOSFETs. Superimposed on the device characteristics is a straight line that represents the constraint imposed by the bias circuit—namely, Eq. (5.137). The intersection of this straight line

with the i_D-v_{GS} characteristic curve provides the coordinates (I_D and V_{GS}) of the bias point. Observe that compared to the case of fixed V_{GS} , here the variability obtained in I_D is much smaller. Also, note that the variability decreases as V_G and R_S are made larger (thus providing a bias line that is less steep).

Two possible practical discrete implementations of this bias scheme are shown in Fig. 5.48(c) and (e). The circuit in Fig. 5.48(c) utilizes one power-supply V_{DD} and derives V_G

⁷The action of R in stabilizing the value of the bias current I is not unlike that of the resistance R which we included in the source lead of a CS amplifier in Section 5.3.4. In the latter case also, R_S works to reduce the change in iD with the result that the amplifier gain is reduced.

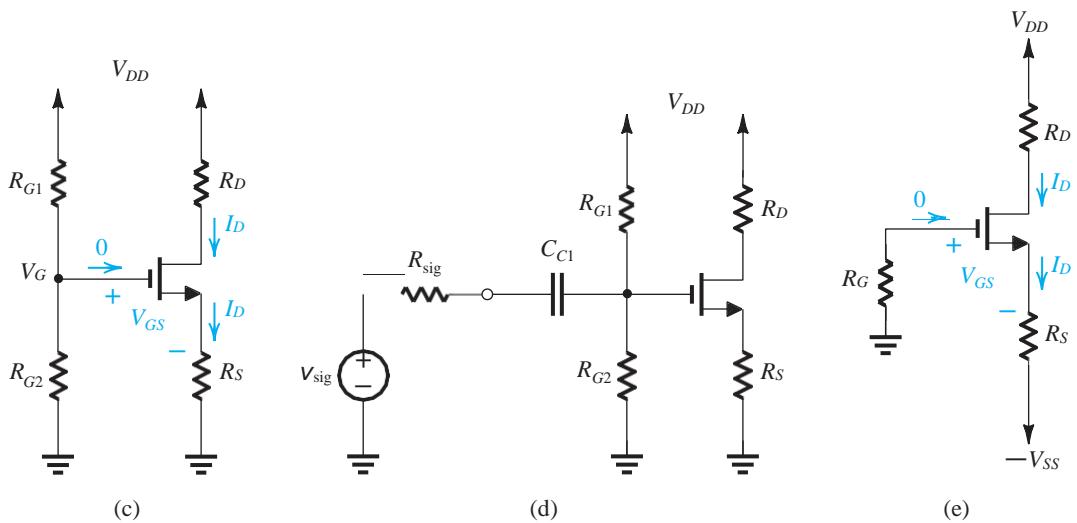
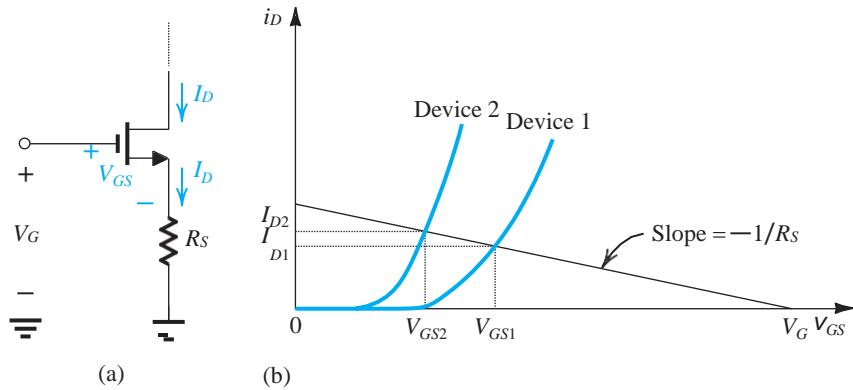


Figure 5.48 Biasing using a fixed voltage at the gate, V_G , and a resistance in the source lead, R_s : (a) basic arrangement; (b) reduced variability in I_D ; (c) practical implementation using a single supply; (d) coupling of a signal source to the gate using a capacitor C_{C1} ; (e) practical implementation using two supplies.

