THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 51 November, 1972 Number 9

The Static Stability of Half-Bubbles

By W. J. DeBONTE

(Manuscript received May 16, 1972)

A variational model is used to calculate the static stability limits and equilibrium properties of "half-bubbles," magnetic domains residing on one surface of a magnetic bubble material platelet. Stability is achieved through the presence of gradients in the domain wall energy density and/or saturation magnetic moment. The model evidences two distinct types of instability behavior separated by a critical value of the wall energy density gradient. Unlike the standard cylindrical domain, the half-bubble has a minimum stable value of the ratio of domain diameter to height.

The half-bubble is shown to possess a number of properties which make it potentially useful for device applications. It is self-biased by its closure wall and in some cases is stable in zero external bias. Bias margins are of the same order as those for the standard cylindrical domain. It stabilizes on only one platelet surface, and its properties are independent of both material thickness and of minor irregularities on the second surface. In addition, its structure may be advantageous in avoiding the undesirable properties of hard bubbles.

I. INTRODUCTION

The current interest in cylindrical magnetic domains and their device applications has stimulated the search for such domains in a wide variety of materials produced by a number of growth techniques. In

some samples, Bobeck[1] has observed domains of reversed magnetization which apparently do not penetrate through the entire thickness of the sample (Fig. 1). These domains, referred to as "half-bubbles," have four features which make them potentially attractive for device applications. First, they contact only one surface of the platelet in which they reside so that their properties should not be as critically dependent on surface preparation as those of the usual cylindrical domains. Second, as shown in this paper, their properties are independent of platelet thickness. Third, this article also shows that their top closure wall produces a self-biasing effect which makes external biasing unnecessary in some cases. Finally, Bobeck et al.[2] have observed in two-layer films that domains having a closure wall do not exhibit the undesirable properties of hard bubbles, and it is a reasonable presumption that this property also pertains to half-bubbles.

Bobeck[1] has suggested that half-bubbles may be stabilized by the presence of gradients normal to the material surface in one or more of the material characteristics (e.g., domain wall energy density σ_w and/or saturation magnetic moment M_s). In this paper we use a variational model to analyze the static stability of half-bubbles in materials having such gradients. Our model for the half-bubble and its instability modes is introduced in Section II. In Section III, we examine the stability of the model when the gradient in M_s is zero. While no stability occurs in this simple case, the calculation of Section III is a useful preliminary to the more general calculation of Section IV, where the gradient in M_s is assumed to be non-zero. The results of the stability calculations and their interpretation are discussed in Section V.

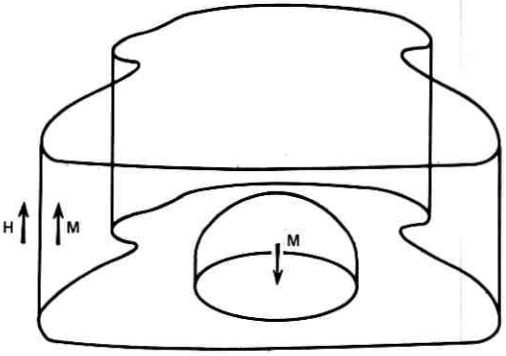


Fig. 1—Half-bubble magnetic domain configuration.

II. HALF-BUBBLE MODEL

In constructing our model we idealize the half-bubble in two ways. First, since the details of the shape of the half-bubble are not known, we approximate the actual shape of the domain by a right circular cylinder of radius r_o and height h (h assumed to be less than the platelet thickness), as shown in Fig. 2. This approximation gives the model domain a form identical to that of the standard cylindrical domain, the stability of which has been calculated by Thiele.[3] We shall exploit this identity in calculating the half-bubble energy and its derivatives. Second, we defer all consideration of the microscopic structure of the domain wall of the half-bubble and assume that the wall energy density σ_w varies linearly with distance through the platelet.

The magnetostatic energy of our half-bubble model possesses an important invariance property. As we show in Appendix A for the case in which the M_s gradient is zero, the magnetostatic energy of a cylindrical domain of fixed size located on the bottom surface of a platelet is independent of the position of the top surface of the platelet. This invariance is equally valid if M_s is a function of z, or if the domain assumes a more general shape. Thus, if we conceptually move the upper platelet surface down to the top of the model half-bubble in Fig. 2, we see that the magnetostatic energy of the half-bubble is equal to that of a standard cylindrical domain of the same size.

This invariance property has two important consequences. First, in the limit where the M_s gradient is zero, it allows us to obtain the magnetostatic energy of our model half-bubble from the results of

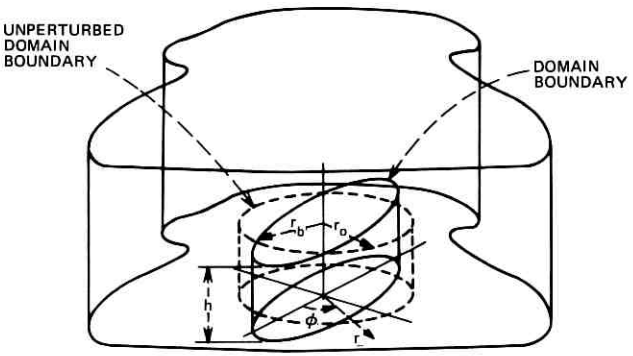


Fig. 2—Model for the half-bubble magnetic domain.

Thiele[3] and Thiele et al.[4] for the standard cylindrical domain. Second, it indicates that the platelet thickness is not a fundamental length of the half-bubble as it is for the standard cylindrical domain.

A convenient alternative standard of length, easily defined in terms of material characteristics, is provided by the material length at the bottom surface of the platelet ($z = 0$):

$$l = \sigma_w(0)/4\pi M_s^2(0). \tag{1}$$

In defining the σ_w and M_s gradient parameters γ_w and γ_M, we then scale the variations in wall energy density and saturation magnetic moment to this material length:

$$\sigma_w(z) = \sigma_w(0)\left(1 + \frac{\gamma_w z}{l}\right), \tag{2a}$$

$$M_s(z) = M_s(0)\left(1 + \frac{\gamma_M z}{l}\right). \tag{2b}$$

Experimental observation[1] indicates that there are three modes of instability which are of importance for the half-bubble: collapse, run-out, and "run-through". The general character of the collapse and run-out modes is already familiar from Thiele's stability calculation[3] for the standard cylindrical domain; the half-bubble collapse is not purely radial, however, but involves axial collapse as well. The term "run-through" is used to describe the transition from a half-bubble to a standard cylindrical domain which contacts both surfaces of the platelet.

As indicated in Fig. 2, the size of our model half-bubble is specified by the radius r_o and the height h. In considering stability against run-out, we include an elliptic distortion r_2 and specify the radial boundary of the domain by

$$r_b(\varphi) = r_o + r_2 \cos 2\varphi. \tag{3}$$

The model will be in equilibrium if the first derivatives of the total energy vanish. In the absence of in-plane anisotropy, $\partial E_T/\partial r_2 \equiv 0$ at $r_2 = 0$ by symmetry (cf. Appendix B). The equilibrium conditions are then

$$\frac{\partial E_T}{\partial r_o} = 0, \qquad \frac{\partial E_T}{\partial h} = 0. \tag{4a,b}$$

The stability of the resultant equilibrium is determined from the second derivatives, stability occurring when

$$\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.}\left(\frac{\partial^2 E_T}{\partial h^2}\right)_{eq.} - \left(\frac{\partial^2 E_T}{\partial r_o\, \partial h}\right)^2_{eq.} > 0, \tag{5}$$

$$\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.} > 0, \tag{6}$$

and

$$\left(\frac{\partial^2 E_T}{\partial r_2^2}\right)_{eq.} > 0. \tag{7}$$

If the inequality in eq. (5) is replaced by the corresponding equality, then the half-bubble is on the verge of either collapse or run-through. If $(\partial E_T/\partial r_2^2)_{eq.} = 0$, then the model is about to undergo run-out.

III. HALF-BUBBLE STABILITY WITH CONSTANT M_s

We begin the actual calculation of the stability limits for the half-bubble by considering the limiting case of constant M_s. This case lacks practical interest since we show that it leads to no region of stability for half-bubbles. In this limit, however, the problem loses much of the analytic complexity which characterizes the general case in which the gradient parameter $\gamma_M \neq 0$. Our method of solution–treating the second derivatives of the energy as polynomials in h/l with coefficients depending only on $2r_o/h$–thus becomes clearly visible here.

As noted in the preceding section, our half-bubble model is markedly similar in form to the model used by Thiele[3] and Thiele et al.[4] to treat the standard cylindrical domain. For this reason, in the present calculation we adopt the notation of these papers wherever practical.

The total energy of the half-bubble, measured relative to that of the uniformly-magnetized platelet, is

$$E_T = E_w + \Delta E_H + \Delta E_M \tag{8}$$

where E_w is the wall energy, (including the energy of the top closure wall), ΔE_H is the energy of interaction with the externally applied bias field, and ΔE_M is the magnetostatic energy. Since the half-bubble is independent of the magnetic material above it, we can describe our model in terms of the unit step function u as

$$\vec{M} = \hat{\imath}_z[1 - 2u(r_b(\varphi) - r)]u(z)u(h - z) \tag{9}$$

where $r_b(\varphi)$ is as defined in eq. (3).

The wall energy, consisting of integrals of $\sigma_w(z)$ over the side and top surfaces of the half-bubble, assumes the form

$$E_w = \int_0^h \sigma_w(z)\,dz \int_0^{2\pi} \left\{ r_b^2(\varphi) + \left(\frac{\partial r_o}{\partial \varphi}\right)^2 \right\}^{\frac{1}{2}} d\varphi + \sigma_w(h) \int_0^{2\pi} \frac{r_b^2(\varphi)}{2}\,d\varphi. \quad (10)$$

From eq. (10) it is clear that different wall energy densities for side and top walls could easily be included in our model. While such a distinction has intuitive appeal, experimental evidence on this point is lacking. We thus omit this added degree of freedom in the present calculation.

The energy of interaction with the external bias field $\vec{H} = \hat{\imath}_z H$, relative to that for the uniformly magnetized plate, is

$$\Delta E_H = -2 \int_{\text{h.}-\text{b.}} \vec{M}\cdot\vec{H}\,dV = M_s H h \int_0^{2\pi} r_b^2(\varphi)\,d\varphi. \quad (11)$$

If the elliptical distortion r_2 is set equal to zero, E_w and ΔE_H become

$$E_w = 2\pi r_o h \sigma_w(0)\left(1 + \frac{\gamma_w h}{2l}\right) + \pi r_o^2 \sigma_w(0)\left(1 + \frac{\gamma_w h}{l}\right) \quad (12a)$$

$$= 2\pi r_o h \sigma_w(h/2) + \pi r_o^2 \sigma_w(h), \quad (12b)$$

$$\Delta E_H = 2\pi r_o^2 h M_s H. \quad (13)$$

It should be noted that the interaction energy ΔE_H is proportional to the domain volume $\pi r_o^2 h$. This characteristic is shared by the gradient term in the energy of the top wall [eq. (12)], leading us to view this term as arising from an effective bias field. The non-gradient part of the top wall energy lacks any h-dependence, being proportional to πr_o^2. It may be thought of as coming from an h-dependent bias field analogous to that introduced by Liu[5] in connection with sputtered thin films which are exchange-coupled to a magnetic bubble material.

