Solution of a Transistor Transient Response Problem*

J. R. MACDONALD†

Introduction

GOOD TRANSIENT response is desirable in many high-frequency and switching applications of transistors. A single transient response measurement can, in principle, yield frequency response information over a range of frequencies of width determined by the short-time resolution of the transient response measurement and the duration of the measurement. Further, measurement of the response to a unit step or unit impulse (delta function) of voltage or current applied at the input of the device allows its response to any other driving curve-shape to be calculated.[1]

Calculation of transistor transient response is complicated by the fact that a transistor is a distributed active element requiring an infinite number of passive elements (resistances and capacitances) and a finite number of current and/or voltage generators for its complete description by means of an equivalent circuit. Even transistor equivalent circuits having a finite number of elements can be excessively complicated.[2] The situation is further complicated because transistors are bilateral instead of unilateral elements; hence, changes in the output circuit are reflected in the input and *vice versa*.

The present work was undertaken to investigate the utility and practicality of calculating transistor transient response by bypassing both the equivalent-circuit transistor representation with an infinite number of elements or an approximate form of it with a finite number of elements. Transient response is calculated by dealing directly with the meromorphic functions arising from the solution of the small-signal linearized differential equations pertaining to transistor operation.[3] Moll has treated the large-signal transient response of junction transistors by approximate methods of accuracy sufficient for many purposes.[4] In addition, he has calculated the small-signal transient response by making use of the usual approximate expression for the short-circuit current transfer ratio α. For a single mode of operation, the present work shows to what order of accuracy this approximation is valid and also yields more accurate results suitable for a quick and approximate oscilloscope determination of transistor frequency response and material constants by means of transient response measurements.

In his work, Moll has taken transistor emitter efficiency, γ, as unity and neglected its frequency response. Although this approximation is not necessary, it is usually a good one and simplifies the analysis greatly. Therefore, it will be made here as well. Then, the current transfer ratio α may be written for a p-n-p junction transistor as

$$\alpha(p) = \beta(p) = \operatorname{sech}\left[\frac{W}{L_b}(1 + p\tau_b)^{1/2}\right], \qquad (1)$$

where W is the base width, τ_b the lifetime of holes in the base, $L_b = (D_p\tau_b)^{1/2}$ is the base diffusion length, and D_p is the hole diffusion constant in the base. The quantity p is the complex frequency variable $\sigma + i\omega$, where σ is a constant. When the argument of the hyperbolic function is small compared to unity, $\alpha(p)$ becomes, to first order

$$\alpha(p) \cong \frac{(1 + W^2/2L_b^2)^{-1}}{1 + p(2D_p/W^2 + \tau_b^{-1})^{-1}} = \frac{\alpha_0}{1 + p/\omega_c}. \qquad (1')$$

This approximation yields $\omega_c = 2D_p/W^2 + \tau_b^{-1}$ for the α-cutoff frequency; better approximations for this quantity have been discussed by Rittner[5] and Pritchard.[6,7] Eq. (1′) was used by Moll in his small-signal transient response analysis.

Transient Response Calculation

For simplicity's sake and because it is sufficient to illustrate the method, we shall consider only grounded base operation of a junction transistor. Taking $\gamma = 1$ and assuming with Moll that $(r_c^{-1} + \omega_c C_c)(r_b + R_c) \ll 1$, the current transfer ratio h_{21} reduces to its short-circuit value $-\alpha$, with α given by (1). If ω_c is replaced by ω in the above inequality, it becomes obvious that it cannot hold at arbitrarily high frequencies. Hence, taking h_{21} as $-\alpha$ for all frequencies is equivalent to neglecting the effect of the collector capacitance C_c at the very high frequencies, say ω_∞ (assumed much higher than ω_c), where it becomes important. Such neglect causes the transient response to be in error in the region of very short times comparable to or less than ω_∞^{-1}. Again, although it would be possible to take such quantities as r_b and C_c into account explicitly, the analysis would be considerably complicated without

* Manuscript received by the PGCT, December 8, 1955.
† Texas Instruments Inc., Dallas, Texas.

