for voltage and current readings, and a 0-50 volt *d-c* power supply with 1% ripple.

Conclusion

As can be seen from *Fig. 3*, the maximum error that would be introduced by the circuit components at the high and low ends of the expanded scale is 1.25%. On a production basis it would be highly practical to select matched Zener diodes and very probable that this source of error can be eliminated.

The series diodes D_1 and D_2, because of their operation in the forward direction, provide some temperature compensation of the Zener diodes. D_1 and Z_1, D_2 and Z_2 can be matched to provide temperature coefficients as low as 0.05 $mv/°C$.

A slight modification to the circuit readily adapts it for use with an *a-c* rectifier type meter. *Fig. 4* shows replacement of diodes D_1 and D_2 with Zeners Z_3 and Z_4 of the same type as Z_1 and Z_2. D_4 has been added parallel to D_3 in the reverse direction to perform the same function as D_3 on the alternate half cycle.

Transistor Switching Analysis

DR. C. A. MEAD*

Part 3

Transistor Switching Performance

It was recognized early in the development of the junction transistor that the symmetry of the device implied the unique ability to *saturate*. Since both the collector and emitter junctions are capable of emitting minority carriers into the base region when forward biased, and since the diffusion current across the base is due to the difference in minority carrier density between the two junctions, it is clear that when the transistor is saturated *current may flow in either direction*, depending upon which density is the larger. In the saturated condition the transistor closely resembles a closed switch, and dynamic resistances of a few tenths of an ohm are easily obtained with small units. If, on the other hand, both junctions are reverse biased, the transistor is cut off and only the very small junction reverse currents flow. Hence the transistor resembles an open circuit, impedances of several megohms being common. When a transistor spends the majority of its time in one of two states, fully saturated or fully cut off, passing through the normally biased region only to get from one state to the other, it is said to be operating as a *switch*. A distinction should be drawn at this point between true switching service and non-saturating service. A so-called non-saturating switch is one where the transistor may become either cut off or normally biased (usually with a rather low collector voltage), but not saturated. Such operation is more properly termed Class *C* pulse amplifier service. When the transistor is not caused to saturate, its operation may be analyzed with adequate accuracy by the use of small signal equivalent circuits. However, when the transistor is caused to saturate, the analysis becomes quite complicated and has traditionally been avoided in circuit work. Here the lumped model comes into its own, since it transforms the difficult problem of transistor saturation and storage into one of simple *R-C* circuit analysis. The advantage of this approach for the circuit engineer is obvious.

Quasi Static Performance. Consider a transistor connected in the circuit shown in *Fig. 12*. For this analysis, we shall use the lumped model of *Fig. 9*. However, we may ignore C_1 and C_2 since we are only interested in slowly varying *d-c* quantities. The three regions of transistor operation will now be considered:

(a) Cut-Off. In the cut-off state, $\rho_1 = \rho_2 = -p_n$. Therefore

$$i_c = p_n G_2 = \frac{i_{co}(1+\beta)}{1+\beta+\beta_i}$$

which becomes

$$i_c \approx \frac{i_{co}}{1+\beta_i/\beta} \quad \text{if} \quad \beta >> 1 \tag{7}$$

In the usual case, i_{co} is very small and hence the collector voltage is approximately equal to the supply voltage

$$v_c \approx V$$

The base current

$$i_b = -p_n(G_1+G_2) \approx -i_{co} \quad \text{if} \quad \beta >> 1 \tag{8}$$

In this state the only significant contribution to the power dissipation of the transistor is from the collector.

$$P \approx i_c V = \frac{i_{co} V}{1+\beta_i/\beta} \tag{9}$$

* Electrical Engineering Department, California Institute of Technology, Pasadena, California

(b) Active Region. In the active region

$$\rho_2 = -p_n \approx 0 \quad \text{if} \quad i_c >> i_{co}$$

The collector excess density may be considered zero provided the collector current is large compared with i_{co}. Thus

$$i_c = \beta i_b$$
$$v_c = V - \beta i_b R$$

Since $\beta >> 1$ and the collector voltage is larger than the base voltage (since the transistor is not yet saturated), the power dissipated due to base current is negligible compared with that due to the collector current. Thus

$$P \approx \beta V i_b - \beta^2 R i_b^2 \qquad (10)$$

which reaches a maximum when $v_c \approx V/2$.