The relative magnetostatic energy is obtained from the general expression

$$E_M = \frac{1}{2} \int_v dV \int_{v'} dV' \frac{(\nabla\cdot\vec{M})(\nabla'\cdot\vec{M}')}{|\vec{r} - \vec{r}'|} \quad (14)$$

in the form

$$\Delta E_M = M_s^2 \int_0^\infty r\,dr \int_0^{2\pi} d\varphi \int_0^\infty r'\,dr' \cdot \int_0^{2\pi} d\varphi' [kk' - 1][\rho^{-1} - (\rho^2 + h^2)^{-\frac{1}{2}}] \quad (15)$$

where

$$k(r, r_b(\varphi)) = 1 - 2u(r_b(\varphi) - r), \quad (16)$$

$$\rho^2 = r^2 + r'^2 - 2rr' \cos(\varphi - \varphi'). \tag{17}$$

This integral is given in closed form by Thiele et al.[3] for the case $r_2 = 0$:

$$\Delta E_M = -2\pi h^3 \cdot 2\pi M_s^2 I(x), \tag{18}$$

$$I(x) = -\frac{2x}{3\pi}\left\{x^2 + \left(1 + \frac{1}{x^2}\right)^{\frac{1}{2}}\left[(1 - x^2)E\left(\frac{x^2}{1 - x^2}\right) - K\left(\frac{x^2}{1 - x^2}\right)\right]\right\} \tag{19}$$

where $x = 2r_o/h$ and K, E are complete elliptic integrals.

It is now a straightforward matter to differentiate E_T with respect to r_o, h, and r_2. Using the definitions for the normalized bias field $\tilde{H}$ and Thiele's[3] wall-force function F

$$\tilde{H} = H/4\pi M_s, \tag{20}$$

$$F(x) \equiv \frac{dI(x)}{dx} = \frac{2x^2}{\pi}\left[\left(1 + \frac{1}{x^2}\right)^{\frac{1}{2}}E\left(\frac{x^2}{1 - x^2}\right) - 1\right], \tag{21}$$

we may conveniently write the equilibrium conditions $\partial E_T/\partial r_o = 0$, $\partial E_T/\partial h = 0$ in normalized form as

$$\frac{l}{h}\left(1 + \frac{\gamma_w h}{2l}\right) + \frac{1}{2}x\frac{l}{h}\left(1 + \frac{\gamma_w h}{l}\right) + x\tilde{H} - F(x) = 0, \tag{22a}$$

$$\frac{l}{h}\left(1 + \frac{\gamma_w h}{l}\right) + \tfrac{1}{4}x\gamma_w + \tfrac{1}{2}x\tilde{H} - \frac{3}{x}I(x) + F(x) = 0. \tag{22b}$$

To show the biasing effect of the top wall we write eq. (22a) as

$$\frac{l(h/2)}{h} + x\left[\tilde{H} + \frac{1}{2}\frac{l(h)}{h}\right] - F(x) = 0 \tag{23}$$

where $l(h/2)$ and $l(h)$ are the material lengths as measured at the middle and at the top of the domain, respectively, and are defined by the obvious generalization of eq. (1). Equation (23) should be compared with the force equation for the standard cylindrical domain [Ref. 2, eq. (69)]

$$\frac{l}{h} + x\tilde{H} - F(x) = 0. \tag{24}$$

Equations (22a) and (22b) may also be used to obtain expressions for the values of γ_w and $\tilde{H}$ required to obtain an equilibrium state characterized by specified values of l/h and x:

$$\gamma_w = -\frac{2}{3}\frac{l}{h}\left(1 - \frac{x}{2}\right) + \frac{4}{x}I(x) - 2F(x) \tag{25a}$$

$$\tilde{H} = -\frac{1}{3}\frac{l}{h}\left(\frac{2}{x} + 1 + \frac{x}{2}\right) - 2\left(\frac{1}{x} + \frac{1}{x^2}\right)I(x) + \left(1 + \frac{2}{x}\right)F(x). \quad (25b)$$

The stability or instability of a given equilibrium point is determined by examining the second derivatives of the energy. Calculating these derivatives and removing a factor of $\beta^{-1} = 4\pi h \cdot 4\pi M_s^2$ we find

$$\beta\frac{\partial^2 E_T}{\partial r_o^2} = \frac{1}{2}\frac{l}{h} + \tfrac{1}{2}\gamma_w + \tilde{H} - \frac{dF(x)}{dx}, \quad (26a)$$

$$\beta\frac{\partial^2 E_T}{\partial h\,\partial r_o} = \frac{1}{2}\frac{l}{h} + \frac{1}{2}\gamma_w\left(1 + \frac{x}{2}\right) + \frac{1}{2}x\tilde{H} - F(x) + \frac{x}{2}\frac{dF(x)}{dx}, \quad (26b)$$

$$\beta\frac{\partial^2 E_T}{\partial h^2} = \frac{1}{4}x\gamma_w - \frac{3}{2}I(x) + xF(x) - \frac{x^2}{4}\frac{dF(x)}{dx}, \quad (26c)$$

$$\beta\frac{\partial^2 E_T}{\partial r_2^2} = \frac{l}{h}\left(\frac{1}{4} + \frac{2}{x}\right) + \gamma_w\left(\frac{1}{4} + \frac{1}{x}\right) + \frac{1}{2}\tilde{H} - \frac{4}{x^2}I(x) + \frac{2}{x}F(x) - \frac{1}{2}\frac{dF(x)}{dx}. \quad (26d)$$

Operations with these second derivatives are simplified if we define

$$G(x) \equiv -I(x) + xF(x) - \frac{x^2}{2}\frac{dF(x)}{dx}. \quad (27)$$

Using the relation[6]

$$\frac{2}{x}I(x) - F(x) = \tfrac{1}{4}(S_0(x) + 3S_2(x)) \quad (28)$$

we may write $G(x)$ in terms of the stability functions $S_0(x)$ and $S_2(x)$ of Ref. 3:

$$G(x) = \tfrac{3}{8}x(S_0(x) - S_2(x)). \quad (29)$$

After consulting plots[3] of $S_0(x)$ and $S_2(x)$, it is easily seen that $G(x) \geqq 0$ for $x \geqq 0$.

If we use eqs. (25a, b) to eliminate γ_w, $\tilde{H}$ from eqs. (26a–d), we obtain expressions for the second derivatives of E_T at equilibrium in terms of l/h and x only

$$\beta\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.} = -\frac{1}{3}\frac{l}{h}\left(\frac{2}{x} + \frac{1}{2}\right) + \frac{2}{x^2}G(x), \quad (30a)$$

$$\beta\left(\frac{\partial^2 E_T}{\partial h\,\partial r_o}\right)_{eq.} = -\frac{1}{6}\frac{l}{h}(1 + x) - \frac{1}{x}G(x), \quad (30b)$$

$$\beta\left(\frac{\partial^2 E_T}{\partial h^2}\right)_{eq.} = -\frac{1}{6}\frac{l}{h}x\left(1 - \frac{x}{2}\right) + \frac{1}{2}G(x), \quad (30c)$$

$$\beta\left(\frac{\partial^2 E_T}{\partial r_2^2}\right)_{eq.} = \frac{l}{h}\left(\frac{1}{4}+\frac{1}{x}\right)+\frac{1}{x^2}G(x). \tag{30d}$$

The derivatives are now in a form suitable for comparison with the stability conditions, eqs. (5–7). Inspection of eq. (30d) shows that $\beta(\partial^2 E_T/\partial r_2^2)_{eq.} > 0$ for all positive values of l/h so that eq. (7) is satisfied for this system. Equation (6) is satisfied if

$$\frac{l}{h} < \frac{6}{x^2}G(x)\left(\frac{2}{x}+\frac{1}{2}\right)^{-1}. \tag{31}$$

Finally, we require

$$\beta^2\left[\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.}\left(\frac{\partial^2 E_T}{\partial h^2}\right)_{eq.} - \left(\frac{\partial^2 E_T}{\partial r_o\,\partial h}\right)^2_{eq.}\right]$$
$$= \frac{1}{12}\left(\frac{l}{h}\right)^2\left(1-x-\frac{x^2}{2}\right)-\frac{l}{h}\left(\frac{1}{x}+\frac{1}{4}\right)G(x) > 0 \tag{32}$$

which may be factored[7] into $l/h > 0$ and

$$\frac{l}{h} > \left(\frac{12}{x}+3\right)G(x)\left(1-x-\frac{x^2}{2}\right)^{-1}. \tag{33}$$

However, it is easy to show that eqs. (31), (33) and the condition $l/h > 0$ cannot be simultaneously satisfied. Thus, if $\gamma_M = 0$, the half-bubble model has no stability against collapse and/or run-through.

While its result is negative, this stability calculation illustrates the method used in this paper for separating the l/h- and d/h-dependence of the half-bubble model and for finding its stability region. We first use the equilibrium conditions to remove γ_w and $\tilde{H}$ from the expressions for the second derivatives of E_T. The resultant expressions are essentially polynomials in l/h with coefficients depending only upon $x = d/h$ [cf. the right-hand sides of eqs. (30a–d) and (32)]. For fixed x, the roots of these polynomials provide us with the limits of the stability region (if one exists). In the next section, we generalize this method to include the effects of non-zero γ_M.

IV. HALF-BUBBLE STABILITY WITH $\gamma_M \neq 0$

We now generalize the equations of Section III to calculate the stability of the half-bubble model in the presence of a gradient in the saturation moment M_s. The wall energy E_w is unchanged from the expression given in eq. (10) and the interaction energy ΔE_H only gains a new factor to become

$$\Delta E_H = M_s\left(1 + \frac{\gamma_M h}{2l}\right)Hh \int_0^{2\pi} r_b^2(\varphi)\, d\varphi. \tag{34}$$

The expressions for the magnetostatic energy, however, become substantially more complex than those given in eqs. (15) through (19).

The magnetization distribution of eq. (9) is now replaced by

$$\vec{M} = \hat{\imath}_z M_s(0)\left(1 + \frac{\gamma_M z}{l}\right)[1 - 2u(r_b(\varphi) - r)]u(z)u(h - z). \tag{35}$$

The corresponding magnetostatic charge density is

$$-\nabla\cdot\vec{M} = -M_s(0)[1 - 2u(r_b(\varphi) - r)] \times \left\{\delta(z) - \delta(z - h) + \frac{\gamma_M h}{l}\left[-\delta(z - h) + \frac{1}{h}u(z)u(h - z)\right]\right\}. \tag{36}$$

The first two δ-functions in this expression are the familiar platelet surface charge terms which occur in the stability theory of the standard cylindrical domain [Ref. 3, eq. (19)]. The third δ-function gives the correction to the charge on the upper surface, while the product of step-functions describes the uniform volume charge resulting from the gradient γ_M.

The magnetostatic energy integral of eq. (14) has a quadratic dependence on the charge density $-\nabla\cdot\vec{M}$. As a result, we have terms in ΔE_M up to second order in γ_M. For convenience, we write

$$\Delta E_M = E_M^{(0)} + \gamma_M E_M^{(1)} + \gamma_M^2 E_M^{(2)} \tag{37}$$

where $E_M^{(0)}$ is identical to the magnetostatic energy of the preceding section. The quantity $\gamma_M E_M^{(1)}$ has two parts. The first arises from the interaction of the $\gamma_M = 0$ charge distribution with the uniform volume charge and is zero by symmetry. The second, representing the interaction of the $\gamma_M = 0$ charge distribution with the excess charge on the upper surface, is easily evaluated. We find

$$\gamma_M E_M^{(1)} = \frac{\gamma_M h}{l} E_M^{(0)}. \tag{38}$$

The final term $\gamma_M^2 E_M^{(2)}$ may be written as

$$\gamma_M^2 E_M^{(2)} = \frac{1}{2}\left(\frac{\gamma_M h}{l}\right)^2 M_s^2(0) \int_0^\infty r\, dr \int_0^{2\pi} d\varphi \int_0^\infty r'\, dr' \int_0^{2\pi} d\varphi'(kk' - 1) \times \left\{\rho^{-1} - 2h^{-1}\int_0^h dz[\rho^2 + (z - h)^2]^{\frac{1}{2}} + h^{-2}\int_0^h dz \int_0^h dz'[\rho^2 + (z - z')^2]^{\frac{1}{2}}\right\}. \tag{39}$$

The ρ^{-1} term in the curly brackets corresponds to the self-energy of the excess charge on the upper platelet surface; the z-integral following it arises from the interaction of this excess charge with the uniform volume

charge, and the double-z-integral corresponds to the self-energy of the volume charge. After evaluating the z-integrals, we obtain the following integral expression for the second-order energy:

$$\gamma_M^2 E_M^{(2)} = \left(\frac{\gamma_M h}{l}\right)^2 M_s^2(0) \int_0^\infty r\,dr \int_0^{2\pi} d\varphi \int_0^\infty r'\,dr' \int_0^{2\pi} d\varphi'(kk' - 1) \times \left\{2\rho^{-1} - \frac{\rho - (\rho^2 + h^2)^{\frac{1}{2}}}{h^2}\right\}. \quad (40)$$

Setting $r_2 = 0$, we find

$$\gamma_M^2 E_M^{(2)} = \frac{1}{15}\left(\frac{\gamma_M h}{l}\right)^2 (2\pi M_s^2(0))(2\pi h^3) \cdot \left\{\frac{x^3}{\pi} - (x + x^3)F(x) - (3 - x^2)I(x)\right\}. \quad (41)$$

The derivatives of the total energy E_T are now used to find the range of stability of the half-bubble model. Experimentally, this stable range would be mapped out by selecting materials characterized by given values of γ_w and γ_M and then varying $\tilde{H}$ to find the stable region in the variables d/l and h/l. The analytic form of our equations, however, precludes following such a physical approach in the present calculation. Instead, we use an extension of the method followed in Section III. We use the equilibrium conditions eqs. (4a, b) to eliminate γ_w and $\tilde{H}$ from the second derivatives and then treat the second derivatives as polynomials in $y \equiv h/l$, the coefficients of the polynomials being functions of $x \equiv d/h$ and γ_M. This manifestly unphysical approach gives the desired stable regions with a minimum of algebraic and numerical complexity.