[1] M. F. Gardner and J. L. Barnes, "Transients in Linear Systems," John Wiley & Sons, Inc., New York, N. Y., 1942.

[2] H. Statz, E. A. Guillemin, and R. A. Pucel, "Design considerations of junction transistors at higher frequencies," Proc. IRE, vol. 42, pp. 1620–1628; November, 1954.

[3] W. Shockley, M. Sparks, and G. K. Teal, "*P-n* junction transistors," *Phys. Rev.*, vol. 83, pp. 151–162; July 1, 1951.

[4] J. L. Moll, "Large-signal transient response of junction transistors," Proc. IRE, vol. 42, pp. 1773–1784; December, 1954.

[5] E. S. Rittner, "Extension of the theory of the junction transistor," *Phys. Rev.*, vol. 94, pp. 1161–1171; June 1, 1954.

[6] R. L. Pritchard, "Frequency variations of current-amplification factor for junction transistors," Proc. IRE, vol. 40, pp. 1476–1481; November, 1952.

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

an important gain in accuracy in the most significant transient response region defined by $(5\omega_c)^{-1} < t < (0.3\omega_c)^{-1}$. Here, therefore, we shall only be concerned with response of a system governed by α alone. Using Pritchard's more exact results for h_{21}, more accurate transient response calculations may be carried out when required.

If the transient response of a system to a unit impulse or a unit step applied at its input is known, its response to any other input can be calculated.[8] Although either of these responses is sufficient for the above purpose, they emphasize different regions of the time scale differently, and it is therefore worthwhile to calculate them both. We require the change in output current, $B(t)$, from the transistor produced by a unit impulse (unity charge) of current applied at $t = 0$, or the change in current, $A(t)$, arising from a unit step of current applied at the input at $t = 0$. As is well known, $B(t)$ and $A(t)$ are given by the inverse Laplace transforms of $-\alpha(p)$ and $-\alpha(p)/p$, respectively. For convenience sake, we shall suppress the minus sign.

To yield a basis for comparison, we wish to obtain $A(t)$ and $B(t)$ for both the exact $\alpha(p)$ of (1) and the approximate form (1′). The results for the latter are readily found from tables[1] to be

$$B_a(t) = (2D_p/W^2)e^{-(2W/\pi L_b)^2\tau}e^{-8\tau/\pi^2}, \qquad (2)$$

$$A_a(t) = [1 + W^2/2L_b^2]^{-1}\{1 - e^{-(2W/\pi L_b)^2\tau}e^{-8\tau/\pi^2}\}, \qquad (3)$$

where we have introduced the new time variable $\tau = D_p(\pi/2W)^2 t$ for later convenience. It will be noted that the ratio of the two exponential time constants in the above equations is $W^2/2L_b^2$, which will be considerably less than unity for a good transistor. For many practical purposes, the first exponential term may therefore be set equal to unity.

Calculation of $B(t)$ and $A(t)$ from the exact expression for $\alpha(p)$ is somewhat more complicated but follows the usual treatment of a meromorphic function.[9] The function, either $\alpha(p)$ or $p^{-1}\alpha(p)$, is first written as an infinite product in terms of the roots of its denominator. Then, the infinite product is rewritten in terms of an infinite series of partial fractions involving these roots. The inverse Laplace transform of each of these fractions is easily carried out, yielding the final result as an infinite series of exponentials. For the present case, we find

$$B_e(t) = (\pi D_p/W^2)e^{-(2W/\pi L_b)^2\tau}\sum_0^\infty (-1)^n(2n+1)e^{-(2n+1)^2\tau}, \qquad (4)$$

$$A_e(t) = \operatorname{sech}(W/L_b) - (4/\pi)e^{-(2W/\pi L_b)^2\tau}\sum_0^\infty \frac{(-1)^n(2n+1)e^{-(2n+1)^2\tau}}{\left(\frac{2W}{\pi L_b}\right)^2 + (2n+1)^2}. \qquad (5)$$

These results may also be expressed in terms of Jacobian theta functions[10] and evaluated from tables when available. Since the series are very rapidly convergent even for $\tau = 0.1$, it is easiest to sum them directly, however.