(c) Saturation. When the base current is increased to the point where $v_c = v_b$, the transistor saturates and the collector current becomes substantially independent of further increases in base current. Since both junctions are forward biased, the base and collector voltages will be neglected in comparison with V. Hence

$$i_c \approx V/R$$
$$i_b > V/\beta R$$

We may now solve for the excess densities ρ_1 and ρ_2 in order to obtain the junction voltages v_{be} and v_{bc}. We may write the equations for base and collector current as

$$i_c = (\rho_1 - \rho_2)\, G_d - \rho_2 G_2$$
$$i_b = \rho_1 G_1 + \rho_2 G_2$$

which may be solved for ρ_1 and ρ_2 writing β for G_d/G_1 and β_i for G_d/G_2

$$\rho_1 G_d \approx \frac{(1+\beta_i)\, i_b + i_c}{1 + \frac{\beta_i}{\beta} + \frac{1}{\beta}} \qquad (11)$$

$$\rho_2 G_d \approx \frac{\beta i_b - i_c}{1 + \frac{\beta}{\beta_i} + \frac{1}{\beta_i}} \qquad (12)$$

The combination ρG_d is a convenient quantity with which to deal since it has the dimensions of a current. When $\beta i_b = i_c$, the transistor has just become saturated and

$$i_c = \rho_1 G_d = V/R$$

At higher base currents the collector current remains essentially constant, but $\rho_2 G_d$ increases. Hence $\rho_2 G_d$ is a significant measure of the amount by which the transistor has been driven into saturation. We may now determine the junction voltage since

$$\rho_1 G_d = p_n G_d\, (e^{qv_{be}/kT} - 1)$$
$$\rho_2 G_d = p_n G_d\, (e^{qv_{b_c}/kT} - 1)$$

Therefore

$$v_{bc} = \frac{kT}{q} \operatorname{Ln} \left(\frac{\rho_1 G_d}{p_n G_d} + 1 \right)$$

$$v_{bc} = \frac{kT}{q} \operatorname{Ln} \left(\frac{\rho_2 G_d}{p_n G_d} + 1 \right)$$

where $p_n G_d$ may be determined from equations 5 or 6. Although the i_{co} expression is not as accurate as a measurement of junction voltage and current, it is often convenient for germanium units.

$$v_{be} \approx \frac{kT}{q} \operatorname{Ln} \left[\frac{i_c + (1+\beta_i)\, i_b}{\beta_i i_{co}} + 1 \right] \qquad (13)$$

$$v_{bc} \approx \frac{kT}{q} \operatorname{Ln} \left[\frac{(1+\beta)\, i_b - i_c}{\beta i_{co}} + 1 \right] \qquad (14)$$

The collector saturation voltage is just the difference between the two junction voltages. In most cases the drive current is sufficiently large that ρ_1 and ρ_2 are both much greater than p_n. Thus from equations 13 and 14

$$v_c \approx \frac{kT}{q} \operatorname{Ln} \frac{\rho_1 G_d}{\rho_2 G_d}$$

which from equations 11 and 12 becomes

$$v_c = \frac{kT}{q} \operatorname{Ln} \left(\frac{1 + \frac{1 + i_c/i_b}{\beta_i}}{1 - i_c/\beta i_b} \right) \qquad (15)$$

In the inverted connection the normal and inverse quantities merely exchange places and the junction voltages become

$$v_{bc} = \frac{kT}{q} \operatorname{Ln} \left[\frac{i_e + (1+\beta)\, i_b}{\beta i_{co}} + 1 \right] \qquad (16)$$

$$v_{be} = \frac{kT}{q} \operatorname{Ln} \left[\frac{(1+\beta_i)\, i_b - i_e}{\beta_i i_{co}} + 1 \right] \qquad (17)$$

