

It would be convenient to have some general principles from which it would be possible to estimate response curve shapes from the initial conditions. It will be recalled that this was done in the case of R-C and R-L circuits, where the response consists of a single exponential. Those cases are handled by fitting a suitable exponential curve between initial and final values, where the parameters of the circuit determine the time constant of the exponential.

We would like to do the same thing in this case, but now there are two exponential functions, with time constants $1/s_a$ and $1/s_b$. This makes the problem more difficult, as illustrated by the three examples of Fig. 6-15. Depending on relative signs and sizes of A_1 and B_1 , there are three possible curves. Thus, it is apparent that consideration of Eqs. (6-25) does not yield all the information we need. The reason is that both initial conditions affect the curve shape, whereas in looking at Fig. 6-15 only $x_1(0)$ is in evidence. The need to incorporate both initial conditions naturally suggests looking at the phase plane for the possible answer.

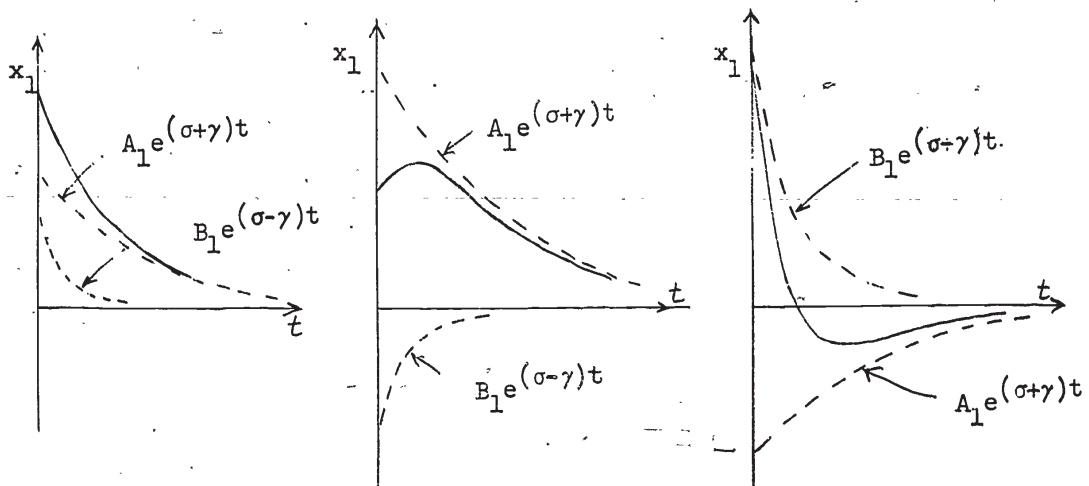


Figure 6-15.

There is an initial state point having coordinates $x_1(0)$ and $x_2(0)$, and a final state point at the origin. First, let us consider the behavior of the trajectory as t approaches infinity. To do this, consider the ratio x_2/x_1 as

⁺The remaining discussion of case (1) can be omitted without loss of continuity.

1 t approaches infinity. It is recalled that $e^{(\sigma-\gamma)t}$ approaches zero faster
 2 than $e^{-(\sigma+\gamma)t}$. Therefore, for very large t we have (approximately)

$$\frac{x_2}{x_1} = \frac{A_2 e^{(\sigma+\gamma)t}}{A_1 e^{-(\sigma+\gamma)t}} = \frac{A_2}{A_1}$$

This ratio is negative, and given by

$$\frac{A_2}{A_1} = \frac{\sigma-\gamma}{\omega_0} \quad (6-26)^+$$

3 In other words, the trajectory must approach the origin in such a way that
 4 the ratio x_2/x_1 will be nearly constant, which means the state point approaches
 the origin so as to be tangent to a radial line of slope $(\sigma-\gamma)/\omega_0$.

This result is portrayed in Fig. 6-16a. Some ambiguity remains, however.
 Equation (6-26) gives no information for making a choice among the four cases
 shown in Fig. 6-16a, all of which are tangent to the diagonal line at the origin.

Further information depends upon investigating properties of the trajectory
 at finite values of t . Help is obtained by returning to the original differential
 equations, to determine those points in the phase plane at which the velocity
 of the state point is toward the origin.

Referring to Fig. 6-17, it is seen that this will occur when

$$\frac{dx_2}{dx_1} = \frac{x_2}{x_1}$$

⁺From Eqs. (6-23) and (6-24)

$$\frac{A_2}{A_1} = \frac{\omega_0 x_1(0) - (\sigma-\gamma)x_2(0)}{(\sigma+\gamma)x_1(0) - \omega_0 x_2(0)} = \left(\frac{\sigma-\gamma}{\omega_0} \right) \left[\frac{(\frac{\omega_0}{\sigma-\gamma})x_1(0) - x_2(0)}{(\frac{\sigma+\gamma}{\omega_0})x_1(0) - x_2(0)} \right]$$

The quantity in brackets is unity, as may be seen by referring to Eq. (6-17)
 which shows that $\omega_0^2 = \sigma^2 - \gamma^2$, or $\omega_0/(\sigma-\gamma) = (\sigma+\gamma)/\omega_0$.

From Eqs. (6-5) and (6-6),

$$\frac{dx_2}{dx_1} = - \frac{\omega_0 x_1}{\alpha x_1 + \omega_0 x_2} = - \frac{\omega_0}{\alpha + \omega_0 \frac{x_2}{x_1}}$$

Thus, for the condition specified

$$- \frac{\omega_0}{\alpha + \omega_0 \frac{x_2}{x_1}} = \frac{x_2}{x_1}$$

or

$$\omega_0 \left(\frac{x_2}{x_1} \right)^2 + \alpha \left(\frac{x_2}{x_1} \right) + \omega_0 = 0$$

which has the two solutions

$$\begin{aligned} \frac{x_2}{x_1} &= - \frac{\alpha}{2\omega_0} \pm \frac{1}{\omega_0} \sqrt{\frac{\alpha^2}{4} - \omega_0^2} \\ &= \frac{\sigma+\gamma}{\omega_0} \quad \text{and} \quad \frac{\sigma-\gamma}{\omega_0} \end{aligned}$$

These are constant values of the ratio x_2/x_1 , which correspond to values on the radial straight lines shown in Fig. 6-16b. One of these is the line previously shown in Fig. 6-16a.

A trajectory cannot cross either of these lines. To do so would imply a contradiction, since it would require a velocity component normal to the line. This observation leads to the further conclusion that for any initial state point lying within the shaded sectors, the entire trajectory must lie within the sector, and have a form of the general nature illustrated by the two examples in Fig. 6-16b.[†]

If the initial state point is in one of the regions shown shaded in Fig. 6-16c, it is clear from observation of the figure that the trajectories must be as indicated. All other possibilities would either cross one of the diagonal lines, or would not approach the origin tangentially to the heavy line.

[†]It can be shown that $dx_2/dx_1 > x_2/x_1$ at any point within the shaded sectors of Fig. 6-16b.

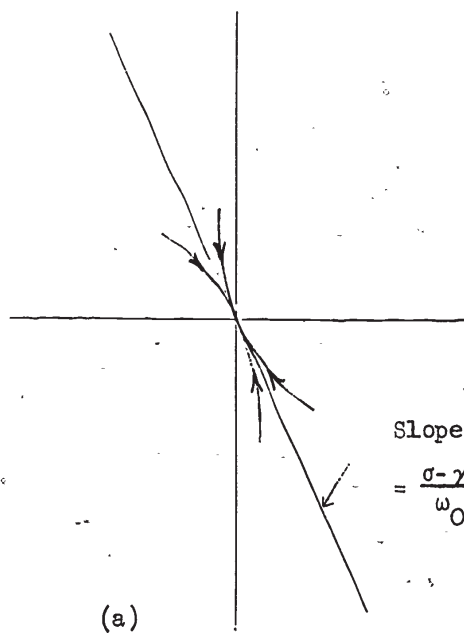
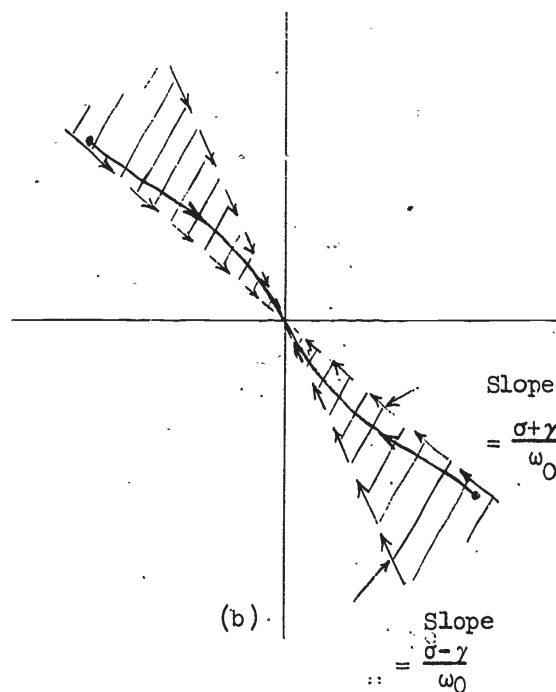
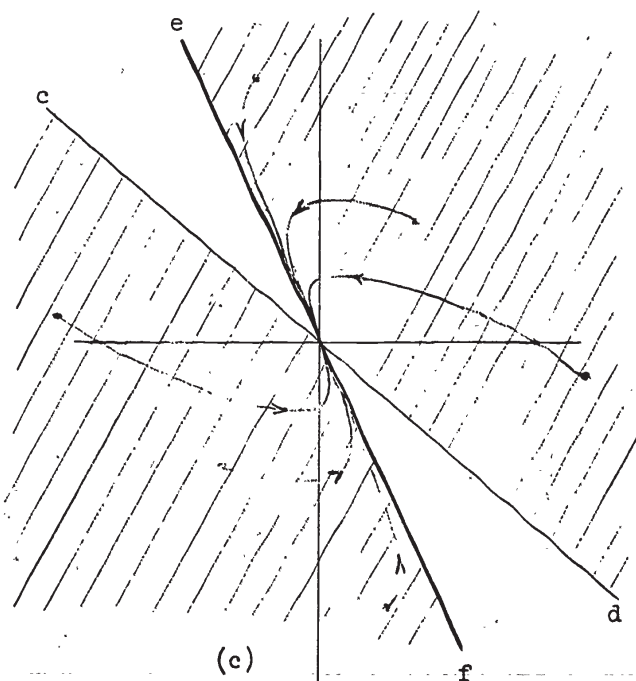


Figure 6-16.

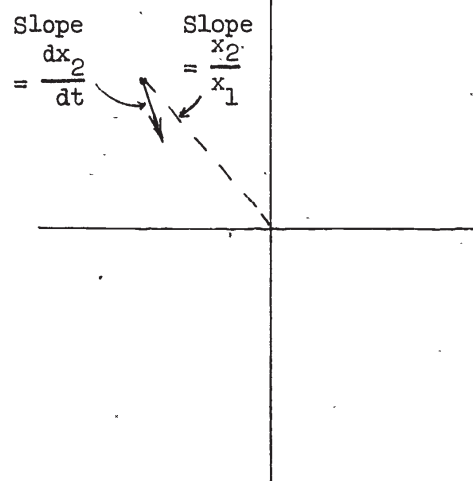


Figure 6-17.

Of course, this information does not determine the trajectory with precision. However, one further clue is obtained by observing that the trajectory must be normal to the x_2 axis as it crosses that axis. This is because $dx_2/dt = \omega_0 x_1 = 0$ at points on this axis.

These various trajectories can be used, in the customary manner of projecting onto the axes, to estimate shapes of the x_1 and x_2 curves. The existence of two sets of "forbidden" lines, which cannot be crossed by the trajectory gives a vivid confirmation that the responses obtained in the case of real eigenvalues cannot change sign more than once.

Case (2): $s_a = s_b$

The solution for this case is obtained from the previous one, by investigating the limit as γ approaches zero. To begin, the case (1) solutions are written explicitly, by multiplying A_1 and A_2 by $e^{(\sigma+\gamma)t}$, and B_1 and B_2 by $e^{(\sigma-\gamma)t}$, where the A's and B's are given by Eqs. (6-23) and (6-24). In the interest of brevity, let the factor $e^{\sigma t}$ be extracted while the remaining terms are combined with $x_1(0)$ and $x_2(0)$ as factors.

The results are

$$x_1 = \frac{e^{\sigma t}}{2\gamma} \left\{ x_1(0) [(\sigma+\gamma)e^{\gamma t} - (\sigma-\gamma)e^{-\gamma t}] - x_2(0) [\omega_0 e^{\gamma t} - \omega_0 e^{-\gamma t}] \right\}$$

$$x_2 = \frac{e^{\sigma t}}{2\gamma} \left\{ x_1(0) [\omega_0 e^{\gamma t} - \omega_0 e^{-\gamma t}] - x_2(0) [(\sigma-\gamma)e^{\gamma t} - (\sigma+\gamma)e^{-\gamma t}] \right\}$$

or

$$x_1 = e^{\sigma t} \left[\sigma x_1(0) \left(\frac{e^{\gamma t} - e^{-\gamma t}}{2\gamma} \right) + x_1(0) \left(\frac{e^{\gamma t} + e^{-\gamma t}}{2} \right) - \omega_0 x_2(0) \left(\frac{e^{\gamma t} - e^{-\gamma t}}{2\gamma} \right) \right] \quad (a)$$

(6-27)

$$x_2 = e^{\sigma t} \left[\omega_0 x_1(0) \left(\frac{e^{\gamma t} - e^{-\gamma t}}{2\gamma} \right) - \sigma x_2(0) \left(\frac{e^{\gamma t} - e^{-\gamma t}}{2\gamma} \right) + x_2(0) \left(\frac{e^{\gamma t} + e^{-\gamma t}}{2} \right) \right] \quad (b)$$

Now let γ approach zero. The quantity $(e^{\gamma t} + e^{-\gamma t})/2$ approaches unity. However, $(e^{\gamma t} - e^{-\gamma t})/2\gamma$ becomes indeterminate. One way to obtain the limit is to use l'Hospital's rule, differentiating numerator and denominator with respect to γ . The result is

$$\frac{te^{\gamma t} + te^{-\gamma t}}{2} = t$$

1 in the limit as γ goes to zero. Thus, when $\gamma = 0$, Eqs. (6-27) become

$$x_1 = e^{\sigma t} [\sigma t x_1(0) + x_1(0) - \omega_0 t x_2(0)]$$

$$2 \quad x_2 = e^{\sigma t} [\omega_0 t x_1(0) - \sigma t x_2(0) + x_2(0)]$$

This can be further simplified, since in this case

$$3 \quad \gamma = \sqrt{\frac{\alpha^2}{4} - \omega_0^2} = 0$$

Thus,

$$\omega_0 = \frac{\alpha}{2} = -\sigma$$

4 (rejecting $-\alpha/2$ because $\omega_0 = 1/\sqrt{LC}$ must be positive). The final solutions are

$$x_1 = [x_1(0) + x_2(0)] (\sigma t) e^{\sigma t} + x_1(0) e^{\sigma t} \quad (a)$$

(6-28)

$$5 \quad x_2 = [-x_1(0) + x_2(0)] (\sigma t) e^{\sigma t} + x_2(0) e^{\sigma t} \quad (b)$$

6 Although the equations are different from those for case (1), the curves of x_1 and x_2 are similar. (The phase plane interpretation of Fig. 6-16 applies, but lines c-d and e-f now coincide, abolishing one of the regions.)⁺

Case (3): s_a and s_b are complex.

7 For this case, γ is imaginary. Therefore, it is convenient to change notation to

$$s_a = \sigma + j\omega$$

(6-29)

$$s_b = \sigma - j\omega$$

where

$$\omega = -j\gamma = \sqrt{\omega_0^2 - \frac{\alpha^2}{4}} \quad (6-30)$$

⁺This statement refers to material which you may have omitted, in accordance with a previous footnote.

The algebra leading up to Eqs. (6-27) remains valid if γ is imaginary. Therefore, the required solutions are obtained immediately by replacing γ by $j\omega$, giving

$$x_1 = e^{\sigma t} \left[\sigma x_1(0) \left(\frac{e^{j\omega t} - e^{-j\omega t}}{2j\omega} \right) + x_1(0) \left(\frac{e^{j\omega t} + e^{-j\omega t}}{2} \right) - \omega_0 x_2(0) \left(\frac{e^{j\omega t} - e^{-j\omega t}}{2j\omega} \right) \right]$$

$$x_2 = e^{\sigma t} \left[\omega_0 x_1(0) \left(\frac{e^{j\omega t} - e^{-j\omega t}}{2j\omega} \right) - \sigma x_2(0) \left(\frac{e^{j\omega t} - e^{-j\omega t}}{2j\omega} \right) + x_2(0) \left(\frac{e^{j\omega t} + e^{-j\omega t}}{2} \right) \right]$$

However, the exponential combinations represent trigonometric functions, as follows:

$$\frac{e^{j\omega t} - e^{-j\omega t}}{2j} = \sin \omega t$$

$$\frac{e^{j\omega t} + e^{-j\omega t}}{2} = \cos \omega t$$

Thus, the solutions become

$$x_1 = \left\{ \left[\frac{\sigma x_1(0) - \omega_0 x_2(0)}{\omega} \right] \sin \omega t + x_1(0) \cos \omega t \right\} e^{\sigma t} \quad (a)$$

$$x_2 = \left\{ \left[\frac{\omega_0 x_1(0) - \sigma x_2(0)}{\omega} \right] \sin \omega t + x_2(0) \cos \omega t \right\} e^{\sigma t} \quad (b)$$

The two trigonometric functions within the brackets combine to yield sinusoidal waves of angular frequency ω . It is recalled from Eq. (6-31) that

$$\omega = \sqrt{\omega_0^2 - \frac{\alpha^2}{4}}$$

but, since $\sigma = -\alpha/2$, this is also

$$\omega = \sqrt{\omega_0^2 - \sigma^2} = \omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} \quad (6-32)$$

This is an interesting result, giving an analytic proof of the fact mentioned in connection with Fig. 6-7, that ω , the natural angular frequency of the wave, is less than ω_0 .

For interpretation of these equations, it is convenient to combine the trigonometric functions. The quantities in brackets are steady state sinusoids, to which phasor ideas apply, with the results⁺

$$x_1 = X_0 \cos \left[\omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} t - \theta_1 \right] e^{\sigma t} \quad (a) \quad (6-33)$$

$$x_2 = X_0 \cos \left[\omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} t - \theta_2 \right] e^{\sigma t} \quad (b)$$

where

$$X_0 = \sqrt{\frac{\omega_0^2 \{ [x_1(0)]^2 + [x_2(0)]^2 \} - 2\omega_0 \sigma x_1(0)x_2(0)}{\omega_0^2 - \sigma^2}} \quad (6-34)$$

$$\theta_1 = \arctan \left[\frac{\sigma - \omega_0 \frac{x_2(0)}{x_1(0)}}{\sqrt{\omega_0^2 - \sigma^2}} \right] \quad (a) \quad (6-35)$$

$$\theta_2 = \arctan \left[\frac{\omega_0 \frac{x_1(0)}{x_2(0)} - \sigma}{\sqrt{\omega_0^2 - \sigma^2}} \right] \quad (b)$$

⁺Using x_1 as an example, since $\sin \omega t = \cos(\omega t - \frac{\pi}{2})$, the phasors for the $\sin \omega t$ and $\cos \omega t$ terms are, respectively,

$$-j \left[\frac{\sigma x_1(0) - \omega_0 x_2(0)}{\omega} \right] \quad \text{and} \quad x_1(0)$$

Accordingly, the phasor for the sum is

$$\begin{aligned} & \sqrt{x_1(0)^2 + \frac{\sigma^2 [x_1(0)]^2 - 2\sigma \omega_0 x_1(0)x_2(0) + \omega_0^2 [x_2(0)]^2}{\omega^2}} \angle -\arctan \left[\frac{\sigma - \omega_0 \frac{x_2(0)}{x_1(0)}}{\omega} \right] \\ &= \sqrt{\frac{\omega_0^2 \{ [x_1(0)]^2 + [x_2(0)]^2 \} - 2\sigma \omega_0 x_1(0)x_2(0)}{\omega_0^2 - \sigma^2}} \angle -\arctan \frac{\sigma - \omega_0 \frac{x_2(0)}{x_1(0)}}{\sqrt{\omega_0^2 - \sigma^2}} \end{aligned}$$

Both waves are of the general form shown in Fig. (6-18). Their amplitudes are the same, but their initial angles are different. The effect of the exponential in reducing the magnitudes of successive peaks is evident in the figure.

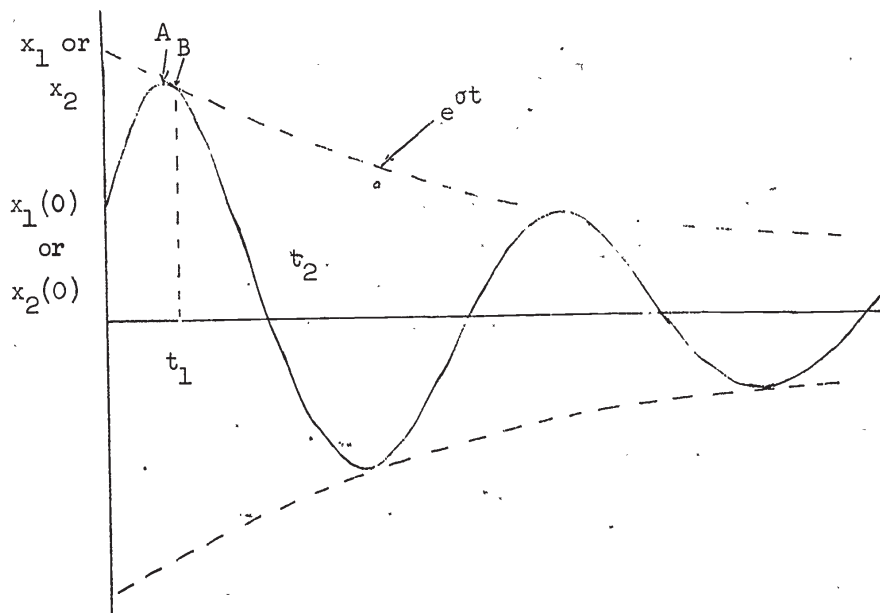


Figure 6-18.

By differentiating the equations it can be shown that points where the undamped sinusoid is maximum correspond to points (like B in Fig. 6-18) where the damped wave is tangent to the exponential envelope. The actual peak values (like A) occur slightly earlier. The separation between these two points becomes larger as the rate of decay increases. The first tangent point (B) occurs when the argument of the cosine is zero; that is, when $t = t_1$ is a solution of

$$\omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} t_1 = \theta_1 \quad \text{or} \quad \omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} t_1 = \theta_2$$

respectively, for the x_1 and x_2 curves. Also, the first zero occurs at t_2 , where

$$\omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} t_2 = \theta_1 + \frac{\pi}{2} \quad \text{or} \quad \omega_0 \sqrt{1 - \left(\frac{\sigma}{\omega_0}\right)^2} t_2 = \theta_2 + \frac{\pi}{2}$$

- 1 With the aid of the envelope curves, and the reference points indicated in Fig. 6-18, it is relatively easy to sketch the response curves obtained for given initial conditions. However, it is well to recall the simplicity of viewpoint provided by the phase plane. Sketches like those in Fig. 6-19 quickly provide information about the x_1 and x_2 curves for various positions of the initial state point.
- 2

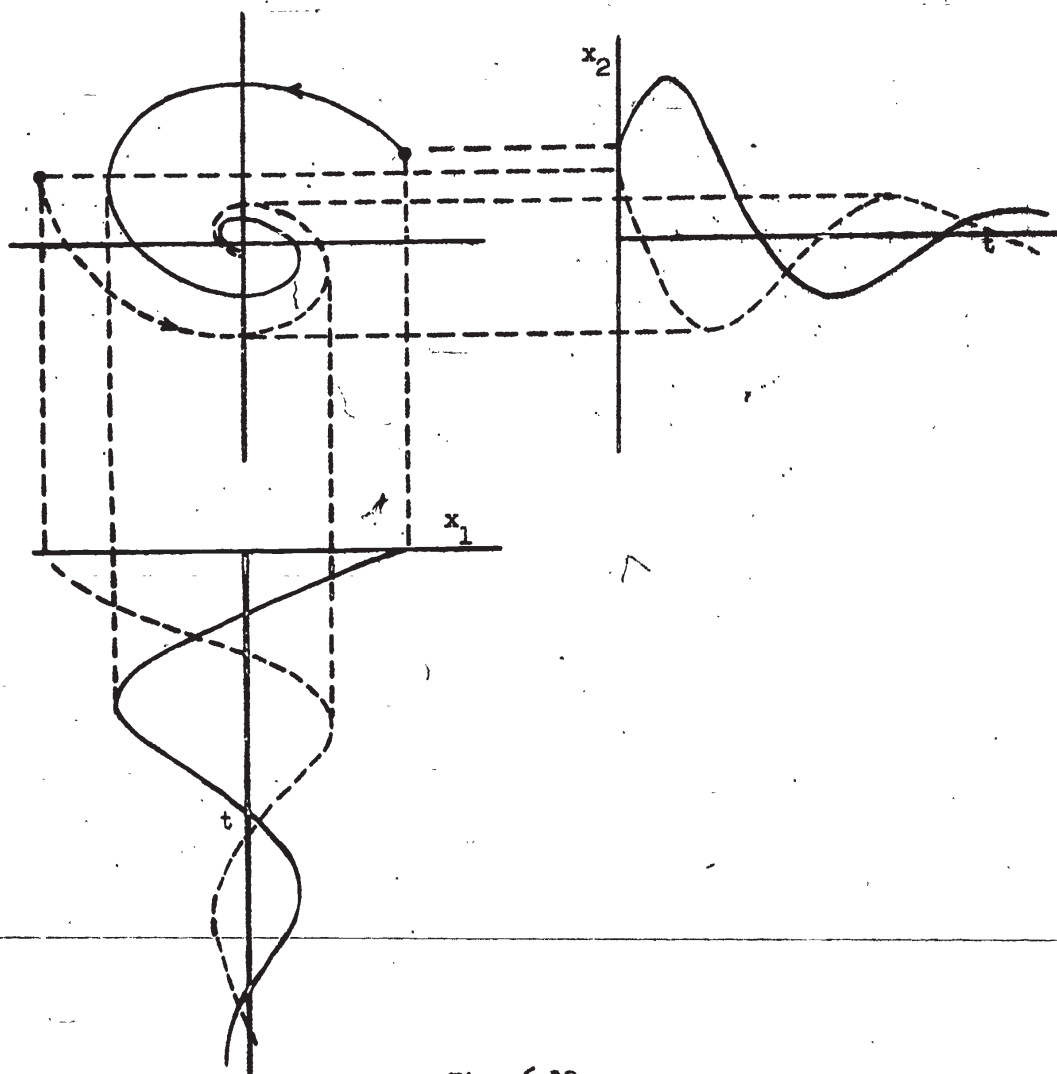


Fig. 6-19

Chapter 8

ELECTRIC MOTORS

1 Introduction:

It is generally true that any electrical machine that can operate as a generator can also operate as a motor*. When such a machine is connected in a circuit and mechanically connected to a mechanical load, or possible source of driving torque, whether the machine operates as a motor or generator depends on conditions external to it. For example, the electric machine (motor or generator) attached to the axle of an electric locomotive is acting as a motor when it is receiving electric power from overhead lines and delivering mechanical power to drive the locomotive up a hill. Conversely, if the locomotive is going down hill, the direction of power flow through the machine can be reversed. It can be driven mechanically, and deliver electrical power to the overhead line. No electrical connections are changed in the process. Whether gravity is pulling back or forward determines the direction of the mechanical torque at the machine shaft, and hence the direction of power flow.

These principles are most easily seen in the case of a d-c machine. In Fig. 8-1a, a d-c machine with separately excited field (to make the flux independent of speed) is connected to a driving engine of some sort, which can drive the machine at various speeds, or be driven by it.

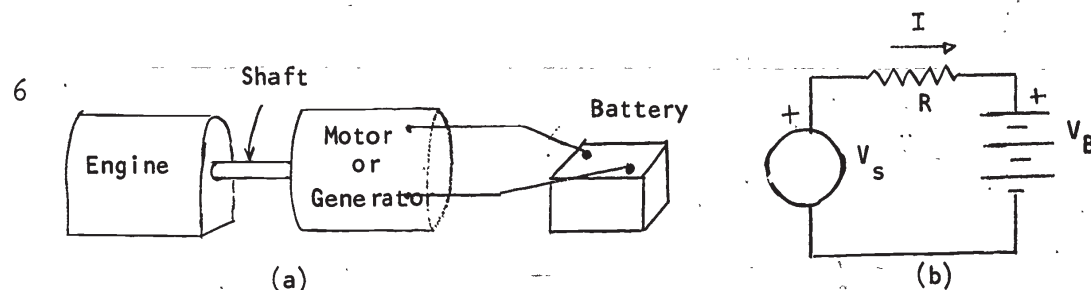


Figure 8-1.

The equivalent circuit of the electric machine is shown in Fig. 8-1b.

Resistor R includes the resistance in the equivalent circuits of both the generator and the battery, and also the connecting wires. The potential source V_s is equal to the induced emf of the machine, and is proportional to its rotational speed ω_m , thus:

$$V_s = k \omega_m \quad (8-1)$$

where k is a constant of proportionality.

*An exception to this statement is discussed in Sec. 8-6.

1 Solution of Fig. 8-1b readily yields

$$I = \frac{k \omega_m - V_B}{R} \quad (8-2)$$

2 If I is positive, electrical power leaves the machine and enters the battery. That is to say, the machine acts as a generator. If I is negative, electrical power leaves the battery and enters the machine, and so it is then acting as a motor.

3 If we use the example of the locomotive, suppose it is going up a hill at a speed ω_{m1} . The electric machine must be acting as a motor, and so in order to have a negative I , we must have

$$V_B > k \omega_{m1}$$

4 This condition can be accomplished by adjusting the supply voltage V_B , or by adjusting the field strength (which affects k) to make $k \omega_{m1}$ sufficiently small. Now, after passing the top of a hill, in beginning to coast down, ω_{m2} will increase. At any speed ω_{m2} where

$$V_B < k \omega_{m2}$$

5 the machine will act as a generator, because the current will have reversed.

