# An Approximation to Alpha of a Junction Transistor*

R. D. MIDDLEBROOK† AND R. M. SCARLETT‡

*Summary*—**In this paper, a new approximation for the frequency dependence of the short-circuit current gain of a theoretical junction triode is derived, which is a rational function of frequency and convenient to use. It is shown that the approximation is in excellent agreement with the frequency response of the theoretical expression for alpha in both magnitude and phase to above the alpha cutoff frequency. The approximation is also considered in the time domain, where it is in good agreement except for small values of time corresponding to frequencies well above alpha cutoff.**

## Introduction

PRESENT knowledge of transistors is in about the same stage as that of vacuum tubes thirty years ago. The flow of electrons from cathode to plate had been studied, leading to equations describing the action of a tube. Also, experiment had shown that the small-signal action could be described by a simple equivalent circuit, the change in behavior at high frequencies being described by suitable capacitances. This equivalent circuit was substantiated both qualitatively and quantitatively by the theory, and thus the theory was able to predict the behavior of a tube throughout its useful range.

With a transistor, this last step is somewhat more difficult to accomplish, partly because transit time effects exert a controlling influence throughout a large part of the useful range. As a result, the equivalent circuit is inherently more complicated than that for a tube, and various approximations must be made in order that useful results can be achieved.

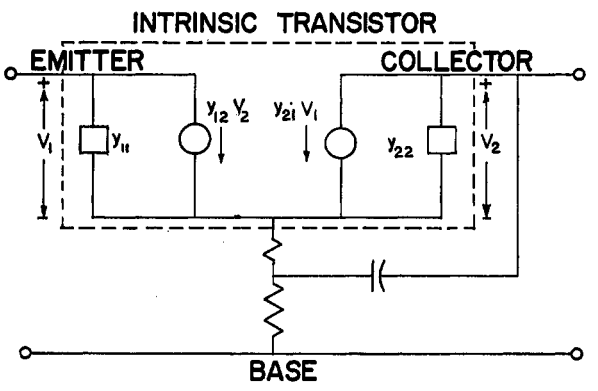


Fig. 1—Equivalent circuit of the complete junction transistor.

Although the theoretical derivation of an equivalent circuit proceeds most conveniently in terms of the short-circuit admittance, or $y$ parameters (shown for the intrinsic transistor in Fig. 1, other sets of parameters which can be calculated from the $y$'s may be more desirable for some circuit applications. In particular, the short-circuit current gain alpha, defined as the ratio of output (collector) to input (emitter) current when the output is short-circuited, is a most useful and fundamental transistor parameter.[1,2,3] This paper describes a new approximation to the frequency dependence of alpha which is valid somewhat beyond the alpha cutoff frequency. The approximation in the time domain is shown to be good, except for small values of time.

## Previous Work

The fundamental equation, from which stems the quantitative treatment of a junction transistor, is the continuity equation for minority carriers in the base region, which is analogous to Bernouilli's equation for laminar fluid flow. For hole minority carriers in the base region of $p$-$n$-$p$ transistor, the continuity equation can be written:

$$\frac{\partial p}{\partial t} = \frac{p_n - p}{\tau_p} + D_p \frac{\partial^2 p}{\partial x^2}, \tag{1}$$

where $p$ is the instantaneous hole density in the $n$-type base, $p_n$ is the equilibrium hole density in the base, and $\tau_p$ and $D_p$ are the lifetime and the diffusion constant, respectively, for holes in the base region.

The solution of (1) for $p$ as a function of $x$ and $t$ requires two constants, which are obtained from a knowledge of the boundary conditions at the emitter and collector junctions. The collector boundary condition takes into account the variation of base width with collector voltage. From the resulting equation for $p$, the hole currents at the emitter and collector junctions may be found, which lead to expressions for four admittances for small ac signals superimposed on suitable biases. The transistor (excluding base resistance and collector space-charge capacitance) can be represented as the intrinsic transistor in Fig. 1, where equations for the four $y$'s have been given by Early.[4] These are in terms of physical parameters of the transistor materials, the bias currents, and hyperbolic functions of frequency.