The equilibrium conditions corresponding to eqs. (4a, b) are conveniently written in the following normalized forms:

$$\frac{l}{h}\left(1 + \frac{x}{2}\right) + \frac{\gamma_w}{2}(1 + x) + x\tilde{H}\left(1 + \frac{\gamma_M h}{2l}\right) - \left(1 + \frac{\gamma_M h}{l}\right)F(x) + \frac{1}{3}\left(\frac{\gamma_M h}{l}\right)^2\left[\frac{x^2}{\pi} - (1 + x^2)F(x) + xI(x)\right] = 0, \quad (42a)$$

$$\frac{l}{h} + \gamma_w\left(1 + \frac{x}{4}\right) + \frac{1}{2}x\tilde{H}\left(1 + \frac{\gamma_M h}{l}\right) + F(x) - \frac{3}{x}I(x) + \frac{\gamma_M h}{l}\left[F(x) - \frac{4}{x}I(x)\right] - \left(\frac{\gamma_M h}{l}\right)^2\frac{1}{x}I(x) = 0. \quad (42b)$$

These equations may be solved to find the appropriate values of γ_w and $\tilde{H}$ for an equilibrium state characterized by given values of γ_M, x, and $y \equiv h/l$. For brevity we write $F(x)$ and $I(x)$ as F and I respectively, and define a denominator D by

$$D = y\left[3 + \gamma_M\left(1 - \frac{x}{2}\right)y\right]. \tag{43}$$

We then find

$$\begin{aligned}\gamma_w = D^{-1}\Big\{&(x-2) + (\gamma_M x - 6F + 12x^{-1}I)y + \gamma_M(-10F + 22x^{-1}I)y^2 \\ &+ \frac{2}{3}\gamma_M^2\left[\frac{x^2}{\pi} - (7+x^2)F + \left(\frac{18}{x} + x\right)I\right]y^3 \\ &+ \frac{2}{3}\gamma_M^3\left[\frac{x^2}{\pi} - (1+x^2)F + \left(\frac{3}{x} + x\right)I\right]y^4\Big\},\end{aligned} \tag{44a}$$

$$\begin{aligned}\tilde{H} = D^{-1}\Big\{&-\left(\frac{2}{x} + 1 + \frac{x}{2}\right) + \left[\left(\frac{6}{x} + 3\right)F - \left(\frac{6}{x^2} + \frac{6}{x}\right)I\right]y \\ &+ \gamma_M\left[\left(\frac{6}{x} + 3\right)F - \left(\frac{8}{x^2} + \frac{8}{x}\right)I\right]y^2 \\ &- \gamma_M^2\left[\frac{1}{3}\left(\frac{4}{x} + 1\right)\left(\frac{x^2}{\pi} - (1+x^2)F + xI\right) + \frac{2}{x^2}(1+x)I\right]y^3\Big\}.\end{aligned} \tag{44b}$$

The denominator D in both of these expressions is always greater than zero and finite in the region of interest, and thus contributes no singularities or zeros to the calculation.

The second derivatives of E_T are developed in much the same fashion as in Section III. They are cast in dimensionless form by extracting a factor of $\beta^{-1} = 4\pi h \cdot 4\pi M_s^2(0)$, and γ_w and $\tilde{H}$ are then eliminated using eqs. (44a, b). The resultant expressions assume the form of polynomials in y after being multiplied by D. Because of the length of the expressions, we give only the final polynomial forms:

$$\begin{aligned}D\beta\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{\text{eq.}} = &-\left(\frac{2}{x} + \frac{1}{2}\right) + \left(\frac{6}{x^2}G - \frac{\gamma_M}{x}\right)y \\ &- \gamma_M\left[\left(4 - \frac{x}{2}\right)\frac{dF}{dx} - \left(\frac{9}{x} - \frac{1}{2}\right)F + \frac{11}{x^2}I\right]y^2 \\ &+ \gamma_M^2\left[\frac{2}{3}\frac{x}{\pi} - \left(2 - \frac{x}{2} + x^2\right)\frac{dF}{dx}\right. \\ &\left. + \left(\frac{13}{3x} - \frac{1}{2} + \frac{x}{3}\right)F - \left(\frac{6}{x^2} + \frac{1}{3}\right)I\right]y^3 \\ &+ \frac{\gamma_M^3}{3}\left[-\frac{x^2}{2\pi} + \left(\frac{x}{2} - 1\right)(x^2+1)\frac{dF}{dx}\right. \\ &\left. + \left(\frac{2}{x} - \frac{1}{2} + x\right)F - \left(\frac{3}{x^2} + 1\right)I\right]y^4\end{aligned} \tag{45a}$$

$$D\beta\left(\frac{\partial^2 E_T}{\partial r_o\, dh}\right)_{eq.} = -\frac{1}{2}(1+x) - \left[\frac{3}{x}G + \frac{\gamma_M}{2}\left(1+\frac{x}{2}\right)\right]y$$
$$+ \gamma_M\left[\frac{x}{2}\left(4-\frac{x}{2}\right)\frac{dF}{dx} - \left(\frac{9}{2}-x\right)F + \left(\frac{4}{x}-\frac{3}{2}\right)I\right]y^2$$
$$+ \gamma_M^2\left[-\frac{x^2}{3\pi} + \frac{x}{2}\left(1-\frac{x}{2}\right)\frac{dF}{dx}\right.$$
$$\left. - \left(\frac{5}{3}-\frac{5x}{4}-\frac{x^2}{3}\right)F + \left(\frac{1}{x}-2-\frac{x}{3}\right)I\right]y^3$$
$$+ \frac{\gamma_M^3}{3}\left[-\frac{x^2}{\pi} - \left(\frac{1}{2}-\frac{3x}{4}-x^2\right)F - \left(\frac{3}{2}+x\right)I\right]y^4 \quad (45b)$$

$$D\beta\left(\frac{\partial^2 E_T}{\partial h^2}\right)_{eq.} = -\frac{x}{2}\left(1-\frac{x}{2}\right) + \left[\frac{3}{2}G - \frac{\gamma_M x}{4}\left(1-\frac{x}{2}+\frac{x^2}{4}\right)\right]y$$
$$+ \gamma_M\left[-x^2\left(1-\frac{x}{8}\right)\frac{dF}{dx} + \left(\frac{15}{4}x - \frac{x^2}{8}\right)F - \frac{23}{4}I\right]y^2$$
$$+ \gamma_M^2\left[\frac{x^3}{6\pi} - \frac{x^2}{4}\left(1-\frac{x}{2}\right)\frac{dF}{dx}\right.$$
$$\left. + \left(\frac{11x}{6}-\frac{3x^2}{8}-\frac{x^3}{6}\right)F - \left(4-\frac{x}{2}-\frac{x^2}{6}\right)I\right]y^3$$
$$+ \frac{\gamma_M^3}{12}\left[-\frac{x^4}{2\pi} + \left(3x-x^2+\frac{x^4}{2}\right)F - \left(9-3x+\frac{x^3}{2}\right)I\right]y^4 \quad (45c)$$

$$D\beta\left(\frac{\partial^2 E_T}{\partial r_2^2}\right)_{eq.} = \left(\frac{3}{x}+\frac{3}{4}\right) + \left[\frac{3}{x^2}G + \frac{3\gamma_M}{2x}\right]y$$
$$+ \gamma_M\left[-\left(2-\frac{x}{4}\right)\frac{dF}{dx} + \frac{5}{4}\left(\frac{2}{x}-1\right)F + \left(\frac{1}{2x^2}+\frac{2}{x}\right)I\right]y^2$$
$$+ \gamma_M^2\left[-\frac{x}{5\pi} - \left(1-\frac{x}{4}+\frac{x^2}{2}\right)\frac{dF}{dx}\right.$$
$$\left. + \left(\frac{7}{10x}-\frac{5}{4}+\frac{7x}{10}\right)F + \left(\frac{13}{5x^2}+\frac{2}{x}-\frac{7}{10}\right)I\right]y^3$$
$$+ \gamma_M^3\left[(16+7x)\frac{x}{60\pi} - \frac{1}{6}\left(1-\frac{x}{2}+x^2-\frac{x^3}{2}\right)\frac{dF}{dx}\right.$$
$$\left. + \frac{1}{5}\left(\frac{1}{3x}+\frac{17}{12}-\frac{x}{2}-x^2\right)F + \frac{1}{5}\left(\frac{7}{2x^2}+\frac{2}{x}+\frac{1}{2}+x\right)I\right]y^4 \quad (45d)$$

The problem of finding the ranges of parameters yielding stability of the half-bubble model now becomes one of numerical analysis. For given values of x and γ_M we find the zeros of the three polynomials

$$D^2\beta^2\left[\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.}\left(\frac{\partial^2 E_T}{\partial h^2}\right)_{eq.} - \left(\frac{\partial^2 E_T}{\partial r_o\,\partial h}\right)^2_{eq.}\right],$$

$$D\beta\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.}, \quad \text{and} \quad D\beta\left(\frac{\partial^2 E_T}{\partial r_2^2}\right)_{eq.}.$$

A stable region corresponds to a range of y beginning and ending with one of these zeros and having the property that all three of the polynomials are positive for values of y on the interior of the interval.

The results of this procedure are shown for $\gamma_M = -0.025$, -0.05, and -0.1 in Figs. 3–5. The stable regions are plotted in the x, $1/y$ plane; $1/y = l/h$ was used as a coordinate in order to facilitate comparison with the results of Thiele for the standard cylindrical domain (cf. Ref. 3, Fig. 3). In the interior of the stable region we have provided two families of curves which describe the behavior of a stable half-bubble. The families γ_w = constant and $\tilde{H}$ = constant are obtained from the polynomials of eqs. (44a, b) after the boundaries of the stability region have been found. Since h is not fixed in the half-bubble, the domains do not follow lines of l/h = constant as $\tilde{H}$ is varied. Rather, they follow the curves γ_w = constant which characterize the given material.

V. DISCUSSION

We have explored the static stability of the half-bubble magnetic domain using a variational model. In spite of the simplicity of our model, the generality of some of the principles underlying our calculation lead us to expect that our results have at least semi-quantitative validity (cf. Appendices A and B).

