Distortion in Complementary-Pair Class-B Amplifiers

In which the author develops, among other things, a new treatment of crossover distortion

By B. M. Oliver

THERE ARE TWO PRINCIPAL TYPES OF DISTORTION in class-B amplifiers using complementary pairs of transistors. One is caused by β difference between the two transistors; the other, known as crossover distortion, is caused by differences in the slope of the transfer characteristic near the operating point as compared with the slope at high current levels. We shall consider both types using a general approach applicable to a wide variety of particular circuit configurations and shall study the effectiveness of negative feedback in suppressing the distortion. In all cases we assume a sinusoidal input and take as the measure of distortion the ratio of total harmonic power to fundamental power in the output.

β-Difference Distortion

Consider the circuit of Fig. 1, in which two transistors having different β's are biased by an appropriate means and have their common emitter node connected to a load resistor, R_2. The source is assumed to have an internal resistance, R_1. There are, of course, little resistances like R_b, the base ohmic resistance, and R_e, the emitter resistance, but for the moment we shall consider these to be included in R_1 and R_2. Since we shall be considering large signals here, we shall neglect the junction-law distortion and assume the junction resistance to be zero for forward bias and infinite for reverse bias.

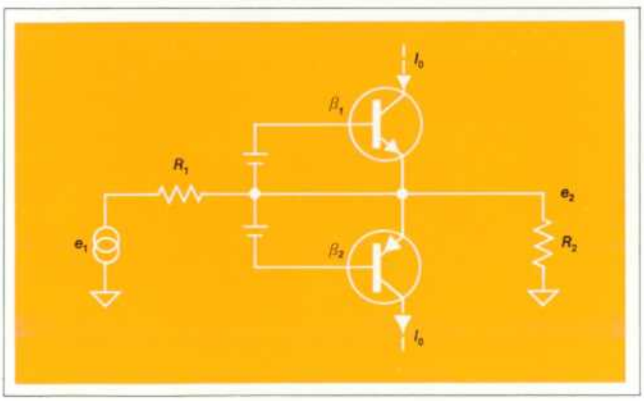

Fig. 1. *Two transistors of different β feed common load.*

For e_1 positive, the transmission is given by

$$k_1 = \frac{e_2}{e_1} = \frac{R_2}{(1-\alpha_1)\,R_1 + R_2} = \frac{\beta_1 R_2}{\alpha_1 R_1 + \beta_1 R_2} \quad (1a)$$

where α_1 is the α of the upper transistor. Similarly for negative e_2,

$$k_2 = \frac{e_2}{e_1} = \frac{R_2}{(1-\alpha_2)\,R_1 + R_2} = \frac{\beta_2 R_2}{\alpha_2 R_1 + \beta_2 R_2}\cdot \quad (1b)$$

The ratio of these transmissions is

$$\frac{k_1}{k_2} = \frac{\beta_1}{\beta_2}\,\frac{\alpha_2 R_1 + \beta_2 R_2}{\alpha_1 R_1 + \beta_1 R_2}. \quad (2)$$

If R_1 is infinite (current source of magnitude $\frac{e_1}{R_1}$) then $\frac{k_1}{k_2} = \frac{\alpha_2\beta_1}{\alpha_1\beta_2} \approx \frac{\beta_1}{\beta_2}$ while if R_1 is much less than βR_2, then $k_1/k_2 \approx 1$. The ratio of R_2/R_1 determines the amount of local feedback or degeneration and affects the ratio of the transmission for positive signals to that for negative signals. β-difference distortion thus depends upon the circuit configuration and will in general be low if the pair is voltage driven and high if the pair is current driven. The point is that, for any configuration, one can always find the appropriate k_1 and k_2 and from these compute two normalized slopes

$$m_1 = \frac{k_1}{(k_1+k_2)/2} \text{ and } m_2 = \frac{k_2}{(k_1+k_2)/2} \quad (3a)$$

which have the properties

$$\frac{m_1+m_2}{2} = 1 \text{ and } \frac{m_1}{m_2} = \frac{k_2}{k_1}. \quad (3b)$$

We are now in a position to represent the actual amplifier, whatever its configuration, by a linear amplifier followed by a non-linear device having the transfer characteristic shown in Fig. 2, where w is the input amplitude and v the output amplitude. We neglect all curvature at the origin, present in actual devices, electing instead to simplify the analysis and obtain a slightly pessimistic result.