8-1. Electromotive Force of a d-c Machine

6 As indicated in the introduction, the conditions relevant to induced emf depend only on the motion of armature conductors in a magnetic field. Such motion is present in either a motor or generator, and so the principles which govern induced emf are the same for both.

7 As you have seen in the consideration of induced emf (see programmed text) the emf depends on the rotational speed, the distribution of flux around the airgap, and the number of turns of wire on the armature, and the number of commutator segments. For a given machine, the armature winding is constant, and the flux distribution can be assumed to be reasonably constant, although the total flux distribution may vary slightly.

8

- 1 If E is the emf of a given machine, it is therefore true that it is proportional to speed and to the airgap flux ϕ , and thus can be written

$$V_s = E = k_E \phi \omega_m \quad (8-3)$$

- 2 where k_E is a proportionality constant which will vary from machine to machine. (What does k_E depend upon?)

- Referring to Fig. 8-2, the flux ϕ in this expression is the total flux through a plane P-P' located as indicated in the figure. For the flux and rotational directions shown in Fig. 8-2, the emf will be into the paper (indicated

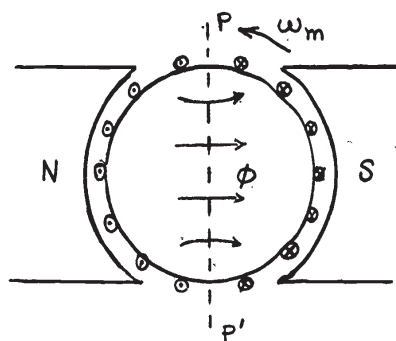


Figure 8-2.

- by the X symbol) for all conductors to the right of P-P', and will be out of the paper (indicated by the · symbol) for all conductors to the left of P-P'. The armature connections are such that half of these conductors are in series between the brushes, to produce emf E . Note that no reference direction is shown on the diagram for E , because the end connections of the armature conductors and commutator are not shown, making it impossible to define a reference direction for E for this particular diagram.

8-2 Torque of a d-c Machine

- When the machine portrayed symbolically in Fig. 8-2 acts as a generator, the currents in the armature conductors are in the same directions as the emfs, as indicated in that figure. If the machine is acting as a motor, the currents are reversed, as discussed in the introduction. Thus, Fig. 8-3a would represent

- 1 relative directions of rotation, flux, and current, for a motor. The current
 2 reference directions shown will be for armature current I_a , which is
 also shown on the equivalent circuit in Fig. 8-3b. Thus, motor action will
 be associated with positive values of I_a . (This is opposite to the reference
 3 direction in Fig. 8-1, the change having been made because we want to
 4 concentrate on operation of a motor.) By making this change we will avoid

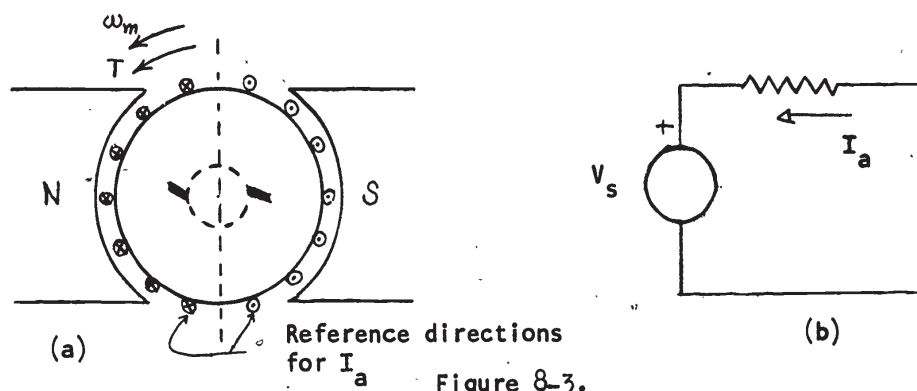


Figure 8-3.

- 5 having always to consider I_a as negative. However, this change was quite
 arbitrary, and was unnecessary.)

- The current in each armature conductor will be $I_a/2$ (Why?). Thus,
 considering a pair of diametrically opposite conductors, as in Fig. 8-4,
 6 each will experience a force as indicated. Specifically, if B_{m1} is the

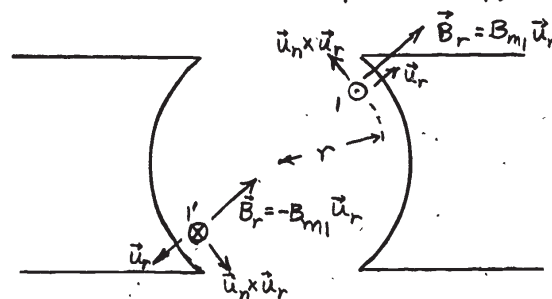


Figure 8-4.

8 magnitude of flux density at each point, $\vec{B}_r$ has the values indicated in
 the figure. Let $\vec{u}_n$ be a unit vector out of the paper. The force on

1 conductor (1) will be

$$F_1 = \frac{B_{m1} I_a \ell}{2} (\vec{u}_n \times \vec{u}_r)$$

2 and the force on conductor (1') will be

$$\begin{aligned} F_{1'} &= \frac{(-B_{m1})(-I_a)\ell}{2} (\vec{u}_n \times \vec{u}_r) \\ &= \frac{B_{m1} I_a \ell}{2} (\vec{u}_n \times \vec{u}_r) \end{aligned}$$

3 The directions of the unit vector products are indicated in the figure.

4 It is evident that these forces are additive in producing torque about the axis of rotation, giving a torque

$$T_{11'} = B_{m1} I_a \ell r \quad (8-4)$$

5 Each pair of conductors will produce a similar effect, and so the total torque will be a sum

$$\begin{aligned} T &= T_{11'} + T_{12'} + \dots \text{etc.} \\ &= I_a r \ell (B_{m1} + B_{m2} + \dots) \end{aligned}$$

6 The sum in parentheses is proportional to the total flux ϕ passing through the rotor. (Why is it proportional rather than equal to the flux?) Thus, we can absorb r and ℓ in proportionality constant which depends on the machine's dimensions, to yield

$$T = k_T I_a \phi \quad (8-5)$$

7 as a relation that shows how torque varies with the essential variables I_a and ϕ .

8 T is a scalar quantity in this expression, having the reference direction shown. That is, when T is positive it will be in the direction of rotation, confirming the earlier finding that positive values of I_a will correspond to motor action.

1 8-3 Speed-Torque Curves

Much information about the operation of motors is obtained by a consideration of how speed varies with torque. This is because a mechanical load is characterized by a driving torque, which also is a function of speed. Thus, if speed torque curves of a motor and a load are plotted on the same set of axes, they will intersect at the operating point.

The two equations

$$V_s = k_E \phi \omega_m \quad (a) \quad (8-6)$$

$$T = k_T \phi I_a \quad (b)$$

and the equivalent circuits shown in Fig. 8-5 provide the information needed to predict how speed of a motor will vary with torque.

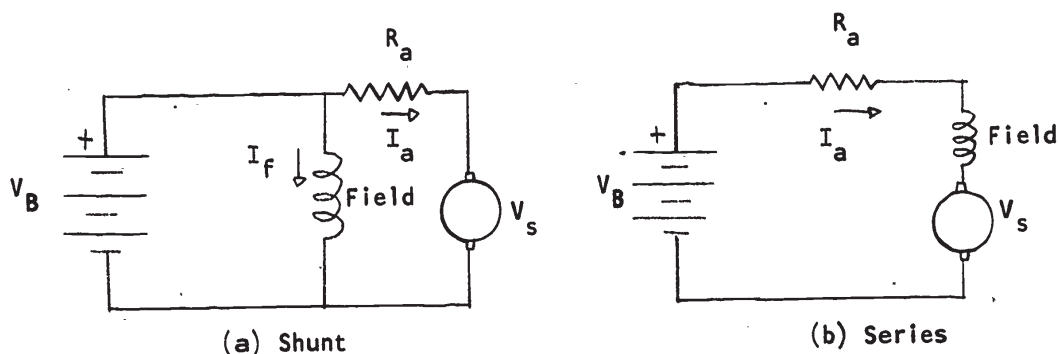


Figure 8-5.

7 8-4 Shunt Motor

Figure 8-5a illustrates a shunt motor, the word "shunt" referring to the fact that the field coil is "shunted across" the armature. The field current can be adjusted by means of a field rheostat R_f (a means of speed control, as we shall see) but in the absence of such an adjustment, ϕ remains constant.

From the equivalent circuit

$$V_s = V_B - I_a R_a$$

or

$$k_E \phi \omega_m = V_B - I_a R_a$$

- 1 Also, from Eq. (8-6b), $I_a = T/k_T \phi$ giving

$$k_E \phi \omega_m = V_B - \left(\frac{R_a}{k_T \phi} \right) T$$

- 2 or

$$\omega_m = \frac{V_B}{k_E \phi} - \left(\frac{R_a}{k_E k_T \phi^2} \right) T \quad (8-7)$$

A set of ω_m vs. T curves for several values of ϕ (or I_f) is shown in Fig.

- 3 8-6.* Curves of I_a vs. T are also shown by the dashed lines. Labels 1, 2, and 3 indicate progressively increasing values of ϕ (or increasing I_f).

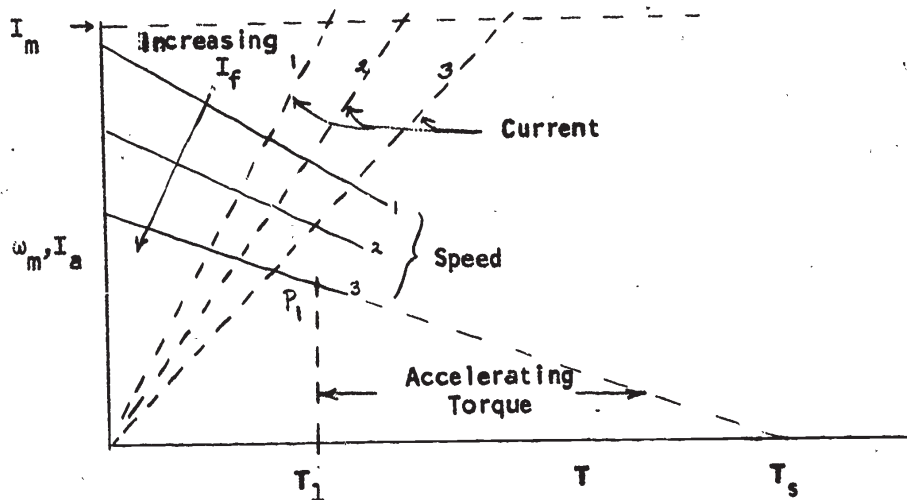


Figure 8-6.

- 7 Much can be learned from these curves about the operation of a shunt motor. Suppose it is required to start such a motor when its shaft is connected to a load for which the torque is independent of speed (say starting a locomotive on a hill), and that the field strength corresponds to curve 3.
- 8 The constant torque is represented by the dashed vertical line through T_1 . Since the initial speed is zero, the initial torque provided by the motor will be the intercept T_s . The difference $(T_s - T_1)$ is available for acceleration. This would seem to be satisfactory. With increasing speed there would gradually be less accelerating torque, as indicated in the figure, and steady speed would be reached at point P_1 . However, for any motor there is a value of I_a which

*The straight-line curves assume ϕ is independent of I_a , which is not exactly true due to armature reaction (see programmed text).

- 1 should not be exceeded, to avoid damage due to overheating. Suppose I_m
 2 is such a value. From the dashed curve (3) it is evident that the current
 limit would be exceeded at the start. Thus, although the motor would run
 satisfactorily at point P_1 under this specified load, it would not start
 properly.

- Starting can be accomplished within the current limitation by temporarily
 inserting additional resistance in the armature circuit. As an explanation,
 from Eq. (7) it is seen that increasing R_a will increase the slope of the
 3 curve, as shown by the curve labeled "increased R_a " in Fig. 8-7. With this
 increased R_a , steady speed will be reached at P_2 . If the added resistance
 is then removed, operation will shift to the original curve, and the motor
 torque will jump to T_0 , and so the difference $T_0 - T_1$ is available for
 4 acceleration. Acceleration will stop at point P_1 .

- This is a normal procedure for starting a shunt motor. As shown in Fig.
 8-8, a high current-capacity resistor is provided with a series of taps which
 are contacted sequentially by a sliding arm. The number of taps depends on
 5 the characteristics of the load. For example, whereas one step was sufficient
 for a load torque T_1 , for a greater torque, say T_2 , two steps would be required.

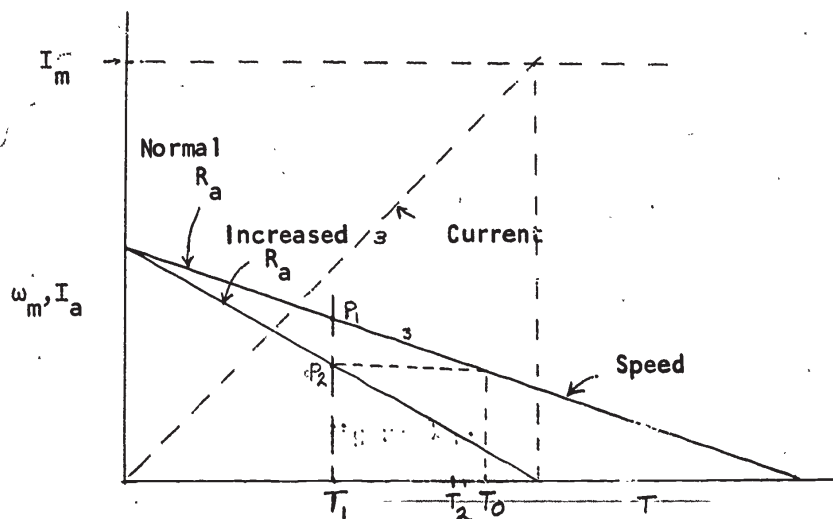


Figure 8-7.

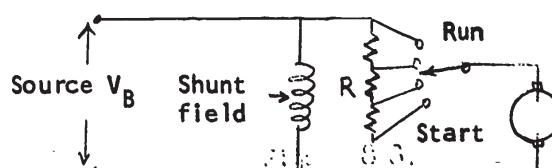


Figure 8-8

- 1 A still larger number of steps would be required to start along curves (2) or (3) of Fig. 8-6, which correspond to a weaker field. (Do you see why?)

In reference to Fig. 8-8, it should be observed that the field coil is directly across the source. That is, the starting resistor is not placed in the line to the motor. (Why not?)

- 2 By using the example of a constant torque (independent of speed) we have seen how the operating point at constant speed is determined by the intersection of the speed torque curves of the motor and the load. The constant torque curve (vertical line) typifies situations where the load torque is due to the pull of gravity, as in the previously cited example of a locomotive on a constant slope incline, or a building elevator.

- 3 Of course, friction always provides some of the load torque, and this is not independent of speed. Dry friction is nearly constant, but decreases somewhat after motion starts. A fan is a totally different type of load, having a torque which is nearly proportional to the square of speed. These three examples are illustrated in Fig. 8-9.

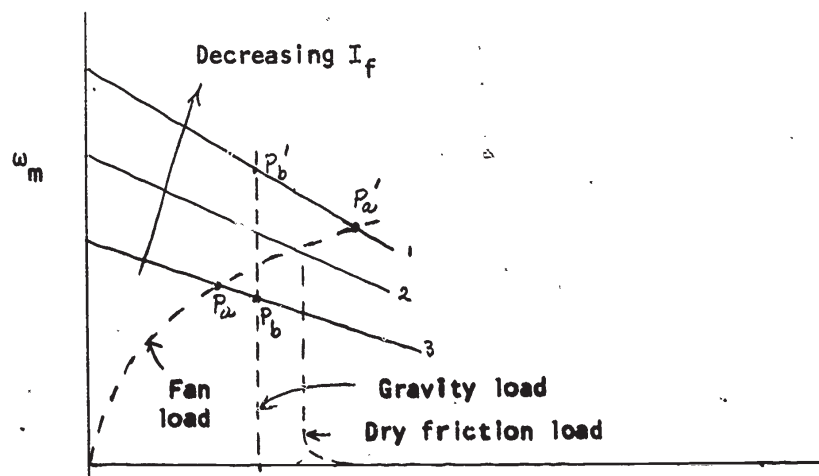


Figure 8-9

- 8 The effect of using a field current adjustment to increase speed is seen to be related to the characteristic of the load. For example, on curve (3) the fan and gravity loads will be driven at constant speeds represented respectively by points P_a and P_b . If the field current is decreased, shifting the motor characteristic to curve (1), operation will be respectively at points P_a' and P_b' .

- 1 The speed of the fan will increase less than the speed of the gravity load, because the torque requirement increases with speed.

The characteristic curves of shunt motors display the property of relatively small variation of speed over the operating range, for fixed field current. They are therefore useful in driving loads which should maintain nearly constant speed.

- 2 On the other hand, shunt motors have the disadvantage of not being efficient for starting loads which have high torque demands at low speed, such as gravity loads. As an explanation of this, refer to Fig. 8-10, which includes the motor and load speed-torque curves, the I_a vs. torque curve, with the indication of the maximum safe starting current I_m . Suppose this maximum starting current is twice the rated current for operation at constant load, which we shall call I_r . Furthermore, suppose the load torque is such that at constant speed the motor will operate at rated power output, which means that I_r will be the value of I_a . This constant speed operating condition is indicated by point P in Fig. 8-10.

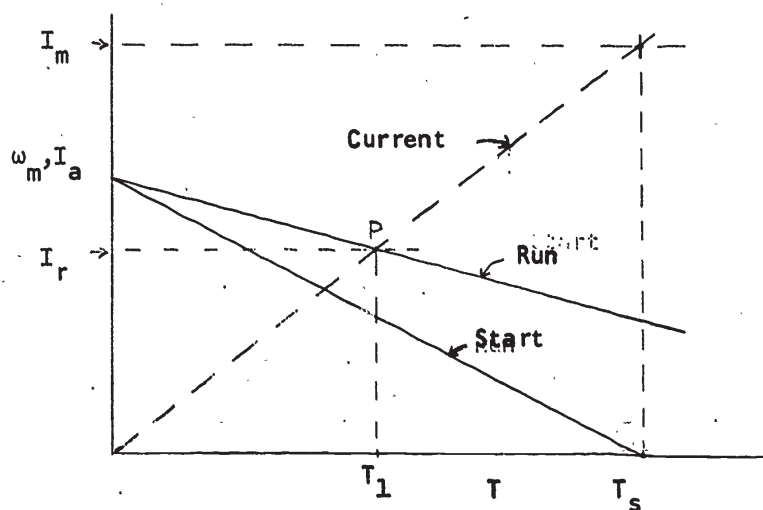


Figure 8-10.

We are interested in the maximum starting torque. This is T_s , which is twice T_1 , as a result of torque being proportional to current. Thus, for this particular case, the torque available for acceleration is $T_s - T_1 = T_1$.

- 9 This amount of accelerating torque might be unduly small, resulting in an undesirably slow start. For loads having small starting torques as, for

- 1 example, a fan, the small accelerating torque would be less of a disadvantage.
 In summary, we have shown that for a shunt motor starting up under a constant
 torque load, near its maximum torque (as determined by I_r as the maximum
 armature current), the torque available for acceleration will be equal to
 2 this maximum torque.

8-5 Series Motor

- If the field excitation is obtained by connecting a field coil in series
 with the armature, as in Fig. 8-11, the resulting machine is called a series
 3 motor. Compared to the shunt case, the series field coil will have a relatively
 small number of turns, because the armature current can be large compared with
 the current in a shunt coil.

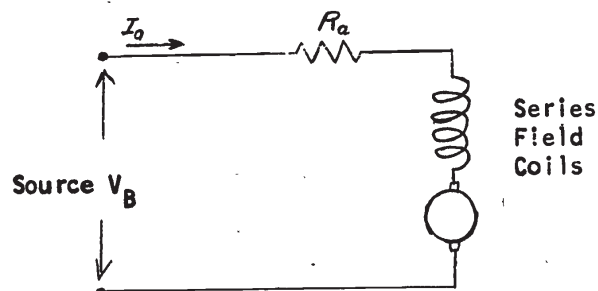


Figure 8-11.

6 The general properties of the speed-torque curves can be obtained from the
 two basic equations

$$k_E \phi \omega_m = V_B - I_a R_a$$

$$T = k_T \phi I_a$$

by making ϕ proportional to I_a . To do this is somewhat of an approximation,
 because in reality ϕ is related to I_a in accordance with a nonlinear magnetization
 8 curve. However, in the interest of simplicity, and since we are not interested
 in quantitative details, we shall use the linear approximation

$$\phi = k_\phi I_a \quad (8-8)$$

- 9 where k_ϕ is a constant which depends upon the number of turns on the field
 coil, and dimensions of the magnetic circuit.

1 Substituting Eq. (8-8) for ϕ in the above equations yields

$$(k_E k_\phi) I_a \omega_m = V_B - I_a R_a \quad (a)$$

$$T = (k_T k_\phi) I_a^2 \quad (b)$$

(8-9)

2 For a given machine, the quantities in parentheses are constants. Equation (8-9a) can be written

$$(k_E k_\phi) \omega_m = \frac{V_B}{I_a} - R_a$$

3 and from Eq. (8-9b), $I_a = \sqrt{T/(k_T k_\phi)}$. Therefore,

$$k_E k_\phi \omega_m = \sqrt{\frac{k_T k_\phi}{T}} V_B - R_a$$

4 or

$$\omega_m = \sqrt{\frac{k_T}{k_E^2 k_\phi}} \frac{V_B}{\sqrt{T}} - \frac{R_a}{k_E k_\phi}$$

5 Of course, since the field coils are in series with the armature, in the above equations R_a is the sum of the armature resistance and the resistance of the field coils. Now that we are ready to interpret the above formula, let us simplify it by writing

$$k_1 = \sqrt{\frac{k_T}{k_E^2 k_\phi}} \quad \text{and} \quad k_2 = \frac{R_a}{k_E k_\phi}$$

so that

$$\omega_m = k_1 \frac{V_B}{\sqrt{T}} - k_2 \quad (8-10)$$

expresses the desired relationship between speed (ω_m) and torque (T).

This formula shows the interesting feature of infinite speed when torque is zero, and has the general form shown in Fig. 8-12, for two values of V_B . Regardless of the value of V_B , current and torque are related by the same curve, shown dashed in the figure. The zero torque condition is never actually realized because there is always some friction in the machine, which has been neglected here.

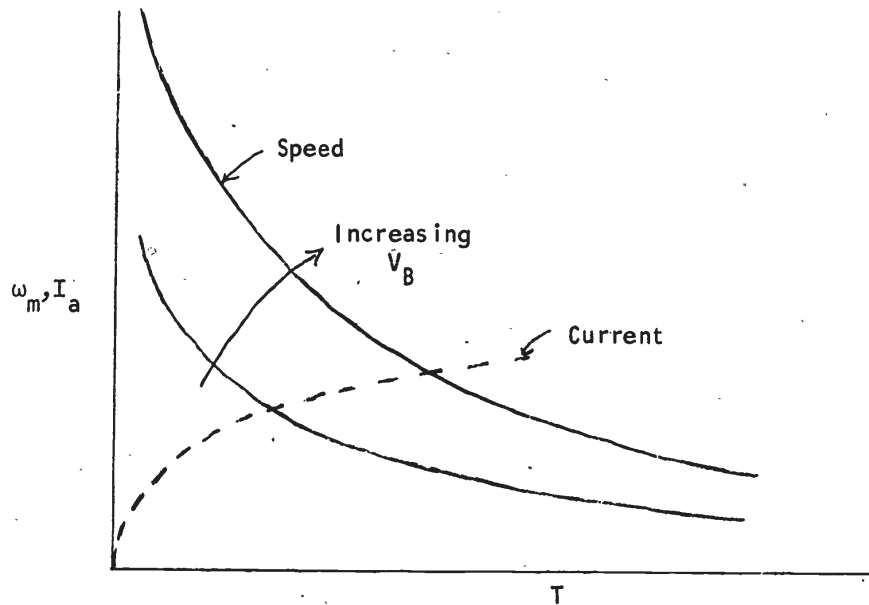


Figure 8-12.

Series connected motors inherently have more widely varying speed, as a function of torque, than shunt motors. Accordingly, they are useful in applications where load torque remains relatively constant over long periods of time, and where the load can never be removed. The latter condition is necessary to prevent "run away" at no load. Most small series motors will have enough losses due to friction to prevent damage if they are accidentally unloaded, but a large motor (1 hp. or more) can be very dangerous if allowed to run away, because centrifugal force can cause it to "explode". Railway traction motors are of the series type.

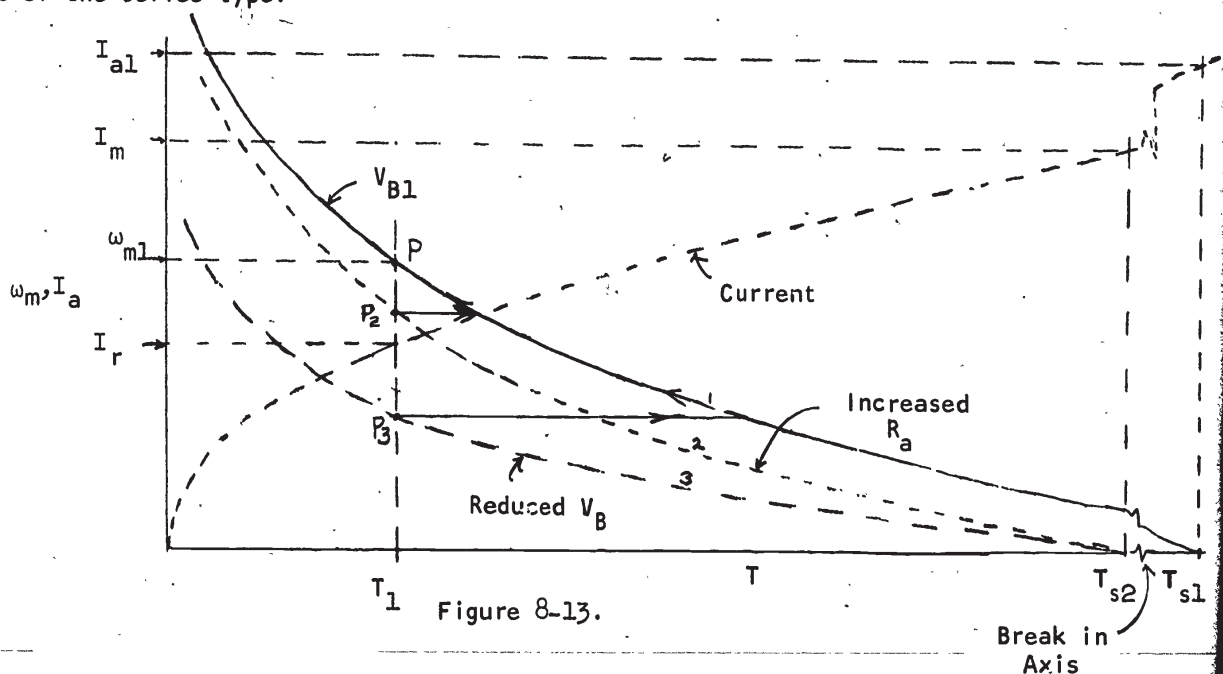


Figure 8-13.

Speed control, and starting conditions will be discussed in terms of Fig. 8-13. Let P represent a desired operating point for a load requiring a torque T_1 to be driven at a speed ω_{m1} . This will determine the particular value of source voltage V_{B1} , this being the voltage for the characteristic curve passing through point P.

The condition portrayed is the most severe possible in regard to starting, because the operating point is the full capacity of the machine, by which we mean that the armature current (as obtained from the I_a vs. T curve) is the maximum allowable value (I_r) for continuous running. As before, we shall assume that during the temporary starting interval the armature current can be allowed to reach a value, $I_m = 2I_r$.

If the motor were to be started "across the line", the starting torque would be T_{s1} , with a corresponding current I_{a1} , as indicated, which is in excess of I_m . The allowable starting torque T_{s2} can be obtained by projecting down from the current-torque curve at the point corresponding to I_m . Two curves, labeled (2) and (3) are shown which yield T_{s2} at zero speed. Curve (2) is obtained by shifting curve (1) downward a uniform amount throughout its extent. Referring to Eq. (8-10) we see that this change is accomplished by increasing k_2 (or increasing R_a). Curve (3) is obtained by multiplying $\omega_m + k_2$ by a constant factor. This explains why the reduction in going from curve (1) to curve (3) is greater at high values of ω_m . Referring to Eq. (8-10) again, it is seen that the change to curve (3) is obtained by reducing the source voltage V_B .

Thus, it is seen that there are two methods of holding the armature current within tolerable limits when the speed is zero, on starting. One method is to insert additional resistance in series with the motor, the other is to reduce the supply voltage.* Although curves (2) and (3) are slightly different, either one will provide adequate starting. The motor

*You may wonder why reduction of supply voltage was not mentioned in connection with starting a shunt motor. This would be possible if two sources were available, a fixed voltage for the field and a variable one to supply armature current. Reducing the voltage to both would result in weakening the field and perhaps would reduce the torque to the point where the machine would not start.

- 1 will be allowed to come up to constant speed (at points P_2 or P_3), and then the additional resistance will be taken out, or full voltage will be applied, at which time progress to the final operating point would be as indicated by the arrows.
- 2 We now make an important observation. Since torque is proportional to the square of current, the starting torque is $T_{s2} = 4T_1$, and hence the torque available for acceleration is $4T_1 - T_1 = 3T_1$. This is the worst possible condition, where T_1 corresponds to rated current, and is the same case
- 3 considered for the shunt motor, where it was found that the torque available for acceleration was T_1 . Thus, the series motor supplies three times as much accelerating torque as the shunt motor, when started under conditions of constant load torque of maximum allowable value. For lighter starting load torque, the
- 4 advantage of the series motor is slightly greater. The difference is, of course, due to the difference between the linear (T proportional to I_a) relation for the shunt motor and square law (T proportional to I_a^2) for the series motor.*

- From the standpoint of getting the motor rotating, it makes little
- 5 difference whether series resistance is inserted, on the supply voltage is reduced. Conditions affecting the choice can be appreciated when we consider speed control. It is evident from Fig. 8-13 that curves (2) and (3) both provide reduced speed for a given load, and therefore that the two methods of
- 6 starting also comprise two methods of speed control.