Before comparing (2), (4), (3), and (5) in detail, several facts are immediately obvious. First, $B_a(0) = 2D_p/W^2$, while $B_e(0)$, the exact result, is zero. Second, $\lim \tau \to \infty$ $[B_e(\tau)/B_a(\tau)] = (\pi/2)e^{-0.1894\tau}$. Thus, the final decay of the exact impulse response is somewhat more rapid than that of the approximate response. Next, $A_e(0)$ equals $A_a(0)$, since the series in (5) with $\tau = 0$ may be shown to sum to sech (W/L_b). Finally, $A_e(\infty)$ and $A_a(\infty)$ are somewhat unequal unless $W^2/2L_b^2$ is vanishingly small.

It may also be noted that the complete zero-order exponential decay constant occurring in A_e and B_e is $(\tau/t)[1 + (2W/\pi L_b)^2] = \tau_b^{-1} + (\pi/2)^2D_p/W^2$. This quantity is the exact zero-order α-cut-off radial frequency ω_c and may be compared to the previous approximate result where 2 appeared instead of $(\pi/2)^2 = 2.4674$. With the omission of the τ_b^{-1} term, Pritchard and Rittner obtained by a different method approximately the present result, but with 2.434 instead of 2.4674. The exact zero-order (lowest) α cut-off frequency f_c^0 may now be written as

$$f_c^0 = \omega_c^0/2\pi = (2\pi\tau_b)^{-1} + \pi D_p/8W^2 = 0.1592\tau_b^{-1} + 0.3927D_p/W^2. \qquad (6)$$

The first term, arising from recombination, will usually be negligible in a high-α_0 transistor.

The above calculation of ω_c by taking it as the smallest exponential decay constant appearing in the transient response is not exactly equivalent to the usual definition of this quantity determined from the relation $\alpha_0/|\alpha(\omega_c)| = \sqrt{2}$. The difference arises from the fact that the transient response actually involves an infinite number of time constants. This is our reason for denoting the above cut-off frequency as the "zero-order" value. It depends on α_0 implicitly through the relation $\alpha_0 = \operatorname{sech}(W/L_b)$. Using this relation, the zero-order α cutoff radial frequency may be written

$$\omega_c^0 = [(\pi/2)^2 + (\operatorname{sech}^{-1}\alpha_0)^2](D_p/W^2). \qquad (7)$$

In this form, its dependence on α_0 is evident. For convenience, the magnitude of the term in square brackets is presented, for three values of α_0, in the third column of Table I. These results may be compared with the similar dependence of the conventional α cutoff frequency on α_0 (or β_0) given by Pritchard.[7]

Fig. 1 is a log-log comparison of the exact and approximate transient response for the limiting case $\alpha_0 = 1$. The approximate curves are those dotted. The impulse response curves are normalized as shown. It will be noted that the deviations between the exact and approximate curves are largest for small values of τ. Figs. 2 and 3 present the

[8] Gardner and Barnes, *op. cit.*, pp. 234, 262.
[9] Gardner and Barnes, *op. cit.*, p. 241.
[10] E. T. Wittaker and G. N. Watson, "A Course of Modern Analysis," Cambridge University Press, Cambridge, England, 1935.