In the feedback amplifier of Fig. 3, which incorporates our non-linear device, we have for positive signals

$$v^{+} = \frac{m_1\mu}{1 - m_1\mu\beta} s^{+} \tag{4}$$

and similarly, for negative signals:

$$v^{-} = \frac{m_2\mu}{1 - m_2\mu\beta} s^{-}. \tag{5}$$

$\mu\beta$ is the loop gain if $v = u$ (unity slope). This β is the traditional feedback ratio. Nothing to do with transistors!

If m were unity and we wished an output

$$y = a \sin \phi, \tag{6}$$

we would need an input

$$w = \frac{1 - \mu\beta}{\mu} a \sin \phi; \tag{7}$$

with this input and the actual transfer characteristic, the output will be

$$\begin{aligned} y &= as_1 \sin \phi, \quad y \geq 0 \\ y &= as_2 \sin \phi, \quad y \leq 0 \end{aligned} \tag{8}$$

where

$$s_i = \frac{1 - \mu\beta}{1 - m_i\mu\beta} m_i, \quad i = 1,2. \tag{9}$$

s_1 and s_2 are, of course, the transfer characteristic slopes as modified by the feedback. Both approach unity as $\mu\beta \rightarrow \infty$ unless m_1 or m_2 is zero.

The amplitude of the fundamental component in the output is:

$$a_1 = \frac{\int_{-\pi}^{\pi} y \sin \phi \, d\phi}{\int_{-\pi}^{\pi} \sin^2 \phi \, d\phi} = a\frac{s_1 + s_2}{2}. \tag{10}$$

The mean square departure from the fundamental is

$$\overline{\delta a^2} = \frac{\int_{-\pi}^{\pi} (y - a_1 \sin \phi)^2 d\phi}{\int_{-\pi}^{\pi} d\phi} = \frac{a^2}{8}(s_1 - s_2)^2. \tag{11}$$

The distortion, d, is given by

$$d = \frac{\textit{rms departure}}{\textit{rms fundamental}} \frac{\sqrt{\overline{\delta a^2}}}{\frac{a_1}{\sqrt{2}}} = \frac{|s_1 - s_2|}{s_1 + s_2} \tag{12}$$

$$= \left|\frac{m_1 - m_2}{2}\right| \frac{1}{1 - m_1 m_2 \mu\beta}.$$

The distortion is reduced by feedback but the effective loop gain is not $\mu\beta$ but $m_1 m_2 \mu\beta$. Thus, if either m_1 or m_2 is zero, the feedback is useless. On the other hand, if $m_1 = 1 + \Delta$ and $m_2 = 1 - \Delta$ where $\Delta << 1$, then $m_1 m_2 = 1 - \Delta^2 \approx 1$, and we can consider the normal feedback as being effective. We note that, for the sharp-cornered transfer characteristic we have assumed, the β-difference distortion is independent of amplitude.

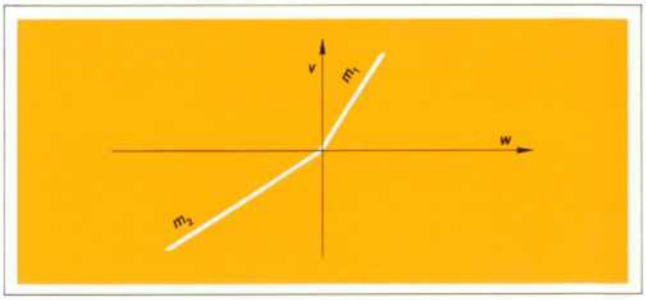

Fig. 2. *Amplifier transfer characteristic, input w vs. output v.*

Crossover Distortion

Fig. 4 shows a typical emitter follower class-B output stage. The transistors are forward biased by a current I flowing through two diodes and a resistor. R_s is the total resistance of this diode string. We include a resistor R_s in the base lead of the upper transistor to keep the circuit symmetrical. Each emitter is coupled to the output through a resistor R_c.