The forward short-circuit current gain $\alpha$ is given by

$$\alpha = -\frac{y_{21}}{y_{11}}. \tag{2}$$

* Received by the PGED, April, 1955. This work was performed at Stanford University, and supported in part by the U. S. Army (Signal Corps), the U. S. Air Force, and the U. S. Navy (Office of Naval Research) under Contract N6onr25107.

† Elec. Eng. Dept., California Inst. of Tech., Pasadena, Calif.
‡ Electronics Research Laboratory, Stanford University, Stanford, Calif.

[1] W. Shockley, M. Sparks, and G. K. Teal, "P-n junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; July, 1951.
[2] R. L. Pritchard, "Frequency variations of current amplification factor for junction transistors," PROC. IRE, vol. 40, pp. 1476–1481; November, 1952.
[3] E. L. Steele, "Theory of alpha for p-n-p diffused junction transistors," PROC. IRE, vol. 40, pp. 1424–1428; November, 1952.
[4] J. M. Early, "Design theory of junction transistors," *Bell Syst. Tech. J.*, vol. 32, pp. 1271–1312; November, 1953.

If the electron currents are neglected, which is equivalent to taking the emitter efficiency and the collector multiplication factor each equal to unity, (2) may be written as[1]

$$\alpha = \operatorname{sech}(cw_0/L_p), \tag{3}$$

where $w_0$ is the base width, and $L_p$ is the diffusion length for hole minority carriers in the base region. At zero frequency $c = 1$ and $\alpha = \alpha_0$ ; therefore

$$\alpha_0 = \operatorname{sech}(w_0/L_p). \tag{4}$$

Eq. (3) is valid for all frequencies. In order to obtain a simple rational function representation, some approximation is necessary. One procedure[4] has been to take the first two terms in the McLaurin expansion of the hyperbolic sine and cosine functions: this led to simplified expressions for the four admittances and for alpha, which is

$$\alpha = \frac{\alpha_0}{1 + j\omega/\omega_\alpha}, \tag{5}$$

where

$$\omega_\alpha = \frac{2D_p}{w_0^2}. \tag{6}$$

This expression is valid only for $\omega/\omega_\alpha \ll 1$. Comparison of (5) with (3) indicates that there is considerable discrepancy in the phase angle of alpha at the 3-db point: (5) gives a phase angle of 45°, whereas the correct value is approximately 57°.

It should be pointed out that the value of $\omega_\alpha$ given by (6) is not the actual measured 3-db-down point. This has been discussed by Pritchard,[2] and is mentioned by Early.[4] A modified form of (6) is given later in this paper. If the modified form were used in (5), the phase angle of alpha would agree better with the theoretical values, whereas the approximation for the magnitude would be worse.

In the following section the derivation of a new approximation for alpha is described, which is valid up to and beyond the alpha cutoff frequency.

## Form of the Approximation for Alpha

The equations for the two pertinent short-circuit admittances of the transistor may be written in the form[4]

$$y_{11} = Pc \coth(cw_0/L_p) \tag{7}$$

$$y_{21} = -Pc \operatorname{cosech}(cw_0/L_p) \tag{8}$$

$$c^2 = 1 + j\omega\tau_p$$

where the nonfrequency-varying parameters are included in the constant $P$. It is required to find approximations to these expressions so that the approximation to alpha, as given by (2), is satisfactory up to the alpha cutoff frequency.