- A choice depends on considerations of efficiency. A series resistance absorbs power, and so variation of supply voltage is a more efficient way to control speed than by inserting a variable resistance. Thus, in applications
- 7 where speed control is needed continuously, the increased efficiency of the variable voltage method may be definitive in indicating a choice of that method, although it involves a more expensive installation in the form of a separate motor generator set.**

- 8
-
- * It is to be observed that the numbers computed above depend upon the arbitrary statement that starting current could be allowed to be twice the maximum allowable running current. If this factor had been three, the accelerating torque would be $3T_1 - T_1 = 2T_1$, and $9T_1 - T_1 = 8T_1$ respectively for the shunt and series cases.
- 9 ** Or possibly a rectifier with controllable output voltage, if the d-c supply is obtained by rectifying an a-c source.

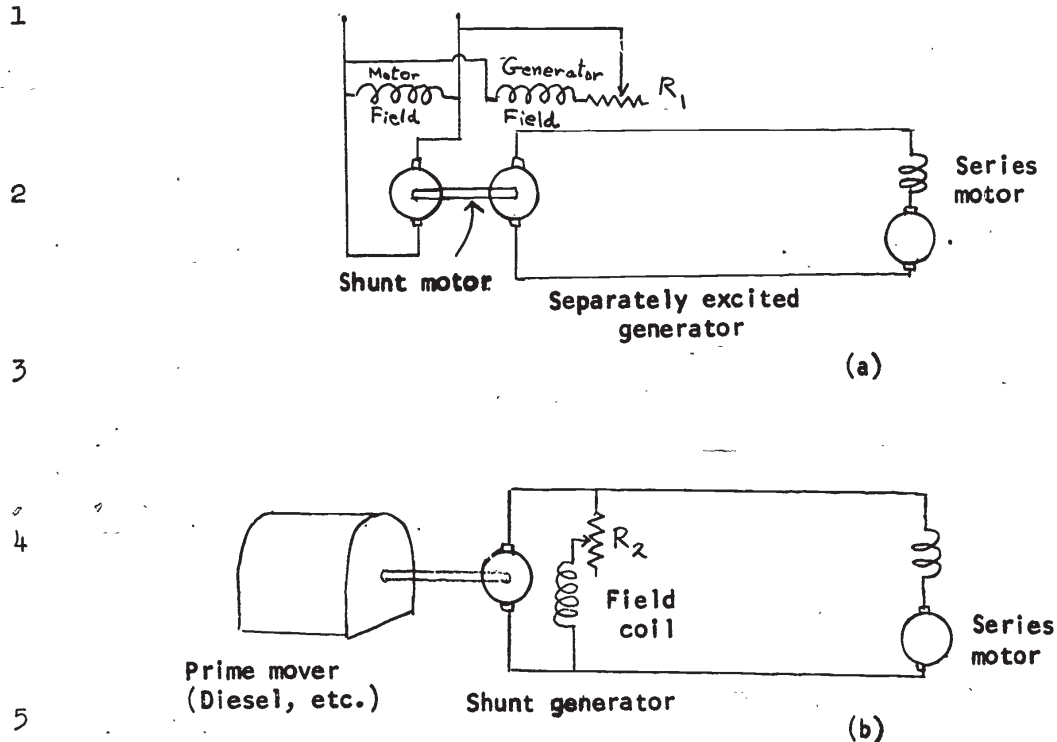


Figure 8-14.

Two examples of variable voltage speed control are shown in Fig. 8-14. The system at (a) is used where d-c power is available from a constant voltage source. The source is connected to a nearly constant-speed shunt motor which drives a separately excited generator (meaning that the field current is independent of its own armature voltage). The strength of the field of this generator is controlled by a field rheostat R_1 which, therefore, also controls the voltage at the terminals of the generator armature. This is also the supply voltage for the series motor. A system like this would be applicable, for example, to the drive for a building elevator. The continuous use of such a system would warrant the initial expense of the motor-generator set which would not be needed if resistance control were used.

The system shown in Fig. 8-14b embodies the rudiments of a diesel-electric locomotive. In this case the generator voltage is determined by appropriate setting of the generator speed (as controlled by the engine throttle) and the setting of a field rheostat R_2 .

1 8-6 Commutator a-c Motors

In the shunt and series d-c motors described in the previous sections it can be seen that the direction of rotation is independent of the polarity of the source. A change in source polarity will change the field flux direction and also the armature current direction. Reference to Fig. 8-4 will show that when both of these are reversed, the torque will still be in the same direction.

One might therefore expect that such motors would also operate satisfactorily on alternating current. This expectation is correct for a series motor, but not for a shunt motor. Let us first briefly consider why the shunt connection will not operate on a-c. The field coil and armature circuit, which are in parallel across the supply voltage, will each be inductive and carry currents which will not be in phase with the supply voltage, nor necessarily in phase with each other. On the other hand, in order to produce torque, it is necessary that the field current (and hence flux) and armature current shall be in phase (so that the product ΦI_a shall always be positive). Although this might be arranged in the shunt connection for a certain operating condition, any circuit change for the purpose of speed control or starting would involve changing armature or field resistances, and would disturb the phase relationship between flux and current. Furthermore, since the field coil has a high inductance, both armature and field currents would lag the voltage by nearly 90° . The result would be operation at low power factor, which means an excessively large current for the amount of power delivered. Thus, for these practical reasons, a shunt connection is not satisfactory for a-c operation.

The series motor does not suffer from these difficulties for the reasons that the series connection ensures that the field and armature current will be in phase, and the relatively few turns of a series field coil yields a low enough inductance so the power factor will not be excessively low.

This discussion is not meant to imply that any d-c series motor can be operated on a-c. The main difference is that for a-c operation the entire magnetic circuit must be laminated to reduce eddy current losses associated with the continual reversal of the magnetic field. In a d-c machine, the flux reverses in the rotor only, and so it must be laminated, but the stationary part of the magnetic circuit (yoke and pole pieces) can be solid iron. Thus, an a-c series motor can be operated on d-c, but not vice-versa.

- 1 Commutator a-c motors are particularly common in small sizes; being
the typical motors for low power household appliances like vacuum cleaners,
mixers, sewing machines, and the like.

- It should be pointed out that a-c commutator motors are an exception to
2 the general statement made in the introduction that there is no difference
between a motor and a generator. If mechanical power is applied to drive an
a-c commutator motor, it will produce d-c.

8-7 The Induction Motor

- 3 Applications requiring nearly constant speed a-c motors (similar to
the d-c shunt motor) are served by a-c induction motors, which we shall now
consider briefly. Unfortunately, an induction motor is much more difficult
to analyze than its d-c counterpart, but at least enough theory can be
4 presented here to discuss its important characteristics.

- As a first step, consider Fig. 8-15a, which is a reproduction of Fig. 41
of the programmed text, with the slight modification of showing pairs of coil
sides placed in slots. Figure 15b shows the same arrangement, uncluttered
5 by the end connections. These figures represent the stator winding of an
induction motor. Again referring to Fig. 15a, assume an a-c source is
connected to terminals 1 and 2, and let i_1 have the reference direction
shown. An alternating current will flow in the individual coil sides shown
6 in Fig. 15b, with reference directions as indicated by the X and $\cdot$ symbols.

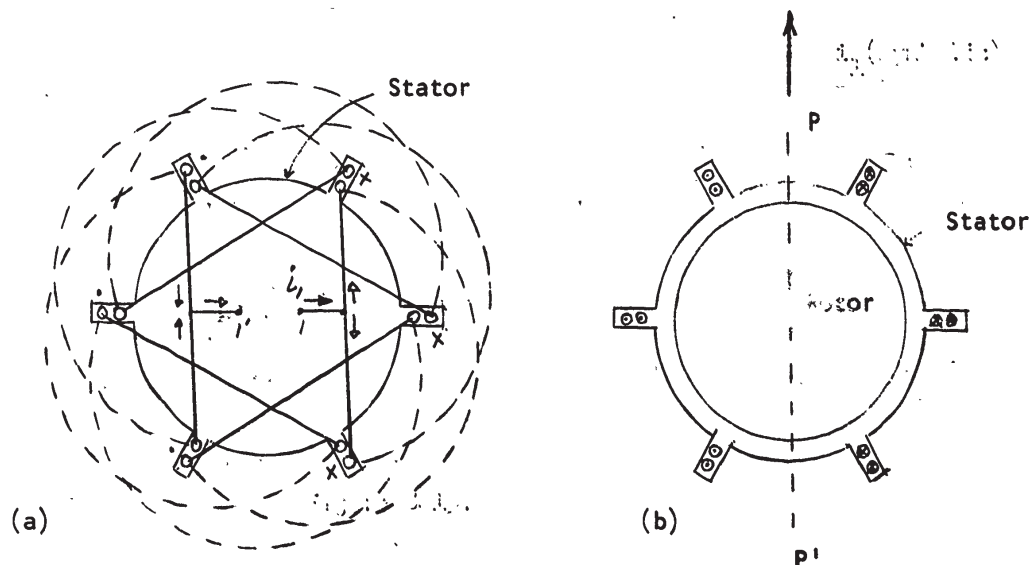


Figure 8-15.

This figure serves to remind you that coil connections are such that at any instant of time currents to the right and left of line P-P' always flow in opposite directions, and that the whole winding acts like a coil having P-P' as an axis and producing flux having a reference direction indicated by the arrow.

Twenty or more coils would be a more reasonable number for a practical machine. Figure 8-16a shows the coil sides for a ten coil winding. This figure is introduced to develop the notion that in an actual machine coil sides are separated by a small angle δ , and therefore that the actual condition is not much different from the idealized continuous distribution of current shown in Fig. 8-16b. This idealization is introduced because it makes the subsequent analysis simpler than it would be if we were to bother with the details of individual coils. In other words, we are proceeding on the basis of the hypothesis that this approximation will lead to an adequate theory, with justification depending upon agreement with experimental observation.

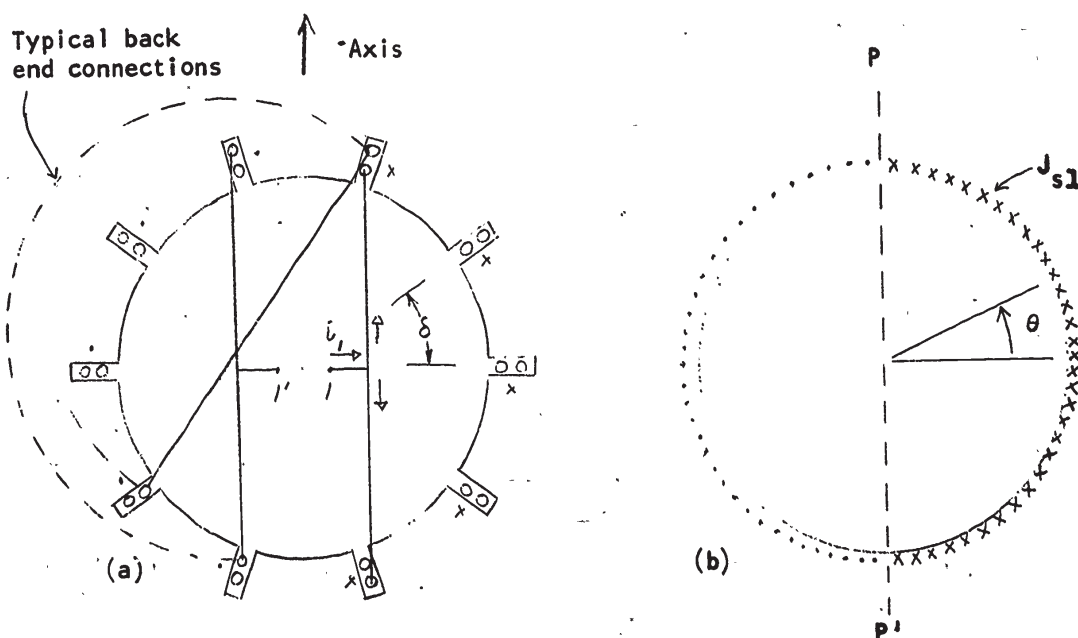


Figure 8-16.

The symbol shown for current in Fig. 8-16b is J_{s1} . This is a surface current density, measured in amperes per meter of circumference. To explain what we mean by this, suppose the winding of Fig. 8-16a has N coils and C turns per coil. There will be NC conductors on one side of line P-P', and each conductor carries a current $i_1/2$ (Why?). Thus, the total current on one side will be $NCi_1/2$. For Fig. 8-16b to be equivalent, it must have the same total current on each side of line P-P'. This is $\pi r J_{s1}$, and hence we see explicitly how to find J_{s1} ; namely,

$$J_{s1} = \frac{NCi_1}{2\pi r} \quad (8-11)$$

Having shown how machine parameters determine J_{s1} , we shall use J_{s1} as the prime quantity from which to start the analysis. Meanwhile, we must not forget that i_1 is an alternating current, and therefore that J_{s1} is also a sinusoidal quantity. It is also a function of θ , and can be written

$$J_{s1} = (J_m \cos \omega_e t) f(\theta) \quad (8-12)$$

where the $J_m \cos \omega_e t$ factor accounts for the time variation and $f(\theta)$ describes the variation with θ . For Fig. 8-16b, $f(\theta)$ is the square wave shown by the solid line in Fig. 8-17.

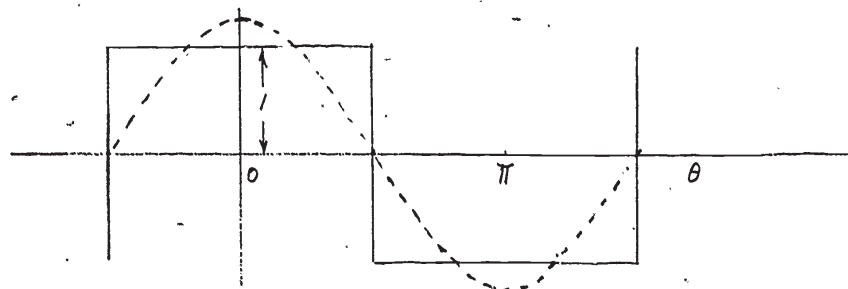


Figure 8-17.

At this point we make a further approximation, by using the dotted cosine function shown in the figure, to approximate the square wave.* Thus, we shall

*This approximation is not necessary, but permits a simplified treatment in this text which does not presuppose a knowledge of Fourier Series. A more complete analysis would retain the square wave which can be represented by the Fourier series

$$\frac{4}{\pi} \left(\cos \theta - \frac{1}{3} \cos 3\theta + \frac{1}{5} \cos 5\theta + \dots \right)$$

Thus it is seen that the subsequent treatment will use only the first term of this series.

use

$$J_{s1} = \frac{4J_m}{\pi} \cos \omega_e t \cos \theta \quad (8-13)$$

and, in view of Eq. (8-11), it is also given by

$$J_{s1} = \frac{NCI_m}{\pi^2 r} \cos \omega_e t \cos \theta \quad (8-14)$$

where I_m is the maximum input current.

The next step is to use the identity $\cos x \cos y = (1/2)[\cos(x-y) + \cos(x+y)]$ to write Eq. (8-13) as

$$J_{s1} = \frac{2J_m}{\pi} [\cos(\omega_e t - \theta) + \cos(\omega_e t + \theta)] \quad (8-15)$$

This is an important step because the dependence of J_{s1} on t and θ has been converted from a product of a function of t by a function of θ to a sum of two functions in which t and θ are combined in the arguments $\omega_e t - \theta$ and $\omega_e t + \theta$.

We shall interpret the term $\cos(\omega_e t - \theta)$ in terms of Fig. 8-18a which portrays a fictitious cylindrical distribution of constant currents which is rotating at angular velocity ω_e . The position of this rotating system of currents is defined by the position of its axis of symmetry O-A. In time t this axis has rotated through an angle $\omega_e t$, as indicated. With respect to angle θ' measured from an axis fixed in the rotating cylinder, J_{s1} is a function of θ' only, namely

$$J_{s1} = \frac{2J_m}{\pi} \cos \theta'$$

This relationship is suggested by the varying sizes of the circles which represent currents. At a fixed observation point, at angle θ on the stator, we see from the diagram that $\theta' = \omega_e t - \theta$, and so the current density observed will be

$$\frac{2J_m}{\pi} \cos \theta' = \frac{2J_m}{\pi} \cos(\omega_e t - \theta)$$

which is the first term of Eq. (8-15). Accordingly, that part of the variation of J_s due to $\cos(\omega_e t - \theta)$ is accurately portrayed by the conceptual picture of a rotating system of currents which are constant in time. The time variation is provided by the rotation. Similarly, the $\cos(\omega_e t + \theta)$ term can be viewed as due to a similar set of currents rotating in the opposite direction, as suggested by Fig. 8-18b.

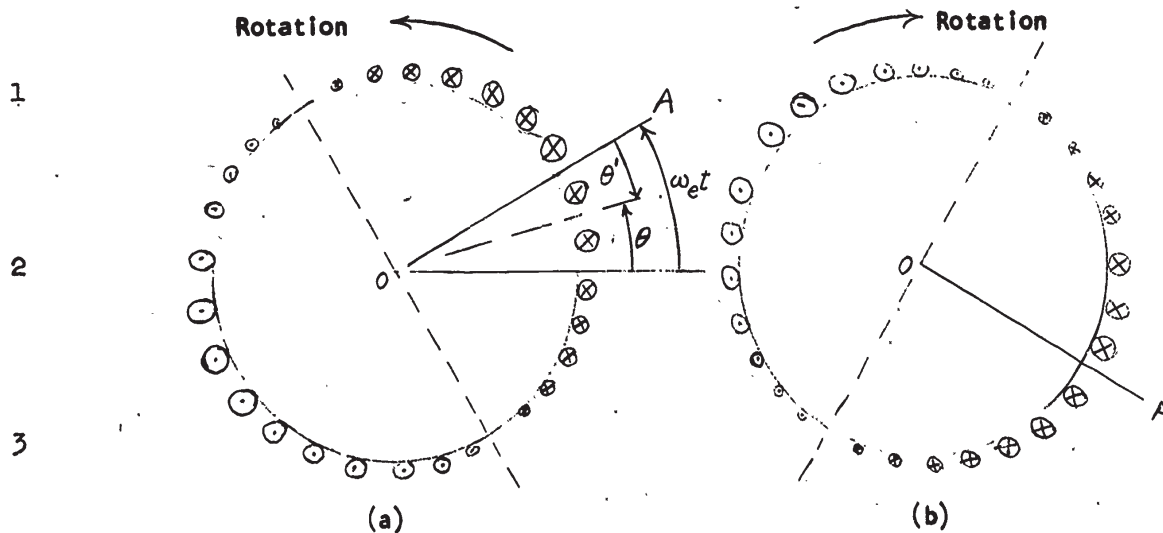


Figure 8-18.

We have now reached the important conclusion that a sinusoidally time varying current distribution in a fixed winding (as represented by Fig. 8-16b) can be replaced conceptually by the superposition of two oppositely rotating constant distributions, as in Fig. 8-18.

This analysis permits us to give an intuitive description of how an induction motor operates. A rotor carrying a winding short-circuited upon itself is placed inside the stator we have just described. The winding may consist of coils placed in slots, like the stator winding, or it may be a "squirrel cage" winding as portrayed in perspective view in Fig. 8-19a, in which a series of equally spaced bars are placed in rotor slots, and are connected at the ends by a pair of conducting rings.

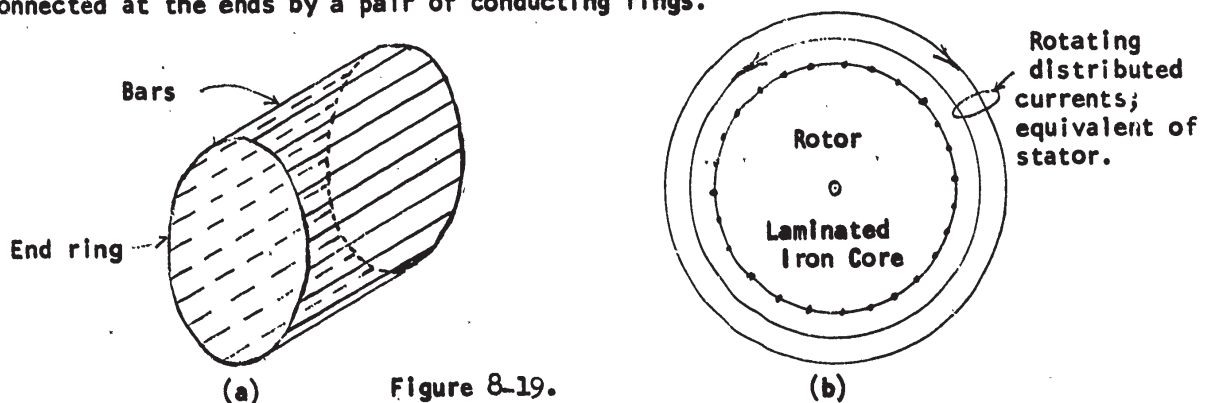


Figure 8-19.

Figure 8-19b describes the state of affairs so far established. The stator is equivalent to two oppositely rotating current distributions, as if there were two oppositely rotating systems of coils. In view of the opposing directions, the rotor will not be affected when stationary, but if it is started

- 1 in either direction, rotor currents will be induced which will react strongly
 with the current distribution rotating in that direction, and weakly with the
 current distribution rotating in the opposite direction. This will result in
 a net torque in the direction of rotation, which will keep it going. Such a motor
 2 will run in whatever direction it is started, but will not be self starting.

The above facts are crucial, because they indicate what must be done
 in order to make an induction motor that will be self starting and will run
 in only one direction. It is necessary to eliminate one of the counter-
 3 rotating current distributions. There is more than one way to do this,
 one of them being illustrated by Fig. 8-20, which shows a reproduction of
 Fig. 8-16a in solid lines, and a second equivalent set of coil slots drawn
 dotted. The second set of coils are shifted 90° in a counter clockwise
 4 direction in space, and carries a current which lags i_1 by 90° . That is,

$$i_2 = I_m \cos(\omega_e t - \frac{\pi}{2})$$

This can be approximately described by a second equivalent continuous current
 5 distribution

$$J_{s2} = \frac{4J_m}{\pi} \cos(\omega_e t - \frac{\pi}{2}) \cos(\theta - \frac{\pi}{2}) \quad (8-18)$$

$$= \frac{2J_m}{\pi} [\cos(\omega_e t - \theta) + \cos(\omega_e t + \theta - \pi)]$$

However, $\cos(\omega t + \theta - \pi) = -\cos(\omega t + \theta)$, and so the above reduces to

$$J_{s2} = \frac{2J_m}{\pi} [\cos(\omega_e t - \theta) - \cos(\omega_e t + \theta)] \quad (8-19)$$

If both windings are simultaneously energized, the total equivalent current
 distribution is

$$J_s = J_{s1} + J_{s2} \\ = \frac{4J_m}{\pi} \cos(\omega_e t - \theta) \quad (8-20)$$

which also represents a fixed distribution rotating in only one direction
 9 (counter-clockwise).

2
8-24

1

2

3

4

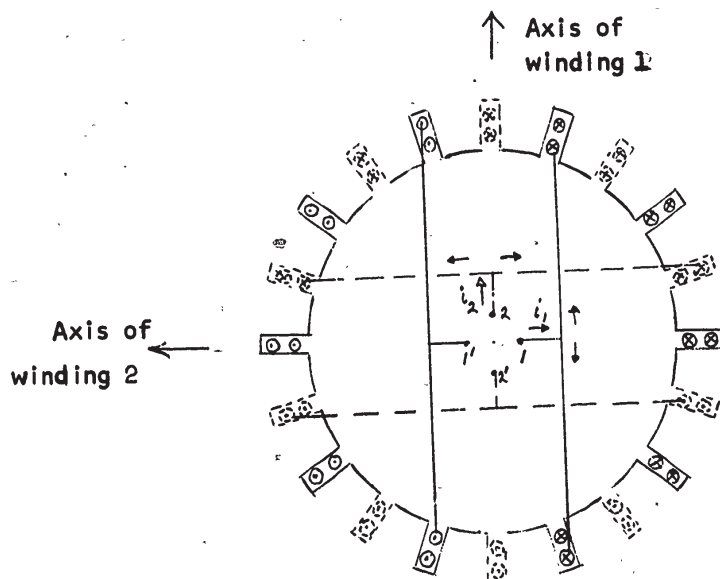


Figure 8-20.

5

A similar analysis will show that if current i_2 were to lead i_1 by 90° , rotation would be in the clockwise direction. This fact indicates how such a motor can be reversed. A change from lead to lag in winding (2), compared to winding (1), can be accomplished merely by interchanging its two terminal leads.

6

Such a motor is called a two-phase motor, because it has two separate windings supplied by two separate sources. In practice further modifications are made, as described below.

7

8-8 Single-phase motors

In small and medium sizes (usually a fraction of a horsepower) it is convenient to have motors that do not require the complexity of two separate power supplies. It will be recalled that a single winding motor (called a single-phase motor) will run once it is started. Such a motor can be made self starting by adding a second winding, which is used only during the starting period. This raises the question of how to arrange for the 90° phase difference in currents i_1 and i_2 , when supplying them from the same source.

9

There are two practical solutions to this problem. One is to make the second winding of high resistance, by constructing it of smaller wire, and to connect the two windings in parallel across the source terminals, as shown in Fig. 8-21a. Current i_1 flows in a circuit of high L and low

10

- 1 R, and will lag the voltage by nearly 90° . On the other hand, winding (1) is of high resistance and so its current will be nearly in phase with the voltage. Thus, i_2 will lead i_1 by something less than 90° (perhaps 60°), as suggested by the phasor diagram in Fig. 8-21a, and so the condition
- 2 required by the theory will be approximated, to a sufficient degree that the motor will start if the required starting torque is not too great.* After the motor has started, a centrifugal switch opens the starting winding circuit. Continuous operation with the two windings would not be efficient because
- 3 of the high resistance of the starting winding, and also because as the motor speeds up winding (1) becomes less inductive and the phase difference would become much less than 90° .

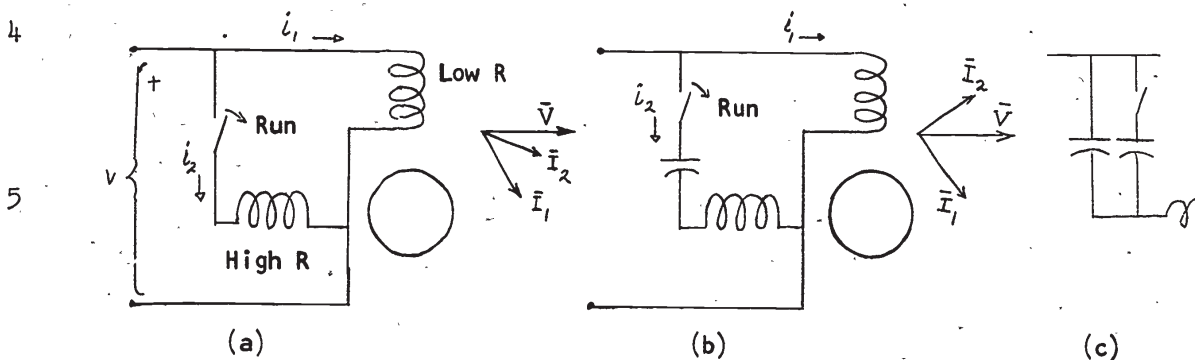


Figure 8-21.

The above describes the original design for single phase motors (1/8 to 1/2 hp.) until the advent of electrolytic capacitors which made it possible to use a capacitor in series with the second winding, as shown in Fig. 8-20b. By suitable choice of the capacitor size, current i_2 can be made to lead i_1 by exactly 90° at the start, since with a sufficiently large capacitor i_2 can actually lead the voltage, as shown in the phasor diagram of Fig. 8-21b.

* Observe that when we introduced the two-phase motor it was indicated that i_2 should lag i_1 by 90° , whereas this description shows it leading by nearly 90° . This is incidental, since it only affects the direction of rotation, and can be compensated for by a simple interchange of connections to winding (2).

- 1 Under running conditions, the winding (2) circuit can be opened by a centrifugal switch, as in the previous case. However, improved performance can be obtained by retaining the second winding under running conditions. In that case, under running conditions, i_1 will be more nearly in-phase with the voltage which means that i_2 must lead the voltage by a larger angle on "run" than on "start". This change can be accomplished by having a smaller series capacitor for the "run" condition. This can be accomplished as shown at (c) in Fig. 8-21. Two capacitors are in parallel on "start", and then as the motor comes up to speed a centrifugal switch opens, removing one of them from the circuit. The capacitor-start capacitor-run motor just described is a popular type of motor for applications such as domestic refrigerators. It has the advantage of providing more torque for a given size, because the crossed field arrangement of a two-phase motor is retained. It has the further advantage of operating at nearly unity power factor, by virtue of the fact that the current in one winding leads the voltage and therefore can cancel the lagging component of current in the other winding.
- 5 Both types are called split-phase motors. They operate from a single-phase source, but this source is "split" to yield effectively two phases within the motor.

- 6 A third type of single-phase induction motor is worth mentioning because it is quite popular in small sizes (small fan motors, phonograph turntable drives, etc.). Such a motor does not have a distributed winding, and has definite pole pieces, much as in the d-c case. However, as indicated in Fig. 8-22, part of each pole is encircled by a heavy copper band. This is called a shading coil. As the a-c flux changes in that part of the pole encircled by

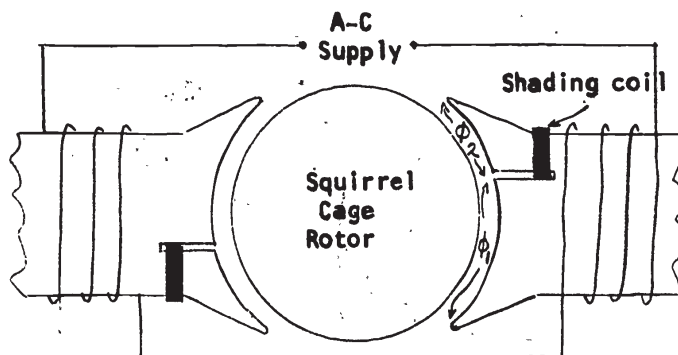


Figure 8-22.