Also, the emitter saturation voltage becomes

$$v_e = \frac{kT}{q} \operatorname{Ln} \left(\frac{1 + \frac{1 + i_e/i_b}{\beta}}{1 - i_e/\beta_i i_b} \right) \qquad (18)$$

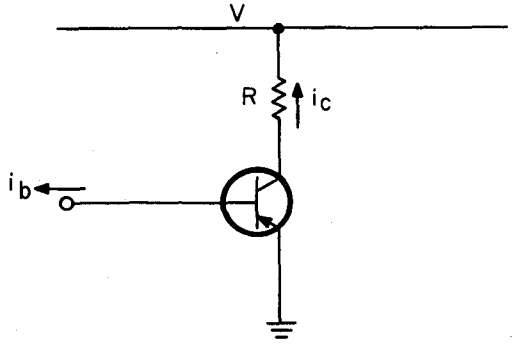

Fig. 12—Elementary transistor switch.

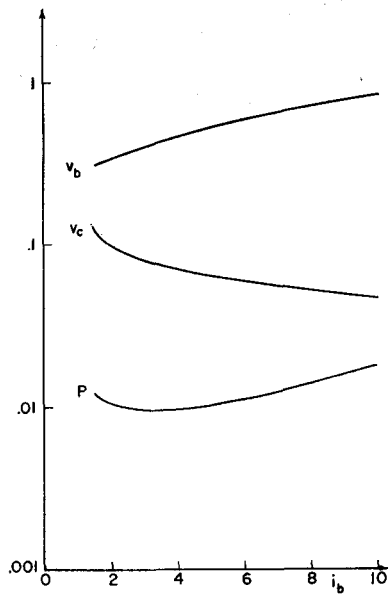

Fig. 13—Voltage and power variation for normally connected transistor.

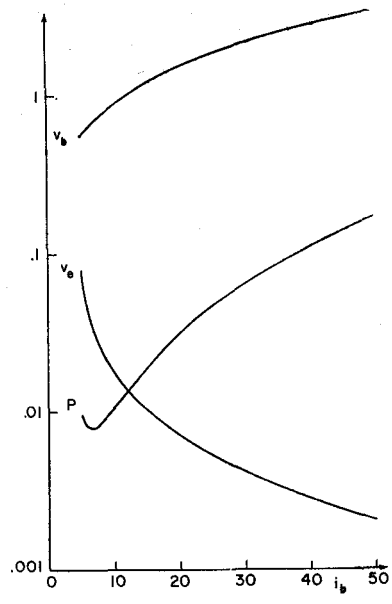

Fig. 14—Voltage and power variation for inversely connected transistor.

Plots of the base input voltage, saturation voltage, and power dissipation for a typical switching transistor operating in the normal and inverse connections are shown in *Figs. 13* and *14*. It is of interest to note the minimum in power dissipation which occurs at some base drive current. Clearly this drive current represents an optimum operating point for the particular transistor and collector current involved. The serious nature of large overdrive currents is quite apparent.

Dynamic Resistance. The resistance of a saturated transistor to small *a-c* signals is often of interest. This dynamic resistance is given by

$$R_s = \frac{\delta v_c}{\delta i_c}$$

for the normal connection and may be obtained from equation 15.

$$R_s = \frac{kT}{q}\left[\frac{1}{(1+\beta)\, i_b - i_c} + \frac{1}{(1+\beta_i)\, i_b + i_c}\right] \qquad (19)$$

A similar expression is obtained for the inverted connection, only the roles of the collector and emitter are interchanged and β is interchanged with β_i. For sufficiently large drive currents

$$R_s \approx \frac{kT}{q}\left[\frac{\beta+\beta_i}{\beta(1+\beta_i)\, i_b}\right] \qquad \beta >> 1 \qquad (20)$$

which shows the saturation resistance inversely dependent on β, β_i and the drive current. The importance of *both* β and β_i *is* dramatically illustrated by this formula.