through a voltage divider (R_{G1}, R_{G2}). Since $I_G = 0$, R_{G1} and R_{G2} can be selected to be very large (in the megohm range), allowing the MOSFET to present a large input resistance to a signal source that may be connected to the gate through a coupling capacitor, as shown in Fig. 7.48(d). Here capacitor C_{C1} blocks dc and thus allows us to couple the signal v_{sig} to the amplifier input without disturbing the MOSFET dc bias point. The value of C_{C1} should be selected large enough to approximate a short circuit at all signal frequencies of interest. We shall study capacitively coupled MOSFET amplifiers, which are suitable only in discrete-circuit design, in Section 5.5. Finally, note that in the circuit of Fig. 5.48(c), resistor R_D is selected to be as large as possible to obtain high gain but small enough to allow for the desired signal swing at the drain while keeping the MOSFET in saturation at all times.

When two power supplies are available, as is often the case, the somewhat simpler bias

arrangement of Fig. 5.48(e) can be utilized. This circuit is an implementation of Eq. (5.137), with V_G replaced by V_{SS} . Resistor R_G establishes a dc ground at the gate and presents a high inputresistance to a signal source that may be connected to the gate through a coupling capacitor.

Example 5.11

It is required to design the circuit of Fig. 7.48(c) to establish a dc drain current $I_D = 0.5$ mA. The MOSFET is specified to have $V_t = 1$ V and $k^* W/L = 1$ mA/V². For simplicity, neglect the channel-length modulation effect (i.e., assume $\lambda = 0$). Use a power-supply $V_{DD} = 15$ V. Calculate the percentage change in the value of I_D obtained when the MOSFET is replaced with another unit having the same $k^* W/L$ but $V_t = 1.5$ V.

Solution

As a rule of thumb for designing this classical biasing circuit, we choose R_D and R_S to provide one-third of the power-supply voltage V_{DD} as a drop across each of R_D , the transistor (i.e., V_{DS}), and R_S . For $V_{DD} = 15$ V this choice makes $V_D = +10$ V and $V_S = +5$ V. Now, since I_D is required to be 0.5 mA, we can find the values of R_D and R_S as follows:

$$R_D = \frac{V_{DD} - V_D}{I_D} = \frac{15 - 10}{0.5} = 10 \text{ k}\Delta$$

$$R_S = \frac{V_S}{I_D} = \frac{5}{0.5} = 10 \text{ k}\Delta$$

The required value of V_{GS} can be determined by first calculating the overdrive voltage V_{OV} from

$$I_D = k^* \frac{(W/L)V_{OV}^2}{2}$$

$$0.5 = \frac{1}{2} \times 1 \times V_{OV}^2$$

which yields $V_{OV} = 1$ V, and thus,

$$V_{GS} = V_t + V_{OV} = 1 + 1 = 2\text{V}$$

Now, since $V_S = +5$ V, V_G must be

$$V_G = V_S + V_{GS} = 5 + 2 = 7\text{V}$$

To establish this voltage at the gate we may select $R_{G1} = 8 \text{ M}\Delta$ and $R_{G2} = 7 \text{ M}\Delta$. The final circuit is shown in Fig. 7.49. Observe that the dc voltage at the drain (+10 V) allows for a positive signal swing of +5 V (i.e., up to V_{DD}) and a negative signal swing of 4 V [i.e., down to $(V_G - V_t)$].

If the NMOS transistor is replaced with another having $V_t = 1.5$ V, the new value of I_D can be found as follows:

$$I_D = \frac{1}{2} \times 1 \times \frac{V_{GS} - 1.5}{k^*}^2 \quad (7.138)$$

$$V_G = V_{GS} + I_D R_S$$

$$7 = V_{GS} + 10 I_D \quad (7.139)$$

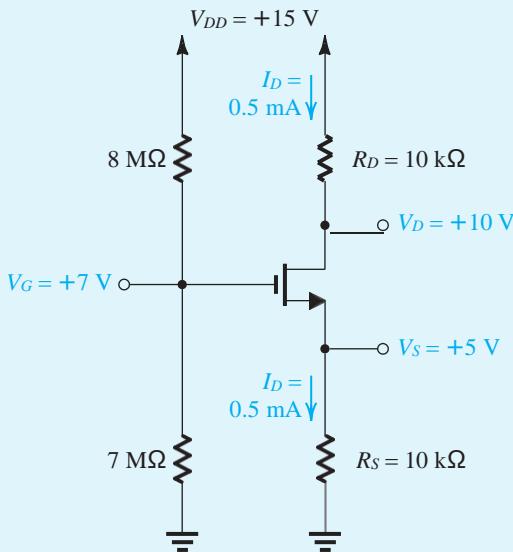


Figure 7.49 Circuit for Example 7.11.