Consider the function of frequency $c\coth(cw_0/L_p)$. Expansion of the hyperbolic function gives

$$c\coth(cw_0L_p) = \coth(w_0/L_p)[1 + (1/3)\tau_p(w_0/L_p)^2 j\omega - \cdots], \tag{9}$$

and similarly

$$c\operatorname{cosech}(cw_0/L_p) = \operatorname{cosech}(w_0/L_p) \cdot [1 - (1/6)\tau_p(w_0/L_p)^2 j\omega + \cdots]. \tag{10}$$

Eqs. (9) and (10) can be written

$$c\coth(cw_0/L_p) \approx [\coth(w_0/L_p)](1 + j\omega T) \tag{11}$$

$$c\operatorname{cosech}(cw_0/L_p) \approx [\operatorname{cosech}(w_0/L_p)](1 - j\omega mT). \tag{12}$$

and from (2)

$$\alpha = \operatorname{sech}(w_0/L_p)\frac{1 - j\omega mT}{1 + j\omega T} = \alpha_0\frac{1 - j\omega mT}{1 + j\omega T}. \tag{13}$$

Now (11) and (12) are valid only for $j\omega T \ll 1$ and $j\omega mT \ll 1$, which in effect limits the validity to frequencies much below the alpha cutoff frequency. However, if some arbitrary values are assigned to $m$ and $T$ so that (13) is correct at the alpha cutoff frequency, then it may be expected that the approximation will be reasonable at frequencies up to $\omega_\alpha$, although of course at low frequencies the error will be greater than if $m$ and $T$ had the values given by the hyperbolic expansions in (9) and (10). It is in this step that the present work differs from previous work. In practical applications, $\alpha_0$ and $\omega_\alpha$ are measured quantities, hence it is desired to find $m$ and $T$ as functions of $\alpha_0$ and $\omega_\alpha$. This is done by matching, at the 3-db-down point, the magnitude and phase of the approximate expression for alpha (13) with these quantities given by the accurate expression for alpha in (3). Thus the approximate expression is required to be identical with the theoretical expression at $\omega = 0$ and at $\omega = \omega_\alpha$.

The method for finding $m$ and $T$ was as follows: an explicit expression for the magnitude of alpha, obtained from the accurate equation (3), was equated to $\alpha_0/\sqrt{2}$ at the 3-db-down frequency $\omega_\alpha$. The resulting equation expressed a relation between $\alpha_0$ and the parameter $\omega_\alpha\tau_p$. An IBM Model II Card-Programmed Calculator was employed to obtain numerical solutions of this equation, which gave a table of values of $\omega_\alpha\tau_p$ as a function of $\alpha_0$. At the same time, values of the phase angle $\phi_\alpha$ at the frequency $\omega_\alpha$ were obtained as a function of $\alpha_0$.

From the approximate equation for alpha (13), explicit expressions for the magnitude and for the phase of $\alpha$ were obtained. Equating the magnitude of alpha to $\alpha_0/\sqrt{2}$ at the frequency $\omega_\alpha$ gave one relation between $m$ and $T$; equating the phase angle at $\omega_\alpha$ with that obtained above gave another relation, and thus enabled values of $m$ and of $\omega_\alpha T$ to be obtained as functions of $\alpha_0$. These results are plotted in Fig. 2. It is seen that both $m$ and $\omega_\alpha T$ vary but slightly over the range $\alpha_0 = 0.80$ to 1.00.

It was pointed out above that the value of $\omega_\alpha$ given by (6) was not the measured 3-db-down frequency. In order to obviate this inconsistency, (6) should be replaced by

$$\omega_\alpha = \kappa\frac{D_p}{w_0^2}, \tag{14}$$

where $\kappa$ is a function of $\alpha_0$. Substitution of $D_p = L_p^2/\tau_p$ into (14) gives

$$\kappa = \omega_\alpha\tau_p(w_0/L_p)^2. \tag{15}$$

The number $\kappa$ relates the alpha cutoff frequency to the physical parameters of the transistor.