We will first analyze this circuit exactly to see if there is an optimum value of R_c so far as crossover distortion is concerned. Assuming the transistors to be identical so that there is no β-difference distortion, all distortion will arise from the non-linearity of the emitter-base junction law. To find an optimum R_c, we need only minimize the *variation* in output resistance with signal current. Accordingly, we ground the input to the stage and apply a voltage to the output as shown in Fig. 5.

Let R be one-half the ohmic resistance around the emitter base-bias loop, referred to the emitter:

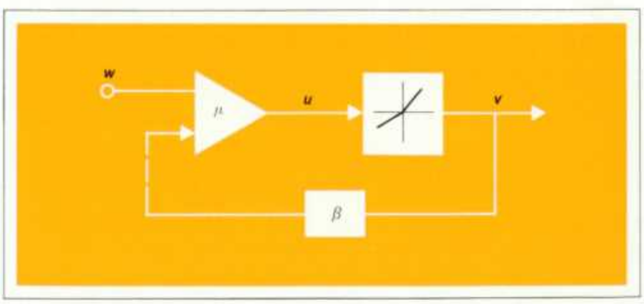

Fig. 3. *Feedback amplifier incorporating nonlinear element.*

$$R = (1 - \alpha)(R_s + R_b) + R_e + R_c. \tag{13}$$

If I_o is the operating current and i_1 and i_2 are the signal currents produced by the voltage e, we then have

$$i = i_2 - i_1 \tag{14}$$

and

$$\frac{kT}{q} ln\left(1 + \frac{I_o + i_1}{I_s}\right) + R\,(I_o + i_1) = V_o - e \tag{15}$$

$$\frac{kT}{q} ln\left(1 + \frac{I_o + i_2}{I_s}\right) + R\,(I_o + i_2) = V_o + e. \tag{16}$$

I_s is the saturation current of each transistor, i and e the output current and voltage. When $i = 0$, $i_1 = i_2 = 0$ and $e = 0$, so

$$V_o = \frac{kT}{q}\, ln\left(1 + \frac{I_o}{I_s}\right) + RI_o. \tag{17}$$

Substitution of this expression into (15) and (16) gives, after multiplication throughout by $\frac{kT}{q}$ and neglecting I_s in the result:

$$ln\left(\frac{i_1}{I_o}\right) + g_o\, R\left(\frac{i_1}{I_o}\right) = -\frac{qe}{kT} \tag{18}$$

$$ln\left(\frac{i_2}{I_o}\right) + g_o\, R\left(\frac{i_2}{I_o}\right) = \frac{qe}{kT} \tag{19}$$

where $g_o = \frac{qI_o}{kT}$ and is the emitter conductance at the operating point.

If we assume a value for $\frac{i_2}{I_o}$, we can easily compute the corresponding value of $\frac{qe}{kT}$. But then to find $\frac{i_1}{I_o}$ from (18) involves the solution of a transcendental equation. No way being known to rewrite (18) explicitly in $\frac{i_1}{I_o}$, the 9100A Calculator was programmed to find this quantity by successive approximations.

Having found $\frac{i_1}{I_o}$, it is then a simple matter to compute the output resistance R_o, or rather the normalized quantity g_oR_o, which consists of the two normalized resistances

$$g_oR_1 = g_oR + \frac{1}{1 + i_1/I_o} \tag{20}$$

and

$$g_oR_2 = g_oR + 1 + \frac{1}{i_2/I_o} \tag{21}$$

in parallel.