The Total Instantaneous Power Output $\Psi(t;0)$ [*See* (55)]

Setting $\tau=0$ in (14) gives, in the notation of this example

$$\Psi(t;0) = 4\int_0^t \phi_\alpha(t-x;x)\psi(x)dx. \quad (105)$$

Substituting (103) and (104) in (105) we find, on performing the integration,

$$\Psi(t;0) = \frac{w_{00}}{4}\alpha\beta\left\{\frac{1}{\alpha+\beta}\left[1-e^{-(\alpha+\beta)t}\right] + \frac{e^{-2\alpha t}}{\alpha-\beta}\left[1-e^{-(\beta-\alpha)t}\right]\right\} \quad (106)$$

$$\rightarrow \frac{w_{00}}{4}\cdot\frac{\alpha\beta}{\alpha+\beta} \quad \text{as} \quad t\rightarrow\infty. \quad (107)$$

A Solution of the Integral Equation for $\psi(x)$

In this section we assume a system of the form shown in Fig. 5. Then in line with the previous discussion, we interpret (105) as an integral equation for $\psi(x)$. It is assumed here that $\Psi(t;\ 0)$ is the known function. In practice it is very easily measured. We now proceed to find the exact solution of (105) for a $\phi(t;x)$ appropriate to the system of Fig. 5.

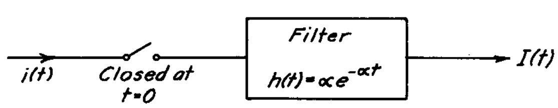

Fig. 5

We write (105) in the form (dropping the subscript α on ϕ)

$$\Psi(t;0) = \int_0^t \psi(x)\left\{4\phi(t-x;x)\right\}dx. \quad (108)$$

Differentiate both sides partially with respect to t and denote partial differentiation by primes; thus

$$\Psi'(t;0) = \int_0^t \psi(x)\left\{4\phi'(t-x;x)\right\}dx \quad (109)$$

$$\text{as} \quad \phi(0;x)=0 \quad \text{by (24)}. \quad (110)$$

But from (104) we have

$$4\phi(t-x;x) = \alpha e^{-\alpha x}\left[1-e^{-2\alpha(t-x)}\right] \quad (111)$$

and thus

$$4\phi'(t-x;x) = \alpha e^{-\alpha x}\cdot 2\alpha e^{-2\alpha(t-x)}. \quad (112)$$

Substituting (112) in (109) gives

$$\Psi'(t;0) = \int_0^t \left\{\alpha e^{-\alpha x}\psi(x)\right\}\cdot\left\{2\alpha e^{-2\alpha(t-x)}\right\}dx. \quad (113)$$

Writing

$$f(x) = \alpha e^{-\alpha x}\psi(x) \quad (114)$$

$$g(x) = 2\alpha e^{-2\alpha x} \quad (115)$$

it is easily seen that (113) may be written in the form of a convolution integral. Thus

$$\Psi'(t;0) = \int_0^t f(x)g(t-x)dx. \quad (116)$$

This equation is readily solved by the Laplace transform method. We define a function $g^*(t)$ such that

$$L_p\{g^*(t)\} = \frac{1}{pL_p\{g(t)\}}\cdot \quad (117)$$

It is then easy to show that the solution of (116) is

$$f(t) = \int_0^t g^*(t-x)\left\{\frac{\partial}{\partial x}+\delta_1(x)\right\}\Psi'(x;0)dx \quad (118)$$

where $\delta_1(x)$ is the one-sided impulse function such that

$$\int_0^\infty \delta_1(x)dx = 1. \quad (119)$$

We may rewrite (118) in the more convenient form

$$f(t) = g^*(t)\Psi'(0;0) + \int_0^t g^*(t-x)\Psi''(x;0)dx. \quad (120)$$

Now using (115) we have:

$$L_p\{g(t)\} = \int_0^t 2\alpha e^{-2\alpha t}\cdot e^{pt}dt = \frac{2\alpha}{p+2\alpha} \quad (121)$$

and thus form (117)

$$L_p\{g^*(t)\} = \frac{p+2\alpha}{2\alpha p} = \frac{1}{2\alpha}+\frac{1}{p}, \quad (122)$$