Typical results of our stability calculation are shown in Figs. 3–5. The boundaries of each of the half-bubble stability regions shown in the figures roughly resemble two connected parabolic line segments, both parabolas opening to the right. The lower segment is determined by the zeros of the polynomial $D\beta(\partial^2 E_T/\partial r_2^2)_{eq.}$ and corresponds to the run-out instability mode. The upper segment arises from the zeros of

$$D^2\beta^2\left[\left(\frac{\partial^2 E_T}{\partial r_o^2}\right)_{eq.}\left(\frac{\partial^2 E_T}{\partial h^2}\right)_{eq.} - \left(\frac{\partial^2 E_T}{\partial r_o\,\partial h}\right)^2_{eq.}\right].$$

This part of the boundary gives the limits of stability against run-through and collapse. Thus, the lines γ_w = constant on the right side

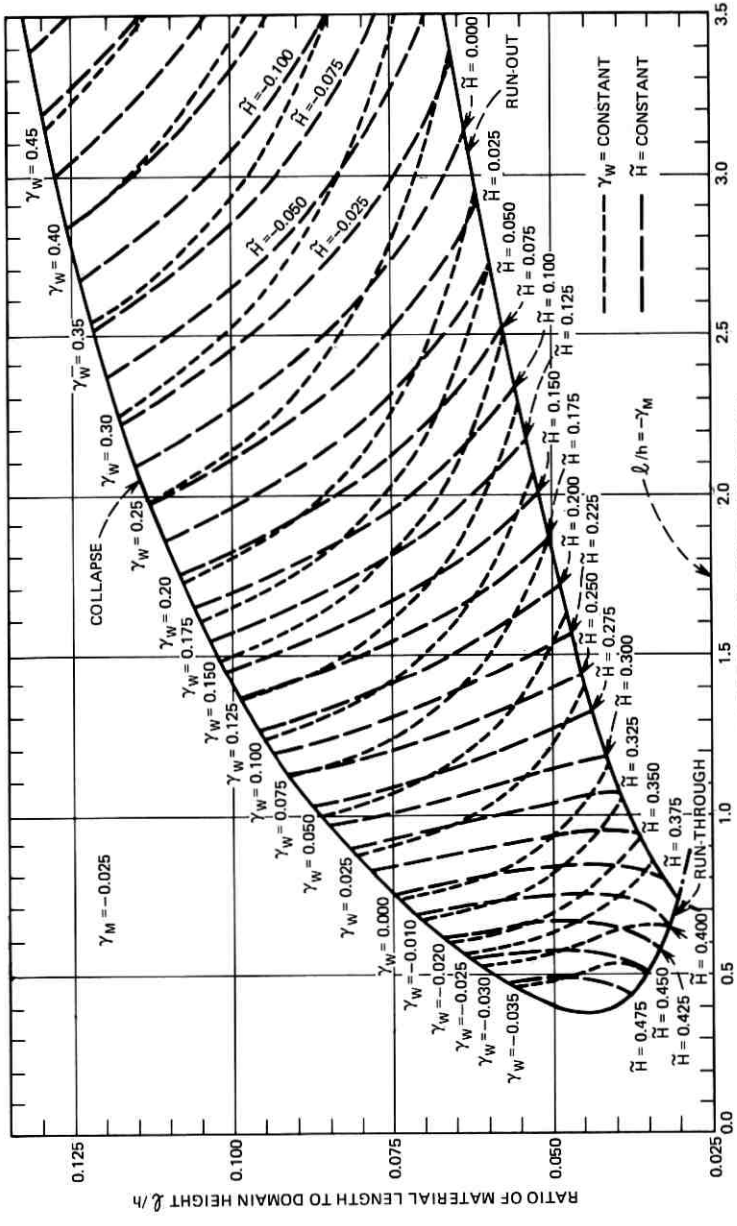


Fig. 3—Static stability chart for the half-bubble model, $\gamma_M = -0.025$.

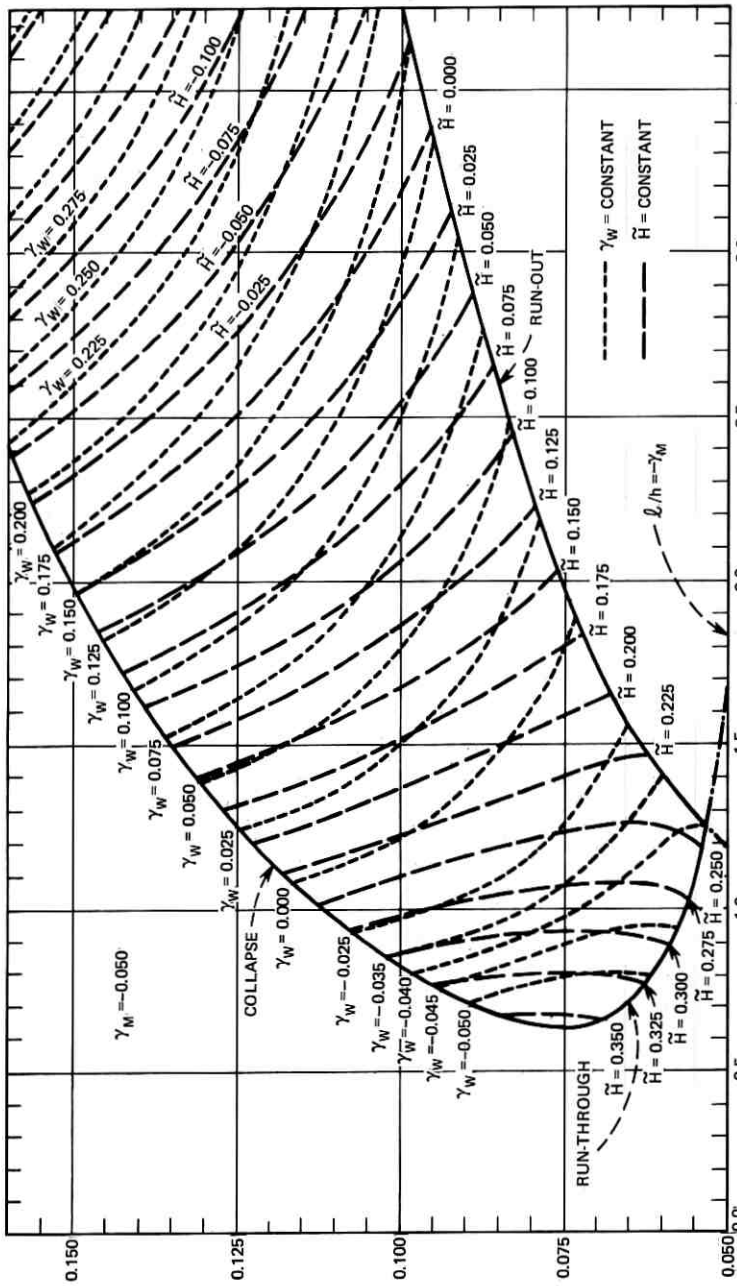


Fig. 4—Static stability chart for the half-bubble model, $\gamma_M = -0.050$.

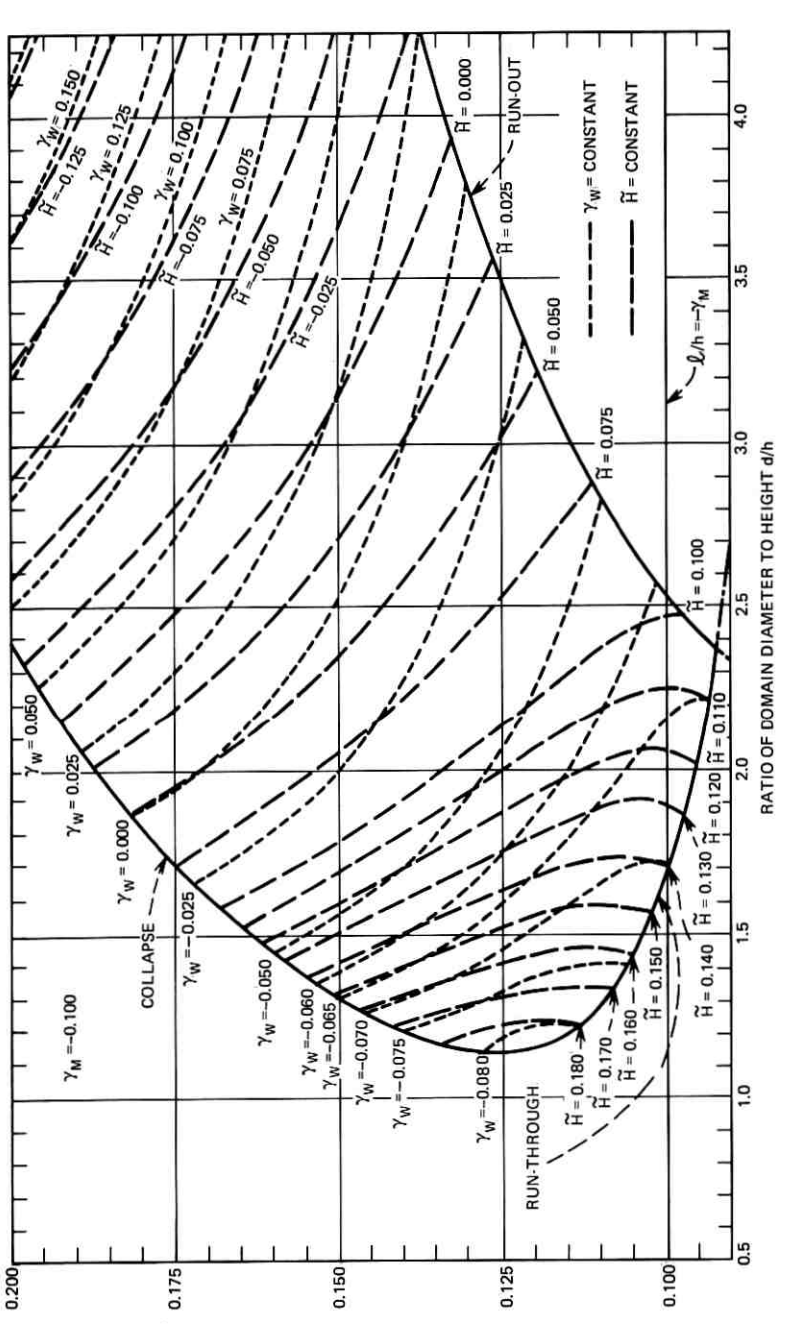


Fig. 5—Static stability chart for the half-bubble model, $\gamma_M = -0.100$.

of the figure go from run-out at the low-$\tilde{H}$ end of the line to collapse on the high-$\tilde{H}$ end. Values of γ_w which are algebraically small enough to correspond to lines on the left end of the stable region follow a different pattern, going from collapse to run-through as $\tilde{H}$ is decreased. Our stability charts then indicate two fundamentally different types of behavior as $\tilde{H}$ is decreased: the half-bubbles may undergo run-through to assume the shape of a standard cylindrical domain (and then presumably run out as $\tilde{H}$ is decreased further),[8] or the half-bubble may run out first (presumably to undergo run-through as $\tilde{H}$ is further lowered and give the usual demagnetized stripe pattern). These two types of behavior are separated by a critical wall energy gradient γ_{wc} which is a function of γ_M, being approximately -0.027 and -0.04 for $\gamma_M = -0.025$ and -0.050, respectively. (For $\gamma_M = -0.1$, the problem is more complicated and will be discussed below.)

Comparison of the half-bubble stability regions with those of the standard cylindrical domain (Ref. 3, Fig. 3) reveals one marked difference between the two systems. In the half-bubble case, there is a minimum stable value of d/h for given γ_M. For $\gamma_M = -0.025$, -0.050, and -0.100, these minimum values are $d/h = 0.39$, 0.64, and 1.14, respectively. Long, narrow bubbles are then not attainable in the half-bubble system.

It should be noted that all three of the stability charts correspond to $\gamma_M < 0$. While a general proof of the instability of half-bubbles with positive γ_M has not been formulated, numerical investigations show no regions of stability for $\gamma_M > 0$. As was shown to be the case for $\gamma_M = 0$, eqs. (5) and (6) could not be satisfied simultaneously for $\gamma_M > 0$. This property of gradient-stabilized half-bubbles is actually an advantage for device work as it implies a useful type of stability. A half-bubble originating on the lower surface of a platelet, after being allowed to undergo run-through, cannot be squeezed down to a half-bubble residing on the upper surface as $\tilde{H}$ is increased again. The domain is thus forced to return to its original state, the half-bubble state on the opposite surface being unstable.