Solving Eqs. (7.138) and (7.139) together yields

$$I_D = 0.455 \text{ mA}$$

Thus the change in I_D is

$$\beta I_D = 0.455 - 0.5 = -0.045 \text{ mA}$$

which is $\frac{-0.045}{0.5} \times 100 = -9\%$ change.

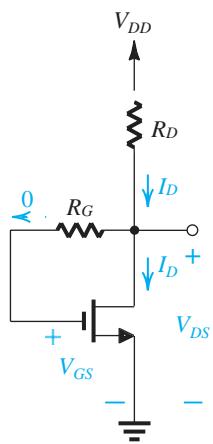


Figure 5.50 Biasing the MOSFET using a large drain-to-gate feedback resistance, R_G .

5.3.2 Biasing Using a Drain-to-Gate Feedback Resistor

A simple and effective discrete- circuit biasing arrangement utilizing a feedback resistor connected between the drain and the gate is shown in Fig. 5.50. Here the large feedback resistance R_G (usually in the megohm range) forces the dc voltage at the gate to be equal to that at the drain (because $I_G = 0$). Thus we can write

$$V_{GS} = V_{DS} = V_{DD} - R_D I_D$$

which can be rewritten in the form

$$V_{DD} = V_{GS} + R_D I_D \quad (5.140)$$

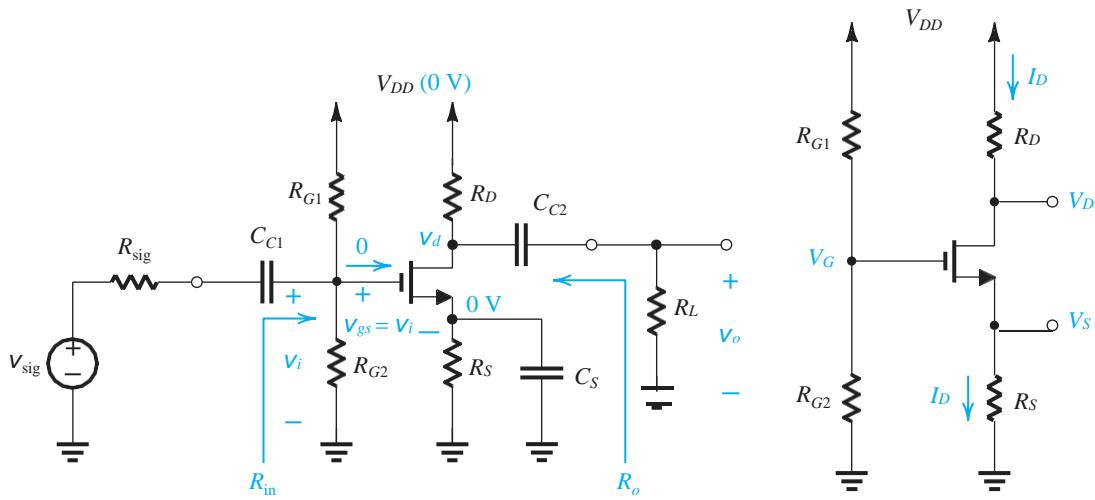
which is identical in form to Eq. (5.137), which describes the operation of the bias scheme discussed above [that in Fig. 5.48(a)]. Thus, here too, if I_D for some reason changes, say increases, then Eq. (5.140) indicates that V_{GS} must decrease. The decrease in V_{GS} in turn causes a decrease in I_D , a change that is opposite in direction to the one originally assumed. Thus the negative feedback or degeneration provided by R_G works to keep the value of I_D as constant as possible.

The circuit of Fig. 5.50 can be utilized as an amplifier by applying the input voltage signal to the gate via a coupling capacitor so as not to disturb the dc bias conditions already established. The amplified output signal at the drain can be coupled to another part of the circuit, again via a capacitor. We considered such an amplifier circuit in Section 5.2 (Example 5.3).