- 1 the shading coil, it induces a current in the shading coil which, by virtue of Lenz's law, tends to oppose this change. The result is that flux ϕ_2 will lag flux ϕ_1 . It is also displaced in space, and hence the combination produces an approximation of the condition of a set of rotating windings. There are
- 2 high losses in the shading coils, which remain continuously operative. Thus, such a motor is inefficient, but its great simplicity makes it attractive for applications where the power requirement is low and where, therefore, a low efficiency is not serious.

3 8-9 Three phase motors

- The split-phase and shaded-pole motors described above generally are inefficient and are larger in size, compared with motors supplied from a two-phase source. Therefore, in sizes larger than about one horsepower, it is
- 4 necessary to have a method of supplying power directly to the second winding. We shall now discuss how this is done in practice. The two coil arrangement of Fig. 8-20 is of tutorial value, as we have used it above, and also is of historical interest, having been a practical solution in some early motors.

- 5 In practice, the second winding of Fig. 8-20 is replaced by two windings, giving a total of three windings and creating what is called a three-phase machine.

- Figure 8-23 symbolically represents the winding arrangement. There are
- 6 three separate windings, labeled (1), (2), and (3), which are insulated from each other. Winding (1), for which circles are used to represent coil sides in slots, is the same as in Fig. 8-16a. Winding (2), represented by squares, has its axis displaced 120° from the axis of winding (1). As you study this
 - 7 figure, observe that every third position belongs to winding (1), and that likewise the coil sides of winding (2) occupy every third position. Also, observe the locations of the X and dot symbols, in relation to the current reference direction at the one terminal shown for each winding. The connections
 - 8 to the second terminals are omitted, although parts of the end connections are shown as dotted lines. (Refer to Fig. 8-16a to refresh your memory as to where the second terminal is connected.) Finally, winding (3) is similar, but has its axis shifted another 120° . The symbolism whereby circles, squares, and triangles represent coil sides is indicated by the insert.
 - 9

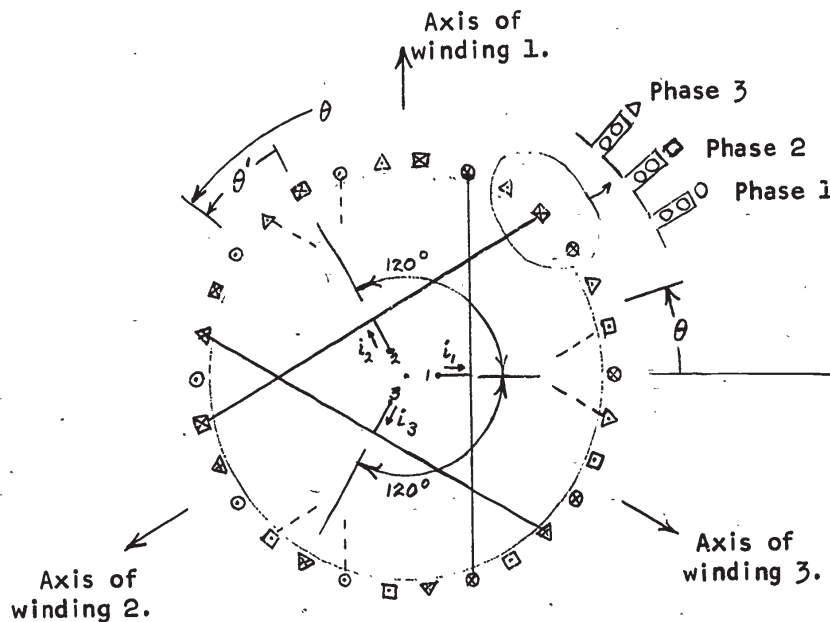


Figure 8-23.

To begin the analysis, you are reminded that winding (1) produces an equivalent current distribution

$$J_{s1} = \frac{2J_m}{\pi} [\cos(\omega_e t - \theta) + \cos(\omega_e t + \theta)]$$

and that winding (2) of Fig. 8-20 introduces a second distribution

$$J_{s2} = \frac{2J_m}{\pi} [\cos(\omega_e t \pm \theta) - \cos(\omega_e t + \theta)]$$

such that the negative sign of the second term cancels the unwanted second term of the expression for J_{s1} . We shall now show how Fig. 8-23 accomplishes a similar cancellation, if the sources to the three windings are adjusted to have equal amplitudes, and phase relations in accordance with the expressions

$$i_1 = I_m \cos \omega_e t \quad (a)$$

$$i_2 = I_m \cos(\omega_e t - \frac{2\pi}{3}) \quad (b) \quad (8-21)$$

$$i_3 = I_m \cos(\omega_e t + \frac{2\pi}{3}) \quad (c)$$

- 1 Each of these will be represented by an approximately equivalent
 2 distribution, to be represented by J_{s2} and J_{s3} . (Observe that the subscript
 2 now refers to Fig. 8-23, rather than Fig. 8-20.) The equivalences are in
 3 accordance with the treatment of Fig. 8-16, the only differences being in
 4 angular orientations of the windings in space, and the phase positions of the
 5 currents. For winding (2) the current distribution is

$$J_{s2} = \frac{4J}{\pi} \cos(\omega_e t - \frac{2\pi}{3}) \cos(\theta - \frac{2\pi}{3}) \quad (8-22)$$

6 The reason for the $(-2\pi/3)$ term in the time varying term is evident, since
 it is the phase angle established in Eq. (8-21b). In regard to $\cos(\theta - 2\pi/3)$,
 observe from the figure that winding 2 is oriented in such a way that its
 7 distribution is $\cos(\theta')$. However, it is evident from Fig. 8-23 that
 8 $\theta' = \theta - 120^\circ$, or $\theta - 2\pi/3$ in radian measure. The identity $\cos x \cos y =$
 $(1/2) [\cos(x-y) + \cos(x+y)]$ is used on Eq. (8-22) to give

$$\begin{aligned} J_{s2} &= \frac{2J}{\pi} \left[\cos(\omega_e t - \theta) + \cos(\omega_e t + \theta - \frac{4\pi}{3}) \right] \\ &= \frac{2J}{\pi} \left[\cos(\omega_e t - \theta) - \frac{1}{2} \cos(\omega_e t + \theta) + \frac{\sqrt{3}}{2} \sin(\omega_e t + \theta) \right] \end{aligned} \quad (8-23)$$

- 9 where the factors $(-1/2)$ and $(-\sqrt{3}/2)$ are the cosine and sine of $4\pi/3$,
 10 respectively.

The third winding is treated in an exactly similar fashion, to yield
 a distribution

$$\begin{aligned} J_{s3} &= \frac{4J}{\pi} \cos(\omega_e t + \frac{2\pi}{3}) \cos(\theta + \frac{2\pi}{3}) \\ &= \frac{2J}{\pi} \left[\cos(\omega_e t - \theta) + \cos(\omega_e t + \theta + \frac{4\pi}{3}) \right] \\ &= \frac{2J}{\pi} \left[\cos(\omega_e t - \theta) - \frac{1}{2} \cos(\omega_e t + \theta) + \frac{\sqrt{3}}{2} \sin(\omega_e t + \theta) \right] \end{aligned} \quad (8-24)$$

- 1 The sum of Eqs. (8-23) and (8-24) is

$$J_{s2} + J_{s3} = \frac{2J_m}{\pi} [2\cos(\omega_e t - \theta) - \cos(\omega_e t + \theta)]$$

- 2 and when this is combined with J_1 , the $\cos(\omega_e t + \theta)$ term cancels to give

$$J_s = J_{s1} + J_{s2} + J_{s3} = \frac{6J_m}{\pi} \cos(\omega_e t - \theta) \quad (8-25)$$

- 3 Thus it has been shown trigonometrically that when all three windings act
at once, the result is a counter-clockwise rotating equivalent current
distribution like Fig. 8-18a. The direction of rotation resulted from our
choice of phase relationships for the currents. If we had used $i_2 =$
4 $I_m \cos(\omega_e t + 2\pi/3)$ and $i_3 = I_m \cos(\omega_e t - 2\pi/3)$, the $\cos(\omega_e t - \theta)$ terms would have
cancelled, leaving $\cos(\omega_e t + \theta)$ in Eq. (8-24). This would represent a rotation
in the clockwise direction.

- Equation (8-25) for the three-phase machine, is to be compared with Eq.
5 (8-20) for the two-phase case. It is evident that the only change is that
the factor 4 has been replaced by 6. The comparatively simple results
expressed by Eqs. (8-20) and (8-25) are important because they show that
6 either the two-phase or the three-phase windings, when properly excited,
are equivalent to a current distribution rotating in one direction only.
Thus, when a rotor is placed under the magnetic influence of such a set of
windings it will experience a starting torque, and will rotate in only one
direction, this direction being determined by the phase relationships among
7 the stator currents.

- A word is in order concerning how power is supplied to a three-phase
motor. In Fig. 8-24a the three motor windings, which are represented
symbolically, are connected by separate circuits to three generators
8 rotating on the same shaft. The common generator shaft is necessary to
ensure the desired 120° phase differences among the currents in the three
circuits.

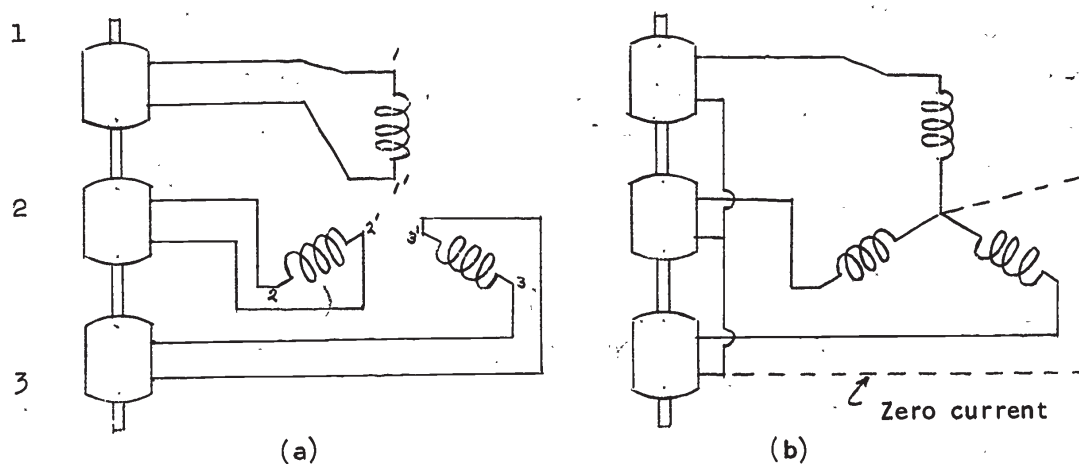


Figure 8-24.

All these wires are not necessary. To see why, observe that the bottom wires of each pair can be combined into a single wire, as in Fig. 8-24b. Each circuit can still operate independently, although the three circuits are now electrically connected on one side.* Furthermore, from Eqs. (8-21) we can write

$$i_1 = I_m \cos \omega_e t$$

$$i_2 = I_m \cos(\omega_e t - \frac{2\pi}{3}) = I_m \left[-\frac{1}{2} \cos \omega_e t + \frac{\sqrt{3}}{2} \sin \omega_e t \right]$$

$$i_3 = I_m \cos(\omega_e t + \frac{2\pi}{3}) = I_m \left[-\frac{1}{2} \cos \omega_e t - \frac{\sqrt{3}}{2} \sin \omega_e t \right]$$

and so it is seen that

$$i_1 + i_2 + i_3 = 0 \quad (8-26)$$

Thus, no current will flow in the common (dashed) wire, and so it can be omitted.

We have arrived at the conclusion that power can be supplied to a three-phase motor through three wires. To round out this brief discussion of three-phase power transmission, it is to be mentioned that the three generators do not need to be physically separate as in Fig. 8-24. They can be incorporated on a single stator, and always are. In fact, the three-phase system of windings we have described for the induction motor will also serve as the set of three windings of a three phase generator.

*The circuits are independent if the impedance of the common wire is negligible.

The principles involved in the three-phase induction motor provide the explanation why all commercial power systems are three-phase. This is why the transmission lines you see over the countryside always have three wires (or multiples of three).

8-10 Induction Motor Speed-Torque Curves

It is beyond the scope of this treatment to derive the equation for the speed-torque characteristic of an induction motor. We therefore give typical examples without derivation, in Fig. 8-25. Curve A is typical for a three-phase motor with squirrel cage rotor. In the running range, speed is relatively constant. However, if the load torque is increased too far, the knee of the curve will be reached (this is called the pull-out torque) and the motor will suddenly come to a stop. This phenomenon is due to changing phase relationships in the rotor currents. Such a characteristic is not satisfactory for starting against a large torque. A squirrel cage induction motor would not be satisfactory in electric traction applications. They are satisfactory, however, in such an application as a lathe drive, where starting is only against the friction of the lathe, the cutting load being applied later.

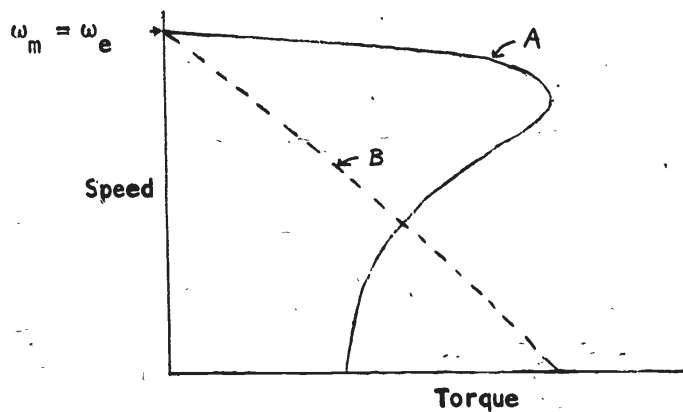


Figure 8-25.

The starting torque of an induction motor can be improved by increasing the electrical resistance of the rotor conductors. For this reason, some motors are made with a wound rotor (with a winding essentially like the stator) with connections to three external variable resistors through a set of slip rings. By increasing this resistance, a characteristic like B can be obtained,

- 1 giving maximum starting torque. After the motor is running, the external
resistance can be removed, causing the characteristic to become like curve
A.

- 2 Concerning current and power, it is recalled that a-c power is given
by $VI \cos \theta$, where θ is the phase difference between voltage and current.
If V and I are for any one phase, the total power is

$$P = 3VI \cos \theta$$

- 3 The current of an induction motor does not vary widely from no load to
full load, as in the case of d-c machines. The no-load current is large,
of the order of half the full-load current. Most of the change in power
comes about through change in phase angle. At no load (running but with
4 no load torque) the current will lag the voltage by nearly 90° , and at
full load an angle of 20° is reasonable.

- 5 These relatively large phase angles are a disadvantage which cause
considerable difficulty to power companies. It may be necessary to install
capacitors in the vicinity of large industrial loads, to compensate for the
low power factor caused by a large number of induction motors.

- 6 In regard to starting, the use of rotor circuit resistance in a wound
rotor machine has been mentioned. Small squirrel cage motors (up to about
2 hp.) can be started at full line voltage. However, it may be necessary
to remove protective devices (fuses or circuit breakers) in order to prevent
their operating on the heavy current that flows for a few cycles. Larger
squirrel cage motors are started by applying reduced voltage, through a
7 bank of transformers.

8-11. Motor Ratings

- 8 In the discussion of starting d-c motors it was pointed out that starting
current can be permitted to exceed rated current for operation at full load.
The name-plate current rating of a machine is determined by the allowable
temperature rise which, in turn is primarily influenced by the insulation
used in the windings.

- 9 Rated current is based on continuous operation. Momentary loads can
cause the current to exceed rated value by considerable amounts, if they
endure only for a few seconds. In order to obtain good service from a motor,
it is necessary that it operated within its name-plate current ratings.

- 1 Operation of a motor at other than its rated voltage is not generally recommended, except in the case of d-c motors which can be operated at reduced voltage. An induction motor should not be operated for long at reduced voltage. This is because its speed is largely determined by the line frequency, and
- 2 therefore with a given load its power output is nearly independent of voltage. If the voltage is reduced, the current must increase, to maintain the power, and the maximum allowable current may be exceeded.

8-12. Number of Poles

- 3 All of the discussion in this text (and the programmed text) assumes machines of two poles. However, four, six, etc. pole machines are possible. The characteristics of d-c machines are independent of the number of poles. The only things affected are the armature winding and voltage and current
- 4 ratings.

- Although poles are not physically present in an induction motor, the equivalent of a pole is defined by the extent of the circumference of the stator subtended by one coil. In our illustrations, this has been 180° , which defines a two-pole motor. When there are four poles, the angle subtended is 90° , etc. In our example, the equivalent current distribution rotates at an angular speed ω_e , and this is the maximum speed of the rotor. With a four pole winding, the rotating current distribution will have two
- 6 cycles of current variation (variation with angle) instead of one cycle as in the two pole case. Functions like

$$\cos(\omega_e t - 2\theta)$$

- 7 will appear. Such a wave moves with angular speed $\omega_e/2$. Likewise, with six poles the speed is $\omega_e/3$, and in general is $\omega_e/(\text{number of pole pairs})$.

- The speeds discussed here are the maximum (or synchronous) speeds. Actual speeds are usually higher than 95% of synchronous speed. At 60 cps.
- 8 $\omega_e = 377$ radians per second or 3600 RPM. This is approximately the speed of a two-pole machine. Four and six pole machines are very popular, and have approximate operating speeds, respectively, of 1800 and 1200 RPM. Thus, if you see a name-plate indicating a speed of 1740 RPM, this is a
- 9 four-pole motor.

BASIC SEMICONDUCTOR THEORY

Each of the following paragraphs presents a few ideas that are essential to an understanding of semiconductor theory. Be sure you know the material in each paragraph before going on to the next.

1. Electric current occurs because of moving charges. Current may flow in a metal, a liquid, a gas, and a vacuum, but the charge carriers may vary from one medium to another. Commonly, charges are carried as a result of the movement of electrons or ions.
2. The structure of an atom can be represented as a positively charged nucleus with electrons moving in various orbits around this nucleus. The electrons in a particular orbit possess a certain level of energy. The smaller the orbit (closer to the nucleus), the lower the energy level.
3. Electrons can jump (make "transitions") from one orbit to another, or from one energy level to another. An electron may acquire additional energy from a source outside the atom (from light, heat, collisions, etc.) and jump to a higher energy orbit. On the other hand, an electron may lose energy (by radiation) and fall to a lower energy state -- make the transition to a smaller orbit.
4. The closer the electrons are to the nucleus, the more tightly they are bound to it; the difference in energy levels between small orbits being greater than the difference between larger orbits. The outermost electrons (called the "valence" electrons) are loosely bound to the nucleus. It would take only a relatively small amount of external energy to free them completely from the nucleus. The energy required to free a valence electron will be different for one type of atom than for another. For the metal germanium, only 0.7ev (electron-volts) is required.
5. The ability of a solid material to conduct electricity depends upon the availability of charge carriers. In some materials, the valence electrons are tightly bound to the nucleus and are not available as charge carriers. These materials conduct electricity poorly and are therefore used as insulators.
6. In other materials, like metals, the valence electrons are so loosely bound that they are practically free of the nucleus. Even a small applied electric field can get these electrons to move through the material. These materials are good conductors.

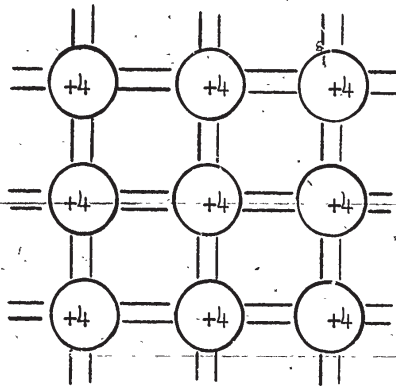
7. Still other materials fall between these two extremes; they are neither good insulators nor good conductors. Examples of such materials are germanium and silicon. They are semiconductors.
8. The number of free electrons in conductors is enormously greater than in semiconductors. At room temperatures:

For conductors, approximately 1 electron per atom.

For germanium, approximately 1 atom in 10^{10} is ionized.

For silicon, approximately 1 atom in 10^{13} is ionized.

9. Pure Ge (Germanium) or Si (Silicon) atoms are arranged in a crystal lattice. Both atoms have 4 valence electrons in the outermost shell. This largest orbit would be complete if it contained eight (8) electrons; there are, therefore, 4 empty places in this outer shell. In a regular (uniform) crystal lattice, each atom "shares" an electron with each of its four neighboring atoms. The atoms that share each other's valence electrons thus form "covalent bonds." In this way each atom comes closer to having its outer shell filled.



A 2-Dimensional Model
of a Crystal Lattice

Figure 9.1 is a two-dimensional presentation of a germanium lattice showing covalent bonds between adjacent atoms. These bonds hold the atoms fixed in space relative to each other and thus the collection of atoms forms a regular crystal lattice.

10. The atoms in a lattice, although fixed in position relative to each other, vibrate around their equilibrium positions. The degree of agitation is a function of temperature. As temperature increases the thermal agitation may be enough to break a covalent bond and an atom can become ionized by freeing an electron. (It takes .7ev of energy to free a valence electron in Ge.) Free electrons formed in this manner leave behind a hole, an empty position for another electron to fill. These holes effectively have a positive charge.

11. As free electrons move about in the crystal, they may encounter a hole and by "captured" by it. The electron then is recombined with an atom. These two tendencies -- ionization and recombination -- are opposite in nature. Under equilibrium conditions at a given temperature the rates of ionization and recombination are the same. In the process of ionization and recombination, as electrons move about, the holes appear to move as well. For all intents and purposes the holes can be considered as mobile charge carriers like electrons except that they have a positive sign.
12. In addition to pure crystals of a semiconductor, it is possible to have semiconductor crystals in which some foreign, impurity atoms are introduced deliberately. It is possible to increase the density of free electrons in a crystal by adding impurity atoms having a valence of +5 (five valence electrons in the outer shell). When this +5 material is incorporated in

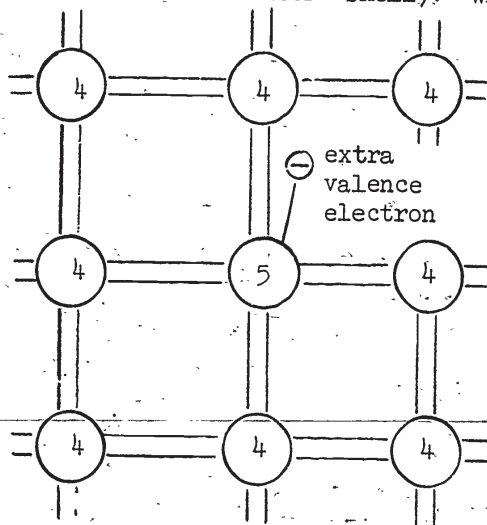


Fig. 12.1
N-Type Material

a germanium crystal, covalent bonds are again formed by sharing electrons with neighboring atoms, but now the fifth valence electron cannot be shared in any covalent bond. Since the outer layers of all atoms are effectively filled, the extra electron is not bound to any nucleus and is thus effectively free. In this way, there are free electrons without corresponding holes. The valence-5 atom "donates" a free electron and so is called a donor atom. (Actually, it requires about .01 eV of energy to

free this electron compared with .7 eV

for pure germanium.) Since the charge carriers having a negative charge (electrons) are in greater number than those carrying a positive charge (holes), a semiconductor material having these characteristics is called N-type (Negative-type). The material is electrically neutral, however; there is no net charge. The process of adding the impurity atoms is called doping. In typical doped semiconductors, the fraction of impurity atoms is of the order of one part in 10^6 .

13. Another alternative is to include some impurity atoms with valence +3 (like gallium or indium). Each such atom again shares electrons with its neighbors

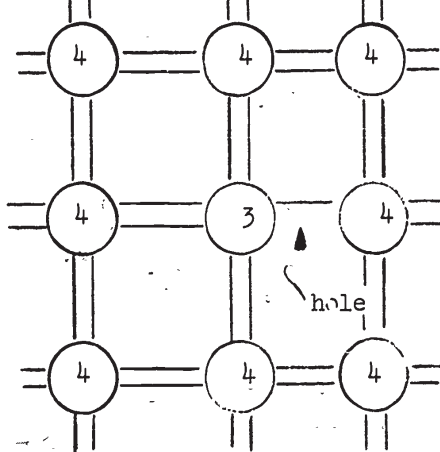


Fig. 13.1
P-Type Material

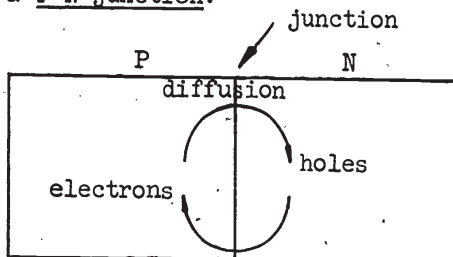
to form covalent bonds. But now there aren't enough electrons to go around, so that a hole is formed. Any free electrons moving by can "fall into" this hole. Since the valence-3 atom "accepts" such electrons, it is called an acceptor atom. Each such impurity atom introduces an extra hole without a corresponding electron. Since the majority charge carriers in this type of material carry a positive charge, the material is called P-type.

14. Charge carriers move through a semiconductor by two mechanisms: by diffusion and by drift. (a) Diffusion is the process connected with random motion due to thermal agitation. (b) Drift motion is due to an externally applied electric field and is superimposed on top of the random motion. At room temperature the diffusion velocity in crystals is of the order of 10^5 meters/sec., which is quite high. On the other hand, drift velocities are much smaller. For example, for N-type germanium the drift velocity is of the order of 1 meter/sec.

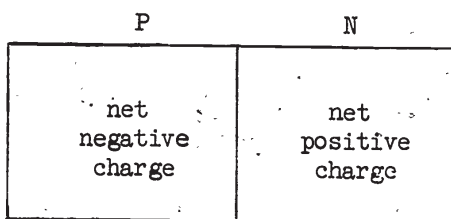
Summary. In doped semiconductor crystals, there are two sources of charge carriers: those due to thermal ionization of any atom in the crystal, and those resulting from the impurity atoms. In typical semiconductor devices, the density of impurity atoms is in the range 10^{15} to 10^{17} per cm^3 . The density of charge carriers due to thermal ionization is only about 10^{13} per cm^3 for germanium and 10^{10} per cm^3 for silicon. Since each impurity atom leads to 1 charge carrier, it is seen that the number of charge carriers due to impurity atoms greatly exceeds that due to thermal ionization. Thus, for P-type material the majority carriers are holes and for N-type materials the majority carriers are electrons.

The P-N Junction Diode

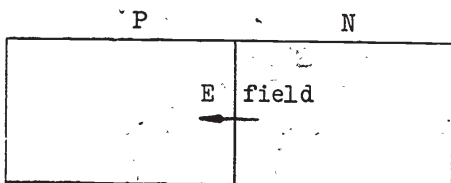
15. A piece of P-type material in contact with a piece of N-type material forms a P-N junction.



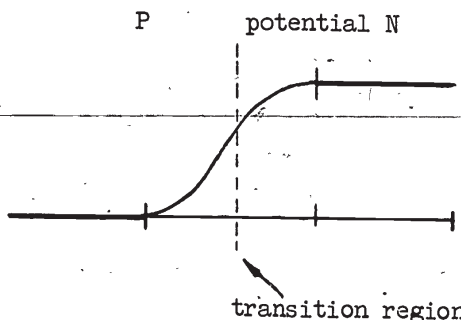
The charge carriers in each type diffuse into the other material. Thus, electrons (majority carriers in N type) diffuse to the left and holes (majority carriers in P-type) diffuse to the right.



Electrons leaving the N region leave it positively charged and at the same time carry an excess negative charge into the P region. Holes leaving the P region add to this effect by leaving the P negatively charged and carrying a positive charge to the N region.

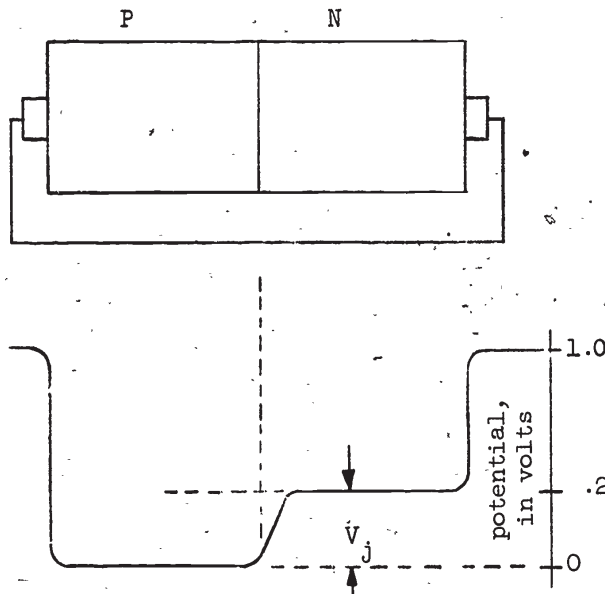


This separation of charges will produce an electric field which tends to inhibit further diffusion of charges.

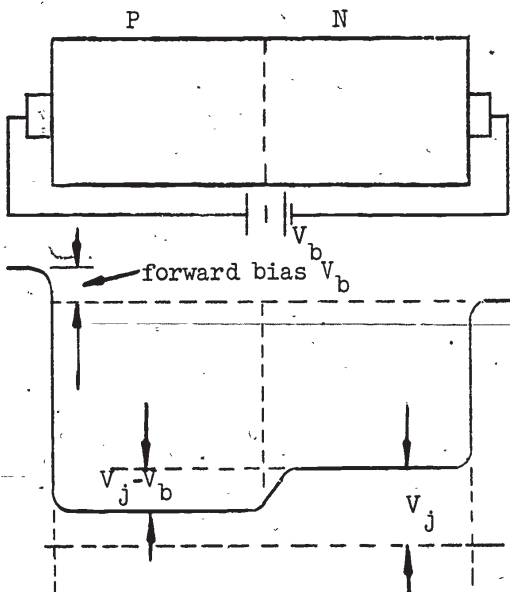


Because of the electric field at the junction, there will be a difference of potential across the junction between the two regions, as illustrated. The potential does not change abruptly when going from one region to the other, but changes gradually throughout a transition region surrounding the junction. (The width of this transition region is typically 1 micron = 10^{-6} meter.) The potential "hill" constitutes a barrier

to the continuation of the diffusion process. The height of the "hill" is determined by an equilibrium between thermal ionization on both sides of the junction and diffusion of carriers across the junction. In equilibrium, there is no net current flow across the junction.