As $\frac{i_2}{I_o} \to \infty$, $\frac{i_1}{I_o} \to -1$. Thus $g_oR_o \to g_oR_2 \to g_oR$. The difference, Δg_oR_o, between g_oR_o and its limiting value for large signal currents is therefore

$$\Delta g_oR_o = g_o\,(R_o - R). \tag{22}$$

The 9100A was used not only to solve for $\frac{i_1}{I_o}$, but to compute and plot the 9125A plotter Δg_oR_o as a function

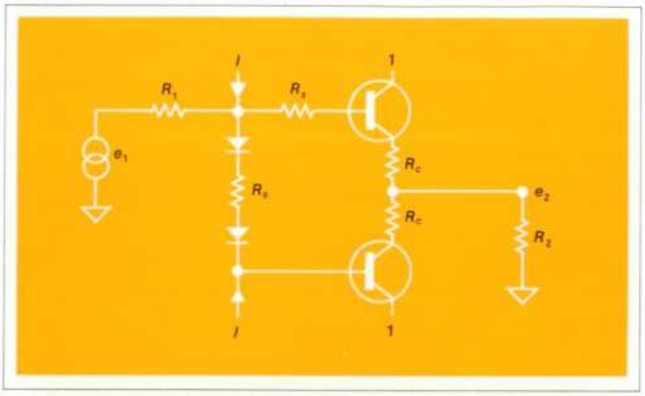

Fig. 4. *Typical emitter follower Class-B stage.*

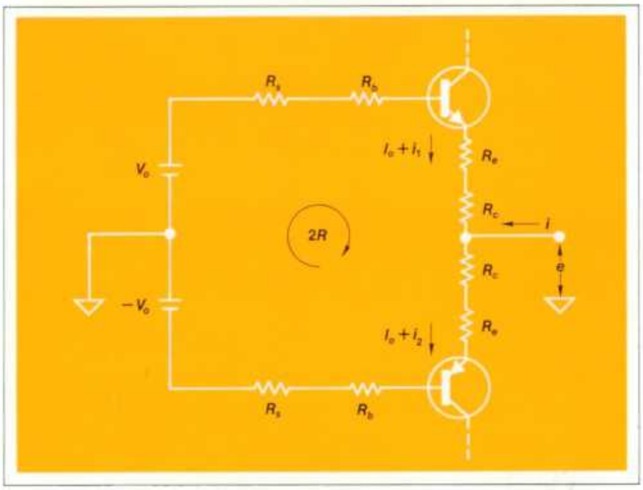

Fig. 5. *To find optimum R_c input is grounded, voltage applied to output.*

of

$$\frac{i}{I_o} = \frac{i_2}{I_o} - \frac{i_1}{I_o}. \tag{23}$$

The results are shown in Fig. 6.

We see that for $g_oR_o = 0$, the resistance falls with increasing signal currents, while for $g_oR > 1$, the resistance rises. With $g_oR = \frac{1}{2}$, the resistance falls monotonically as with $g_oR = 0$, but by only half as much. With $g_oR = 1$, the initial and final values of g_oR_o are the same but there is a bump in between at $\frac{i}{I_o} \approx 4$. Thus a value of R somewhere between $\frac{1}{2g_o}$ and $\frac{1}{g_o}$ is optimum.

However, the drop across this resistance produced by the operating current is only $\frac{I_o}{2g_o} = \frac{1}{2}\frac{kT}{q}$ to $\frac{I_o}{g_o} = \frac{kT}{q}$ or from 13 to 26 millivolts. Over the temperature ranges from 0°C to 100°C, the junction drop of a silicon transistor will change typically by 250 millivolts. Thus, unless the biasing diodes (see Fig. 4) track this change within a few percent, I_o will be very unstable. If the biasing diode were integrated on the same chip with the transistor, accurate enough tracking might be achieved, in which case these results would be of practical interest.

At present the most practical solution to the temperature stability problem appears to be to make R_c many times larger than $\frac{1}{g_o}$ and to rely on negative feedback to reduce the resulting distortion.