All transistors have certain ohmic resistances associated with their collector and base circuits due to the semiconductor material between the active region of the device and its contact to the outside world. For the first approximation we may assume these resistances constant, and if the voltage drops across them are not negligible, we must add them to the appropriate voltages already calculated. For switching transistors made by the alloying process, and many others, the collector series resistance is negligible, but the base resistance R_b should always be taken into account. The total base voltage thus becomes

$$v_b = v_{be} + i_b R_b$$

Since we are considering base currents up to very high values, we may no longer neglect the power dissipation in the base circuit.

$$P = v_c i_c + v_{be} i_b + i_b^2 R_b \qquad (21)$$

As we have noted, a number of simplifying assumptions have been made which under many conditions may be very questionable. However, for practical circuit design one often uses models which are greatly oversimplified, not because of their extreme accuracy but because they provide approximate answers and still allow a qualitative understanding of the problem. For example, the use of small signal equivalent circuits without regard to the magnitude of the signal level is an accepted engineering procedure. The nonlinearities are taken into account qualitatively after the main circuit behavior has been determined from the linear analysis. The lumped model serves in the same capacity for switching problems as a small signal equivalent circuit does for problems where the transistor is normally biased. It provides a straightforward method of obtaining results with reasonable

accuracy in the majority of cases and hence may claim a certain engineering importance. The physical insight gained by the lumped model analysis is often much more valuable than a slight improvement in accuracy, since it enables the analyst to make qualitative statements concerning changes in circuit parameters, a very important step in the design procedure.

Transient Response.[8] Again we shall consider a junction transistor connected as shown in *Fig. 12.*

(a) Turn On. In the cut-off condition $i_b = -i_{co}$ as before. Now let us apply a positive current step of magnitude I_1 (large compared to i_{co}) and calculate the collector current response. As long as the transistor remains normally biased, the collector current may be computed from small signal formulae

$$i_c(s) = \frac{\beta i_b(s)}{1 + \frac{s}{\omega_\beta}} = \frac{\beta\omega_\beta I_1}{s(s + \omega_\beta)}$$

$$i_c = \beta I_1 (1 - e^{-\omega_\beta t})$$

However, the collector remains reverse biased only so long as

$$i_c < V/R$$

Thus the collector current rises toward the asymptote βI_1 with time constant $1/\omega\beta$. If $\beta I_1 > V/R$, the transistor will saturate when the collector current reaches V/R. The "rise time" required is thus

$$t_r = -\frac{1}{\omega_\beta} \ln\left(1 - \frac{V}{\beta I_1 R}\right) \qquad (22)$$

Usually the transistor is driven quite hard in order to minimize the rise time. Under these conditions $\beta I_1 >> V/R$ and we may expand the log, retaining only the first term

$$t_r \approx \frac{V}{\beta\omega_\beta RI_1}$$

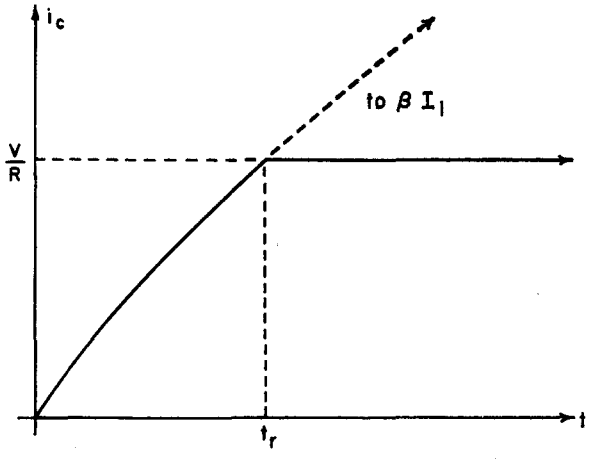

Fig. 15—Rise transient as predicted by transistor lumped model.