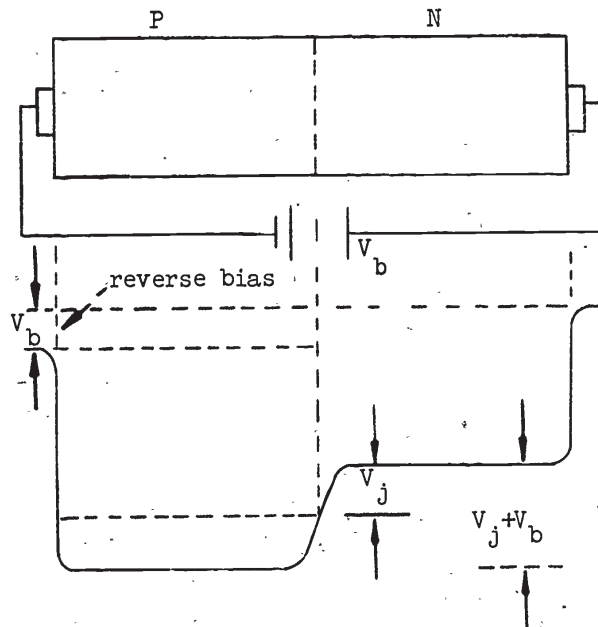


The device formed in this manner is a junction diode. In order to connect the diode in a circuit, metal leads are attached to the two ends. Suppose the ohmic voltages at the lead junctions are negligible. Now let the two leads be short circuited, so that the potentials of the two ends are the same. The diagram shows the resulting variation of potential throughout the device. The values shown are typical. There will be no current flowing in the short circuit.



Now let a battery be connected across the diode with the positive terminal connected to the P-region. The potential of the P-region is increased by the battery voltage V_b while the potential of the N-region remains the same. Thus, the potential hill at the junction is reduced by the voltage V_b , becoming $V_j - V_b$. The equilibrium which kept the net current across the junction is thus disturbed. Since the potential barrier has been lowered, it is easier for the majority carriers (electrons in the N-region and holes in the P-region) to

diffuse in their natural direction. This flow will not now cause an increase in the junction potential since the charge carriers do not accumulate, they continue flowing around the closed circuit. The variation of potential now takes the form shown. The diode is said to be forward-biased.



If the polarity of the biasing voltage is reversed, the diode is said to be reverse-biased. Now the potential of the P-region is reduced by V_b while that of the N-region is unchanged. Hence, the height of the potential hill is increased to $V_j + V_b$. The majority carriers will now be hindered in moving in their natural direction. However, there is always some thermal ionization leading to minority carriers in each region (electrons in P-region, holes in N-region). For these minority carriers, the direction of the electric field is just right for helping their diffusion. Thus,

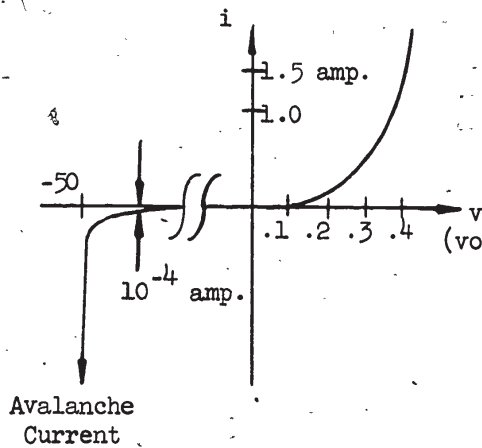
there will be a certain amount of reverse current. The magnitude of the reverse current compared with that of the forward current is approximately as the density of minority carriers to that of majority carriers. This is of the order of 10^{-2} to 10^{-4} for germanium diodes and 10^{-5} to 10^{-7} for silicon diodes. Since the origin of reverse current is thermal ionization, this current is sensitive to temperature.

Zener and Avalanche Effects

16. If the reverse bias applied to a diode is increased (V_b is made larger), the potential barrier will increase in height and so will the magnitude of the electric field. As the field increases it will become so high that electrons can be torn away from the covalent bonds, thus creating hole-electron pairs in the transition region. Although these pairs are not created by thermal ionization, their presence will lead to an increase in reverse current. In fact, this increase will be a large one. This effect is called the Zener effect.

There is another possibility for increasing the reverse current without the breaking of covalent bonds by the Zener effect. The carriers constituting the small reverse current will have collisions with atoms in the

crystal lattice. If the electric field in the transition region becomes high enough, these carriers may gain sufficient energy between collisions to knock electrons out of the covalent bond, again creating an electron-hole pair which contributes to the reverse current. This effect is cumulative since the carriers so created can also lead to ionizing collisions. The result is called the avalanche effect.



Whichever mechanism is involved, the result is a great increase in reverse current which takes place without much further increase in reverse bias voltage.

A sketch of typical diode current against bias voltage is shown. (Note that the scales on the positive axes are different from those on the negative axis.)

A REVIEW TEST

Test yourself on your retention of the material you have just read. If you are unsure of any answers, go back and re-read the related paragraphs.

1. In a P-type semiconductor material consisting of germanium doped with an impurity material, the valence of the impurity atoms is:
 - a) 4
 - b) 3
 - c) 5
 - d) either 3 or 5
 2. In an N-type semiconductor material consisting of germanium doped with an impurity material, the valence of the impurity atoms is:
 - a) 4
 - b) 3
 - c) 5
 - d) either 3 or 5
 3. In a P-type semiconductor the majority carriers are holes.
 - a) True
 - b) False
 4. In an N-type semiconductor the minority carriers are holes.
 - a) True
 - b) False
-
5. A P-type semiconductor has an excess of holes over free electrons. The material as a whole is:
 - a) negatively charged
 - b) neutral
 - c) positively charged

6. An N-type semiconductor has an excess of free electrons over holes. The material as a whole is:
- a) negatively charged.
 - b) neutral.
 - c) positively charged.
7. Charge carriers move through semiconductors as a result of:
- a) drift due to electric fields.
 - b) diffusion due to random processes.
 - c) both a) and b)
 - d) none of the above
8. Forward bias on a P-N diode:
- a) increases the potential hill at the junction.
 - b) decreases the potential hill at the junction.
9. Reverse bias on a P-N diode:
- a) increases the potential hill.
 - b) decreases the potential hill.
10. A P-N diode will not break down, no matter how high the reverse bias.
- a) True
 - b) False
11. The avalanche effect is the result of:
- ~~a) operating the device at improper temperatures.~~
 - b) applying excessive reverse bias.
 - c) doping the germanium incorrectly.

Chapter 9

TRANSISTOR AMPLIFIERS

The Transistor

A transistor is a semiconductor device in which two junction diodes are effectively placed back to back, the net result consisting of three layers of material, as shown in Fig. 1(a).

The center of the "sandwich" is P-type material while the outer layers are N-type. The resulting device is called an N-P-N junction transistor. The center slice is called the base; the lower N-region is called the emitter and the upper N-region, the collector.

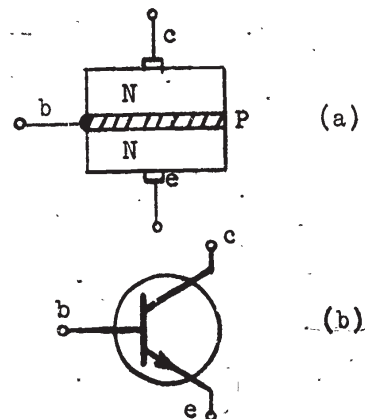


Fig. 9-1 - N-P-N Transistor

Rather than drawing the pictorial representation of a transistor shown in Fig. 1(a), a standard symbol is used as shown in Fig. 1(b). The emitter is distinguished from the collector by means of the arrow. The direction of the arrow is the direction of forward current.

Remember from the discussion of P-N junctions that the junction is forward-biased when the potential of the P-region is made positive relative to that of the N-region; that is, from the base to the emitter. The arrow on the transistor symbol will, thus, have the direction shown.

Forward current across a junction generally has contributions from two sources: holes (which are the majority carriers in the P-region) moving from P to N and electrons (which are the majority carriers in the N-region) moving from N to P. In the junction transistor the forward current actually consists largely of electrons injected from the emitter to the base. (Remember that negative charge flow is opposite to the direction of the current.) This is achieved by making the base region physically very small and also making the

concentration of impurity atoms in the base small compared with that of the emitter. The forward current flow is facilitated by biasing the base-emitter junction (which will be more simply referred to as the emitter junction) in a forward direction; that is, making the base positive relative to the emitter

In the base region (P-type material), the electrons are minority carriers; as they diffuse through the base, some of them recombine with holes, which abound there. If the base layer is thin enough most of the electrons will reach the base-collector junction without recombining. If we now arrange to create a field drawing these electrons across the base-collector junction, we will, in effect, cause an electron flow from emitter, through the base, to the collector (that is, cause a current to flow from collector to emitter). Such a field is obtained by biasing the base-to-collector junction in the reverse direction; that is, making the base negative relative to the collector. The appropriate biases are shown in Fig. 2

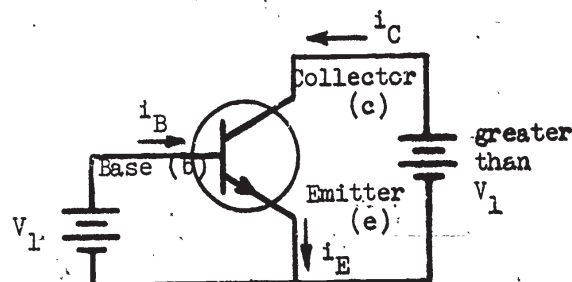


Fig. 9-2

With such an arrangement, the collector current i_c is approximately equal to the emitter current i_e . There is some base current which supplies the few holes that recombine with electrons in the base, and also supplies the holes which diffuse across the emitter junction. By proper choice of the densities of impurities in the base and emitter regions, and by making the base very thin, the hole current across the emitter junction can be made much smaller than the

electron current across this junction. Hence, a small base current will control a large collector current, thus making possible current amplification.

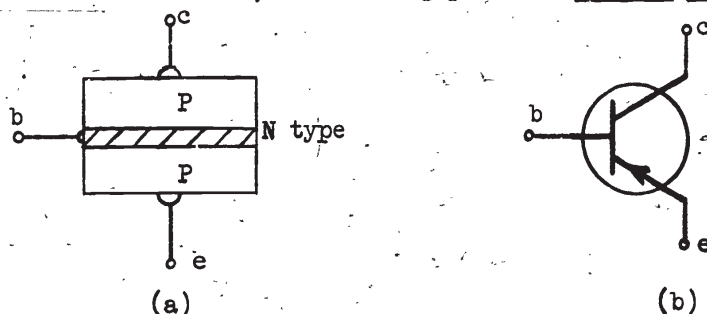


Fig. 9-3. P-N-P Transistor

This is a basic purpose of the transistor.

It is also possible to make a junction transistor of the P-N-P variety by reversing the materials of the N-P-N transistor. A base of N-type semiconductor is sandwiched between two pieces of P-type material, as shown in Fig. 3(a). Forward current now consists of holes moving across the emitter junction from emitter to base. The arrow on the emitter in the symbol for the P-N-P transistor shown in Fig. 3(b) indicates this. Note that the arrow is always directed from P to N, so the direction of the arrow alone is enough to distinguish between N-P-N and P-N-P transistors.

In order to bias the emitter junction of a P-N-P transistor in the forward direction, it is necessary to make the base negative relative to the emitter. Similarly, to give the collector junction a reverse bias it is necessary to make the base positive to the collector. These are just the opposites of what was needed in an N-P-N transistor. In a circuit, a P-N-P transistor can replace a similar N-P-N transistor provided the polarities of all bias batteries are reversed.

The Transistor Amplifier

Figure 4 is a diagram of a basic transistor amplifier. The arrow

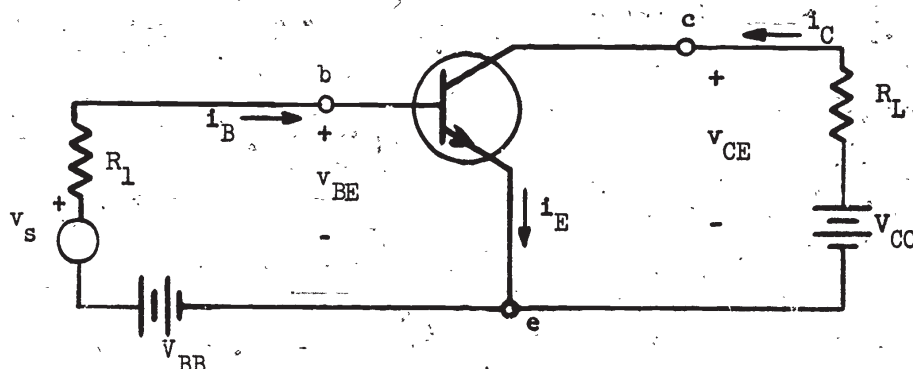


Fig. 9-4

on the emitter shows the transistor to be a N-P-N transistor. The batteries V_{CC} and V_{BB} provides biasing and the source v_s provides the signal which is to be amplified.

The behavior of the transistor can be described in terms of its terminal voltages and currents. By Kirchhoff's voltage law the three possible voltages across the transistor terminals are related by

$$v_{BC} = v_{BE} - v_{CE} \quad (9-1)$$

Thus, giving two of the voltages will determine the third; only two of them are independent. Similarly the three currents are related by

$$i_E = i_B + i_C \quad (9-2)$$

and again only two of the currents are independent.

Now, in order to determine how the two voltages and the two currents are related, we vary some of them and measure the variations in the others.

(It is not important how they are varied; it can be done by changing V_{CC} or V_{BB} .) Typical curves obtained are shown in Fig. 5. To get a complete description of the current-voltage relationships two families of curves are required.

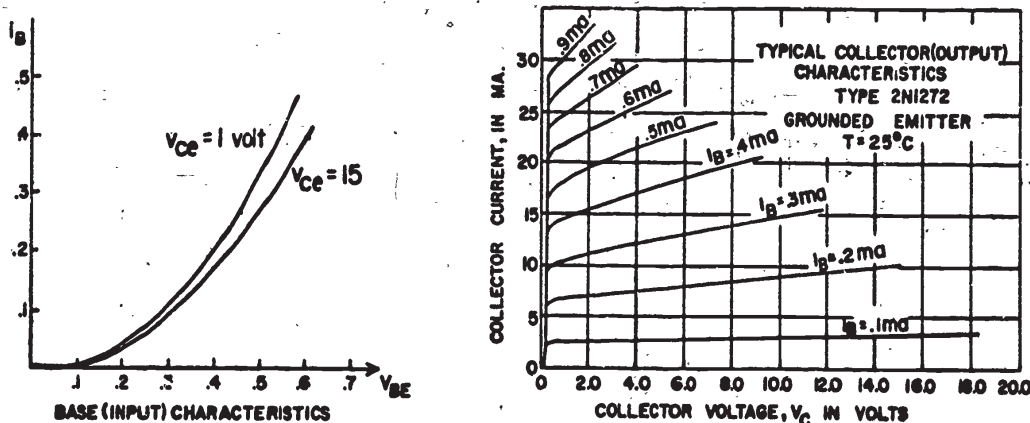


FIG. 5 TRANSISTOR CHARACTERISTICS

In one family of curves the output (collector to emitter) voltage is held constant at some value and a plot of the variation of base current with input (base to emitter) voltage is made. This is repeated with V_{CE} held at other constant values. Figure 5(a) shows that there isn't much variation in this input characteristic as V_{CE} varies from 1 to 15 volts.

The second family of curves is obtained by holding the base current constant and plotting variations of output (collector) current with output (collector to emitter) voltage. The result is shown in Fig. 5(b). Note that even with zero base current there is some output current which is approximately constant for all values of V_{CE} . Note the common features of this family of collector characteristic curves. Except for low values of V_{CE} , the curves are almost horizontal lines with small upward slopes. They are almost evenly spaced for equal increments in base current.

In order to analyze the operation of the amplifier shown in Fig. 4, we write two loop equations around the base-emitter loop and the collector-

9-6

emitter loop.

Thus,

$$v_{BE} - V_{BB} - v_s + R_1 i_B = 0 \quad (9-3)$$

$$V_{CC} - v_{CE} - R_L i_C = 0 \quad (9-4)$$

Solving these for i_B and i_C leads to

$$i_B = \frac{V_{BB} - v_{BE}}{R_1} + \frac{v_s}{R_1} \quad (9-5)$$

$$i_C = \frac{V_{CC}}{R_L} - \frac{v_{CE}}{R_L} \quad (9-6)$$

These are two equations in four unknowns. Two more relationships among

the unknowns are required for a solution. And we do have two such relationships given by the transistor characteristics in Fig. 5. The input characteristic shows that v_{BE} is small; it is normally small (one tenth or less) compared with V_{bb} . Hence, to a first approximation, it can be neglected in Eq.(9-5) and this equation can be rewritten as

$$i_B = \frac{V_{BB}}{R_1} + \frac{v_s}{R_1} = I_B + i_s \quad (9-7)$$

where $i_s = v_s/R_1$ is the signal current and, it is assumed that v_s is a time-varying voltage whose average value is zero; $I_B = V_{BB}/R_1$ is, hence, the average value of the base current.

Equation (9-6) is the equation of a straight line whose variables are

the coordinates of the collector characteristic shown in Fig. 5(b). This straight line can be superimposed on the collector curves, as shown in Fig. 6.

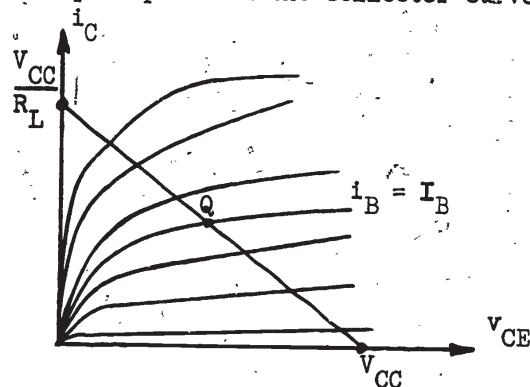


Fig. 9-6

This line is called the load line. (You will recall a similar graphical analysis of a diode circuit.) The signal voltage v_s is time varying and will cause the base current i_B to vary. For any instantaneous value of i_B , the corresponding value of i_C and V_{CC} can be determined by noting the coordinates of the intersection of the load line and the appropriate i_B curve. Thus, as the signal voltage varies, the collector current and v_{CE} will both change, but in such a way that their values will always lie on the load line.

When the signal current ($i_s = V_s/R_L$) is zero, the base current i_B will have its average value I_B (Eq. 9-7). The intersection of the load line with the i_B -curve corresponding to this average value of base current I_B is called the operating point and is labeled Q in Fig. 6.

It is of interest to determine the variations in collector (output) current as a function of the signal current i_s . This can be done by assuming positive and negative values of i_s , adding I_B (the average base current) to get i_B , then reading the corresponding value of i_C from the intersection of the load line with the corresponding curve. The result is graph shown in Fig. 7.

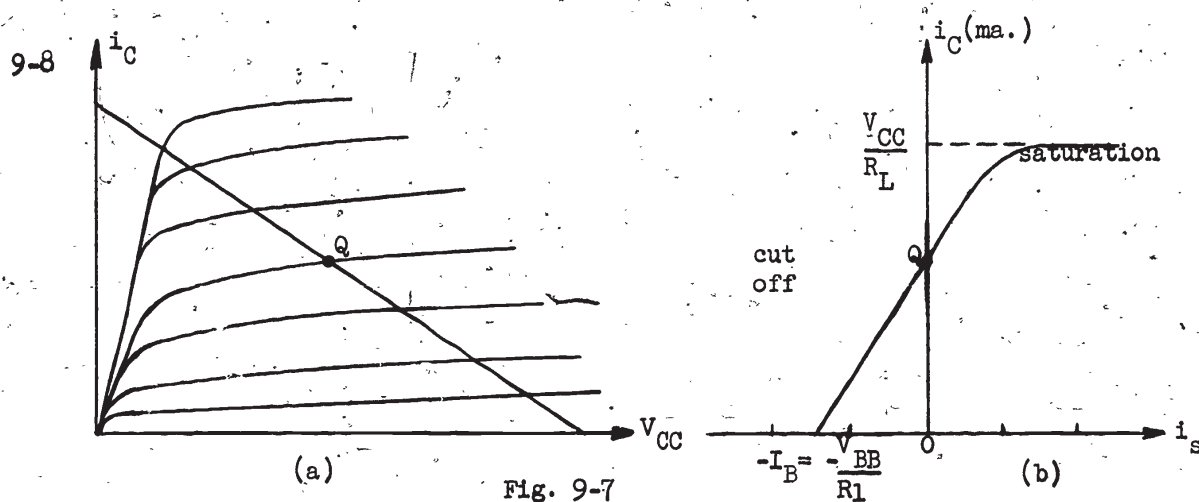


Fig. 9-7

Over a substantial part of the range this curve is approximately linear indicating that the output (collector) current is approximately proportional to the input (signal) current, the proportionality constant being the slope of the curve.

When the signal current becomes sufficiently negative, the collector is cut off and no collector current flows. At the other end, for large positive values of signal current, the collector current reaches a maximum and from then on it is saturated. If an input signal is to be reproduced faithfully, its

amplitude is limited by these considerations. In the design of the amplifier the operating point must be so chosen that the point $i_s = 0$ falls near the center of the linear portion of the curve of Fig. 7, assuming that the signal current has equal positive and negative peaks, like a sinusoid.

Example: Let the curves in Fig. 8 be the approximate collector characteristics of an N-P-N junction transistor to be used in the amplifier circuit of Fig. 9. The bias battery voltages are $V_{CC} = V_{BB} = 15$ volts and $R_1 = 75 \text{ K ohms}$. It is required to find the value of R_L which will cause v_{CE} at the operating point to be 10 volts. It is also required to determine the peak value of signal voltage which can be amplified without distortion.

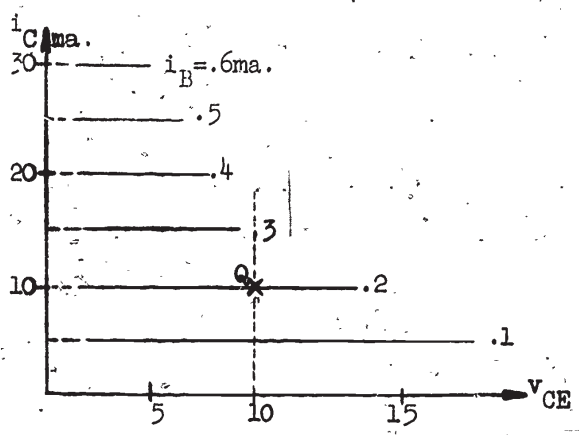


Fig. 9-8

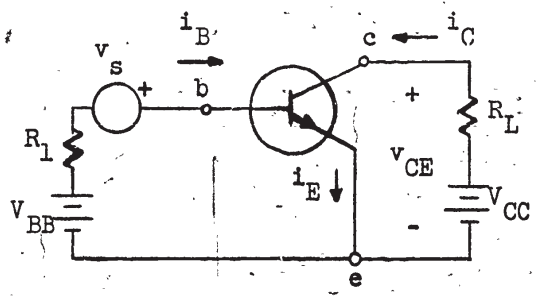


Fig. 9-9

Making the approximation that v_{BE} can be neglected relative to V_{BB} (15 volts), the average value of base current will be $i_B = 15/75000 = 0.2$ ma. The coordinates of the operating point will then be $v_{CE} = 10$, $i_C = 10$ as shown by the X in Fig. 8. Since another point on the load line is $v_{CE} = 15$, $i_C = 0$, the load line can now be drawn, giving an intercept on the i_C axis of 30 ma. The result is shown in Fig. 10. The slope of the load line is found to be $-30 \times 15/15 = -1/500$. From Eq. (9-6) the slope is $-1/R_L$; hence, $R_L = 500$ ohms.

Assume that the input signal v_s has equal positive and negative peaks (for example, the output of a phonograph pickup would be such a symmetrical wave). This signal would cause i_B to vary along the load line, around the operating point Q. If the signal were large enough, the base current could go negative, driving i_C to cut-off (Fig. 7b). Under this condition, the output of the amplifier would be distorted; a portion of the input signal v_s would be lost.

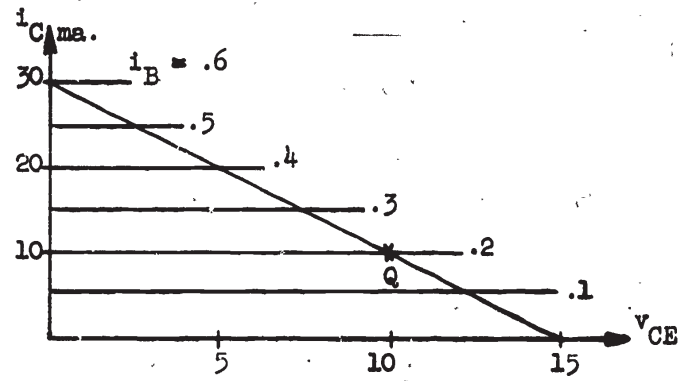


Fig. 9-10

9-10

Since from Eq. (9-7) the total base current equals the average value plus the signal current, the negative peak of the signal current should not exceed 0.2 ma, since $I_B = 0.2$. This means that the largest base current will be .4 ma.

The characteristics show that the output current is linearly related to the base current over a base current range from 0 to .6ma. If the operating point can be moved to the center of this range, then a signal having a larger peak-to-peak value could be amplified without distortion.

Find the value of R_1 and R_L required to move the operating point to

$$I_B = .3 \text{ ma}, V_{CE} = 7.5 \text{ volts}$$

$$\text{Ans: } R_1 = 50K$$

$$R_L = 500 \text{ ohms}$$

Symbols for Circuit Variables

The symbols we have been using for the voltages and currents in a transistor amplifier (such as in Fig. 11, below) have been lower case (small) v and i with upper case (capital) letter subscripts. These capital subscripts refer to the total instantaneous values. In discussing the operation of amplifiers we will have occasion to talk about both the average values of voltages and currents as well as their instantaneous values. Furthermore, these voltages and currents can be written as the sum of their average value plus an alternating component whose average value is zero. So there is need for other symbols to designate these alternating components.

The following conventions will be used.

1. Capital letters with repeated capital subscripts will represent supply or bias sources. Thus, V_{CC} is the collector supply voltage and V_{BB} is the base bias voltage. (Note the two subscripts are the same.)
2. Capital letters with capital subscripts, either single or two different letters, will represent average values. Thus, I_B is the average base current and V_{CE} is the average collector to emitter voltage.
3. Lower case letters with capital subscripts, with single or double, will represent total instantaneous values. Thus, v_{BE} is the total base to emitter voltage; an average value plus an a-c component. (Note, in case of double subscripts, the two subscripts are different.)
4. Lower case letters with lower case subscripts will represent instantaneous values of alternating components. Thus, i_e is the alternating component of emitter current and v_{be} is the alternating component of base-to-emitter voltage (their average values will be zero).

5. Capital letters with lower case subscripts will refer to the rms values of alternating components. Thus, I_e will be the rms value of i_e , and V_{be} will be the rms value of v_{be} .

Self-Test. Can you select the correct answers to the following questions?

There may be more than one correct answer. (Check your answers against those given in the footnote, below.)

1. The term " v_{CE} " designates
 - A. The instantaneous voltage from collector to emitter.
 - B. The sum of V_{CC} and $i_e R_L$.
 - C. The sum of $V_{CE} + v_{ce}$.
2. The term V_{be} refers to
 - A. The average value of base-to-emitter voltage.
 - B. The rms value of base-to-emitter voltage.
 - C. The value of $I_e R_L$.
3. The average value of collector-to-emitter voltage is
 - A. V_{CE} .
 - B. $v_{CE} - v_{ce}$.
 - C. V_{ce} .

Answers: 1. A,C; 2. B,C; 3. A,B.

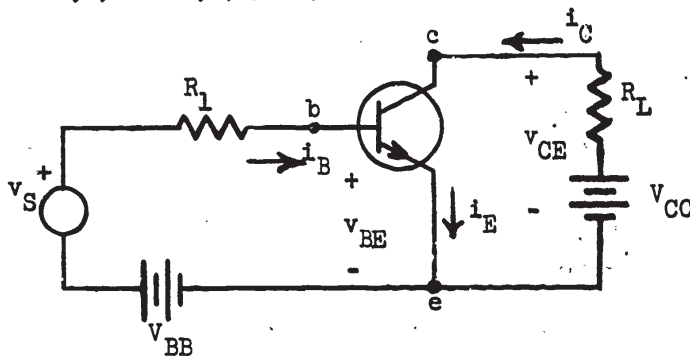


Fig. 9-11

Power in the Collector Circuit

Having established the convention for the notations to be used, let us concentrate on the collector circuit of Fig. 11 ----- the loop formed by the collector and emitter terminals of the transistor, the supply voltage V_{CC} and the load resistance R_L . Interest is to be focussed on three aspects of power, (1) the power supplied by the source, (2) the power dissipated in the transistor, and (3) the power dissipated in the load. We will more specifically, be interested in the average power supplied by or dissipated in each component in the collector circuit.

The first step will be to write expressions for the pertinent voltages and currents in terms of average and alternating components. Thus

$$i_C = I_C + i_c \quad (9-8)$$

$$v_{CE} = V_{CE} + v_{ce} \quad (9-9)$$

We assume that the alternating components have zero average values.