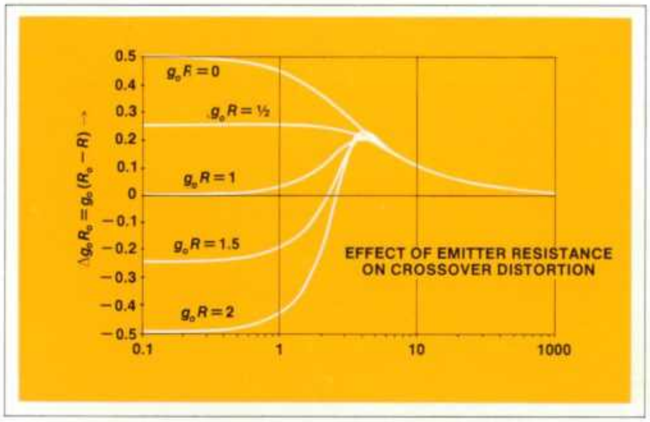

Fig. 6. *Effect of emitter resistance on crossover distortion.*

To avoid wasting signal power in R_c, these resistors may be shunted by diodes as shown in Fig. 7. If the operating current, I_o, produces a voltage drop in R_c greater than half the diode drop required to conduct currents on the order of I_o, then the diode on either side conducts and shunts out its R_c before the opposite transistor cuts off. As a result, the output resistance never rises as in the curves on Fig. 6 for $g_oR > 1$. Rather, neglecting the forward resistance of the diode, it drops by an amount

$$\frac{R_c - R_e - (1-\alpha)(R_b + R_s)}{2}$$

at large signal currents, instead of rising by an amount

$$\frac{R_c + R_e + (1-\alpha)(R_b + R_s)}{2}.$$

Contrary to what one might expect, these power saving diodes do not increase distortion; they may actually reduce it slightly.

When the diode is open, the transmission of the stage, which we will call $m\mu_2$, is

$$m\mu_2 = \left.\frac{e_2}{e_1}\right|_{open}$$

$$= \frac{R_2}{(1-\alpha) + \left(R_1 + \frac{R_s + R_b}{2}\right) + \frac{R_e + R_c}{2} + R_2}. \tag{24}$$

With the diode closed and the opposite transistor not yet open, the resistance of the path through the diode $(1-\alpha)(R_s + R_b) + R_e$, is paralleled by the resistance of the path through the other transistor $(1-\alpha)(R_s + R_b) + R_e + R_c$. Since the latter is ordinarily much larger, we will neglect this intermediate condition and assume the opposite transistor is always open when a diode is closed. The transmission is then

$$\mu_2 = \left.\frac{de_2}{de_1}\right|_{closed} = \frac{R_2}{(1-\alpha)(R_1 + R_s + R_b) + R_e + R_2}. \tag{25}$$

In all these expressions, we have neglected the diode transistor junction resistances. Dividing (24) by (25), we find

$$m = \frac{(1-\alpha)(R_1 + R_s + R_b) + R_e + R_2}{(1-\alpha)\left(R_1 + \frac{R_s + R_b}{2}\right) + \frac{R_e + R_c}{2} + R_2}. \tag{26}$$

One diode or the other closes when the signal voltage at the emitters has the absolute value

$$e_o = \left(1 + \frac{2R_2}{R_c}\right)(V_d - I_oR_c). \tag{27}$$

For higher voltages, the conducting emitter and the output voltage differ by V_d, the diode drop.

Thus we can represent the entire non-linear stage by a linear amplifier of gain μ_2, followed by a non-linear transfer characteristic of slope m up to the breakpoint and slope 1 thereafter. Further, if we agree to normalize all voltages by dividing by e_o, this breakpoint will have the abscissa 1, as shown in Fig. 8.

In Fig. 9, we show this equivalent stage as part of a feedback amplifier whose loop gain (above the breakpoint) is $\mu\beta$. In Figs. 8 and 9

$$w = \frac{e_{in}}{e_o} \tag{28}$$

$$x = \frac{\mu_2 e_1}{e_o} \tag{29}$$

$$y = \frac{e_2}{e_o}. \tag{30}$$

A similar normalization is possible for other circuit configurations, so what follows is generally applicable.