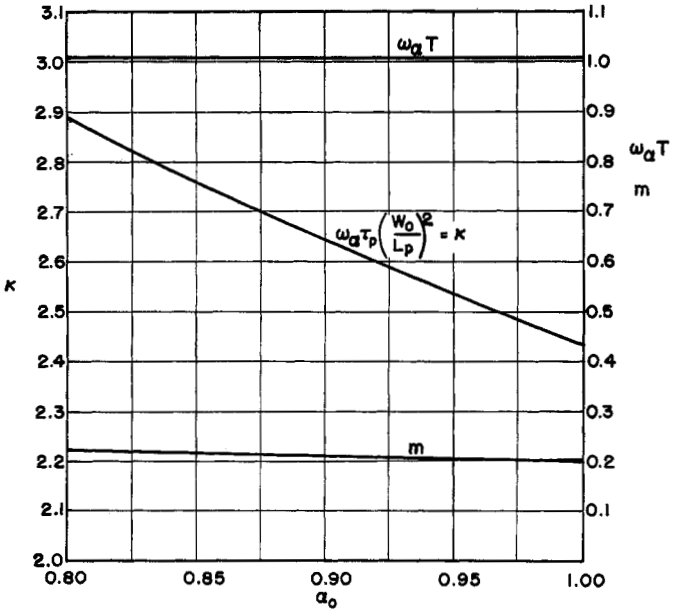


Fig. 2—Values of the parameters $m$, $\omega_\alpha T$, and $\kappa$ as functions of $\alpha_0$.

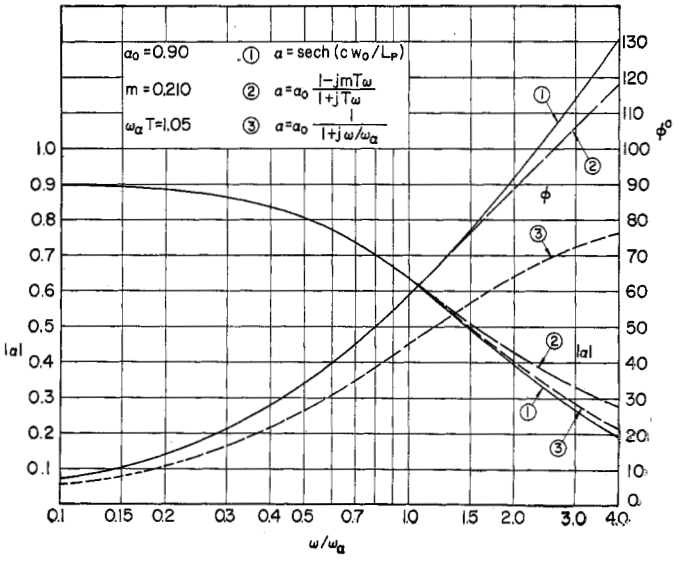


Fig. 4—Comparison of the magnitude and phase of $\alpha$ for $\alpha_0 = 0.90$, as given by the correct expression, the new approximation, and the earlier approximation.

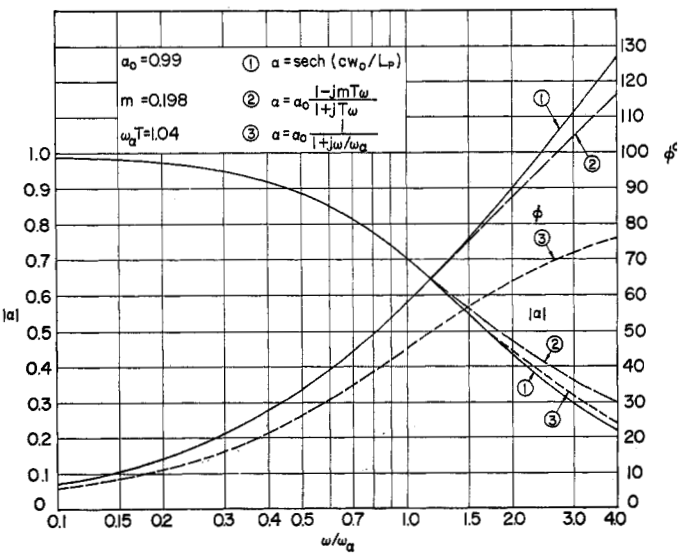


Fig. 3—Comparison of the magnitude and phase of $\alpha$ for $\alpha_0 = 0.99$, as given by the correct expression, the new approximation, and the earlier approximation.