The instantaneous values of the power supplied by the source (p_{CC}), the power dissipated in R_L (p_L), and the power dissipated in the collector (p_C) are

$$p_{CC} = V_{CC} i_C = V_{CC} I_C + V_{CC} i_c \quad (9-10)$$

$$p_L = R_L i_C^2 = R_L (I_C + i_c)^2 = R_L I_C^2 + R_L i_c^2 + 2R_L I_C i_c \quad (9-11)$$

$$\begin{aligned} p_C &= v_{CE} i_C = (V_{CE} + v_{ce}) (I_C + i_c) \\ &= V_{CE} I_C + v_{ce} i_c + V_{CE} i_c + I_C v_{ce} \end{aligned} \quad (9-12)$$

Let us represent the average powers by capital P's and note that the average value of the alternating components are zero. We now take the average value of each of the last three expressions. We find:

$$P_{CC} = V_{CC} I_C \quad (9-13)$$

$$\begin{aligned} P_L &= R_L I_C^2 + \text{average value of } R_L i_c^2 \\ &= R_L I_C^2 + R_L I_c^2 \end{aligned} \quad (9-14)$$

$$\begin{aligned} P_C &= V_{CE} I_C + \text{average value of } v_{ce} i_c \\ &= V_{CE} I_C + V_{ce} I_c \end{aligned} \quad (9-15)$$

Thus, the average power delivered by the supply source is not influenced by the presence or absence of a signal. On the other hand, both the average load power and the average power dissipated in the transistor are dependent on the signal.

Note from Eq. (9-14) that the average load power has contributions from a dc component $R_L I_C^2$ and from an ac component, $R_L I_c^2$. (Remember that I_c is the rms value of the ac component of the collector current.) The same is true of the average collector power (dissipated in the transistor).

The three average powers can be related by noting that

$$V_{CE} = V_{CC} - R_L I_C \quad (9-16)$$

When this is used in Eq. (9-12) instead of the alternative expression for V_{CE}

given in Eq. (9-9), the result for the instantaneous collector power becomes

$$p_C = V_{CC}i_C - R_L i_C^2 = V_{CC}(I_C + i_c) - R_L(I_C + i_c)^2$$

Taking the average value leads to

$$P_C = \underbrace{V_{CC}I_C}_{\substack{\text{average power} \\ \text{supplied by} \\ \text{source}}} - \underbrace{(R_L I_C^2)}_{\substack{\text{dc component} \\ \text{of average} \\ \text{load power}}} + \underbrace{R_L I_c^2}_{\substack{\text{ac component} \\ \text{of average} \\ \text{load power}}} \quad (9-17)$$

or

$$P_C = P_{CC} - P_L \quad (9-18)$$

Under quiescent (no-signal) operating conditions, the average load power is just $R_L I_C^2$. With a signal, this average load power increases to $R_L(I_C^2 + I_c^2)$. Since the average power delivered by the source is constant, this means that, the average power dissipated in the transistor decreased with an increasing signal level. The maximum collector dissipation occurs with no signal. If we label this collector dissipation with no signal as P_{CO} , the collector dissipation can be expressed in terms of:

$$P_C = P_{CO} - R_L I_c^2 \quad (9-19)$$

where

$$P_{CO} = I_C(V_{CC} - R_L I_C)$$

Example

Problem An N-P-N junction transistor is used in the amplifier circuit of Fig. 11. The transistor has a maximum permissible power dissipation of 2.3 watts.

9-16

The source voltages are $V_{CC} = 10$ and $V_{BB} = 14$ volts and the resistors have the values $R_L = 2000$ and $R_L = 10$ ohms. The collector characteristics are approximately horizontal lines with $i_C = 60i_B$. A sinusoidal voltage signal v_s having an amplitude of 12.5 volts is applied. Find the average powers P_C , P_L and P_{CC} both under quiescent conditions and with the signal. Note whether or not the permissible dissipation is exceeded in either case.

Solution.

We first find the average base current to be $I_B = 14/2000 = 7\text{ma}$, assuming v_{BE} to be small compared with 14. The average collector current is then $I_C = 60 \times 7 = 420\text{ma}$.

The average power delivered by the source will be $10(.42) = 4.2$ watts. Under quiescent conditions the power dissipated in the load will be $10(.42)^2 = 1.76$ watts. Hence, the collector dissipation will be $4.2 - 1.76 = 2.44$ watts, which exceeds the maximum permissible.

The rms value of the current in the load resistor will be $\frac{12.5}{\sqrt{2}} \cdot \frac{1}{2000} \cdot 60 = .265$ ma. Hence, the additional average load power will be $10(.265)^2 = .70$ watt giving a total average load power of 2.46. The collector dissipation is now reduced to $4.20 - 2.46 = 1.74$ watts, a safe value.

As the dynamic operating point (to be distinguished from the quiescent operating point, with no signal) moves up and down the load line, the instantaneous power dissipated in the collector changes. It is useful for the circuit designer to have a curve on which the collector dissipation is constant. Such a curve is obtained by setting p_C in Eq. (9-12) equal to a constant. In particular, if the constant is the maximum permissible dissipation, labeled P_{dm} , we will get

$$v_{CE} i_C = P_{dm} \quad (9-20)$$

Since P_{dm} is a constant, this is the equation of a hyperbola. It is shown in Fig. 9-12 superimposed on a set of collector characteristics. For a given

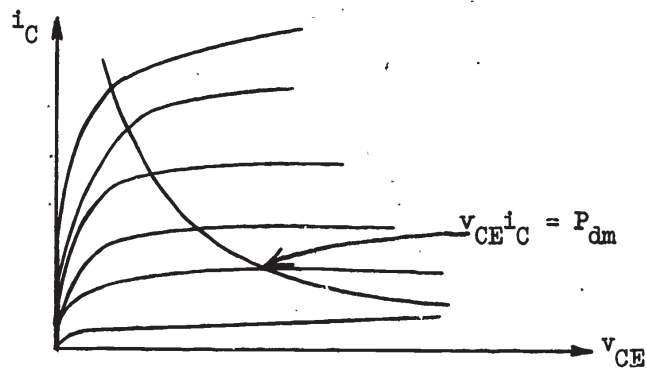


Fig. 9-12

application, the amplifier should be so designed that the load line is tangent to or falls below this maximum dissipation hyperbola.

Models of a Transistor

There are two ways in which one can proceed to obtain models, or equivalent circuits, of a transistor to use in assessing its behavior when it is connected in a network. One approach could be to study the transistor's physical principles and to arrive at equations expressed in terms of its terminal voltages and currents; these equations would contain expressions for such physical parameters as concentrations of holes and electrons in the P-region and N-region, diffusion lengths for holes and electrons, the charge of an electron, etc. These equations could then be related to an equivalent circuit. In this text, we shall not study the pertinent physical principles in such detail as will permit us to carry out such an analysis.

We have, however, discussed some of the elementary principles sufficiently to have a qualitative appreciation of some properties of a transistor. Consider again the diagrams of Fig. 9-1 and Fig. 9-2 and remember that a P-N junction is forward-biased when the P-region is made positive relative to N-region, and is reverse-biased when the opposite is true. With the biases shown in Fig. 9-2, the emitter junction is forward-biased and the collector junction reverse-biased. The electrons which are injected into the base from the emitter diffuse through the base region without much recombination, as already discussed. Because of the reverse-biased collector junction, there is a strong field pulling these electrons from the base into the collector. Thus, the collector current is almost equal to the emitter current.

Let us write

$$i_C = \alpha i_E \quad (9-21)$$

when α is an amplification factor whose value is less than 1, but very close to 1. (A typical value is $\alpha = .98$). Actually, for the amplifier connection we

have been discussing, we are more interested in the amplification relating i_C to i_B . Such a relationship can be obtained by solving Eq. (9-21) for i_E and substituting into Eq. (9-2). The result will be.

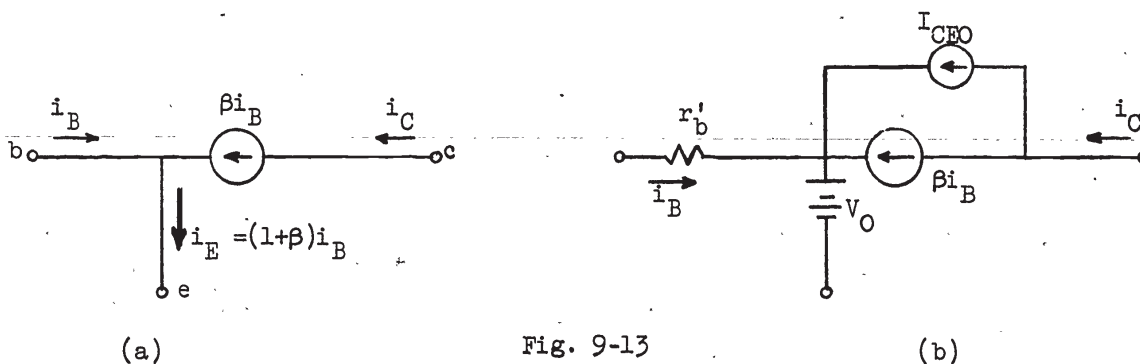
$$i_C = \frac{\alpha}{1-\alpha} i_B = \beta i_B \quad (9-22)$$

where

$$\beta = \frac{\alpha}{1-\alpha} \quad (9-23)$$

is the base-to-collector amplification factor. The closer α is to unity, the greater the value of β . (For $\alpha = .98$, $\beta = 49$).

Indeed, Eq. (9-22) is an analytical expression representing the family of collector characteristic curves. It gives a family of horizontal lines, on the i_C - v_{CE} axes, which are equally spaced for equal changes in i_B . It is thus an idealized characteristic. It is now possible to draw a model representing this behavior of a transistor, as shown in Fig. 9-13a.



The current source whose value is βi_B is a dependent source (also called a controlled source). Such a source is different from the kind of current source we have considered up till now which, by contrast, can be called an independent source. The value of the source current in Fig. 9-13a is dependent on the current somewhere else. Such dependent sources are characteristic of the models for the transistor and other electronic devices.

The circuit of Fig. 9-13a represents only the most important property of the transistor, its current amplification. It does not, however, account for some other effects. Sometimes it is important to take such secondary effects into account. The model in Fig. 9-13b gives a better approximation to the behavior of the transistor. The battery V_0 accounts for the base-to-emitter voltage which is typically less than one volt. (About .2 volt for germanium and .5 volt for silicon transistors.) The resistor r'_b , called the base spreading resistance, accounts for the small voltage from the base terminal of the transistor to the interior of the base region. (A typical value of r'_b is 100 ohms.) And, finally, the current source labeled I_{CEO} is the collector current when the base current is zero. From Fig. 9-13b, the collector current has the value

$$i_C = \beta i_B + I_{CEO} \quad (9-24)$$

The effect of I_{CEO} is to cause the family of collector characteristics to start with a non-zero value of i_C when $i_B = 0$. I_{CEO} is quite sensitive to variations in temperature. It is typically zero at 25°C but becomes 10 ma. at 175°C (for a transistor whose maximum collector current is 30 ma.).

Small-signal Models

Up to this point, the transistor currents and voltages that have been under discussion have been the total instantaneous values. These total values have contributions from two sources, (1) from the bias sources, which serve to fix the quiescent operating point, and (2) from signals applied at the input. Once the quiescent operating point of the device has been fixed, we are subsequently interested in the variations of output that take place as a consequence of the signal. It would, therefore, be convenient to deal only with these incremental changes from the average values; for example, with i_{ce} and v_{ce} rather than i_{CE} and v_{CE} .

The model of a transistor which we shall now develop is obtained by an approach based on empirically obtained measurements of the terminal characteristics of a transistor, rather than from a study of the physical principles of semiconductor junctions.

To start the discussion, note that of the four terminal variables of the transistor (v_{BE} , i_B , v_{CE} , i_C), two can be expressed as functions of the other two. Look at the base and collector characteristics in Fig. 9-5. If we take i_B and v_{CE} as independent variables, then we can write

$$v_{BE} = f_1(i_B, v_{CE}) \quad (9-25)$$

$$i_C = f_2(i_B, v_{CE}) \quad (9-26)$$

where f_1 and f_2 are two functions of i_B and v_{CE} which are represented graphically by the curves in Fig. 9-5. Now suppose that both i_B and v_{CE} take on increments from their operating values. Then both v_{BE} and i_C will take on increments from their respective operating values. To a first approximation, the increments in the dependent variables will be linearly related to those of the independent variables.

Thus, we can write

$$\Delta v_{BE} = a_1 \Delta i_B + a_2 \Delta v_{CE} \quad (a)$$

$$\Delta i_C = a_3 \Delta i_B + a_4 \Delta v_{CE} \quad (b) \quad (9-27)^*$$

* These can be obtained more rigorously by expanding the functions in Eqs. (9-25) and (9-26) in Taylor series about the operating point. Thus

$$v_{BE} = v_{BE} \Big|_{\text{at operating point}} + \frac{\partial v_{BE}}{\partial i_B} di_B + \frac{\partial v_{BE}}{\partial v_{CE}} dv_{CE} + \left\{ \text{terms containing higher derivatives} \right\}$$

The partial derivatives are to be evaluated at the operating point. If higher order terms are neglected, and we note that $v_{BE} - v_{BE} \Big|_{\text{at operating point}}$ is simply the increment in v_{BE} , the result follows.

The coefficients a_1 to a_4 are simply constants. Interpretations for these constants can be obtained by solving for them from the equations themselves. Thus,

$$a_1 = \frac{\Delta v_{BE}}{\Delta i_B} \quad \text{when } \Delta v_{CE} = 0 \quad (a)$$

$$a_2 = \frac{\Delta v_{BE}}{\Delta v_{CE}} \quad \text{when } \Delta i_B = 0 \quad (b)$$

(9-28)

$$a_3 = \frac{\Delta i_C}{\Delta i_B} \quad \text{when } \Delta v_{CE} = 0 \quad (c)$$

$$a_4 = \frac{\Delta i_C}{\Delta v_{CE}} \quad \text{when } \Delta i_B = 0 \quad (d)$$

Further interpretations are obtained graphically from the transistor characteristic curves. Consider Eq. (9-28d) for a_4 . Holding the increment of i_B at zero means the value of i_B stays constant. Thus, let Fig. 9-14a be a collector characteristic curve for a fixed value of i_B and let the coordinates of Q be the operating values of i_C and v_{CE} . We are to hold i_B constant at this value while i_C and v_{CE} change. The ratio of Δi_C to Δv_{CE} is simply the slope of the curve, provided the curve is approximately linear.

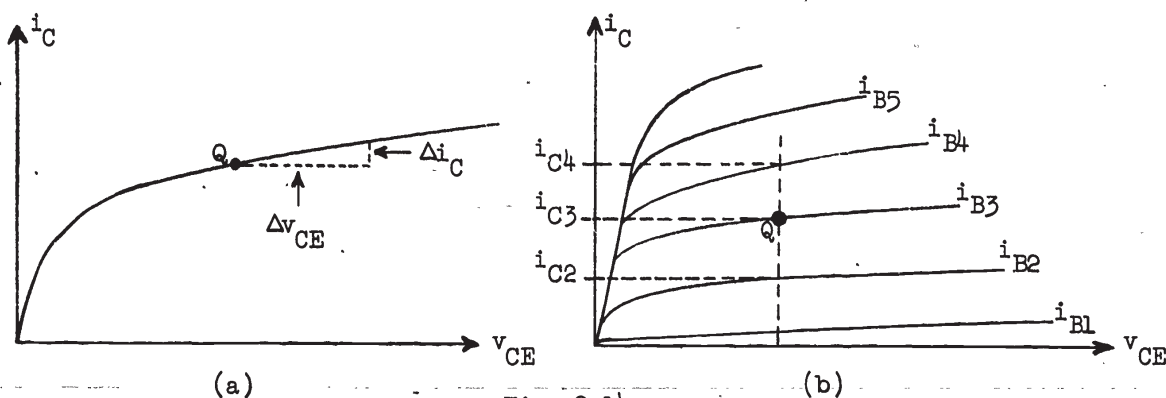


Fig. 9-14

But even if there is some curvature in the curve, this ratio will approximately equal the tangent to the curve at the point Q, provided the increments Δi_C and Δv_{CE} are "small".

Now look at the interpretation of a_3 in Eq. (9-28c). Let Q in Fig. 9-14b be the operating point. $\Delta v_{CE} = 0$ means we are to hold the value of v_{CE} constant. Thus, however i_B and i_C change, the new operating point must lie on the vertical line passing through Q. Thus, if i_B changes from i_{B3} to i_{B4} , then i_C will change from i_{C3} to i_{C4} ; so, a_3 will be the ratio of $i_{C4} - i_{C3}$ to $i_{B4} - i_{B3}$. Similar interpretations can be given for a_1 and a_2 .

Glance back at Eqs. (9-27 or (9-28). The a's do not all have the same dimensions. Thus, a_3 is dimensionless and a_4 is a conductance. It is therefore more appropriate to give them more suggestive symbols. Earlier, we had introduced a notational scheme for expression varying components of the transistor currents and voltage. We will use this scheme and rewrite Eqs. (9-27) as follows.

$$v_{be} = r_{11} i_b + \mu v_{ce} \quad (a) \quad (9-29)$$

$$i_c = \beta i_b + g_{22} v_{ce}$$

Remember that the voltage and current symbols here stand for changes, variations or increments in the instantaneous values. Thus, i_b is not the base current but the incremental base current.

We now have a set of equations; the next task is to draw a model which these equations imply. The first equation is suggestive of Kirchhoff's voltage law; the voltage v_{be} equals the sum of two voltages. One of these two is the drop across a resistor r_{11} ; the other appears to be a dependent source. A similar analysis can be made of the second equation. Remembering that $i_e = i_c + i_b$, the result is the network shown in Fig. 9-15.

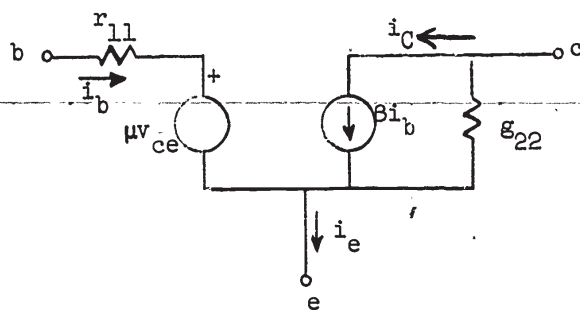


Fig. 9-15

This network is a model of a transistor. Remember the conditions under which it was derived. It does not give relationships among the total currents and voltages but only among changes in these currents and voltages around their operating values. The model includes two dependent sources; one a current source, the other a voltage source. Note that if the collector terminal is shorted to the emitter, then i_c equals βi_b and so β is the short circuit incremental current gain. It has the same value as the previously introduced β . Note also that no bias batteries appear in this incremental model.

On comparing this small-signal model with the model shown in Fig. 9-13, one can note that if r_{11} , μ , and g_{22} are all zero, the small-signal model reduces to the model in Fig. 9-13a. Thus, the basic nature of a transistor is a current amplifier; the other parameters simply measure the extent to which the transistor deviates from this ideal behavior. Because the parameters of this model are dimensionally so heterogeneous, it is called the hybrid model and the four parameters are called the hybrid parameters.

This model of a transistor can be used in the calculation of the gain of an amplifier. Figure 9-16 shows a typical transistor amplifier. It is desired to determine the small-signal voltage gain v_{Ls}/v_s , where v_s is the signal voltage and

v_{Ls} is the component of the voltage across R_L due to the signal. It is also desired to estimate the error in i_b that would result if μ were neglected. The parameters of the model are $r_{11} = 1$ kilohm, $\beta = 60$, $\mu = 4 \times 10^{-4}$, $g_{22} = 25$ micro-mhos; also $R_L = 4$ kilohms.

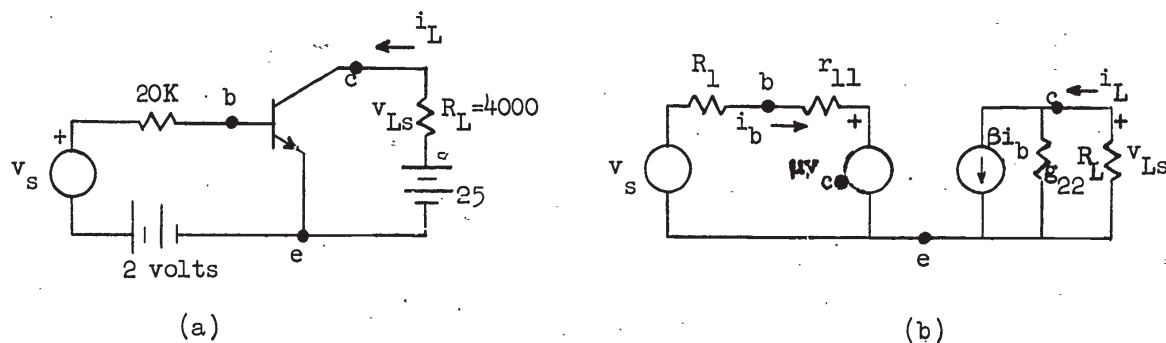


Fig. 9-16

The first step is to draw the small-signal equivalent circuit. This is done by replacing the transistor with its small-signal model, removing all biasing sources and leaving everything else as is. The result is the diagram of Fig. 9-16. Let R be the parallel combination of the two resistors R_L and $1/g_{22}$. That is,

$$R = \frac{R_L}{1 + g_{22}R_L} = \frac{4000}{1 + 4000 \times 25 \times 10^{-6}} = \frac{4000}{1.1} = 3.64 \text{ kilohms.}$$

The incremental load voltage is

$$v_{Ls} = v_{ce} = -\beta i_b R = -.218 \times 10^6 i_b$$

Now this can be substituted for the dependent voltage source in the base circuit.

The voltage of this source becomes $-\mu\beta R i_b$. This is the same voltage that would be obtained if a current i_b were flowing through a negative resistance $-\mu\beta R$. Thus, the base circuit can be redrawn as shown in Fig. 9-17. The negative resistance $-\mu\beta R$ subtracts from $R_1 + r_{11}$.

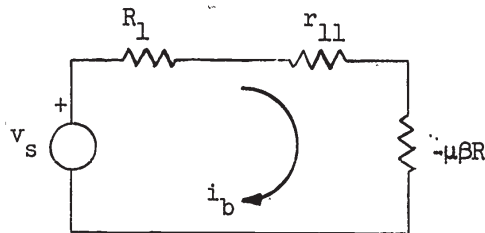


Fig. 9-17

The incremental base current can now be found to be

$$i_b = \frac{v_s}{R_1 + r_{11} - \mu\beta R} = \frac{v_s}{21000 - 87.2} = \frac{v_s}{21000} \frac{1}{1 - \frac{87.2}{21000}}$$

$$= \frac{v_s}{21000} \frac{1}{1 - .00415} = 1.004 \frac{v_s}{21000} \text{ amp.}$$

Note that, if μ were neglected and assumed to be zero, i_b would be simply $v_s/21000$. Thus, the error in neglecting μ would be only about .4 per cent, which is certainly tolerable. (In fact, actual resistors used for R_1 and R_L would have tolerances of 5 to 20 per cent, so that the additional .4 per cent error would be completely negligible.)

To complete the problem, the value of i_b just found is substituted into the previously found expression for load voltage to yield, for the voltage gain,

$$\frac{v_{Ls}}{v_s} = \frac{-\beta R}{R_1 + r_{11} - \mu\beta R} \doteq \frac{-\beta R}{R_1 + r_{11}} = -10.4 \quad (9-30)$$

Note that the gain is a negative number. Thus, if the signal is sinusoidal, the output signal voltage will have a phase angle of 180° relative to the input signal, and its amplitude will be 10.4 times larger.

In this example, all the transistor and circuit parameter values were given and the gain was calculated. However, the expression for the gain can be examined

to see what changes might be made to increase the gain. Of the four quantities on the right side of Eq. (9-30), β and r_{11} are fixed by the transistor. (Their values also depend on the location of the quiescent operating point; thus β depends on the separation of the curves in the family of collector characteristic curves and a glance at Fig. 9-5 shows that this separation depends on where the quiescent point is located. However, if we assume that the quiescent point remains somewhere near the center of the quadrant, the variations of the transistor parameters due to changes in the operating point -- which are determined by the external circuit parameters -- will be secondary. To a first approximation we can claim that β and r_{11} depend only on the transistor.) R_L is an external resistor and R is a combination of an external resistor and the transistor parameter g_{22} . However, for the values in this example, the difference between R and R_L is less than ten per cent. Thus, R can be looked upon as an external parameter.

For a given transistor, then, in order to increase the amplifier gain we should increase R (which means increase R_L) or decrease R_L , or both. But note that R_L and R_L help to establish the load line and fix the quiescent point, so we are not at liberty to change these values at will. A compromise must be reached between desirable quiescent point and large gain.

One additional variable available for achieving desired objectives is the transistor itself. It would be desirable to have transistors with large values of β and small values of r_{11} . The engineer wanting to optimize a circuit must have knowledge of the characteristics of a wide range of transistors; this is a by-product of experience. With the introduction provided here, you can now interpret many transistor characteristic curves, you can design a variety of simple amplifiers, and you can begin to acquire this valuable engineering experience.

Practical Considerations in Amplifier Operation

In the preceding work, consideration was given only to the most basic amplifier configuration. This included two bias sources, one for the collector and one for the base, and two resistors, R_L in the collector circuit and R_1 in the base circuit. Collectively, these quantities served to establish the load line and fix the operating point on the load line.

One hallmark of an engineer is to see if he can accomplish a given task more simply, with fewer components. One candidate for such simplification is the use of two batteries for biasing purposes; can only one battery do the job? The question is answered in Fig. 9-18. The diagram in (a) shows an arrangement which uses only

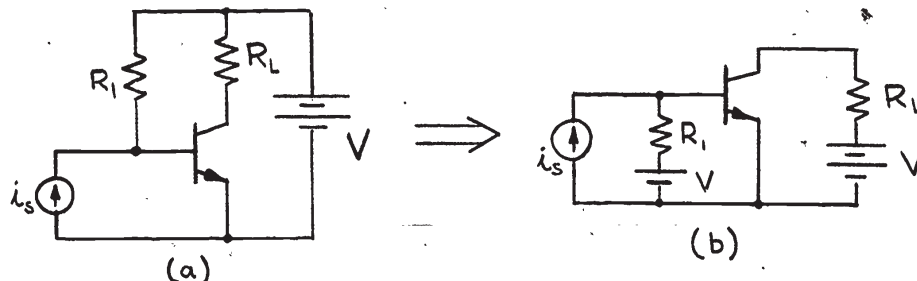


Fig. 9-18

one battery but which is equivalent to the arrangement in (b). (If you don't see this quickly, look back at the discussion of the shifting of sources in Sec. 2-6.) The only difficulty here is that we are constrained to have the same value of biasing source in both collector and base circuits. We need to have some way of adjusting the value of the base biasing battery voltage independently from that of the collector biasing battery.

This objective can be achieved by the arrangement shown in Fig. 9-19a. An extra resistor has been added in the base circuit forming a voltage-divider arrangement. If the structure to the left of the dashed vertical line is replaced by a

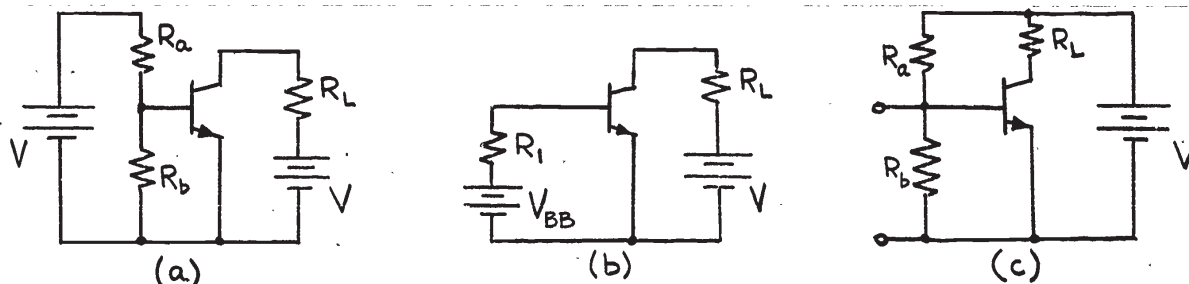


Fig. 9-19

Thevenin equivalent the result is as shown in Fig. 9-19b, where

$$V_{BB} = \frac{R_b}{R_a + R_b} V \quad (a)$$

(9-31)

$$R_1 = \frac{R_a R_b}{R_a + R_b} \quad (b)$$

The final amplifier diagram has the arrangement shown in Fig. 9-19c. Note that the original two batteries and two resistors have been replaced by one battery and three resistors.

In a given problem if the component values are given and it is required to determine the quiescent operating point and the load line, Eqs. (9-31) are used to fix the values in the configuration of Fig. 9-19b. The rest of the procedure is then the same as previously described. On the other hand, if the problem is one of design, and the desired operating characteristics are given -- such as largest signal amplitude with least distortion, which approximately fixes the Q point and the load line -- the problem being to determine appropriate component values, one has to work backwards. Thus, for Fig. 9-19b, the base current can be found, using Eqs. (9-31), to be

$$i_B = \frac{V_{BB}}{R_1} = \frac{V}{R_a}$$

Thus, with V determined from the load line and i_B fixed by the quiescent point, R_a becomes known. There still remains to determine R_b from Eqs. (9-31). Since neither V_{BB} nor R_1 in Fig. 9-19b, nor R_b in Fig. 9-19c, have been fixed by anything specified to this point, only practical considerations and convenience will limit you in selecting their values. For example, you would not want R_b to be either very large (such as several million ohms) or very small (such as only a few ohms) as resistors in these sizes are expensive and fragile. Thus, if you are designing a circuit like Fig. 9-19c, you are free to choose any convenient value of R_b . However, you should be forewarned that other modifications in the amplifier circuit will be introduced, below. They will put further limits on you in selecting your circuit parameters.