Since normally $\beta >> 1$, $\omega_\alpha \approx \beta\omega_\beta$. The rise time may be written

$$t_r \approx \frac{V}{\omega_\alpha RI_1} \qquad (23)$$

It is thus clear that the alpha cut-off frequency is the determining factor in turn-on time and not β or ω_β alone. The conditions during turn-on are illustrated in *Fig. 15.*

(b) Storage. After the transistor has reached the steady state with $i_b = I_2$, let us suddenly reverse the base current to $i_b = -I_3$. As in the case of the diode, the collector junction remains forward biased since ρ_2 cannot change instantaneously. Hence the collector current remains $I_c = V/R$. After a "storage time" t_s, ρ_2 has reached zero and the transistor becomes normally biased.[10] In order to determine ρ_1 and ρ_2 during the storage period, we may determine their initial values from the steady state conditions given by equation 12, assuming $\beta >> 1$

$$\rho_2(0) = \frac{\beta I_2 - I_c}{\frac{\beta}{\beta_i} + 1} \qquad I_c = V/R \qquad (24)$$

The transient densities may be found by superposing the steady state solution above (for $I_c = V/R$, $i_b = I_2$) upon the solution for a negative base current step of magnitude $I_2 + I_3$ and constant collector current $i_c = 0$.

The result of this calculation, assuming $\beta >> 1$ is

$$\rho_2 G_d \approx (I_2 + I_3) \frac{\beta\beta_i}{\beta + \beta_i} e^{-bt} - \frac{\beta_i I_c + \beta\beta_i I_3}{\beta + \beta_i} \qquad (25)$$

where $I_c = V/R$ and

$$b \approx \frac{\omega_\alpha\omega_{\beta i} + \omega_\beta\omega_{\alpha i}}{\omega_\alpha + \omega_{\alpha i}}$$

The "storage time" t_s ends when ρ_2 reaches zero. Hence

$$e^{-bt_s} = \frac{\beta_i I_c + \beta\beta_i I_3}{\beta\beta_i (I_2 + I_3}$$

or

$$t_s = \frac{1}{b} \ln \frac{I_2 + I_3}{I_3 + \frac{I_c}{\beta}} \qquad (26)$$

from which it is clear that the storage time may be reduced by using large turn-off currents. I_2 must be greater than I_c/β for the transistor to be saturated, but the overdrive may be reduced to decrease the storage time. *The controlling time constant for the storage period is* b, hence for small storage times *both* ω_β and $\omega_{\beta i}$ should be large.

If these frequencies are nearly equal

$$b \approx \omega_\beta \approx \omega_{\beta i}$$

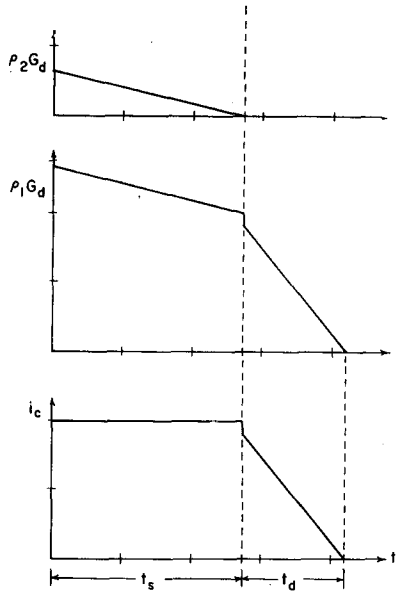

Fig. 16—Storage and decay transients as predicted by transistor lumped model.

Plots of ρ_1, ρ_2 and i_c during the storage period are shown in *Fig. 16.*

(c) Turn Off. When ρ_2 approaches zero, the collector current is made up of two components:

$$I_c = \rho_1 G_d + C_2 \frac{d\rho_2}{dt}$$

However, as the collector junction becomes reverse biased, $d\rho_2/dt$ abruptly becomes zero since ρ_2 cannot become less than $-p_n$ and the lumped model predicts a slight discontinuity in collector current. In reality the actual current changes smoothly, and this is another case where the lumped nature of the model fails to give the completely correct physical picture. However, for purposes of calculating the "decay time" or time required for i_c to reach zero, the lumped model expressions including the discontinuity will be more accurate than the corresponding expressions assuming no discontinuity. The reason is that the true collector current very quickly approaches the predicted value as an asymptote and by the time i_c approaches zero, the lumped model expression is quite accurate. Let us therefore calculate the magnitude of the discontinuity.