In addition to the stability limits imposed by the derivatives of the energy function E_T, we place a lower limit on l/h by requiring $\sigma_w(h) > 0$ and $M_s(h) > 0$. While the motivation for the restriction $\sigma_w(h) > 0$ is physically obvious, the reason for the requirement $M_s(h) > 0$ stems more from the limitations of our model than from physical considerations. Magnetic materials may be compensated so that $M_s(z)$ goes linearly with z from positive to negative values. In the presence of a bias field, however, the magnetization in such a material would be aligned so that

its variation could no longer be described in terms of a linear variation with z. Such a compensated material would then not be susceptible to the present analysis and would presumably give rise to domains with different characteristics than those considered here. Taken together, the conditions $\sigma_w(h) > 0$ and $M_s(h) > 0$ yield the restriction

$$l/h > \max(-\gamma_w, -\gamma_M). \tag{46}$$

In practice, it may be necessary to require that l/h be even greater than the lower limit implied by eq. (46) since $\sigma_w(z) > 0$ must be satisfied not only at the top of the half-bubble ($z = h$), but throughout the region of the platelet above the half-bubble ($z > h$). (By a simple extension of Appendix A, we need not require $M_s(z) > 0$ in the region above the half-bubble.) These restrictions indicate that our calculation is not valid for part of the lower end of the stability region obtained for $\gamma_M = -0.100$ (cf. Fig. 5). As a consequence, the critical wall energy gradient γ_{wc} discussed above cannot be defined in this case.

The normalized bias fields $\tilde{H}$ required to stabilize half-bubbles are somewhat smaller than those needed for standard cylindrical domains as a result of the self-biasing effect provided by the upper closure wall (cf. Section III). Without going to unreasonably large values of γ_w, cases are found which are stable with $\tilde{H} = 0$ or even $\tilde{H} < 0$ ($\vec{H}$ parallel to the magnetization inside the domain). In addition, bias margins for the half-bubble are of the order of those for the standard cylindrical domain, typically falling in the range $0.04 \lesssim \Delta\tilde{H} \lesssim 0.08$. Thus if $4\pi M_s(o)$ is 200 gauss, bias field margins will be of the order 8 to 16 gauss.

To demonstrate the use of the stability charts in obtaining numerical information, we consider the case $\gamma_M = -0.100$ and $\gamma_w = +0.025$, for which a half-bubble is stable in zero bias field. The variations in σ_w and M_s between the bottom and the top of the half-bubble are easily computed after obtaining a value of l/h from the stability chart. In the present case, $l/h = 0.152$ corresponds to zero bias. Evaluating $\gamma_w h/l$ and $\gamma_M h/l$, we find that σ_w increases by 16.4 percent and M_s decreases by 65.8 percent between the bottom and the top of the half-bubble.

The collapse and run-out fields for $\gamma_M = -0.100$ and $\gamma_w = 0.025$ are found by interpolating between the $\tilde{H}$ = constant curves in Fig. 5. We find $\tilde{H}_{\text{coll.}} \cong 0.017$ and $\tilde{H}_{\text{run.}} \cong -0.023$ giving a bias margin of $\Delta\tilde{H} = 0.040$. The size variations of the half-bubble over its stable range are best measured in terms of the $z = 0$ material length defined by eq. (1). For any operating point in the stable range, $h/l = (l/h)^{-1}$ can be found from the vertical coordinate of the point while $d/l = (d/h)(l/h)^{-1}$ can be found as the ratio of the horizontal and vertical coordinates. For

$\gamma_M = -0.100$ and $\gamma_w = +0.025$ we find $(h/l)_{\text{coll.}} = 5.38$, $(d/l)_{\text{coll.}} = 10.85$, $(h/l)_{\text{run.}} = 7.27$, and $(d/l)_{\text{run.}} = 30.91$. Thus, while d increases by a factor of three between collapse and run-out, h increases by only 35 percent over this same range. Since the curves $\tilde{H}$ = constant and γ_w = constant are nearly parallel at the collapse end of the stable range, much of this change in h occurs with a very small change in $\tilde{H}$ near collapse. Thus, the relative variation in h may be decreased considerably by sacrificing a small fraction of the bias margin. The relative smallness of this variation in h is particularly important if half-bubbles on the lower platelet surface are to be propagated by permalloy circuits at the upper surface, as it is to be expected that there is an optimum circuit-domain separation for half-bubbles much as there is for standard cylindrical domains.[9]

In conclusion, we have shown that the half-bubble has a number of properties which are attractive for device applications. Its stability in the absence of external biasing represents a considerable advantage over the externally biased standard cylindrical domain. While some complexity is added by the requirement of gradient materials, the materials problem is simplified on another front by the fact that material thickness is unimportant for half-bubbles. Indeed, the half-bubble stabilizes on only one surface; if propagated from the same surface, irregularities in the second surface are of no consequence. These attractive properties of the half-bubble and its possible utility in avoiding the undesirable properties of hard-bubbles may more than compensate for the added difficulty in growing half-bubble materials.

VI. ACKNOWLEDGMENTS

The author is indebted to A. H. Bobeck, whose experiments provided the inspiration for this work. It is a pleasure to acknowledge many helpful conversations with and suggestions from F. B. Hagedorn and A. A. Thiele.

APPENDIX A

Magnetostatic Energy Invariance

Demagnetization fields equivalent to those of the standard cylindrical domain may be conceptually created through the use of two neutral subsystems of charges. The first consists of two parallel planes with magnetostatic charge density $\pm M_s$ and the second consists of two disks with charge density $\mp 2M_s$ placed on a common axis just inside

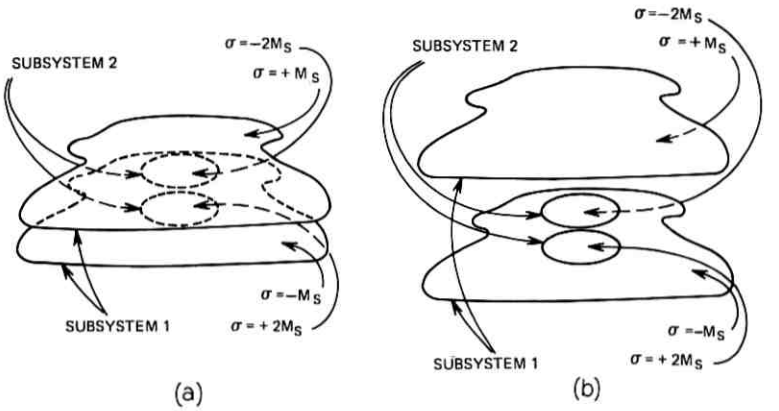


Fig. 6—Source-charge distribution for the demagnetization field of: (a) The standard cylindrical domain. (b) The half-bubble model.

the first subsystem (Fig. 6a). The distances between the upper plane and the upper disk, and between the lower plane and the lower disk are allowed to approach zero. The demagnetization field of the cylindrical domain is the superposition of the fields of the two isolated subsystems:

$$\vec{H}_d = \vec{H}_{d1} + \vec{H}_{d2} \,. \tag{47}$$

The magnetostatic energy of the domain (*including* the energy of the uniformly magnetized platelet) is the sum of the self-energies of subsystems 1 and 2 and their mutual interaction energy:

$$E_M = \frac{1}{8\pi} \int H_d^2 \, dV \tag{48a}$$

$$= \frac{1}{8\pi} \int H_{d1}^2 \, dV + \frac{1}{8\pi} \int H_{d2}^2 \, dV + \frac{1}{4\pi} \int \vec{H}_{d1} \cdot \vec{H}_{d2} \, dV. \tag{48b}$$

The transition from the magnetostatic field configuration of the standard cylindrical domain to that of the half-bubble model is achieved by moving the upper plane of charge upward (Fig. 6b). The field strength of subsystem one is unaltered by this change, while the self-energy of the subsystem increases linearly with the distance moved. The self-energy of subsystem two is, of course, unchanged, but remarkably the mutual interaction energy of the two subsystems is unchanged also. The latter invariance follows from the invariance of the field strength of subsystem

one and of the charge-neutrality of subsystem two (so that all field lines originating in the isolated subsystem two also terminate there). From these two properties it is easily seen that any lines of $\vec{H}_{d2}$ extending above the upper disk give a zero contribution to the mutual interaction energy of the two subsystems. If we now observe that the magnetostatic energy of introducing a domain into a uniformly magnetized platelet is just the sum of the self-energy of subsystem two and the mutual interaction energy, we see that this differential magnetostatic energy is independent of the position of the upper surface of the platelet (provided only that the domain remains inside of the platelet).

It is easily seen that this proof may be generalized to the cases in which M_s is a function of z or in which the domain assumes a more general form.

APPENDIX B

Energy of Half-Bubble Under Elliptical Distortions

The condition $(\partial E_T/\partial r_2)_{r_2=0} = 0$ in the absence of in-plane anisotropy is satisfied not only by our simple half-bubble model but by any convex domain having radial symmetry at equilibrium. For example, consider the generalized half-bubble shape

$$r_b(\varphi, z) = r_o(z) + r_2(z) \cos 2\varphi \tag{49}$$

where $r_2(z) \ll r_o(z)$ and $r_o(h) = 0$ (i.e., the half-bubble is closed at the top).

If there is no in-plane anisotropy, then

$$E_T(r_2) = E_T(-r_2) \tag{50}$$

since the shape function $r_b(\varphi, z) = r_o(z) - r_2(z) \cos 2\varphi$ transforms into that of eq. (49) under rotation. From eq. (50) it follows that

$$\left(\frac{\partial E_T}{\partial r_2}\right)_{r_2=0} \equiv 0. \tag{51}$$

Consequently, cross-derivatives such as $(\partial^2 E_T/\partial r_2\ \partial h)_{r_2=0}$ are also zero for the general domain shape given in eq. (49).

REFERENCES

1. Bobeck, A. H., private communication.
2. Bobeck, A. H., Blank, S. L., and Levinstein, H. J., "Multilayer Epitaxial Garnet Films for Magnetic Bubble Devices—Hard Bubble Suppression," B.S.T.J., *51*, No. 6 (July–August 1972), pp. 1431–1435.

3. Thiele, A. A., "The Theory of Cylindrical Magnetic Domains," B.S.T.J., *48*, No. 10 (December 1969), pp. 3287–3335.
4. Thiele, A. A., Bobeck, A. H., Della Torre, E., and Gianola, U. F., "The Energy and General Translation Force of Cylindrical Magnetic Domains," B.S.T.J., *50*, No. 3 (March 1971) pp. 711–773.
5. Liu, T. W., Bobeck, A. H., Nesbitt, E. A., Sherwood, R. C., and Bacon, D. D., J. Appl. Phys. *42*, (1971), p. 1641.
6. Hagedorn, F. B., private communication.
7. The factorization involving $l/h < 0$ is of no physical significance.
8. The half-bubble stability problem is closely related to the stability problem for the standard cylindrical domain in a platelet with gradients in M_s and σ_w normal to the platelet surface, the latter structure being essentially a half-bubble that has undergone run-through. The properties of this degenerate half-bubble are currently under investigation and will be reported in a separate paper.
9. Chen, Y. S., and Nelson, T. J., unpublished work.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 9, November, 1972
Printed in U.S.A.

Earthquake Environment for Physical Design: A Statistical Analysis

By S. C. LIU and L. W. FAGEL

(Manuscript received April 14, 1972)

A statistical analysis of nationwide seismic activities is made to determine the level and characteristics of earthquake motions to be expected in different geographical locations. Such information is needed to identify the severity of the earthquake threat to telephone facilities and to specify adequate design requirements. This paper describes the essential seismic environment including the expected earthquake-magnitude levels and the corresponding frequency of occurrence for different seismic-risk zones, and the free-field and in-building motions that might be induced by earthquakes. Stochastic models are used to analyze the earthquake-occurrence statistics, and to describe the induced random ground motions. Historical earthquake data are used to verify the theory and to generate information useful for seismic design.