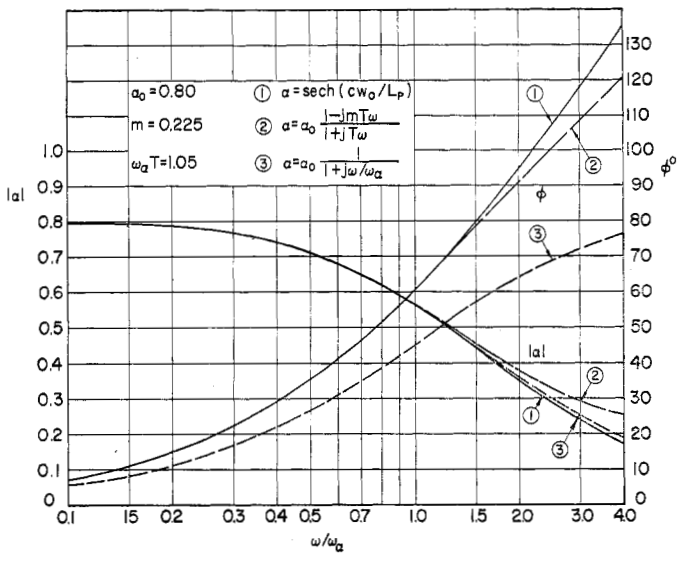


Fig. 5—Comparison of the magnitude and phase of $\alpha$ for $\alpha_0 = 0.80$, as given by the correct expression, the new approximation, and the earlier approximation.

As described above, values of $\omega_\alpha \tau_p$ for different values of $\alpha_0$ = sech $(w_0/L_p)$ were obtained incidentally in the derivation of $m$ and $\omega_\alpha T$, and hence $\kappa$ may be found as a function of $\alpha_0$. Although not necessary for the present work, $\kappa$ is shown plotted in Fig. 2, since previous workers have treated this parameter.[2,3,5] It is seen that $\kappa$ varies considerably over the range $\alpha_0$ = 0.80 to 1.00. Some values of $\kappa$ which agree with those given in Fig. 2 have been obtained by Haneman.[5]

## Numerical Results

To verify the results for the approximate expression for alpha, the magnitude and phase angle of alpha given by (13) are compared in Figs. 3, 4 and 5 with those given by (3), for three values of $\alpha_0$. In plotting the approximate expression, values of $m$ and $\omega_\alpha T$ for each value of $\alpha_0$ were obtained from Fig. 2. Plots of (5) are also given for comparison. The new approximation is indistinguishable from the correct values at all frequencies up to the alpha cutoff frequency, although the magnitude of alpha deviates somewhat more than does the earlier approximation at higher frequencies. The phase angle is approximated much better over the entire range by the new formula than by the earlier approximation.

As mentioned above, $m$ and $\omega_\alpha T$ are practically constant for $\alpha_0$ in the range 0.80 to 1.00. Since practical values of $\alpha_0$ lie between 0.90 and 0.98, little error will be introduced by taking average values of $m$ and $\omega_\alpha T$ in this range. These are found to be $m$ = 0.205, $\omega_\alpha T$ = 1.04.

[5] D. Haneman, "Expression for the 'Alpha cut-off' frequency in junction transistors," Proc. IRE, vol. 42, p. 1808; December, 1954.