At this point, let us use the model of Fig. 9-13 in the amplifier circuit of Fig. 9-19b to find an analytical expression for the dc collector current in order to investigate the factors on which it depends. The resulting network, which applies for the average values I_B and I_C (remember the notation), is shown in Fig. 9-20. By writing a voltage equation around the left-hand loop, I_B can be found. Then,

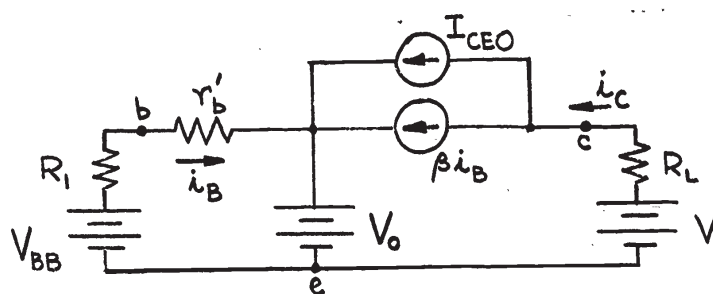


Fig. 9-20

a node equation at the upper right node gives an expression for I_C . Thus,

$$I_B(R_1 + r'_b) = V_{bb} - V_0$$

$$I_C = \beta I_B + I_{CEO} \quad (9-32)$$

$$I_C = \frac{\beta(V_{BB} - V_0)}{R_1 + r'_b} + I_{CEO}$$

In the last expression, the quantities that are transistor parameters are β , V_0 , I_{CEO} and r_b' . It has already been pointed out that I_{CEO} is temperature dependent. The same is true of V_0 . Thus, I_C and the operating point will change with temperature. Furthermore, the value of β for a given type of transistor varies from one individual transistor to another, simply because of manufacturing problems, and this variation can be by a factor of as much as 3 or 4. Since β is a multiplying factor in one term in Eq. (9-32), I_C will change radically with such changes in β . A modification in the amplifier circuit is needed that will tend to compensate for changes in transistor parameters such as β .

A review of the steps in arriving at Eq. (9-32) shows that, although there are two loops in Fig. 9-20, the base current is found from the left-hand loop independently of the collector current. Thus, any changes in the collector current due to changes in β , say, are not reflected back to influence the base current -- we say there is no "coupling" between the two loops. This would not be the case if there were a resistor in the branch between two loops; that is, if a resistor were connected to the emitter in Fig. 9-19. Such a resistor is added in the amplifier circuit and its model shown in Fig. 9-21. This resistor tends to compensate for changes in β .

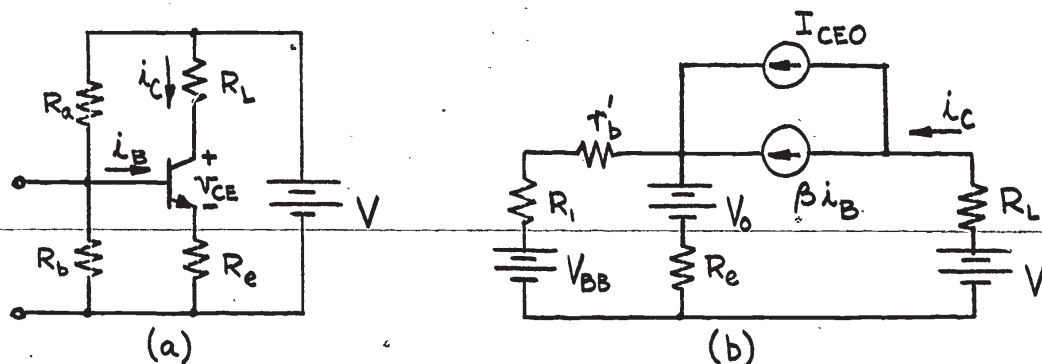


Fig. 9-21

With this change, the equations written earlier now become

$$(r'_b + R_1 + R_e)I_B + R_e I_C = V_{BB} - V_O \quad (a)$$

(9-33)

$$I_C = \beta I_B + I_{CEO} \quad (b)$$

When these are solved for I_B and I_C , the result will be

$$I_B = \frac{V_{BB} - V_O - R_e I_{CEO}}{r'_b + R_1 + (\beta + 1)R_e} \quad (9-34)$$

$$I_C = \frac{\beta(V_{BB} - V_O)}{r'_b + R_1 + (\beta + 1)R_e} + \frac{r'_b + R_1 + R_e}{r'_b + R_1 + (\beta + 1)R_e} I_{CEO} \quad (9-35)$$

(Verify that these reduce to the previous expressions when $R_e = 0$.) Consider the first of these for I_B . Suppose there is a positive increment in either I_{CEO} or β . An increase in β increases the denominator and thus causes a decrease of I_B . Similarly, an increase in I_{CEO} causes a decrease of the numerator and again a decrease of I_B . Thus, an increase in either β or I_{CEO} , or both, tends to reduce I_B . Considering Eq. (9-35b), this decrease in I_B tends to compensate for the increase in β or I_{CEO} , thus tending to keep I_C constant. The operation of the amplifier will thus be greatly improved by the addition of R_e .

Now concentrate on the last equation for I_C (Eq. 9-35). This appears to be a complicated expression, although it is merely the sum of two terms. It is to be compared with Eq. (9-32) for the case of no R_e . In that expression the coefficient of I_{CEO} is 1. Let us examine what value of R_e would be required to make this coefficient in Eq. (9-35) considerably less than 1, for typical values of the other quantities.

Let us call the coefficient of I_{CEO} , A , and solve for R_e . The result will be

$$\frac{r'_b + R_L + R_e}{r'_b + R_L + (\beta+1)R_e} = A$$

or

$$R_e = \frac{(1-A)(r'_b + R_L)}{A(\beta+1) - 1} \quad (9-36)$$

Typical values of β , r'_b and R_L are 50, 100 ohms, and 5-10 kilohms, respectively.

Thus, r'_b is quite negligible compared to R_L . If we wish the coefficient A to be 0.1, then

$$R_e = 1.1 \text{ to } 2.2 \text{ kilohms for } R_L = 5 \text{ to } 10 \text{ kilohms.}$$

These are quite reasonable values of R_e . Thus, the coefficient of I_{CEO} in Eq. (9-35) can be easily made of the order of .1. This means, the contribution of I_{CEO} to I_C can be made small enough to neglect. Thus, Eq. (9-35) reduces to

$$I_C = \frac{\beta(V_{BB} - V_0)}{r'_b + R_L + (\beta+1)R_e} \quad (9-37)$$

This equation is quite useful in determining the collector current at the quiescent point. Of course, this point must also fall on the load line. The equation for the load line, however, will now be somewhat more complicated due to the appearance of the emitter resistor R_e . Thus, from the collector circuit in Fig. 9-21a, the following equation results.

$$V_{CE} = V - (R_L + R_e)i_C - R_e i_B \quad (9-38)$$

Not only does R_e contribute to the coefficient of i_C , it introduces i_B into this equation.

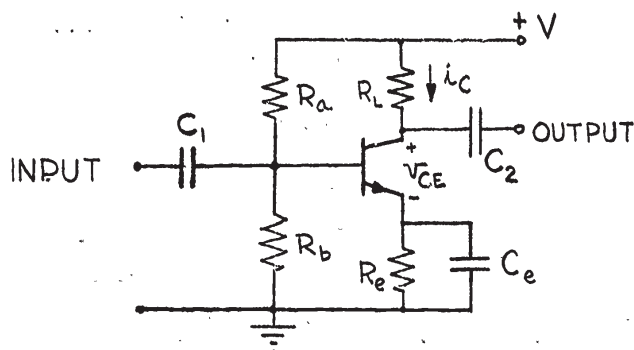
Of course, i_B could be eliminated by writing another equation for the base circuit. However, it would not cause a great error if i_B is neglected completely, since it is quite small compared with i_C , assuming a large enough β . If i_B is neglected, Eq. (9-38) can be solved for i_C to yield

$$i_C = \frac{V}{R_L + R_e} - \frac{1}{R_L + R_e} V_{CE} = \frac{V - V_{CE}}{R_L + R_e} \quad (9-39)$$

Thus, the slope of the load line is determined by the sum of R_L and R_e , rather than just R_L as in the simpler circuit.

The Dynamic Load Line

The introduction of an emitter resistance has been shown to stabilize the value of collector current against changes in the transistor parameters, β and I_{CEO} . The same mechanism that brings about this improvement also brings about a reduction in the amplification of signals. But, we can modify the circuit further, to prevent this reduction. As far as the signal components of collector current are concerned, we would like to remove R_e , yet we would like to keep it there as far as the dc component is concerned. This objective can be achieved by placing a capacitor in parallel with R_e . Such an arrangement is shown in Fig. 9-22. R_e is said to be by-passed and C_e is a by-pass capacitor.



(The purpose of C_1 and C_2 will be discussed in the next few paragraphs.)

Fig. 9-22

Before considering the influence of the capacitors, note that the battery does not appear explicitly. Instead, the terminal to which the positive terminal of the battery should be connected is labeled + V; it is understood that the other terminal of the battery is connected to ground, which is designated by the symbol $\perp$ at the bottom of the diagram.

Now the average value of the current in a capacitor must be zero. Another way of saying this is that a capacitor cannot carry direct current. Hence, for calculation of quiescent values, capacitors behave like open circuits and can simply be removed from the circuit and omitted from any calculations. This means, the presence of C_e in parallel with R_e will not influence the value of I_C , the average collector current, obtained in the last section.

Now let us consider the effect of the parallel combination of R_e and C_e on the signal components of current. Let the impedance of this combination be $\bar{Z}_e$ and assume the signal to be sinusoidal of angular frequency ω . Then

$$\bar{Z}_e = \frac{R_e \frac{1}{j\omega C_e}}{R_e + \frac{1}{j\omega C_e}} = \frac{R_e}{1 + j\omega C_e R_e} \quad (9-40)$$

If $\omega C_e R_e$ is much greater than 1, this expression becomes approximately $1/j\omega C_e$ and the effect of R_e has been eliminated. For example, let $R_e = 500$ ohms and let the lowest frequency signal to be amplified have $\omega = 5000$ rad/sec. What value of C_e will cause $\omega C_e R_e$ to equal, say, 25 and what will be the value of $\bar{Z}_e$ for this value of C_e ? The answer will be

$$C_e = 25/\omega R_e = 10 \text{ mfd}; \text{ and } \bar{Z}_e = -j20$$

Now, 20 ohms is practically a short circuit when compared with the value of $R_e = 500$ ohms. At higher frequencies this same value of capacitance will make the value of $\bar{Z}_e$ even smaller and more like a short circuit. For the signal components, then, the capacitor can be taken as a short circuit, thus eliminating R_e and restoring the signal amplification that we had before adding R_e . Note that a by-pass capacitor must have a large capacitance, such as 10 mfd.

As for the capacitors labeled C_1 and C_2 at the input and output of the

amplifier in Fig. 9-22, they also are effectively an open circuit for dc calculations. At the same time, if their values are made large enough, each will be effectively a short circuit at the signal frequencies. Thus, they have two functions: (1) to couple the time-varying part of the signal into the amplifier and out of it, and (2) to block out any average component it might have. (Why would it be undesirable for a signal with a nonzero average value to be applied to the amplifier?) Thus, these capacitors are called either of two descriptive names, coupling capacitors and also blocking capacitors.

The discussion concerning the effect of the by-pass capacitor has shown that C_e acts as a short circuit to the signal component of emitter current. Thus, the voltage across R_e is caused only by the average value of the emitter current, and therefore remains constant at the value $R_e I_C$. Thus, writing a voltage equation around the collector circuit in Fig. 9-22 and solving for i_C leads to

$$i_C R_L + v_{CE} + I_C R_e - V = 0$$

$$i_C = \left(\frac{V - R_e I_C}{R_L} \right) - \frac{1}{R_L} v_{CE} \quad (9-41)$$

This is the equation of a straight line. It should be compared with the equation of the load line for the circuit without C_e , given in Eq. (9-29). Its slope is different, as are its intercepts. However, it passes through the same quiescent operating point, as can be confirmed by setting i_C in Eq. (9-41) equal to its quiescent value I_C , and noting that the resulting values of v_{CE} in the two equations are the same. This new line is called the dynamic load line in contrast with the previous one which is the static load line. Figure 9-23 shows both lines on a set of collector characteristic curves.

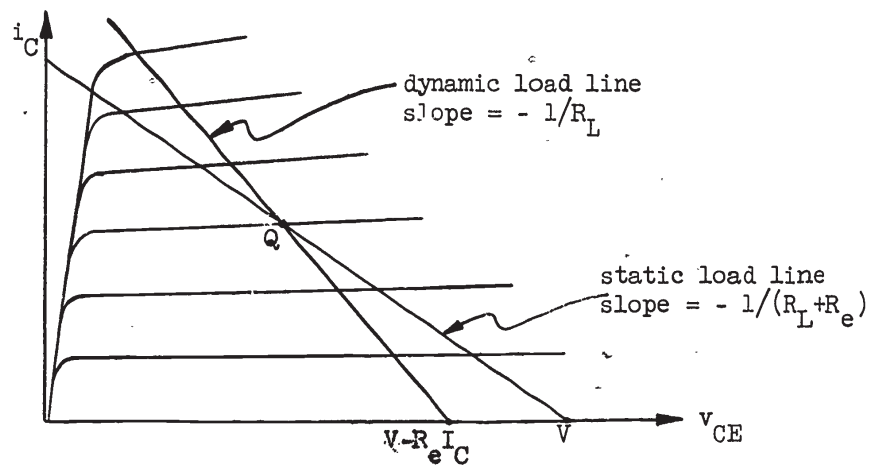


Fig. 9-23

Example

To illustrate the above discussion, consider the amplifier shown in Fig. 9-24.

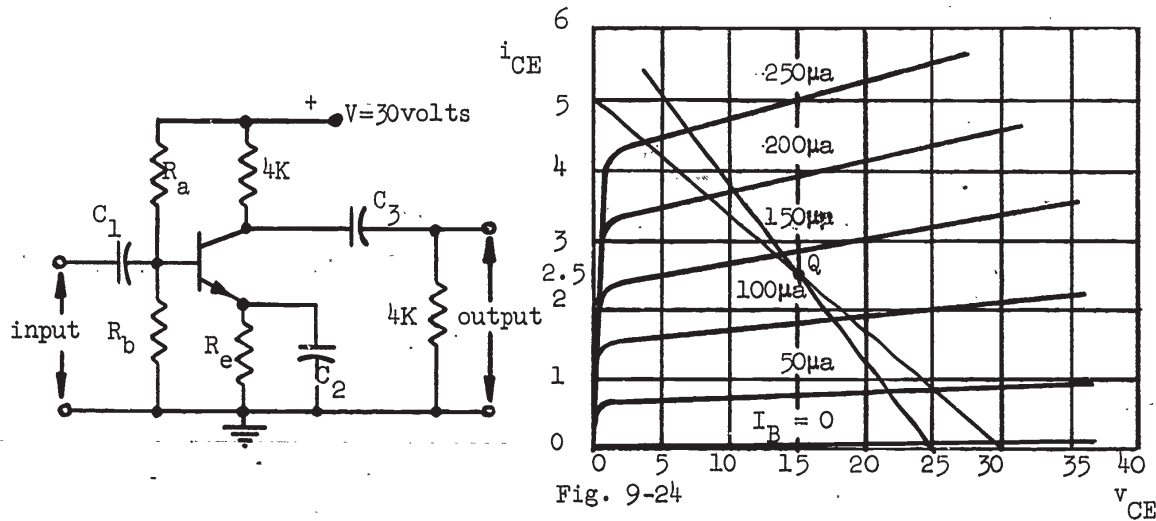


Fig. 9-24

The transistor has a $V_0 = .2$ volts. It is desired to have the quiescent operating point fall at $i_C = 2.5$ mA., $v_{CE} = 15$ volts. Draw the static and dynamic load lines and determine appropriate values for those parameters which are unspecified in the diagram. Use the collector characteristics shown in the figure.

The value of β in the vicinity of the Q point is found from the collector curves to be $\beta \approx 20$. The static load line is easily constructed since it must pass through the v_{CE} axis at 30 volts and through the Q point. It intersects the i_C axis at 5 ma. Hence, the slope will be

$$\text{slope} = - \frac{5 \text{ ma}}{30 \text{ volt}} = - \frac{1}{R_e + R_L}$$

Hence,

$$R_e + R_L = \frac{30}{5} = 6 \text{ kilohms}$$

Since $R_L = 4K$, then $R_e = 2K$.

The dynamic load line can now be drawn. Perhaps the simplest way is to note that it must pass through the Q point and through the point $V - R_e I_C = 30 - 2(2.5) = 25$ volts on the v_{CE} axis.

It remains to find R_a and R_b . Since the quiescent collector current is known, Eq. (9-37) gives a relationship between V_{BB} and R_L . Thus,

$$R_L + 42 = \frac{20}{2.5} (V_{BB} - 2)$$

or

$$R_L = 8V_{BB} - 43.6 \text{ kilohms.}$$

When Eqs. (9-31) are used for R_L and V_{BB} , the result will become a relationship between R_a and R_b . Thus,

$$\frac{R_b(240 - R_a)}{R_a + R_b} = 43.6$$

From this expression it is clear that R_a must be less than 240K. There are no other restrictions on R_a and R_b . Hence, let us arbitrarily choose $R_a = 150K$. Then, from the last expression we find $R_b = 141K$.

Thus we have "designed" a transistor amplifier with predetermined characteristics. It would still be necessary to build this design and try it out in the laboratory. Some of the values may need to be adjusted slightly in order to get just the desired results, but your analysis has done two things: (1) you have saved a lot of time and money in getting a design that should work, and (2) you will understand the effect of making small adjustments in the values of the different components, to get just the performance you want.

Let us continue with this example and now concentrate on finding the small signal gain of the amplifier assuming a signal current i_1 (milliamps) is applied. For this purpose, remember that the capacitors C_1 , C_2 and C_3 are assumed to be short circuits and biasing batteries are removed (assumed shorted). Thus, the resulting network is shown in Fig. 9-25a. Note that the 4K output resistor and R_E are now

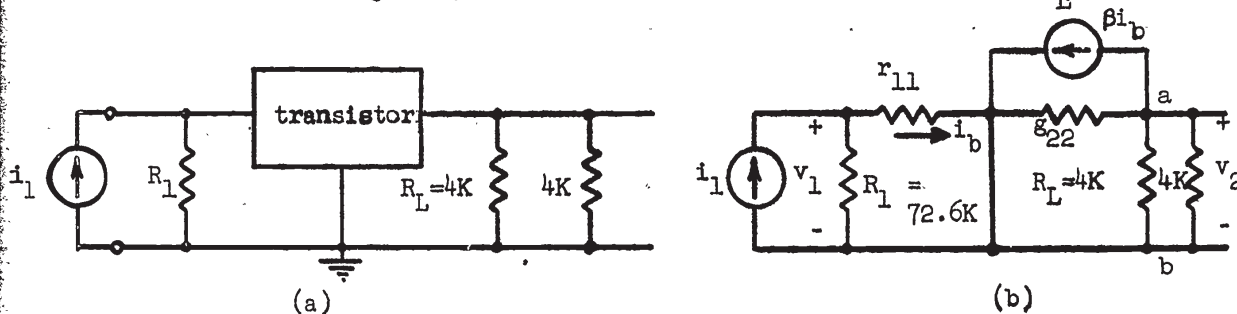


Fig. 9-25

in parallel since C_3 is a short circuit, and R_E has been shorted by C_2 . (R_a and R_b are also in parallel to yield R_1 .) The transistor is now replaced by its small signal model from Fig. 9-15 to yield the result shown in Fig. 9-25b. It has here been assumed that μ is small enough to neglect. β has already been found to be $\beta \approx 20$. From Fig. 9-24, g_{22} is found to be $g_{22} \approx .03$ mmho. Although there is no data from which r_{11} can be determined, it is clear that if it has a typical value in the neighborhood of 1K, it will be overshadowed by R_1 . Let us assume $r_{11} = .4K$ in the interest of convenience and convert i_1 in parallel with R_1 to a voltage

source equivalent and similarly with βi_b in parallel with g_{22} . (Caution: g_{22} is a conductance.) The result is the network shown in Fig. 9-26.

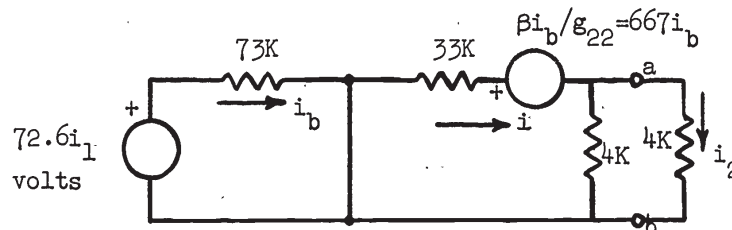


Fig. 9-26

This is relatively easy to solve because each of the two loops, considering the two 4K resistors to be combined, can be handled separately.

$$i_b = \frac{72.6}{73} i_1 \doteq i_1 \text{ ma.}$$

$$i = \frac{-667 i_b}{33 + 2} = -19 i_1$$

Finally, the output current i_2 can be found from the current divider relationship to be

$$i_2 = \frac{4}{4 + 4} i = -9.5 i_1 \text{ ma.}$$

giving a current gain of

$$\frac{i_2}{i_1} = -9.5$$

In addition to the gain, there are other quantities of interest in the small-signal operation of transistor amplifiers. One of these is the impedance at the input terminals. You will recall that the impedance of a network at a pair of terminals is the ratio of the phasor voltage at those terminals to the phasor current. This assumes that the signals applied to the amplifier are sinusoids. If the network in question has no reactive elements (capacitors or inductors), the angle of all sinusoids (currents or voltages) will be the same; there will be no phase

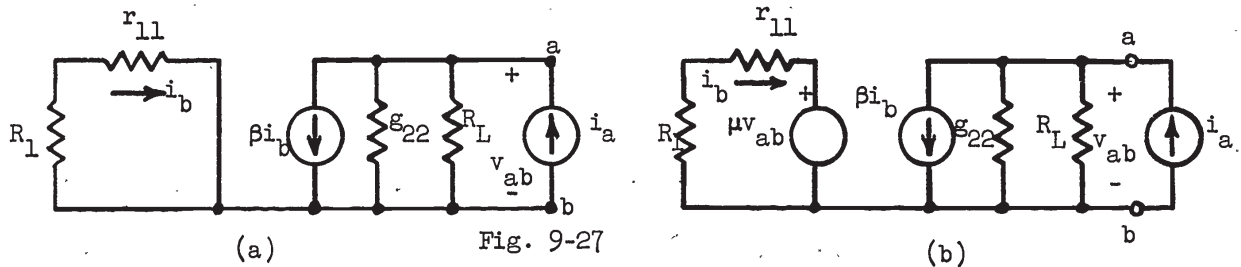
difference, only the amplitudes will be different for different variables. In such a case, all phasors can be made real. Hence, it will be permissible to deal with the sinusoids themselves instead of their phasors. The input impedance is purely real and has no imaginary component.

Thus, for the amplifier network in Fig. 9-25b, the input impedance at the terminals of the amplifier will be v_1/i_1 . From the figure it is seen that r_{11} and R_1 are in parallel and their voltage and current are not influenced by the collector circuit, because μ was assumed negligible. Thus, the input impedance is the resistance which is the equivalent of r_{11} and R_1 in parallel. This will be approximately equal to r_{11} since R_1 is so much larger than r_{11} . But r_{11} is quite small (we took it to be 400 ohms). This is typical of transistor amplifiers in this configuration.

Another quantity of interest is related to the Thevenin equivalent of the amplifier at its output terminals. Consider forming the Thevenin equivalent looking to the left from terminals a-b in either Fig. 9-25b or Fig. 9-26. Interest will be focussed on the Thevenin equivalent impedance. You will recall that this is the impedance at the terminals when the sources in the network are deactivated. Since we now have a type of source -- a dependent source -- which was not introduced when the Thevenin theorem was discussed, the question arises as to whether or not dependent sources are also deactivated when the Thevenin impedance is calculated. Note that in Fig. 9-25b the dependent source current is proportional to i_b . If this source is deactivated, βi_b , and hence also i_b , must be set equal to zero. But i_b is the current in r_{11} ; it may or may not be zero. If i_b should turn out to be zero, then the dependent source current will be zero; otherwise not.

In the present case, let us deactivate the signal current i_1 (make it zero) since it is an independent source and then apply a current source i_a at the terminals a-b, our objective being to find v_{ab} so that the ratio v_{ab}/i_a can be calculated.

The situation is illustrated in Fig. 9-27a, where the network has been redrawn in



order to show more clearly that there is no coupling back from the collector circuit to the base circuit. Thus, even though a source i_a is applied, there will be no base current and hence βi_b will be zero. Hence, the Thévenin impedance will be the parallel combination of g_{22} and R_L . It is common terminology to call this Thévenin equivalent impedance, the output impedance. In this case, the output impedance will be $R_L / (1 + g_{22} R_L) = 3.57K$.

Suppose that the parameter μ in the transistor model is not neglected. Then, Fig. 9-27a becomes modified as in Fig. 9-27b. Now there is coupling from the collector circuit to the base circuit. Hence, i_b will not be zero; it will be

$$i_b = - \frac{\mu v_{ab}}{R_1 + r_{11}}$$

From the collector circuit we find v_{ab} to be

$$v_{ab} = \frac{R_L}{1 + g_{22} R_L} (i_a - \beta i_b) = \frac{R_L}{1 + g_{22} R_L} \left(i_a + \frac{\mu \beta v_{ab}}{R_1 + r_{11}} \right)$$

The last step is obtained by substituting i_b from the preceding equation. Solving for v_{ab} leads to

$$\frac{v_{ab}}{i_a} = \frac{R_L}{1 + (g_{22} - \mu \beta) R_L} \quad (9-42)$$

which reduces to the preceding expression for the output impedance when $\mu = 0$.

Although this example does not prove the general case, it is nevertheless a general result that when finding the Thévenin impedance of a network containing

dependent sources (as well as independent ones), only the independent sources are deactivated -- not the dependent sources. We shall not prove this result here.

The Emitter Follower

The amplifier configuration we have been discussing is called a common emitter connection, because the emitter terminal is common between input and output. We have seen that this amplifier has a relatively low input impedance (several hundred ohms) and a relatively high output impedance (several kilohms). In many applications just the opposite is required: relatively high input impedance (several tens of thousands of ohms) and a relatively low output impedance (in the hundreds or less). We see from Eq. (9-42) that the output impedance depends strongly on R_L ; it will become zero if $R_L = 0$. But in this case there will be no output voltage either, so nothing is gained by setting $R_L = 0$.

As another point of departure, note from Fig. 9-25b that the input impedance will increase if the resistance R_e is not bypassed so that it appears in the small signal model. These thoughts lead to consideration of the amplifier configuration shown in Fig. 9-28a.

The two modifications here are that R_L has been set equal to zero and R_e is not bypassed with a capacitor. Since there is no load resistance, the output must be taken somewhere else; it is taken across R_e . The resulting small-signal equivalent circuit is shown in Fig. 9-28b. This model is redrawn in a more convenient form in Fig. 9-28c, to bring out more clearly the output terminals. Note that R_e and g_{22} are in parallel; let us denote by R the parallel combination:

$$R = \frac{R_e}{1 + g_{22}R_e} \quad (9-43)$$

There are three quantities of interest in the operation of the emitter-follower: the current gain, the input impedance and the output impedance. To find the gain and the input impedance, suppose a current source i_s is applied, leading to Fig. 9-29. If v_2 is found, then the current through R_e will be simply

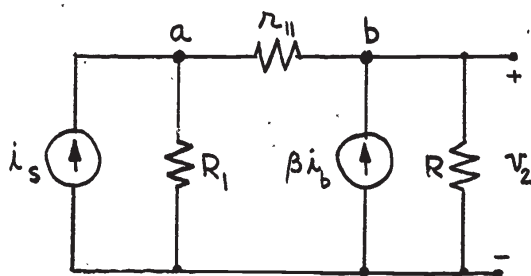


Fig. 9-29

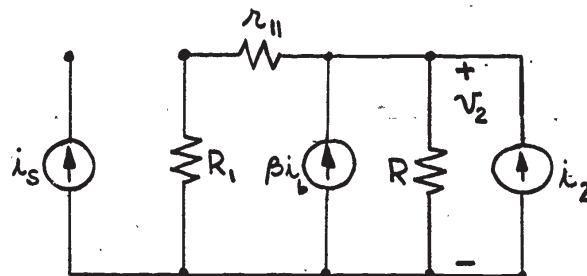


Fig. 9-30

v_2/R_e . Similarly, if v_1 is found, then the input impedance will be v_1/i_s . To find both of these voltages, one procedure is to write node equations. If this is done for the nodes labeled a and b in the diagram, the result will be

$$\begin{aligned} v_1 \left(\frac{1}{R_1} + \frac{1}{r_{11}} \right) - \frac{1}{r_{11}} v_2 &= i_s \\ -\frac{1}{r_{11}} v_1 + \left(\frac{1}{r_{11}} + \frac{1}{R} \right) v_2 - \beta i_b &= 0 \end{aligned} \quad (9-44)$$

But i_b can be found in terms of v_1 and v_2 as

$$i_b = \frac{v_1 - v_2}{r_{11}} \quad (9-45)$$

This is now substituted in the previous equation and the node equations are solved for v_1 and v_2 . The current gain and input impedance are found to be

$$\begin{aligned} \text{current gain} = G_i &= \frac{\text{current in } R_e}{i_s} = \frac{(1 + \beta)R_1}{(R_1 + r_{11})(1 + g_{22}R_e) + (1 + \beta)R_e} \\ &\doteq \frac{(1 + \beta)R_1}{R_1 + (1 + \beta)R_e} \end{aligned} \quad (9-46)$$

$$\begin{aligned} \text{input impedance} = Z_i &= \frac{v_1}{i_s} = \frac{R_1[r_{11}(1 + g_{22}R_e) + (1 + \beta)R_e]}{(r_{11} + R_1)(1 + g_{22}R_e) + (1 + \beta)R_e} \\ &\doteq \frac{R_1(1 + \beta)R_e}{R_1 + (1 + \beta)R_e} \end{aligned} \quad (9-47)$$

These expressions appear to be fairly complicated but they can be simplified as shown if all the resistors are assumed to have typical values. ($r_{11} \doteq 500$ ohms, $g_{22} \doteq 10^{-4}$ mhos, $\beta = 50$, R_1 in kilohms and $R_e \doteq 1$ K.) To get an impression of the values and how they compare with the previously discussed common emitter amplifier, let $R_1 = 72.6$ K, $R_e = 2$ K and $\beta = 50$ as in the example of the last section. Then,

$$G_i = 22.2$$

$$Z_i = 42.5\text{K}$$

It is noteworthy that the gain is about a factor of 2 greater than in the common emitter connection. Furthermore, it is positive, indicating that the input and output signals will be in phase.