5.4 Common-Source (CS) Amplifier using MOSFETs

As mentioned in Section 5.3, the common-source (CS) configuration is the most widely used of all MOSFET amplifier circuits. A common-source amplifier realized using the bias circuit of Fig. 5.48(c) is shown in Fig. 5.55(a). Observe that to establish a **signal ground**, or an **ac ground** as it is sometimes called, at the source, we have connected a large capacitor, C_S , between the source and ground. This capacitor, usually in the microfarad range, is required to provide a very small impedance (ideally, zero impedance—i.e., in effect, a short circuit) at all signal frequencies of interest. In this way, the signal current passes through C_S to ground and thus *bypasses* the resistance R_S ; hence, C_S is called a **bypass capacitor**. Obviously, the lower the signal frequency, the less effective the bypass capacitor becomes. This issue will be studied in Section 10.1. For our purposes here we shall assume that C_S is acting as a perfect short circuit and thus is establishing a zero signal voltage at the MOSFET source.

To prevent disturbances to the dc bias current and voltages, the signal to be amplified, shown as voltage source v_{sig} with an internal resistance R_{sig} , is connected to the gate through a large capacitor C_{C1} . Capacitor C_{C1} , known as a **coupling capacitor**, is required to act as a perfect short circuit at all signal frequencies of interest while blocking dc. Here again, we note that as the signal frequency is lowered, the impedance of C_{C1} (i.e., $1/j\omega C_{C1}$) will increase and its effectiveness as a coupling capacitor will be correspondingly reduced. This problem, too, will be considered in Section 10.1 in connection with the dependence of the amplifier



5.4.1.1 (b)

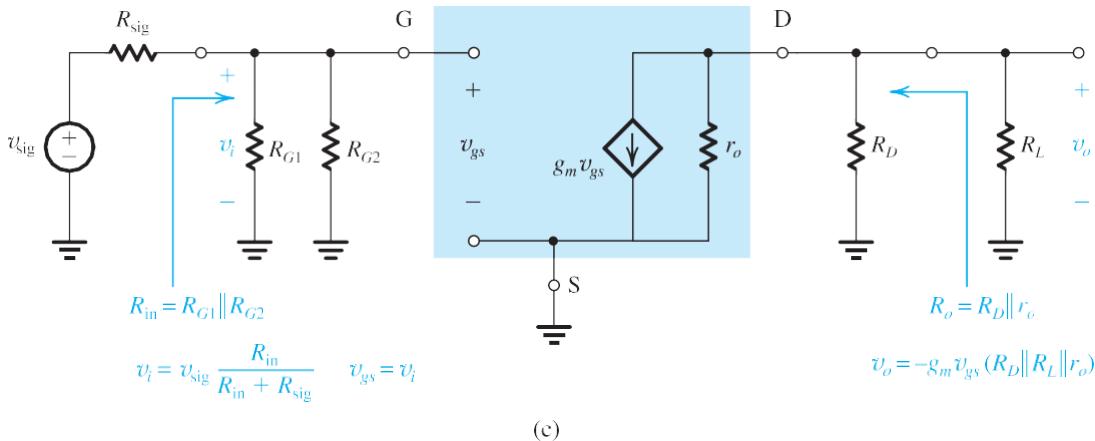


Figure 5.55 (a) A common-source amplifier using the classical biasing arrangement of Fig. 5.48(c).
Circuit for determining the bias point. (c) Equivalent circuit and analysis.

operation on frequency. For our purposes here we shall assume that C_{C1} is acting as a perfect short circuit as far as the signal is concerned.

The voltage signal resulting at the drain is coupled to the load resistance R_L via another coupling capacitor C_{C2} . We shall assume that C_{C2} acts as a perfect short circuit at all signal frequencies of interest and thus that the output voltage v_o v_d . Note that R_L can be either an actual load resistor, to which the amplifier is required to provide its output voltage signal, or it can be the input resistance of another amplifier stage in cases where more than one stage of amplification is needed. (We will study multistage amplifiers in Chapter 9).

Since a capacitor behaves as an open circuit at dc, the circuit for performing the dc bias design and analysis is obtained by open-circuiting all capacitors. The resulting circuit is shown in Fig. 5.55(b) and can be designed as discussed in Section 5.4.1.