In practice, therefore, it is not even necessary to obtain $m$ and $\omega_\alpha T$ for different values of $\alpha_0$ from Fig. 2. One equation for alpha as a function of $\alpha_0$ and $\omega_\alpha$ will suffice:

$$\alpha = \alpha_0 \frac{1 - j(0.205 \times 1.04)(\omega/\omega_\alpha)}{1 + j(1.04)(\omega/\omega_\alpha)} \tag{16}$$

$$= \alpha_0 \frac{1 - j0.214\, \omega/\omega_\alpha}{1 + j1.04\, \omega/\omega_\alpha}. \tag{17}$$

Eq. (17) has a phase angle of 58.3° at the 3-db-down point, compared with values of 59.9° to 57.9° for $\alpha_0 = 0.80$ to $\alpha_0 = 0.99$, obtained from (3).

To indicate the error introduced by taking average values of $m$ and of $\omega_\alpha T$, (17) (for $\alpha_0 = 0.90$ to 0.98) is compared in Fig. 6 with (3) (for $\alpha_0 = 0.80$), for which case the error is greatest. It is seen that the agreement is still almost as good as when the correct values of $m$ and $\omega_\alpha T$ are taken (Fig. 5). In fact, if (17) were plotted together with (13) for the correct values of $m$ and $\omega_\alpha T$ for $\alpha_0 = 0.80$, the two sets of curves would be almost indistinguishable.

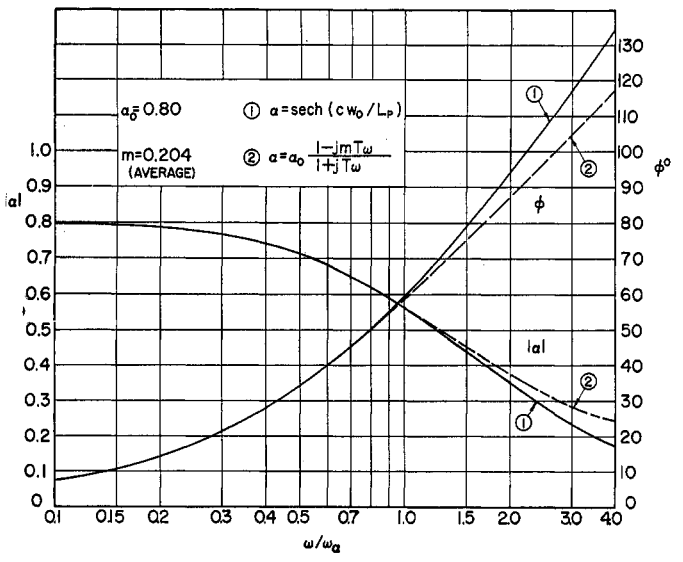


Fig. 6—Comparison of the magnitude and phase of $\alpha$ as given by the correct expression, with those given by the new approximation using average values of $m$ and $\omega_\alpha T$.

The average values of $m$ and $\omega_\alpha T$ obtained above may be inserted in (11) and (12) to obtain approximations for $y_{11}$ and $y_{21}$. However, although the ratio of these is a good approximation to alpha, they are individually rather poor approximations; hence if expressions for these quantities are required, it is desirable to treat them separately.

It should be pointed out that $\alpha_0$ and $\omega_\alpha$ are obtained by measurement of the short-circuit current gain of the complete transistor. This commonly differs slightly from the alpha of the intrinsic transistor due to the effects of collector space-charge layer capacitance and base spreading resistance. The principal effect is to lower the apparent cutoff frequency, as discussed by Pritchard.[6] Thus $\omega_\alpha$ may be somewhat higher than the measured current gain cutoff frequency.