As for the input impedance, the 42.5K is to be compared with the several hundred ohms represented by r_{11} .

The final quantity of interest is the output impedance. As discussed in the last section, to find this impedance only independent sources are deactinated. In Fig. 9-30, the signal source i_s has been open circuited and an external current source i_2 applied. It is required to find v_2 . Applying Kcl at the upper node leads to

$$i_2 + \beta i_b = \frac{v_2}{R} + \frac{v_2}{R_1 + r_{11}} \quad (9-48)$$

Furthermore,

$$i_b = - \frac{v_2}{R_1 + r_{11}} \quad (9-49)$$

Substituting this into Eq (9-48) and rearranging leads to

$$\begin{aligned} \text{output impedance } Z_2 = \frac{v_2}{i_2} &= R_1 + \frac{R(R_1 + r_{11})}{R_1 + r_{11} + (\beta + 1)R} \\ &= \frac{R_1 R_e}{R_1 + (\beta + 1)R_e} = \frac{Z_1}{1 + \beta} \end{aligned} \quad (9-50)$$

Using the same numerical values as before, finally gives:

$$Z_2 = 830 \text{ ohms}$$

Equation (9-50) shows that large Z_1 and small Z_2 are conflicting requirements since Z_2 is proportional to Z_1 . It is seen, however, that the larger the value of β , the better for achieving both objectives. Furthermore, a larger β also means a larger gain. Thus, one of the goals of transistor manufacturing is to make the β 's as large as possible.

Chapter 10

Magnetic Coupling

Introduction

By virtue of the phenomena of magnetic flux and Faraday's law it is possible to transfer energy from one circuit to another, even though the circuits are not connected by wires. The practical device based on this fact is called a "transformer". In some cases, as in communication circuits, two coils are located near each other, in air, as in Fig. 10-1. In power applications, the coils are wound on an iron core, as in Fig. 10-2. The

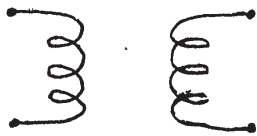


Figure 10-1.

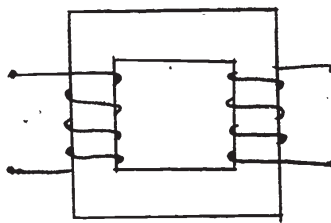


Figure 10-2.

principles of analysis are essentially the same in both cases, with the exception that when there is an iron core it may be necessary to take non-linear effects into consideration.

10-1. Flux Relationships

Consider the pair of coils in Fig. 10-3, which are drawn in a specific way to facilitate the writing of equations. It is assumed that each coil is connected to a circuit which is not shown, so that currents i_1 and i_2 can flow. In general, these currents are varying with time.

Fluxes ϕ_1 and ϕ_2 respectively link coils 1 and 2. For the moment, assume there is no iron present, and that the resistance of each coil is negligible.

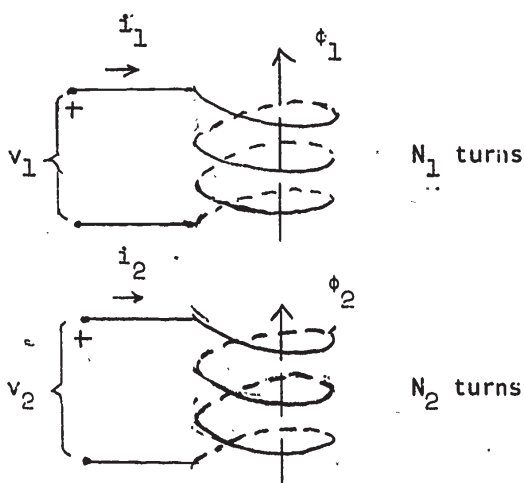


Figure 10-3.

There will be an emf induced in each coil which makes voltages v_1 and v_2 appear at the terminals. These are related to ϕ_1 and ϕ_2 by

$$v_1 = N_1 \frac{d\phi_1}{dt}, \quad v_2 = N_2 \frac{d\phi_2}{dt} \quad (10-1)$$

Next we consider how ϕ_1 and ϕ_2 are related to i_1 and i_2 . Without bothering with a specific case, which could in fact be quite complicated, it is possible to arrive at the required general relationship by the following arguments. If i_2 were zero, ϕ_1 would be proportional to i_1 , and we could use a symbol k_{11} as the proportionality constant, to write

$$\phi_1 = k_{11}i_1$$

Furthermore, it would be possible that i_1 could be zero, and i_2 not zero. Due to proximity of the two coils, there could still be a flux ϕ_1 , which would now be proportional to i_2 . If k_{12} is the proportionality constant, this would be written

$$\phi_1 = k_{12}i_2$$

We have stipulated that no iron is present, so that linearity can be assumed, and therefore that superposition can be applied. Thus, with i_1 and i_2 each non-zero, we will have a general formula

$$\phi_1 = k_{11}i_1 + k_{12}i_2 \quad (10-2)$$

A similar treatment can be applied to coil 2, giving

$$\phi_2 = k_{21}i_1 + k_{22}i_2 \quad (10-3)$$

In doing this, four unknown constants ($k_{11}, k_{12}, k_{21}, k_{22}$) have been introduced. However, these will presently be eliminated, their purpose being merely to show how ϕ_1 and ϕ_2 vary with the two currents.

Now we can use the above equations in Eqs. (10-1) to give

$$v_1 = k_{11}N_1 \frac{di_1}{dt} + k_{12}N_1 \frac{di_2}{dt} \quad (10-4)$$

$$v_2 = k_{21}N_2 \frac{di_1}{dt} + k_{22}N_2 \frac{di_2}{dt}$$

Again temporarily consider the special case where i_2 is identically zero (or constant) so that $di_2/dt = 0$. Then the first of the above equations becomes

$$v_1 = k_{11}N_1 \frac{di_1}{dt}$$

However, we would also have

$$v_1 = L_1 \frac{di_1}{dt}$$

where L_1 is the self inductance of coil 1. Therefore, $k_{11}N_1 = L_1$. Likewise, consideration of coil 2 gives $k_{22}N_2 = L_2$. Therefore, the equations can be written

$$v_1 = L_1 \frac{di_1}{dt} + k_{12}N_1 \frac{di_2}{dt} \quad (10-5)$$

$$v_2 = k_{21}N_2 \frac{di_1}{dt} + L_2 \frac{di_2}{dt}$$

It can be shown that*

$$k_{12}N_1 = k_{21}N_2$$

Let this common value be represented by the letter M. We then have

$$M = k_{12}N_1 = \left(\frac{k_{12}}{k_{11}}\right)L_1$$

$$M = k_{21}N_2 = \left(\frac{k_{21}}{k_{22}}\right)L_2$$

or

$$M = \sqrt{\left(\frac{k_{12}}{k_{11}} \frac{k_{21}}{k_{22}}\right) L_1 L_2}$$

It is customary to use the symbol k to represent the first radical, namely

$$k = \sqrt{\frac{k_{12}}{k_{11}} \frac{k_{21}}{k_{22}}} \quad (10-6)$$

It is called the coefficient of coupling, and we finally have

$$M = k \sqrt{L_1 L_2} \quad (10-7)$$

M is called the mutual inductance between the coils. k is always less than 1.

*Assume i_1 and i_2 are sinusoidal of frequency ω (certainly an acceptable special case). They will be represented by phasors $\bar{I}_1$ and $\bar{I}_2$. Recalling that differentiation produces a multiplication by $j\omega$ in phasor notation, we have

$$\bar{V}_1 = j\omega(L_1\bar{I}_1 + k_{12}N_1\bar{I}_2)$$

$$\bar{V}_2 = j\omega(k_{21}N_2\bar{I}_1 + L_2\bar{I}_2)$$

Recall that power in the sinusoidal case is $\text{Re}(\bar{V}\bar{I}^*)$. The powers into coils (1) and (2) are respectively

$$P_1 = \text{Re}(\bar{V}_1\bar{I}_1^*) = \text{Re}[j\omega(L_1\bar{I}_1\bar{I}_1^* + k_{12}N_1\bar{I}_2\bar{I}_1^*)]$$

$$P_2 = \text{Re}(\bar{V}_2\bar{I}_2^*) = \text{Re}[j\omega(k_{21}N_2\bar{I}_1\bar{I}_2^* + L_2\bar{I}_2\bar{I}_2^*)]$$

However, the product of a complex number by its conjugate is real. Hence $j\omega L_1\bar{I}_1\bar{I}_1^*$ and $j\omega L_2\bar{I}_2\bar{I}_2^*$ are each pure imaginaries and have zero real parts. Also, if $\bar{A}$ is a complex number, $\text{Re}(j\bar{A}) = -\text{Im}(\bar{A})$. Hence

$$P_1 = -\text{Im}(\omega k_{12}N_1\bar{I}_2\bar{I}_1^*)$$

$$P_2 = -\text{Im}(\omega k_{21}N_2\bar{I}_1\bar{I}_2^*) = \text{Im}(\omega k_{21}N_2\bar{I}_2\bar{I}_1^*)$$

(continued on pg. 5)

Thus, it is seen that the original set of four k's have disappeared, having been replaced by three measurable parameters L_1 , L_2 , and k (or M). Note that four parameters reduced to three because one of the original four was not independent, since $k_{12}N_1 = k_{21}N_2$.

10-2. The General Equations

Having developed the concepts of mutual inductance, Eqs. (10-5) can now be written

$$\begin{aligned} v_1 &= L_1 \frac{di_1}{dt} + M \frac{di_2}{dt} \\ v_2 &= M \frac{di_1}{dt} + L_2 \frac{di_2}{dt} \end{aligned} \quad (10-8)$$

A word about signs is in order. These equations were written for Fig. 10-3, in which winding directions were explicitly shown and could be related to the reference directions of current, voltage, and flux. Suppose one of the winding directions (say on coil 2) had been reversed. A little thought will show that this is equivalent to changing the reference directions of i_2 and v_2 . The above equations would become

$$\begin{aligned} v_1 &= L_1 \frac{di_1}{dt} - M \frac{di_2}{dt} \\ -v_2 &= M \frac{di_1}{dt} - L_2 \frac{di_2}{dt} \end{aligned}$$

* (continued).

where in the second equation we have used the fact that $\bar{i}_1 \bar{i}_2^* = (\bar{i}_2 \bar{i}_1^*)^*$. From physical considerations, since neither coil has resistance, the total power ($P_1 + P_2$) must be zero, giving

$$\text{Im} [(-k_{12}N_1 + k_{21}N_2) \bar{i}_2 \bar{i}_1^*] = 0$$

and hence

$$k_{12}N_1 = k_{21}N_2$$

or

$$\begin{aligned} v_1 &= L_1 \frac{di_1}{dt} - M \frac{di_2}{dt} \\ v_2 &= -M \frac{di_1}{dt} + L_2 \frac{di_2}{dt} \end{aligned} \quad (10-9)$$

Thus, the sign associated with M becomes negative in both equations. In certain cases, where coils are changing relative position (say one is rotating) it is sometimes convenient to take this effect into account by allowing M itself to become negative, in which case Eqs. (10-8) are perfectly general. We shall always regard M as positive in this text, however.

The question arises as to how to know whether to use + or - signs, when coils are merely shown on a circuit diagram in symbolic form. This is done by the convention of using a pair of dots, as in Fig. 10-4). The two cases shown at (a) will yield Eqs. (10-8), and Eqs. (10-9) apply to the cases shown at (b).

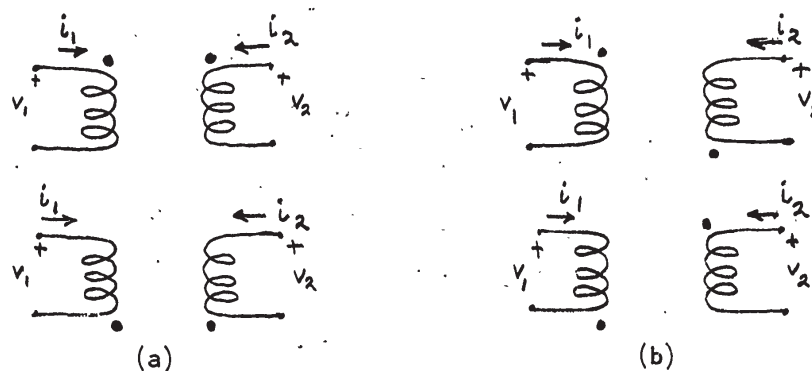


Figure 10-4.

Finally, observe that these equations and examples have all been given for cases where voltage + references appear at terminals where the current reference arrow enters. A change in any one of these will result in corresponding sign changes in the equations. For example, if the direction of i_2 were reversed in Fig. 10-4a, the equations would be

$$\begin{aligned} v_1 &= L_1 \frac{di_1}{dt} - M \frac{di_2}{dt} \\ v_2 &= M \frac{di_1}{dt} - L_2 \frac{di_2}{dt} \end{aligned}$$

Observe that this is different than the effect of changing the winding direction (because only one reference was changed).

With this information, you should be able to write the equations relating voltages and currents for a pair of mutually coupled coils suitably specified in the manner of Fig. 10-4. Admittedly, it is easy to get confused on signs, especially when the reference directions are not like those in Fig. 10-4. Until you acquire skill in this matter, it is not a bad idea to write the equations first for these reference directions, and then change signs in terms containing any variables which have different references.

10-3. Effect of Resistance

The voltages we have been using are terminal voltages, in the absence of resistance, and are really emf's. For the current reference directions we have been using, the terminal voltage is the emf plus the Ri product. Thus, when coils have respective resistances R_1 and R_2 , the complete equations are

$$\begin{aligned} v_1 &= R_1 i_1 + L_1 \frac{di_1}{dt} \pm M \frac{di_2}{dt} \\ v_2 &= \pm M \frac{di_1}{dt} + R_2 i_2 + L_2 \frac{di_2}{dt} \end{aligned} \quad (10-10)$$

10-4. Measurement of M

One way to determine M is to connect the two coils in series, as in Fig. 10-5.

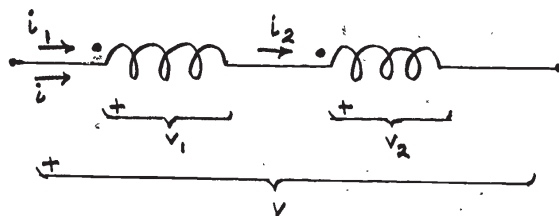


Figure 10-5.

The currents are the same ($i = i_1 = i_2$) and the voltage across the combination is $v = v_1 + v_2$. Thus, Eqs. (10-10) become

$$v_1 = R_1 i + (L_1 + M) \frac{di}{dt}$$

$$v_2 = (L_2 + M) \frac{di}{dt} + R_2 i$$

and so

$$v = (R_1 + R_2) i + (L_1 + L_2 + 2M) \frac{di}{dt}$$

From this equation we reach the conclusion that the inductance of the combination is

$$L_{eq.} = L_1 + L_2 + 2M \quad (10-11)$$

Since L_1 , L_2 , and $L_{eq.}$ can each be measured, M can be found. (What would the last equation be if one of the dots in Fig. 10-5 had been on the other end?)

10-5. Sinusoidal Solutions

If all voltages and currents are sinusoidal, namely

$$\begin{aligned} v_1 &= \sqrt{2} V_1 \cos(\omega t + \theta_{v1}) \\ i_1 &= \sqrt{2} I_1 \cos(\omega t + \theta_{i1}) \\ v_2 &= \sqrt{2} V_2 \cos(\omega t + \theta_{v2}) \\ i_2 &= \sqrt{2} I_2 \cos(\omega t + \theta_{i2}) \end{aligned} \quad (10-12)$$

equations can be written in terms of the phasor defined from the above, which are

$$\begin{aligned} \bar{V}_1 &= V_1 e^{j\theta_{v1}} & \bar{I}_1 &= I_1 e^{j\theta_{i1}} \\ \bar{V}_2 &= V_2 e^{j\theta_{v2}} & \bar{I}_2 &= I_2 e^{j\theta_{i2}} \end{aligned} \quad (10-13)$$

Recalling the principles of sinusoidal circuit analysis, it is recalled that if

$$\bar{I}_1 = I_1 e^{j\theta_{11}} \text{ is the phasor for } i_1 = \sqrt{2} I_1 \cos(\omega t + \theta_{11})$$

then

$$j\omega \bar{I}_1 \text{ is the phasor for } \frac{di_1}{dt}$$

(Refer to the chapter on sinusoids for the derivation of this.) Thus, referring to Eqs. (10-10), in the sinusoidal case only, these can be replaced by

$$\begin{aligned} \bar{V}_1 &= (R_1 + j\omega L_1) \bar{I}_1 \pm j\omega M \bar{I}_2 \\ \bar{V}_2 &= \pm j\omega M \bar{I}_1 + (R_2 + j\omega L_2) \bar{I}_2 \end{aligned} \quad (10-14)$$

These equations make it possible to draw an equivalent circuit diagram in which the effect of M is replaced by a pair of dependent voltage sources, as shown in Fig. 10-6b.

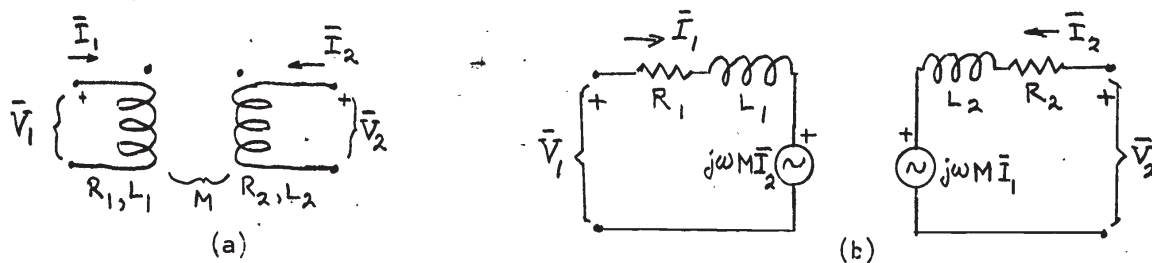


Figure 10-6.

The original circuit is shown at (a). In Fig. 10-6b there is no mutual coupling; the two parts are entirely separate. But it will be observed that the dependent source in the number (1) circuit depends on $\bar{I}_2$, and vice-versa. If you will write Kirchhoff's equations for Fig. 10-6b, you will obtain the version of Eqs. (10-14) having the + signs. (How would Fig. 10-6b be changed if one of the dots in the original circuit had been at the bottom? Does it make any difference which one?)

This equivalent circuit is not necessary, it provides no information not given by Eqs. (10-14). However, it gives exactly the same information as the equations, and may be easier to remember. Also, when the pair of coils is connected to other circuits, the equivalent circuit can sometimes be a trifle simpler to use. The example given in the next section is a case in point. (See if you can do it without the equivalent circuit.)

10-6. Example of a Solution of a Circuit Problem

Referring to Fig. 10-7a, assume that we are to find the power delivered to the resistor R_L . The equivalent circuit is shown in Fig. 10-7b. (Why is the polarity on the dependent source reversed compared with Fig. 10-6b?)

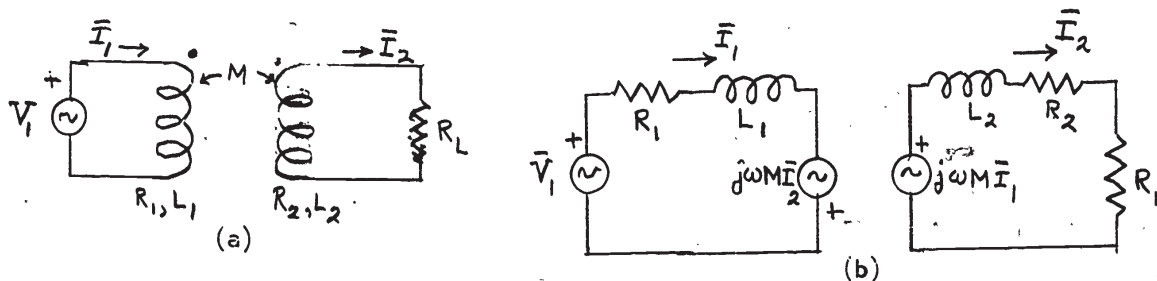


Figure 10-7.

Two equations are obtained from Fig. 10-7b, as follows:

$$\begin{aligned}\bar{V}_1 + j\omega M \bar{I}_2 &= (R_1 + j\omega L_1) \bar{I}_1 \\ j\omega M \bar{I}_1 &= (R_2 + R_L + j\omega L_2) \bar{I}_2\end{aligned}\quad (10-15)$$

The second of these gives

$$\begin{aligned}\bar{I}_1 &= \left(\frac{R_2 + R_L + j\omega L_2}{j\omega M} \right) \bar{I}_2 \\ &= \left(\frac{L_2}{M} - j \frac{R_2 + R_L}{\omega M} \right) \bar{I}_2\end{aligned}$$

This can be substituted into the first of Eqs. (10-15), and after some purely algebraic manipulation, the result is

$$\bar{V}_1 = \frac{1}{M} \left\{ R_1 L_2 + (R_2 + R_L) L_1 + \frac{j}{\omega} [\omega^2 (L_1 L_2 - M^2) - R_1 (R_2 + R_L)] \right\} \bar{I}_2$$

This of course can be solved for $\bar{I}_2$, giving

$$\bar{I}_2 = \frac{M \bar{V}_1}{R_1 L_2 + (R_2 + R_L) L_1 + \frac{j}{\omega} [\omega^2 (L_1 L_2 - M^2) - R_1 (R_2 + R_L)]}$$

To find the power, we want $|\bar{I}|^2$, which is $M^2 |\bar{V}_1|^2$ divided by the square of the absolute value of the denominator. The latter is the sum of the squares of the real and imaginary parts. Thus,

$$P = \frac{M^2 |\bar{V}_1|^2 R_L}{[R_1 L_2 + (R_2 + R_L) L_1]^2 + \left[\omega (L_1 L_2 - M^2) - \frac{R_1 (R_2 + R_L)}{\omega} \right]^2} \quad (10-16)$$

Although this is a long formula, it is simple enough to calculate. Observe that it shows that no power will be transmitted if $M = 0$, which is an expected result since this is the condition of zero coupling.

10-7. The Iron Core Transformer

The equations we have derived apply to a transformer with an iron core, provided it is valid to assume a linear magnetization curve. However, it is worthwhile to give some special consideration to a simplified version which gives good approximate answers when k is very small, as is usually the case for an iron core. This treatment brings the turns ratio N_1/N_2 into evidence as a fundamental transformer parameter.

Refer to Fig. 10-8, which may be considered the equivalent of Fig. 10-3, except that an iron core is shown explicitly. For simplicity, we shall assume zero resistance in each winding. Dots are shown on the windings, even though they are redundant in view of the fact that winding directions are shown explicitly.

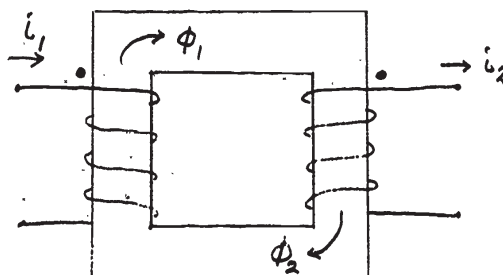


Figure 10-8.

The reference direction of i_2 is reversed compared with Fig. 10-3. This change is not necessary, but is made because the direction in Fig. 10-8 is the one customarily made in the analysis of iron core transformers.

In this case, Eqs. (10-2) and (10-3) become

$$\phi_1 = k_{11}i_1 - k_{12}i_2, \quad \phi_2 = k_{21}i_1 - k_{22}i_2$$

We now make the crucial assumption that ϕ_1 and ϕ_2 differ by a negligible amount. That is, we assume flux is confined to the core. Then, by equating the above, we get

$$k_{11}i_1 - k_{12}i_2 = k_{21}i_1 - k_{22}i_2$$

or

$$(k_{11} - k_{21})i_1 = (k_{12} - k_{22})i_2$$

(10-17)

This equation must be true for all values of i_1 and i_2 . For example, i_1 could be a direct current, and i_2 could be zero; there is no physical reason why this could not be true. Under what condition can the two sides of Eq. (10-17) be equal even for all possible values of i_1 and i_2 ? Only if each quantity in parentheses is zero, giving

$$\frac{M}{N_2} = k_{21} = k_{11} = \frac{L_1}{N_1} \quad (a)$$

$$\frac{M}{N_1} = k_{12} = k_{22} = \frac{L_2}{N_2} \quad (b)$$

(10-18)

Referring to Eq. (10-6) we see now that (for the assumption that $\phi_1 = \phi_2$)

$$k = 1$$

and furthermore, from Eq. (10-7),

$$M = \sqrt{L_1 L_2} \quad (10-19)$$

Now, let us consider ϕ again, which, in view of these results, becomes

$$\begin{aligned} \phi &= k_{11} i_1 - k_{12} i_2 \\ &= \frac{L_1}{N_1} i_1 - \frac{M}{N_1} i_2 \\ &= \frac{L_1}{N_1} i_1 - \frac{\sqrt{L_1 L_2}}{N_1} i_2 \end{aligned}$$

This gives a relationship between i_2 and i_1 as follows:

$$i_1 = \sqrt{\frac{L_2}{L_1}} i_2 + \frac{N_1}{L_1} \phi$$

However, Eq. (10-18b) gives

$$\frac{\sqrt{L_1 L_2}}{N_1} = \frac{L_2}{N_2}$$

or

$$\sqrt{\frac{L_2}{L_1}} = \frac{N_2}{N_1} \quad (10-20)$$

so that

$$i_1 = \frac{N_2}{N_1} i_2 + \frac{\phi}{k_{11}} \quad (10-21)$$

The parameter k_{11} is a function of properties of the magnetic circuit. Specifically, if l is its length, A its cross sectional area, and k_m the relative permeability, solution of the magnetic circuit (considering coil 1 only) gives

$$\phi = \frac{k_m \mu_0 A N_1^2 i_1}{l}$$

and so

$$k_{11} = \frac{k_{12} \mu_o AN_1}{\ell} \quad (10-22)$$

If k_{11} is sufficiently large, or ϕ is sufficiently small, the ϕ/k_{11} term in Eq. (10-21) can be negligible. In that case, we get the approximate equation

$$i_1 = \frac{N_2}{N_1} i_2 \quad (10-23)$$

In other words, the current ratio i_1/i_2 is approximately the inverse of the turns ratio.

10-8. The Iron Core Transformer in the Sinusoidal Steady State

We can get a more satisfying answer as to whether ϕ/k_{11} is negligible in Eq. (10-21) when all quantities vary sinusoidally. Assume again that all voltages and currents are sinusoidal. Flux will then also be sinusoidal and will be represented by the phasor $\bar{\phi}$. In Eq. (10-21) we can replace each instantaneous quantity (sinusoidal case only) by the corresponding phasor. Also, we shall use N_1/L_1 in place of $1/k_{11}$, giving

$$\bar{i}_1 = \frac{N_2}{N_1} \bar{i}_2 + \frac{N_1}{L_1} \bar{\phi}$$

However, we also have

$$\bar{v}_1 = j\omega N_1 \bar{\phi} \quad (10-24)$$

which combines with the above to give

$$\bar{i}_1 = \frac{N_2}{N_1} \bar{i}_2 + \frac{\bar{v}_1}{j\omega L_1} \quad (10-25)$$

Thus, in this case we see that

$$\frac{\bar{i}_1}{\bar{i}_2} = \frac{N_2}{N_1} \quad (10-26)$$

if

$$\frac{\bar{V}_1}{j\omega L_1}$$

is small. This quantity, is called the magnetizing current. It is the current in winding 1 when it is connected to a source of voltage $\bar{V}_1$ and when winding 2 is an open circuit.

Corresponding to Eq. (10-24), we also have

$$\bar{V}_2 = j\omega N_2 \bar{\Phi}$$

and so we see that under the assumptions being made (zero resistance and no leakage flux)

$$\frac{\bar{V}_1}{\bar{V}_2} = \frac{N_1}{N_2} \quad (10-27)$$

If winding resistance is introduced, Eq. (10-27) becomes an approximation.

Equation (10-24) provides the point of departure for the last consideration.

Let us write it as an absolute value relation (using symbols without bars for absolute values). Thus, we can write

$$\Phi = \frac{V_1}{\omega N_1} \quad (10-28)$$

This is crucial because there is a limit to the value that Φ can attain for approximate linear operation in a transformer with iron core.

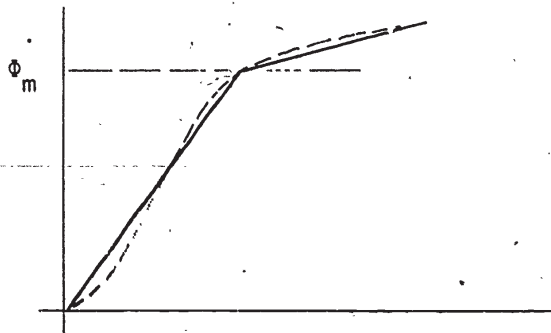


Figure 10-9.

In Fig. 10-9 is shown a linearized approximation of the magnetization curve of the magnetic circuit. It will not be permissible to exceed a certain value of flux, say Φ_m . This means that the Φ in Eq. (10-28) must be less than $\Phi_m/\sqrt{2}$ (Why?). In terms of voltage, frequency, and number of turns, this gives

$$V_1 \leq \frac{\Phi_m}{\sqrt{2}}$$

This relationship has useful and important implications concerning the use of a transformer on other than the voltage or frequency for which it was designed. Suppose it was designed so that at rated voltage and frequency, the above relationship was just barely satisfied. If the voltage is raised, the limit will be exceeded. It will also be exceeded if the frequency is decreased, while the voltage remains unchanged. The only way to operate a transformer at reduced frequency is to reduce the voltage in proportion.

Equation (10-28) also shows why it is expensive to make "audio transformers" that will operate well at low frequencies. They must be provided with massive magnetic circuits that will permit the large Φ demanded by Eq. (10-28) when ω is small. Parenthetically, it might be suspected that good performance at low frequencies would be provided by making N_1 large. This would be true, except that such transformers must also operate at high frequencies, and when N_1 is made too large interwinding capacitance becomes detrimental at high frequencies. Design of transformers that will operate over a wide range of frequencies is not a job for the uninitiated.