I. INTRODUCTION

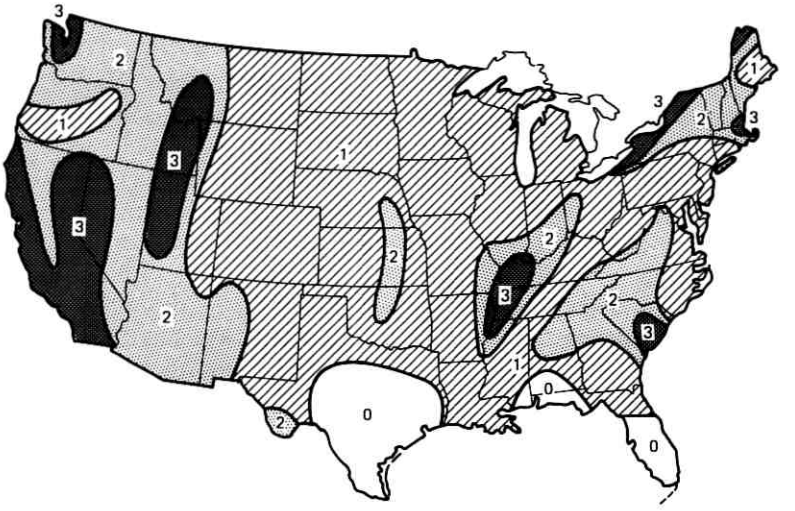
Telephone communications are so vitally important to public health and safety that special efforts must be made to prevent disruption of these services by a destructive earthquake.[1] For this reason, and in order to protect extremely expensive telephone plant investments, it is important to incorporate earthquake-resistant design into telephone facilities. The urgency of safeguarding communications systems against earthquake hazards has become even more evident as a result of experience gained from the recent San Fernando earthquake.[2] At present, earthquake design loads for telephone plants are estimated through use of structural design procedures such as those outlined in the Uniform Building Code (UBC).[3] These procedures may be generally satisfactory for building design, but are not appropriate for determining seismic loads for equipment housed in central office buildings.[4] Consequently, earthquake-resistant guidelines should be established for the physical design of communications facilities, and these guidelines should be

based on realistic estimates of the seismic environments to which telephone plants may be exposed. These environments and the necessary theoretical background for defining them are presented in this paper. The dynamic behavior of telephone structures in these environments and design criteria are not considered.

A statistical study approach is adopted here and historical earthquake data[5] are used both to verify the theoretical analysis and to identify the earthquake environments of engineering significance. The objective in Section II is to derive two types of statistics: macrophenomena, which deal with the single highest value of random parameters that characterize the earthquake occurrence; and microphenomena, which describe the local earthquake motion. In other words, it is a study of the statistical distribution of the magnitude (or intensity) of the largest earthquake in a given area, as well as the distribution of the largest amplitude of the nonstationary, earthquake-induced ground motion. In Section III, the theory is then applied to analyze available earthquake data and to generate information that is important from the viewpoint of seismic design. Geographic regions in this country that have had moderate-to-high levels of seismicity are selected (Fig. 1),[6] and the distribution of the intensity of the maximum annual earthquake for each region is then computed. Earthquake accelerograms along with the expected response spectra for given magnitude levels are presented which can be used for dynamic response analyses of structures.

II. STATISTICAL ANALYSIS OF EARTHQUAKES

Statistical techniques have been applied to many aspects of seismology and earthquake engineering; for example, in the data processing that involves estimating the travel time of seismic waves and locating their origin (epicenter), or in the analysis of seismic actions from a given area. In this section, emphasis is on this latter aspect, i.e., the statistical analysis of seismic action in a certain area from both the global and local standpoints. These two are distinguished by comparing the time scales they involve. The global (or macro-) analysis deals with the earthquake occurrence over a long period of time measured in a multiple of years. In this case, the earthquake-occurrence times and magnitudes (energies) are treated as random sequences. The local (or micro-) analysis deals with the free-field seismic wave motion itself, which generally lasts less than a minute. For each particular event, the local ground motion varies randomly with respect to time and constitutes a stochastic process. The objective of this section is to develop stochastic



ZONE 0 — NO DAMAGE

ZONE 1 — MINOR DAMAGE; DISTANT EARTHQUAKES MAY CAUSE DAMAGE TO STRUCTURES WITH FUNDAMENTAL PERIODS GREATER THAN 1.0 SECOND; CORRESPONDS TO INTENSITES V AND VI OF THE M. M. SCALE

ZONE 2 — MODERATE DAMAGE; CORRESPONDS TO INTENSITY VII OF THE M. M. SCALE.

ZONE 3 — MAJOR DAMAGE; CORRESPONDS TO INTENSITY VIII AND HIGHER OF THE M. M. SCALE.

THE PROBABLE FREQUENCY OF OCURRENCE OF DAMAGING EARTHQUAKES IN EACH ZONE WAS NOT CONSIDERED IN ASSIGNING RATINGS TO THE VARIOUS ZONES.

Fig. 1—Earthquake risk map by the National Oceanic and Atmospheric Administration[6] and UBC.[3]

models that are suitable for seismic-risk study, use them to interpret existing earthquake data, and extract information that can be used to establish loading environments for telephone structures located anywhere in the United States.

2.1 *Stochastic Model for Earthquake Occurrence*

Earthquakes can be considered to be a series of events randomly distributed on a real line (representing time), and the sequence of original times $\{t_n\}$ forms a point process.[7] It is further assumed that the joint statistics of the respective number of shocks in any set of intervals are invariant under a translation of these intervals; this implies that $\{t_n\}$ is a stationary point process. The stationary point process generalizes certain aspects of renewal processes; in particular,

the interval lengths $\tau_k = t_k - t_{k-1}$ between successive events need not be independently or identically distributed.

The simplest stationary point process is the Poisson process. Intuitively, the process $\{t_n\}$ can be approximated as a Poisson process if it represents rare events.* More rigorously it requires that τ_k be independently and identically distributed and follow a negative exponential function. Because of the simplicity and fairly general assumptions of the Poisson process, many studies have been made to test its fitness to the earthquake sequences.[9-16] Unfortunately, results of these tests in most cases indicate that the Poisson process is inadequate to explain the time distribution of low-magnitude shocks. The main deficiency of the simple Poisson model is its inability to describe the aftershocks which are often triggered by a large main shock. To account for the occurrence of aftershocks, the simple Poisson model for the main events is generalized to allow for the occurrence of more than one shock in a time unit, and a rate function $g(t)$ defined for $t \geqq 0$ and normalized to unit area is introduced into the generalized model to describe the distribution of the total number of aftershocks.[9] The function $g(t)$ can be either a decaying exponential or an inverse-power relation depending on the magnitude, depth, and location of the main event.[8,16] This model is known as the trigger or cluster model for earthquake processes.

However, for most practical engineering purposes, the simple Poisson model for earthquakes appears to be adequate. In practice, an engineer is concerned with the earthquake risk of structures located in some specific geographic areas. The risk depends heavily on the statistics of large earthquakes of these areas, and the omission of small earthquakes or aftershock processes is relatively unimportant in terms of earthquake risk. For this reason, the simple Poisson process[17] combined with the extreme-value theory[18] is used in this study to derive the distribution function of the magnitude (or intensity, peak ground accleration, etc.) of the largest shock in a time interval.

Let $N(m)$, the expected number of earthquakes in an area per unit time whose magnitude M exceeds m, be given by Gutenberg and Richter's familiar equation[19]

$$\log_{10} N(m) = \mathrm{a} - \mathrm{b}\, m, \qquad m > 0 \tag{1}$$

or, equivalently,

$$N(m) = \alpha e^{-\beta m} \tag{2}$$

* Rare events may be justifiable if we consider only deep earthquakes or very large earthquakes.[8]

where $\alpha = \exp(a \ln 10)$, $\beta = b \ln 10$, and a and b are constants. Then

$$P'(m) = \text{prob}\,(M > m) = \frac{N(m)}{N(0)} = e^{-\beta m} \tag{3}$$

and

$$P(m) = \text{prob}\,(M \leqq m) = 1 - e^{-\beta m}, \qquad m > 0. \tag{4}$$

Therefore, the probability density function of the earthquake magnitude is given by

$$p(m) = dP(m)/dm = \beta e^{-\beta m}, \qquad m > 0 \tag{5}$$

which indicates that the magnitudes of earthquakes are independently and negative-exponentially distributed with a parameter β. From eq. (5), the mean magnitude of all earthquakes of magnitude $m > 0$ and its variance are, respectively, $1/\beta$ and $1/\beta^2$. The mean return period, i.e., the mean interval in years between earthquakes having magnitude $M > m$, is

$$T_m = \frac{1}{N(m)} = (e^{\beta m})/\alpha. \tag{6}$$

From eq. (4) and under the assumption that the number of earth-earthquakes in an interval t follows a Poisson distribution

$$p(n, t) = (\lambda t)^n \exp(-\lambda t)/n!$$

with an average rate $\lambda > 0$ (note that $\alpha = \lambda t$), it can be shown that the magnitude of the largest shock, denoted by y, in the interval has the distribution

$$F(y, t) = \text{prob}\,(\max m \leqq y) = \sum_{n=0}^{\infty} p(n, t)P(y)^n$$

$$= \exp\{-\lambda t[1 - (1 - e^{-\beta y})]\} = \exp(-\lambda t e^{-\beta y}), \qquad y > 0 \tag{7}$$

which is called the distribution of largest values of the first kind. From eq. (7), the probability density function of y in t years becomes:

$$f(y, t) = dF(y, t)/dy = \lambda t \beta \exp(-\beta y - \lambda t e^{-\beta y}), \qquad y > 0. \tag{8}$$

A concise yet detailed analysis of the above extreme value model and its application in earthquake statistics has been given by Epstein and Lomnitz.[17] It is noted that the most significant parameters in the postulated earthquake occurrence model are λ and β. Based on historical earthquake magnitude data, the values of λ and β can be estimated by

a least-squares regression procedure or by using Gumbel's extreme-value plot, on which $F(y, t)$ should asymptotically fit a straight line. This latter approach is used in Section III to analyze data from various active seismic regions in the country.

2.2 *Stochastic Model for Earthquake Ground Motion*

Attention is now directed to the microanalysis of earthquake statistics; a description of a stochastic model which can be used for this analysis follows. The objective is to derive the probability distribution function of the single highest amplitude of the earthquake ground motion having a given magnitude, as well as statistics of its induced structural response.

2.2.1 *Multiplicative Process Model*

Various stationary and nonstationary models have been proposed to describe the earthquake ground motion.[20] The approach is to adopt here the multiplicative process representative for the ground acceleration $x(t)$ given by

$$x(t) = \phi(t)n(t), \tag{9}$$

where $\phi(t)$ is a deterministic envelope function and $n(t)$ is a Gaussian stationary process with power spectral density $S_n(\omega)$ and autocorrelation function $R_n(\tau)$. The following expression taken from Jennings, et al.[21] represents the envelope function (see Fig. 2):

$$\phi(t) = \begin{cases} (t/t_1)^2 & 0 \leqq t \leqq t_1 , \\ 1 & t_1 \leqq t \leqq t_2 , \\ e^{-\mu(t-t_2)} & t_2 \leqq t \leqq t_3 , \end{cases} \tag{10}$$

where $\mu > 0$ is a constant and t_3 is the total duration of the accelerogram.

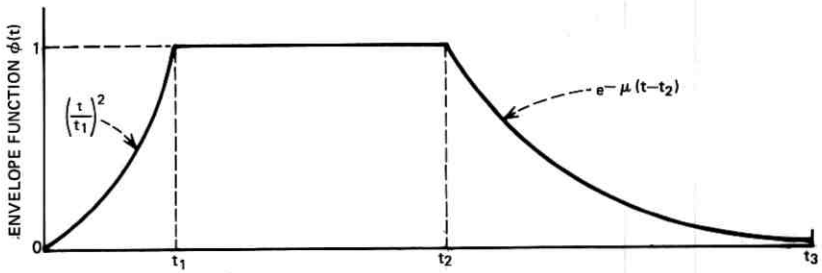


Fig. 2—Envelope function for earthquake accelerogram.