## Transient Response of Alpha

It is instructive to compare in the time domain the foregoing approximation with the theoretical expression for alpha. By the transient response of alpha is meant the form of the short-circuit collector current in response to a unit step function in time of the emitter current. It is assumed that this step function is small so that the transistor remains in the linear small-signal region. In terms of the complex frequency variable $s$, the theoretical expression (from 3) is:

$$\alpha(s) = i_c(s) = \frac{1}{s} \text{sech}\, [(1 + s\tau_p)^{1/2} w_0/L_p]. \tag{18}$$

The time response is found by taking the inverse Laplace transform:

$$\alpha(t) = \mathcal{L}^{-1}\left\{\frac{1}{s} \text{sech}\, [(1 + s\tau_p)^{1/2} w_0/L_p]\right\}. \tag{19}$$

As $t \to \infty$, $\alpha$ approaches the final value

$$\alpha_0 = \text{sech}\, (w_0/L_p). \tag{4}$$

It is convenient to compute the time response in terms of the 3-db-down frequency $\omega_\alpha$ given by (14):

$$\omega_\alpha = \kappa \frac{D_p}{w_0^2}, \tag{14}$$

where $\kappa$ is a slowly varying function of $(w_0/L_p)$, or of $\alpha_0$ as shown in Fig. 2. The transient response can then be written:

$$\alpha(t) = \mathcal{L}^{-1}\left\{\frac{1}{s} \text{sech}\left[\left(\frac{w_0}{L_p}\right)^2 + \frac{\kappa s}{\omega_\alpha}\right]^{1/2}\right\}. \tag{20}$$

Upon taking the inverse transform[7,8] there results:

$$\alpha(t) = \sum_{n=1}^{\infty} \left\{ \exp\left[\frac{-(2n-1)w_0}{L_p}\right] \cdot \text{erfc}\left[\frac{2n-1}{2}\left(\frac{\kappa}{\omega_\alpha t}\right)^{1/2} - \frac{w_0}{L_p}\left(\frac{\omega_\alpha t}{\kappa}\right)^{1/2}\right] + \exp\left[\frac{(2n-1)w_0}{L_p}\right] \cdot \text{erfc}\left[\frac{2n-1}{2}\left(\frac{\kappa}{\omega_\alpha t}\right)^{1/2} + \frac{w_0}{L_p}\left(\frac{\omega_\alpha t}{\kappa}\right)^{1/2}\right]\right\}[(-1)^{n+1}], \tag{21}$$

where

$$\text{erfc}\, x = 1 - \text{erf}\, x = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy.$$

The error integral erf $x$ is a tabulated function.[9]

[6] R. L. Pritchard, "Frequency variations of junction transistor parameters," Proc. IRE, vol. 42, pp. 786–799; May, 1954.

[7] J. S. Schaffner and J. J. Suran, "Transient response of the grounded base transistor amplifier with small load impedance," *Jour. Appl. Phys.*, vol. 24, pp. 1355–1357; November, 1953.

[8] R. F. Shea *et al.*, "Principles of Transistor Circuits," John Wiley and Sons, Inc., New York, 1953.

[9] E. Jahnke and F. Emde, "Tables of Functions," Dover Publications, New York, 1945.

It should be noted that the factors $(w_0/L_p)$ and $\kappa$ depend upon $\alpha_0$ through (4) and the curve given in Fig. 2. Thus the shape of the response (21) will be expected to depend on $\alpha_0$, although as will be seen, the dependence is slight.

Eq. (21) was computed as a function of $\omega_\alpha t$ for two values of $\alpha_0$, 0.80 and 0.99. The series was found to converge quite rapidly, not more than three terms being required for values of $\omega_\alpha t$ up to 4. Results are shown in Fig. 7 where it can be seen that curves are very nearly same.