and hence we find

$$g^*(t) = \frac{1}{2\alpha}\delta_1(t)+1, \quad t>0. \quad (123)$$

Using (123) in (120) we have

$$f(t) = \left\{1+\frac{1}{2\alpha}\delta_1(t)\right\}\Psi'(0,0) + \int_0^t\left\{1+\frac{1}{2\alpha}\delta_1(t-x)\right\}\Psi''(x;0)dx \quad (124)$$

$$= \left\{1+\frac{1}{2\alpha}\delta_1(t)\right\}\Psi'(0,0) + \frac{1}{2\alpha}\Psi''(t;0) + \int_0^t \Psi''(x;0)dx \quad (125)$$

For many purposes, exact fitting of the experimental and theoretical curves is unnecessary. Table I presents some results useful in obtaining approximate values of α cutoff and of the transistor constants W and τ_b. The peak of the $B_e(\tau)$ curve occurs very nearly at $\tau = 0.4$ for all α_0 values in the range from 0.8 to 1.0. Column 4 of Table I presents the normalized peak height of the $B_e(\tau)$ curve for three α_0 values in this range. Intermediate values can readily be obtained to within sufficient accuracy by linear or graphical interpolation. If the strength of the input impulse can be established, the peak height of the output pulse will yield W^2/D_p. The value obtained in this way may be compared to that found from the time at which the pulse reaches its maximum to give both a check and a possible determination of α_0 if this quantity is unknown.

TABLE I

α_0	W/L_b	$\omega_c^0 W^2/D_p$	$W^2/\pi D_p\ B_e(0.4)$	τ_1
1.0	0	2.467	0.589	0.94
0.90	0.467	2.686	0.568	0.89
0.80	0.693	2.948	0.544	0.84

Column 2 of Table I shows how W/L_b depends on α_0. Finally, column 5 is calculated from the implicit relation

$$A(\tau_1)/A(\infty) = A(\tau_1)/\alpha_0 = 1/2.$$

The quantity τ_1 is therefore the normalized time at which the step response reaches half of its final value. Since the half height position is readily measured and is independent of input amplitudes, the corresponding time t_1 may be used to calculate D_p/W^2 from the relation $D_p/W^2 = (2/\pi)^{\frac{1}{2}}(\tau_1/t_1)$. When W^2/D_p is known, the conventional α cut-off radial frequency ($2.434D_p/W^2$ for $\alpha_0 = 1$) may be immediately calculated.

Finally, for the convenience of those who wish to plot the entire $A(\tau)$ or $B(\tau)$ curves for comparison with experiment, we present in Table II the quantities $A_e(\tau)$ and $(W^2/\pi D_p)B_e(\tau)$ in the τ range of interest.

TABLE II

τ	$A_e(\tau)$			$(W^2/\pi D_p)B_e(\tau)$		
	$\alpha_0 = 1$	$\alpha_0 = 0.9$	$\alpha_0 = 0.8$	$\alpha_0 = 1$	$\alpha_0 = 0.9$	$\alpha_0 = 0.8$
0	0	0	0	0	0	0
0.1	8.89×10^{-4}	8.88×10^{-4}	8.87×10^{-4}	.0461	.0458	.0452
0.2	.0260	.0256	.0252	.3561	.3499	.3425
0.3	.0851	.0839	.0823	.5420	.5284	.5106
0.4	.1581	.1542	.1495	.5886	.5681	.5444
0.5	.2325	.2257	.2198	.5730	.5484	.5200
0.7	.3685	.3547	.3389	.4911	.4616	.4285
1.0	.5317	.5062	.4774	.3675	.3364	.3025
2.0	.8277	.7755	.7023	.1353	.1134	.0917
3.0	.9366	.8553	.7704	.0498	.0486	.0278
5.0	.9914	.8949	.7973	.00674	.00433	.00255