Under the assumption that the process $n(t)$ is the response of a linear ground filter with characteristic constants ω_g (representing natural frequency) and ξ_g (representing damping) to a Gaussian white noise with uniform power spectral density S, it can be shown[22] that the expected number of overshoots of $x(t)$ per unit time over the double barrier $x = \pm a$ is

$$m_{|a|}(t) = \frac{1}{\pi\sigma_n\phi^2} \exp\left(-\frac{a^2}{2\sigma_n^2\phi^2}\right)\left[\sigma_{\dot{n}}\phi^2 \exp\left(-\frac{a^2\dot{\phi}^2}{2\sigma_{\dot{n}}^2\phi^4}\right) + \frac{\pi}{2} a\dot{\phi} \operatorname{erfc}\left(\frac{a\dot{\phi}}{\sqrt{2}\,\sigma_{\dot{n}}\phi}\right)\right], \quad (11)$$

where σ_n and $\sigma_{\dot{n}}$ are standard deviations of $n(t)$ and $\dot{n}(t) = dn/dt$, respectively.

It is also possible to derive the distribution $F(z, t)$ of z, the largest value of $x(t)$ in a duration t, by again assuming that the probability distribution for n crossings in t is Poisson with a nonhomogeneous rate $N_{|a|}(t) = \int_0^t m_{|a|}(\tau)\, d\tau$. By this assumption, which is good for high crossing levels, it can be easily found that $F(z, t) = \exp[-N_{|z|}(t)]$ and the density function is $f(z, t) = dF/dt$. It should be noted that the analytical solution to the probability density function $f(z, t)$ is extremely complex even when $\phi(t)$ has a very simple expression. However, the mean and mean-square values of z in an interval t and its variance can be evaluated numerically without much difficulty. Note that these quantities are time-dependent and are functions of S, the spectral density of the white noise that produces $n(t)$. Some useful approximate solutions of the above statistical quantities are given in Appendix A.

2.3 *Distribution of the Maximum Structural Response*

The effect of the above earthquake process on structures can be represented by its response spectra, the maximum responses of single-degree-of-freedom systems with natural frequency ω_o and damping ξ_o. The primary interest is to find the expected response spectra and the associated variances. Since the input process $x(t)$ is nonstationary, the response process $u(t)$ is also nonstationary. The exact statistical behavior of $u(t)$ is, therefore, very difficult to obtain. However, results of practical value can reasonably be obtained by considering either a stationary approximation or a Monte Carlo approach. The latter is more accurate, of course, and is used in this study; however, the stationary approximation method is sometimes preferred because it can provide quicker and less-expensive numerical results.

2.3.1 *Stationary Approximation*

This method provides the approximation of the expected response spectrum to an input earthquake process and follows directly from the same approximate treatment as that used in arriving at eqs. (13) to (17) of Appendix A. Here it is assumed that the initial build-up and final decaying portions of $x(t)$ do not significantly affect the structural response $u(t)$, and therefore, $u(t)$ can also be treated as stationary. On this basis, eqs. (13) to (17) will also be valid to approximate the corresponding quantities for the response process after replacing z by $\max \mid u \mid$, σ_n by σ_u, which is the standard deviation of $u(t)$, and ν_n by $\nu_u = \sigma_{\dot{u}}/(\pi\sigma_u)$ = the zero crossing rate of $u(t)$ from both above and below. The values of σ_u and $\sigma_{\dot{u}}$ (the standard deviation of $\dot{u}(t) = du/dt$) can be easily determined by integrating the power spectral density functions $S_u(\omega)$ and $S_{\dot{u}}(\omega) = \omega^2 S_u(\omega)$ respectively. The function $S_u(\omega)$ is given by $S_u(\omega) \mid H_s(\omega) \mid^2$, where $H_s(\omega) = (\omega_o^2 - \omega^2 + 2i\xi_o\omega_o\omega)^{-1}$ is the transfer function of the structure and based on the ground motion model, $S_n(\omega) = S/[(\omega^2 - \omega_g^2)^2 + 4\xi_g\omega^2\omega_g^2]$. The results for the statistics of $u(t)$ in terms of given input and system parameters are given in Appendix B.

2.3.2 *Monte Carlo Computation*

The second approach to establish the distribution of the maximum structure response is Monte Carlo computation; the procedure is demonstrated below. For each magnitude level M, the expected spectral intensity $SI_{0.2} = \int_{0.1}^{2.5} S_v(T_o, 0.2)\, dT_o$, representing the damage (response) potential of an earthquake,[23] is obtained from Fig. 3, in which 0.2 is the associated damping ratio, $T_o = 2\pi/\omega_o$ = the natural period

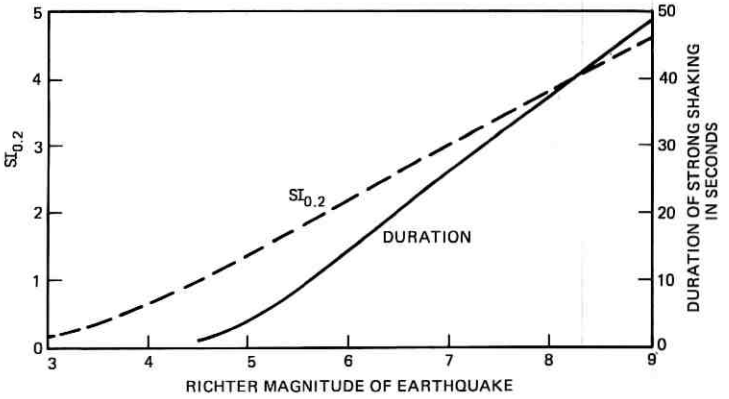


Fig. 3—$SI_{0.2}$ Intensity and duration of strong shaking versus Richter magnitude.

of the structure, and the duration of the strong shocking t_0 is also given in this figure. The desired earthquake process $x(t)$ is obtained by first generating the stationary process $n(t)$ on a computer by the standard method,[24] then shaping $n(t)$ by $\phi(t)$, and then normalizing it to match the expected spectral intensity $SI_{0.2}$. The response spectra $S_v(\omega_o, \xi_o)$ of each of the sample functions of $x(t)$ are then computed and averaged to obtain $E[S_v]$, where E denotes the ensemble average, and σ_{S_v}, the corresponding standard deviation. By this procedure the numerical solutions converge to the true ones rapidly as the sample size increases. With efficient algorithms and computer facilities available, this method should prove satisfactory for practical purposes. This method is used in the next section to generate the desirable ensemble of earthquake accelerograms and the associated response spectra.

III. APPLICATIONS

The statistics of earthquakes that can occur within various specified geographic regions in the United States are next estimated using historical data and the theories given in the previous section.

3.1 *Return Periods of Earthquakes*

The distribution of the magnitude of the largest annual earthquakes that occur within a particular region is given by eq. (7) with $t = 1$ year. Gumbel's probability paper[18] for extreme-value distributions of the first kind is used to plot this data. If the variate m does indeed follow a Poisson distribution, the distribution $F(y, t)$ in eq. (7) will plot in Gumbel's paper as a straight line. Earthquake return periods are equal to $(1 - F(y, t))^{-1}$; therefore, the required information is automatically generated when the Gumbel probability paper is used.

The procedure for constructing an extreme-value plot for a specified region is as follows. The highest annual modified Mercalli Intensity (I) [y in eq. (7)] for n years, as given in the U. S. earthquake catalogue,[5] are tabulated in order of increasing size: $y(1) \leqq y(2) \leqq \cdots \leqq y(n)$. For each $y(i)$, the associated value of $F(y(i), 1) = i/(n + 1)$. The computed $F(y(i), 1)$'s are plotted on an extreme-value graph versus Richter Magnitude (M) by using the relationship $M = 1 + 2I/3$ as proposed by Gutenberg and Richter.[19]

The geographic regions that have been referred to are selected after first dividing a map of the United States into 1-degree-longitude by 1-degree-latitude rectangular areas (each segment is approximately 50 by 70 miles) in Fig. 4. Each rectangle contains two numbers that are

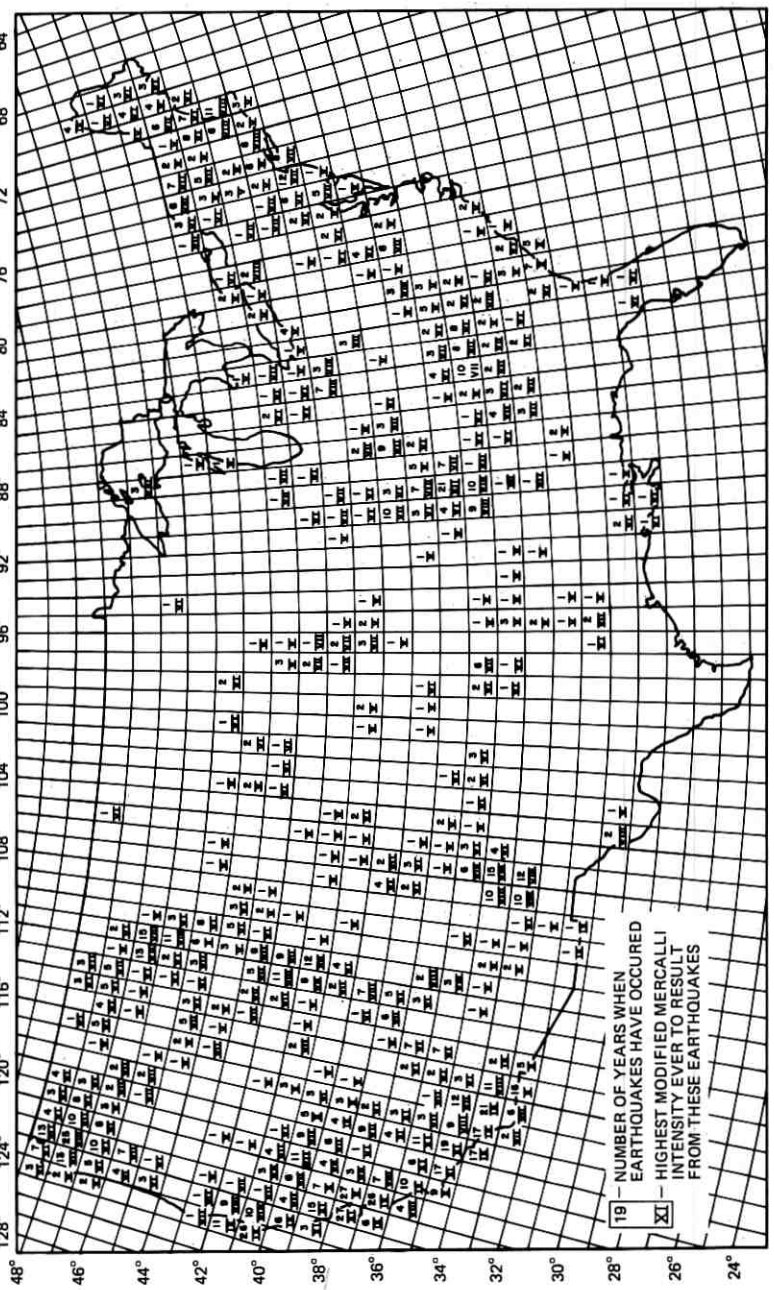
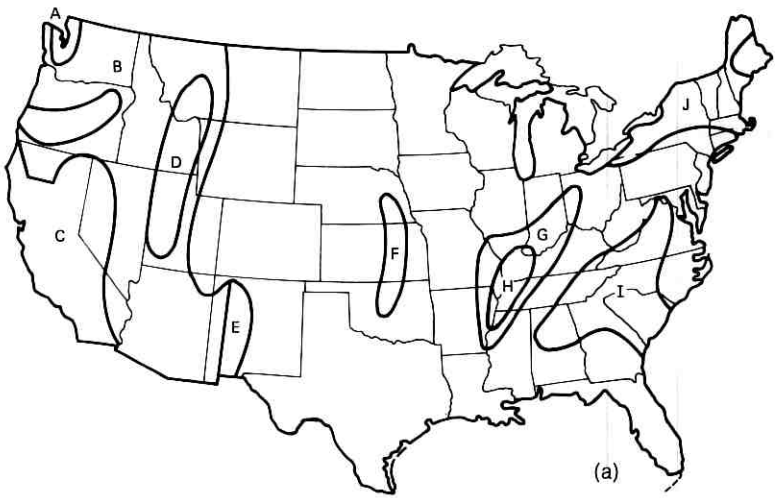