Suppose we wish an output

$$y = a \sin \phi. \tag{31}$$

If the transfer characteristic had unity slope throughout, this would require an input

$$w = \frac{1 - \mu\beta}{\mu} a \sin \phi. \tag{32}$$

With the actual characteristic, the output up to the

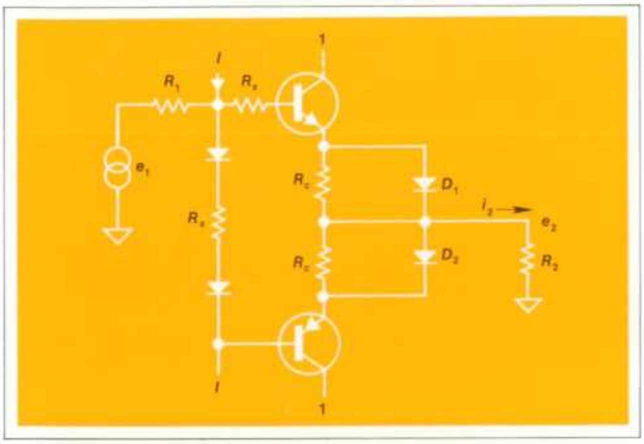

Fig. 7. *Diodes shunt R_c to avoid wasting power in resistors.*

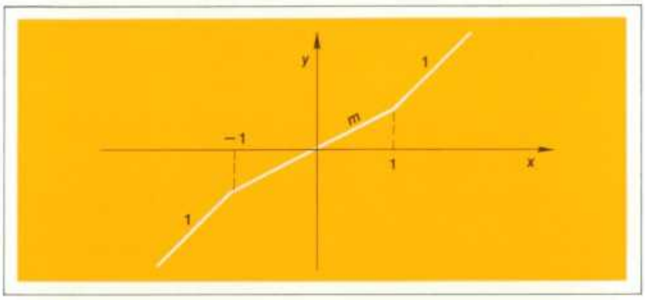

Fig. 8. *Transfer characteristic of entire nonlinear stage.*

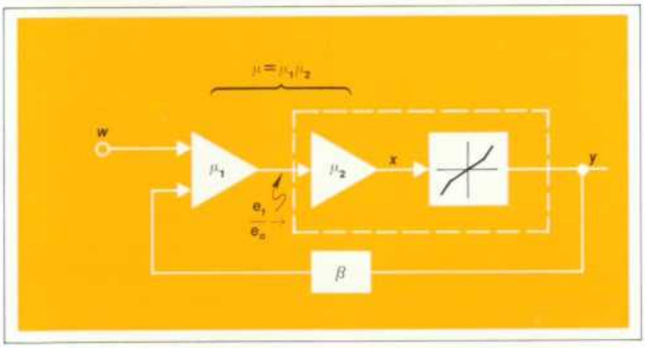

Fig. 9. *Equivalent stage shown as part of an amplifier whose loop gain, above break point, is $\mu\beta$.*

breakpoint is then given by:

$$y = \frac{m\mu}{1-\mu\beta}\, w = sa \sin\phi,\ \phi < \phi_o \qquad (33)$$

where

$$s = m\,\frac{1-\mu\beta}{1-m\mu\beta} \qquad (34)$$

and

$$\phi_o = \sin^{-1}\frac{m}{sa}. \qquad (35)$$

The quantity s will be recognized as the effective slope, as modified by the feedback. Beyond the breakpoint

$$(y - y_o) = \frac{\mu}{1-\mu\beta}\,(w - w_o). \qquad (36)$$

Using (32) and (33), this may be written

$$y = \alpha\,[\sin\phi - (1-s)\sin\phi_o],\ \phi \geq \phi_o. \qquad (37)$$

The fundamental component of the wave described by (33) and (37) is

$$a_1 = \frac{\int_o^{\pi/2} y \sin\phi\, d\phi}{\int_o^{\pi/2} \sin^2\phi\, d\phi} = a\left[1 - (1-s)K\right] \qquad (38)$$

where $$K = \frac{2}{\pi}(\phi_o + \sin\phi_o \cos\phi_o). \qquad (39)$$

The mean square difference between the actual wave and the fundamental is:

$$\overline{\delta a^2} = \frac{\int_o^{\pi/2} (y - a_1)^2\, d\phi}{\int_o^{\pi/2} d\phi} =$$

$$(1-s)^2 a^2 \left\{\frac{2\phi_o}{\pi} + \left(1 - \frac{2\phi_o}{\pi}\right)\sin^2\phi_o - \frac{K(K+1)}{2}\right\}. \qquad (40)$$