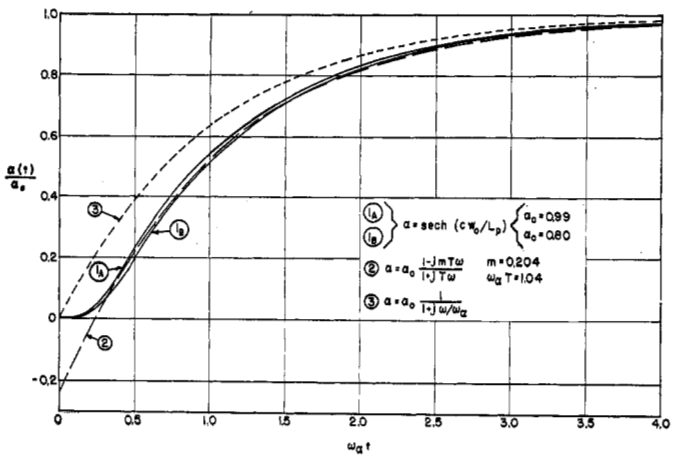


Fig. 7—Comparison of the transient response of $\alpha$ for $\alpha_0 = 0.80$ and for $\alpha_0 = 0.99$, as given by the correct expression, the new approximation and the earlier approximation.

The response has the character of a delay, representing the transit time of the minority carriers across the base region, followed by an approximately exponential rise, representing dispersion of the carriers due to their random diffusive motion. These two properties of delay and dispersion may be treated separately, leading to an approximation for alpha of the form[10]

$$\frac{\alpha(j\omega)}{\alpha_0} = \frac{e^{-j\omega mT}}{1 + j\omega T}, \tag{22}$$

giving a constant delay $mT$ followed by an exponential rise of time constant $T$. From the exact response it is seen that $T$ should be of the order of $1/\omega_\alpha$ and $m$ about 0.2. A more convenient expression results if the first two terms in the series expansion for the exponential are taken:

$$\frac{\alpha(j\omega)}{\alpha_0} = \frac{1 - j\omega mT}{1 + j\omega T}. \tag{23}$$

[10] W. F. Chow and J. J. Suran, "Transient analysis of junction transistor amplifiers," Proc. IRE, vol. 41, pp. 1125–1129; September, 1953.

This is the same form as considered previously in this paper, where by matching the theoretical expression in magnitude and phase at the 3-db-down point, the average values $\omega_\alpha T = 1.04$ and $m = 0.204$ were obtained. The transient response of (23) for these values is:

$$\frac{\alpha(t)}{\alpha_0} = 1 - 1.204e^{-\omega_\alpha t/1.04} \tag{24}$$

which is shown in Fig. 7. The agreement with the theoretical response is excellent, except for the initial negative portion in the region up to $\omega_\alpha t = 0.193$. This corresponds in the frequency domain to frequencies well above the alpha cutoff frequency $\omega_\alpha$, where the approximation is not intended to be valid. A better approximation in the initial region is obtained by defining $\alpha(t)$ to be zero until $\omega_\alpha t = 0.193$, after which time it is given by (24). This is approximately equivalent to (22), although the form (23) is more accurate in its useful region (frequencies up to alpha cutoff) and considerably more convenient to use.

Also shown in Fig. 7 is the response

$$\frac{\alpha(t)}{\alpha_0} = 1 - e^{-\omega_\alpha t} \tag{25}$$

corresponding to the approximation

$$\frac{\alpha(j\omega)}{\alpha_0} = \frac{1}{1 + j\,\omega/\omega_\alpha}, \tag{26}$$

mentioned in (5). The agreement with the theoretical response is rather poor, the difference taking the form of an approximately constant time delay. This comes about because the phase of (26) approaches only 90° at high frequencies, whereas the phase of the theoretical expression (3) continually increases with frequency, the difference between the two being very roughly an angle proportional to frequency, as expressed by (22).

## Conclusion

A new simple approximation for the frequency dependence of alpha has been obtained in terms of measurable quantities on a completed transistor. The approximation is valid beyond the alpha cutoff frequency and for all practical values of the low frequency alpha. It was shown that the transient response of alpha as given by the new approximation agrees well with the theoretical result within expected time limits.

## Acknowledgment

The authors wish to express their thanks to Dr. J. M. Pettit, under whose direction this work was conducted, and to I. W. Yabroff and L. Franks, who programmed the numerical work for the IBM computer.