$$\Delta i = C_2 \frac{d\rho_2}{dt}\bigg|_{t=t_s} = -\frac{C_2}{G_d} b \frac{(\beta_i I_c + \beta\beta_i I_3)}{\beta + \beta_i}$$

Substituting for b and assuming $\beta >> 1$

$$\Delta i = -(I_c + \beta I_3)\left(\frac{\omega_\beta}{\omega_\alpha + \omega_{\alpha i}}\right)$$

Thus under normal circumstances the relative magnitude of the discontinuity is quite small. However, for very large overdrive currents it may become important. Under these conditions

$$\Delta i \approx -\beta I_3 \frac{\omega_\beta}{\omega_\alpha + \omega_{\alpha i}} \qquad (27)$$

After ρ_2 reaches zero, the transistor is again normally biased and we may use the small signal approach, as with the turn-on period. The collector current approaches the asymptote $-\beta I_3$ with a time constant $1/\omega_\beta$. During this period $i_c = \rho_1 G_d$

$$i_c = (I_c - \Delta i)\, e^{-\omega_\beta t} - \beta I_3 (1 - e^{-\omega_\beta t})$$

The "decay time" t_d is the time required for i_c to reach zero

$$t_d = \frac{1}{\omega_\beta} \ln\left(1 + \frac{I_c + \Delta i}{\beta I_3}\right)$$

For large turn-off current $I_3 >> I_c/\beta$ and we may approximate the logarithm. The decay time therefore becomes

$$t_d \approx \frac{1}{\omega_\alpha}\left(\frac{I_c}{I_3} - 1 + \frac{1}{\omega_{\alpha i}/\omega_\alpha}\right) \quad \text{if} \quad \beta >> 1 \qquad (28)$$

Thus, as with the rise time, *the decay time is determined by the magnitude of the drive current and the* α *cut-off frequency,* and not by β or ω_β alone.

After the turn-off period, the transistor is cut off and the collector current resumes its small steady state value as given by equation 7. Comparing the expression for rise time as given in equation 23, we see that the decay time is always less than the rise time for a given base drive current. This is true because recombination is *helpful* during the decay period but *harmful* during the rise period. ■ ■

References

1. Ebers, J. J. and J. L. Moll, "Large Signal Behavior of Junction Transistors," *Proc. IRE,* Vol. 42, (December 1954) p. 1761.
2. Middlebrook, R. D., "An Introduction to Junction Transistor Theory," John Wiley and Sons, Inc., New York, (1957).
3. Linvill, J. G., "Lumped Models of Transistors and Diodes," *Proc. IRE,* Vol. 46, (June 1958) p. 1141.
4. Moll, J. L., "The Evolution of the Theory for the Voltage-Current Characteristics of *P-N* Junctions," *Proc. IRE,* Vol. 46, (June 1958) p. 1076.
5. Kingston, R. H., "Switching Time in Junction Diodes and Junction Transistors," *Proc. IRE,* Vol. 42, (May 1954) p. 829.
6. Early, J. M., "Effects of Space Charge Layer Widening in Junction Transistors," *Proc. IRE,* Vol. 40, (November 1952) p. 1401.
7. Gaicoletto, L. J., "Study of *P-N-P* Alloy Junction Transistors from D-C through Medium Frequencies," *RCA Review,* Vol. 15, (December 1954) p. 506.
8. Moll, J. L., "Large Signal Transient Response of Junction Transistors," *Proc. IRE,* Vol. 42. (December 1954) p. 1773.
9. Hurley, R. B., "Junction Transistor Electronics," John Wiley and Sons, Inc., New York, (1958) Chapt. 19.
10. DeWitt, D. and A. L. Rossoff, "Transistor Electronics," McGraw-Hill Book Co., Inc., New York, (1957) Chapt. 11.