Finally, the distortion is

$$d = \sqrt{\frac{\overline{\delta a^2}}{\frac{a_1{}^2}{2}}} = \frac{\sqrt{\frac{4\phi_o}{\pi} + 2\left(1 - \frac{2\phi_o}{\pi}\right) sin^2\phi_o - K(K+1)}}{\left|\frac{1}{1-s} - K\right|}. \qquad (41)$$

Expressions (38) through (41) are valid only if $\phi_o \leq \frac{\pi}{2}$, i.e., if a $\geq \frac{m}{s} = \frac{1-m\mu\beta}{1-\mu\beta}$. For a $\leq \frac{m}{s}$, there is no distortion although the gain is reduced. Of course, if $m = 0$, the output is zero for a $\leq \frac{1}{1-\mu\beta}$ and in a sense the distortion is large.

We note from (41) that $d \to 0$ as $\phi_o \to 0$ (large signals) and as $\phi_o \to \frac{\pi}{2}$ (unless $s = 0$, in which case $d \to \infty$). There is thus some intermediate value of ϕ_o at which the distortion is worst. However, it is not convenient to find this maximum by differentiating (41). Instead, the 9100A Calculator was programmed to plot d as a function of y for various values of m and $\mu\beta$.

Fig. 10 shows the distortion with no feedback and various values of m, while Figs. 11, 12 and 13 show the same values of m, but with 20 dB, 40 dB and 60 dB of feedback, respectively. The importance of having $m > 0$ is evident, for with $m = 0$ the distortion becomes infinite as a $\frac{E}{V_d} \rightarrow 1$, regardless of the amount of feedback.

Fig. 14 shows the maximum distortion (expressed in dB) as a function of m for the same values of feedback. Note that for $m = 0.6$ the maximum distortion is less than the fundamental by 20 dB + the feedback.