To determine the terminal characteristics of the CS amplifier of Fig. 5.55(a)—that is, its input resistance, voltage gain, and output resistance—we replace the MOSFET with its hybrid- π small-signal model, replace V_{DD} with a signal ground, and replace all coupling and bypass capacitors with short circuits. The result is the circuit in Fig. 5.55(c). Analysis is straightforward and is shown on the figure, thus

$$R_{in} = R_{G1} \parallel R_{G2} \quad (5.149)$$

which shows that to keep R_{in} high, large values should be used for R_{G1} and R_{G2} , usually in the megohm range. The overall voltage gain G_v is

$$G_v = -\frac{R_{in}}{R_{in} + R_{sig}} g_m (R_D \parallel R_L \parallel r_o) \quad (5.150)$$

Observe that we have taken r_o into account, simply because it is easy to do so. Its effect, however, is usually small (this is not the case for IC amplifiers, as will be explained in Chapter 8). Finally, to encourage the reader to do the small-signal analysis directly on the original circuit diagram, with the MOSFET model used implicitly, we show some of the analysis on the circuit of Fig. 5.55(a).

Thus source degeneration increases the output resistance of the CS amplifier from r_o to $(1 + g_m R_s) r_o$, again by the same factor $(1 + g_m R_s)$. In Chapter 11, we will find that R_s introduces negative (degenerative) feedback of an amount $(1 + g_m R_s)$.

5.4.1 The Body Effect

Since in the CS amplifier the source cannot be connected to the substrate, the body effect (see Section 5.4) plays a role in the operation of the CS amplifier. It turns out, however, that taking the body effect into account in the analysis of the CS circuit is a very simple matter. To see how this can be done, refer to Fig. 5.23(a) and recall that the body terminal acts as another gate for the MOSFET. Thus, just as a signal voltage v_{gs} between the gate and the source gives

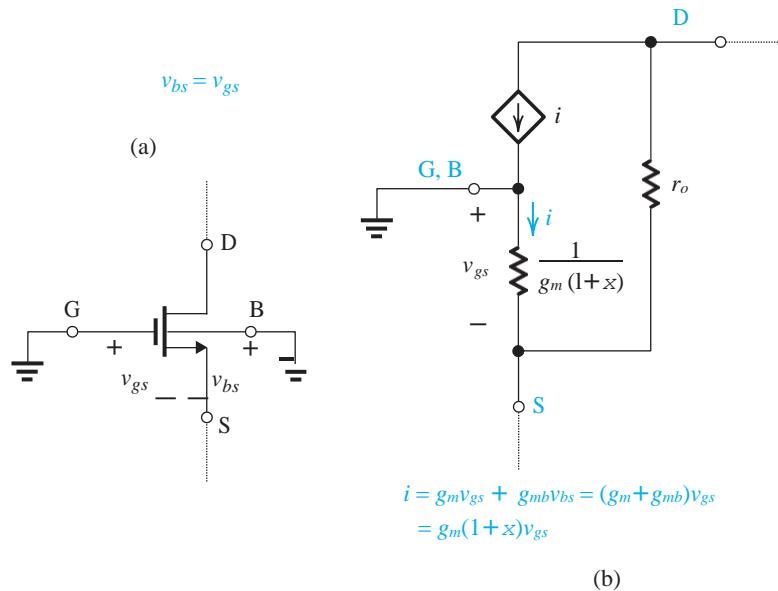


Figure 5.23 The body effect can be easily taken into account in the analysis of the CS circuit by replacing g_m by $(1 + \chi)g_m$, where $\chi = g_{mb}/g_m = 0.1$ to 0.2.

2 Marks

1. Define h parameters?

One of a set of four transistor equivalent-circuit parameters that conveniently specify transistor performance for small voltages and currents in a particular circuit. Also known as hybrid parameter.

2. Give the h_{ie} and h_{oe} equations of BJT?

$h_{11} = h_{ie}$ - The input impedance of the transistor (corresponding to the emitter resistance r_e). Unit ohms Ω .

$$h_{11} = \left. \frac{V_i}{I_i} \right|_{V_o = 0}$$

$h_{22} = h_{oe}$ - The output impedance of transistor. This term is usually specified as admittance and has to be inverted to convert it to impedance. Units' siemens S.

$$h_{22} = \left. \frac{I_o}{V_o} \right|_{I_i = 0}$$

2. Define amplifier

A device which accepts an input signal and produces an output signal proportional to the input is called an amplifier.

3. List out the some features of a differential amplifier?

High differential voltage gain. Low common mode gain.

High CMRR.

Two input terminals. High input impedance.

Large bandwidth.

Low offset voltages and currents.

Low output impedance.

5. Explain the ideal characteristics of voltage amplifier?

Infinite Open Loop Gain, Infinite Input Impedance, Zero Output Impedance, Infinite Bandwidth, Zero Output Offset, and Zero Noise Contribution.

6. Explain about CS Amplifier?

The CS amplifier is a small signal amplifier. For good bias stability, the source resistor voltage drop should be as large as possible. Where the supply voltage is small, V_s may be reduced to a minimum to allow for the minimum level of $V_{ds} \cdot R_2$. R_2 is usually selected as $1M\ \Omega$ or less as for BJT capacitor coupled circuit coupling and bypass capacitors should be selected to have the smallest possible capacitance values. The largest capacitor in the circuit sets the circuit low 3dB frequency (capacitor C_2). Generally to have high input impedance FET is used. As in BJT circuit R_L is usually much larger than Z_o and Z_i is often much larger than

7. What is stability factor?

Stability factor is defined as the rate of change of collector current with respect to the rate of change of reverse saturation current.

8. What is biasing?

To use the transistor in any application it is necessary to provide sufficient voltage and current to operate the transistor. This is called biasing.

9. What are the requirements for biasing circuits?

- The q point must be taken at the Centre of the active region of the output characteristics
- Stabilize the collector current against the temperature variations.
- Make the q point independent of the transistor parameters.
- When the transistor is replaced, it must be of same type.

10. List out the different types of biasing.

Voltage divider bias, Base bias, Emitter feed back bias, Collector feedback bias, Emitter bias.

11. What do you meant by thermal runway?

Due to the self heating at the collector junction, the collector current rises. This causes damage to the device. This phenomenon is called thermal runway.

Essay Questions:

1. Define drain resistance and amplification factor of an FET. Compare BJT and FET?

Draw the diagram,

Express the equations and compare the BJT & FET.

2. Explain the construction and operation of n-channel JFET?

Draw the diagrams,

Explanation of construction and operation.

3. Draw and explain the drain and transfer characteristics of n-channel JFET?

Draw the diagrams,

Explain the characteristics of n-channel JFET.

4. List out different FET biasing methods and then explain the same?

Draw the diagrams,

Explain the different methds of FET.

5. Explain the V_{ds}/I_{ds} characteristics and operation of depletion type MOSFET. With suitable diagram?

Draw the diagrams,

Explantion of characteristics and operation of MOSFET.

MCQ's

1. Which of the following statement is true about FET (b)

- a) It has high output impedance
- b) It has high input impedance
- c) It has low input impedance
- d) It does not offer any resistance

2. Comparing the size of BJT and FET, choose the correct statement? (a)

- a) BJT is larger than the FET
- b) BJT is smaller than the FET
- c) Both are of same size
- d) Depends on application

3. What is the value of current when the gate to source voltage is less than the pinch off voltage? (d)

- a) 1A
- b) 5A
- c) 100A
- d) 0

4. To use FET as a voltage controlled resistor, in which region it should operate

(a)

- a) Ohmic region b) cut off c) Saturation d) cut off and saturation

5. For an n-channel FET, What is the direction of current flow? (b)

- a) Source to drain b) Drain to source c) Gate to source d) Gate to drain

6. For a p-channel FET, What is the direction of current flow? (a)

- a) Source to drain b) Drain to source c) Gate to source d) Gate to drain

7. Aspect ratio of the MOSFET has the units of (a)

- a) No units b) m c) m^2 d) m^{-1}

8. The MOSFET transconductance parameter is the product of (d)

- a) Process transconductance and inverse of aspect ratio
b) Inverse of Process transconductance and aspect ratio
c) Inverse of Process transconductance and inverse of aspect ratio
d) Process transconductance and aspect ratio

9. For a p channel MOSFET which of the following is not true? (b)

- a) The source and drain are a p type semiconductor
b) The induced channel is p type region which is induced by applying a positive potential to the gate
c) The substrate is a n type semiconductor
d) None of the mentioned

10. With the potential difference between the source and the drain kept small (V_{DS} is small), the MOSFET behaves as a resistance whose value varies

with the overdrive voltage (b)

- a) Linearly
b) Inversely
c) Exponentially
d) Logarithmically