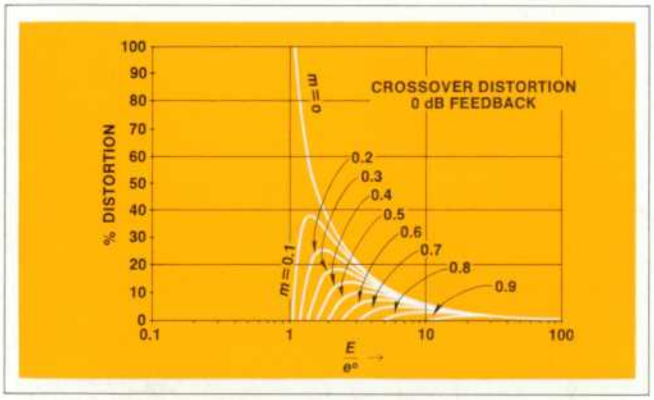

Fig. 10. *Crossover distortion, 0 dB feedback.*

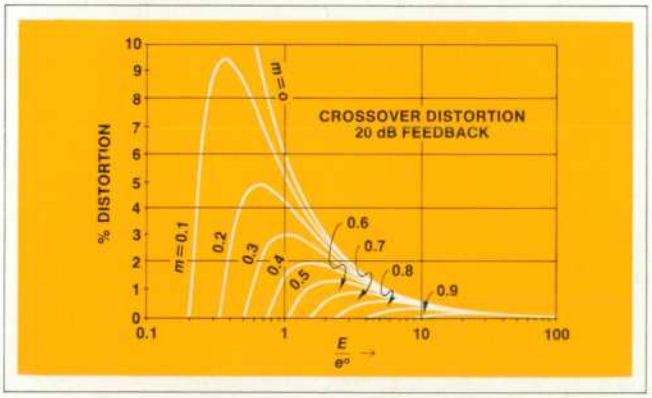

Fig. 11. *Crossover distortion, 20 dB feedback.*

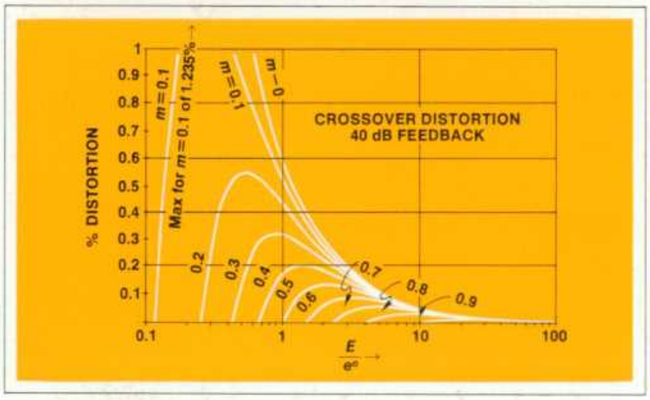

Fig. 12. *Crossover distortion, 40 dB feedback.*

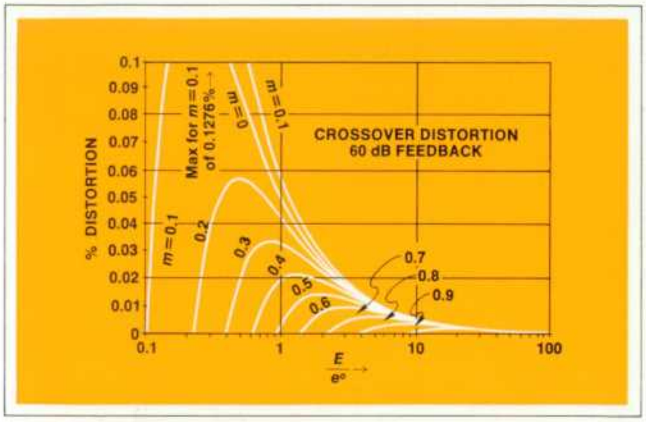

Fig. 13. *Crossover distortion, 60 dB feedback.*

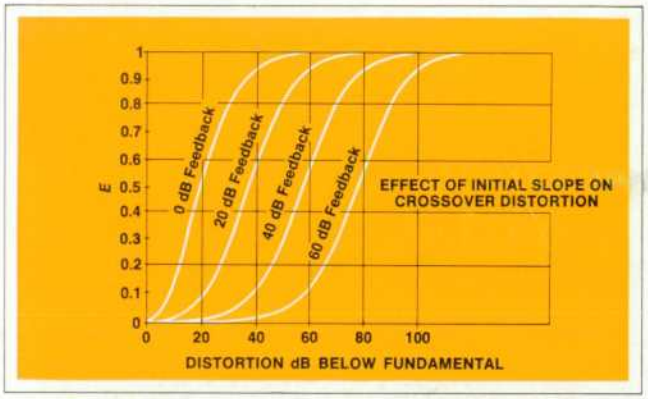

Fig. 14. *Effect of initial slope on crossover distortion.*

Dr. Bernard M. Oliver

Barney Oliver is HP's vice president for research and development. His bachelor's in EE is from Stanford, his MS and PhD from Cal Tech. Before coming to HP he was a member of the technical staff at Bell Labs, making contributions to the development of automatic tracking radar, television, information theory, and efficient coding systems. He holds more than 50 patents, is a Fellow and a past president of the IEEE, and a member of the National Academy of Engineering. He has also served as a lecturer in electrical engineering at Stanford.

At the company's headquarters in Palo Alto, Barney directs the activities of HP Laboratories, a centralized, multi-disciplinary facility which pursues advanced design projects on behalf of all the firm's divisions.

HEWLETT-PACKARD JOURNAL *FEBRUARY 1971 Volume 22 · Number 6*

TECHNICAL INFORMATION FROM THE LABORATORIES OF HEWLETT-PACKARD COMPANY 1501 PAGE MILL ROAD, PALO ALTO, CALIFORNIA 94304 U.S.A.
Hewlett-Packard S.A. 1217 Meyrin – Geneva, Switzerland • Yokagawa-Hewlett-Packard Ltd., Shibuya-Ku, Tokyo 151 Japan