Fig. 4—Earthquake intensity data map of the United States.

obtained from historical data[5] if an earthquake of Modified Mercalli Intensity (I) greater than V had ever occurred within its boundaries. The arabic number represents the number of years when earthquakes have occurred, and the Roman numeral represents the largest intensity associated with any of those events. This information complements the seismic risk map in Fig. 1, and the two are considered together for defining boundaries around areas that appear to have comparable seismic history. These areas and boundaries are shown in Fig. 5a, and graphs showing Richter magnitude versus estimated return periods for earthquakes that may occur within these boundaries are shown in Fig. 5b. The largest magnitude distributions are plotted in Gumbel's paper as straight lines in Fig. 5b with its slope being proportional to the standard deviation of the maximum magnitude levels. Note that these distributions are based on all earthquakes that occurred in the identified areas with the exception of the 1811–1812 New Madrid, Mo. and the 1886 Charleston, South Carolina earthquakes. The return periods for these two violent events are estimated to be thousands of years; consequently, these are not regarded as meaningful data. Notice also that areas A, E, and H (see Fig. 5a), although geographically discontinuous, have very similar and comparable seismicity (Fig. 4) and the calculated distribution data for these areas are represented by a common straight line in the Gumbel's probability paper (Fig. 5b). Although various methods can be used to estimate the parameters of extreme distributions with different degrees of significance, such efforts are not warranted because a slight variation in the estimation is not an overriding concern in engineering risk analysis. Based on the same reason, distributions for areas B, F, G, I, J are also approximated by a common straight line.

Based on the results of Fig. 5b, maps such as that in Fig. 6 showing the Richter magnitude levels associated with different return periods can be constructed. Such maps are of value because they not only show the different degree of severity of earthquake threat in various regions, but also give the information as to how frequently the damaging earthquakes would be expected to occur in these regions. It may be noted that similarities exist between Figs. 1 and 6 because the seismic-risk map was referenced in selecting the boundaries in the return-period study. Some of the dissimilarities are noteworthy, however:

(*i*) The boundaries of the zone-3 area in parts of California and Nevada are not the same as those around the high-Richter-magnitude area in these states. The Great Central Valley below the 40th parallel has had greater seismic activity than most



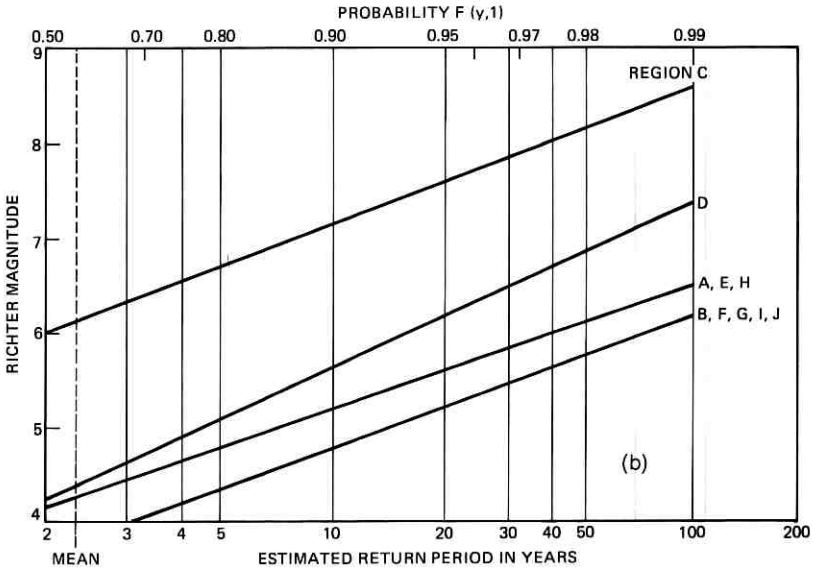


Fig. 5a—Seismic regions and boundaries for magnitude distribution analysis.
Fig. 5b—Earthquake Richter magnitude versus estimated return period.

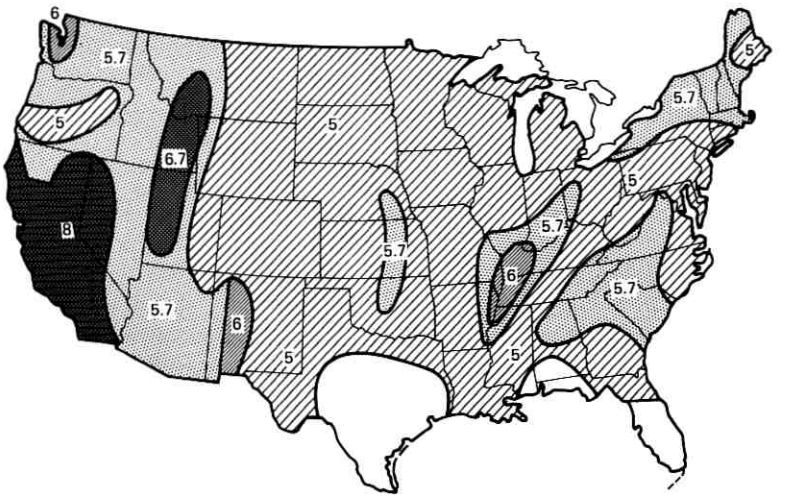


Fig. 6—Estimated maximum earthquake magnitudes for 40-year return period.

other western states, zone-2 areas; therefore, its data is considered with that of the neighboring zone-3 area.

(*ii*) California and Nevada have earthquake histories that indicate that larger earthquakes will occur in this area than in other zone-3 areas for the same return periods. The estimated maximum Richter magnitude is 8, whereas it is 6 for the Seattle area, 6.7 for the Montana-Iowa-Utah zone-3 area, 6 for the St. Louis area, and 5.7 for the Boston and Charleston, S. C. areas. Boston and Charleston do not appear to be more earthquake-prone than their respective surroundings which is in constrast to the information in Fig. 1.

(*iii*) Western New Mexico appears to have more earthquake potential than the rest of the Western United States, zone-2 region.

(*iv*) A small area within the western part of Texas is designated as zone 2 solely because of the $M = 6.4$, 8/13/31 earthquake at Mt. Livermore. Insufficient data prevents an extreme value distribution to be established; therefore, the region is not designated to be different in Fig. 6 from its surroundings.

Ground shaking that may occur in territories that are designated by an $M = 5$ design level in Fig. 6 (zone-1 in Fig. 1) are generally expected to be too small to be of engineering-design interest. Ground motion in

these areas could result from small-magnitude, local tremors or large-magnitude distant earthquakes; however, it does not appear to be a significant threat. The $M = 5$ design level shown for these areas in Fig. 6 should be regarded as an upper bound that is not necessarily applicable, particularly in geographic areas that have never experienced earthquakes (see Fig. 4).

3.2 *Free-Field Ground Motions*

It has not been feasible for building engineers or seismologists to consider geological conditions and then make precise, a priori predictions of ground motions that will occur at any particular point on the earth's surface during an earthquake. This is so because most dependent parameters such as earthquake magnitude, focal depth, epicentral distance, slipped fault length, propagation path, and soil-media properties, are usually not adequately known.

A more direct and realistic approach to defining ground motions is to artificially synthesize them using characteristic parameters from past earthquake accelerogram measurements, such as dominant frequency, amplitude (Fig. 7), spectrum intensity, and duration of ground motion (Fig. 3), etc. Dominant frequencies of an earthquake are defined here as those corresponding to the dominant peak in the Fourier spectrum of the accelerogram. These are known to range between 6 and 60 radians/s with a mean value of approximately 15 radians/s (Fig. 8). Peak acceleration amplitudes can be generally related to magnitude by the curve in Fig. 7. These are upper-bound values for compact alluvium. Accelerations on sites of different geology would fluctuate about the values shown in Fig. 7.

The procedure and references for synthesizing an artificial earthquake

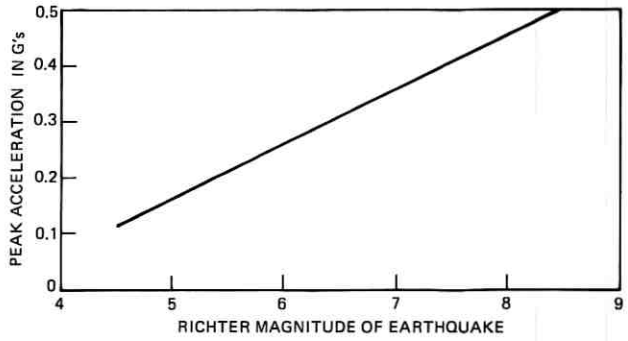


Fig. 7—Expected peak amplitudes of earthquake accelerograms.

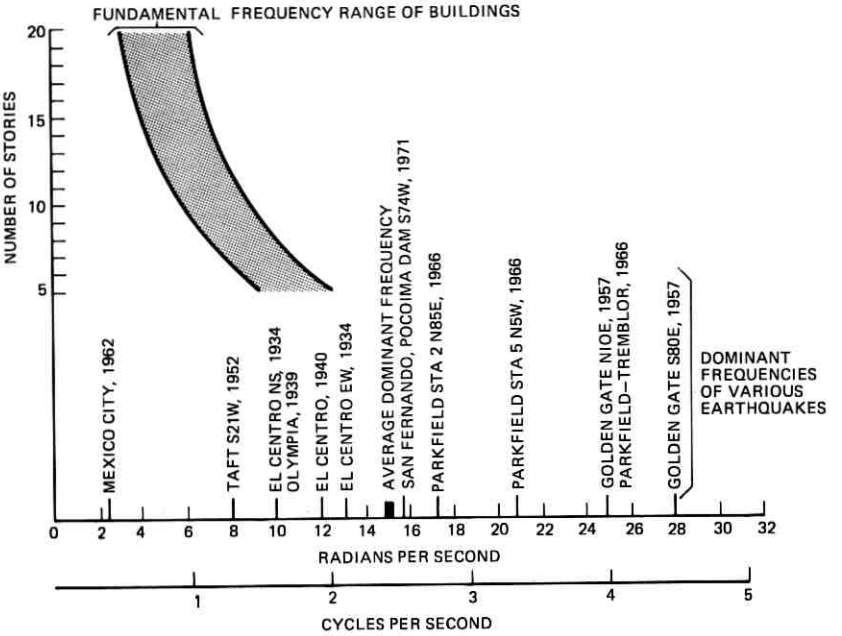


Fig. 8—Typical fundamental frequencies of multistory buildings and dominant frequencies of strong-motion earthquakes.

using the above parameters and a digital computer are given in Refs. 20 and 23. A sample accelerogram for an earthquake of Richter magnitude 6.3 is shown in Fig. 9a. The assumed values of the dependent parameters used to create this waveform are a dominant frequency of 15 radians/s, an $SI_{0.2}$ of 2.3, and an envelope function which is comprised of a 3-second buildup, 9 seconds of strong stationary shaking, and an exponential decay that lasts for 18 seconds. The waveform in Fig. 9a is one of an ensemble of 25 synthetic earthquakes generated according to the Monte Carlo method described earlier, and it corresponds to the average of the peak accelerations from the ensemble. This is 0.28 g's which coincides with the acceleration value given by both Housner[23] and Gutenberg-Richter[19] for an earthquake of this magnitude. The velocity and displacement functions corresponding to this particular accelerogram, obtained by integrations after performing baseline corrections, are shown in Figs. 9b and 9c. The peak amplitudes are 1.5 ft/s and 1.3 ft respectively.

Artificial earthquakes corresponding to different Richter magnitudes and dominant frequencies can be synthesized in the same manner. If the

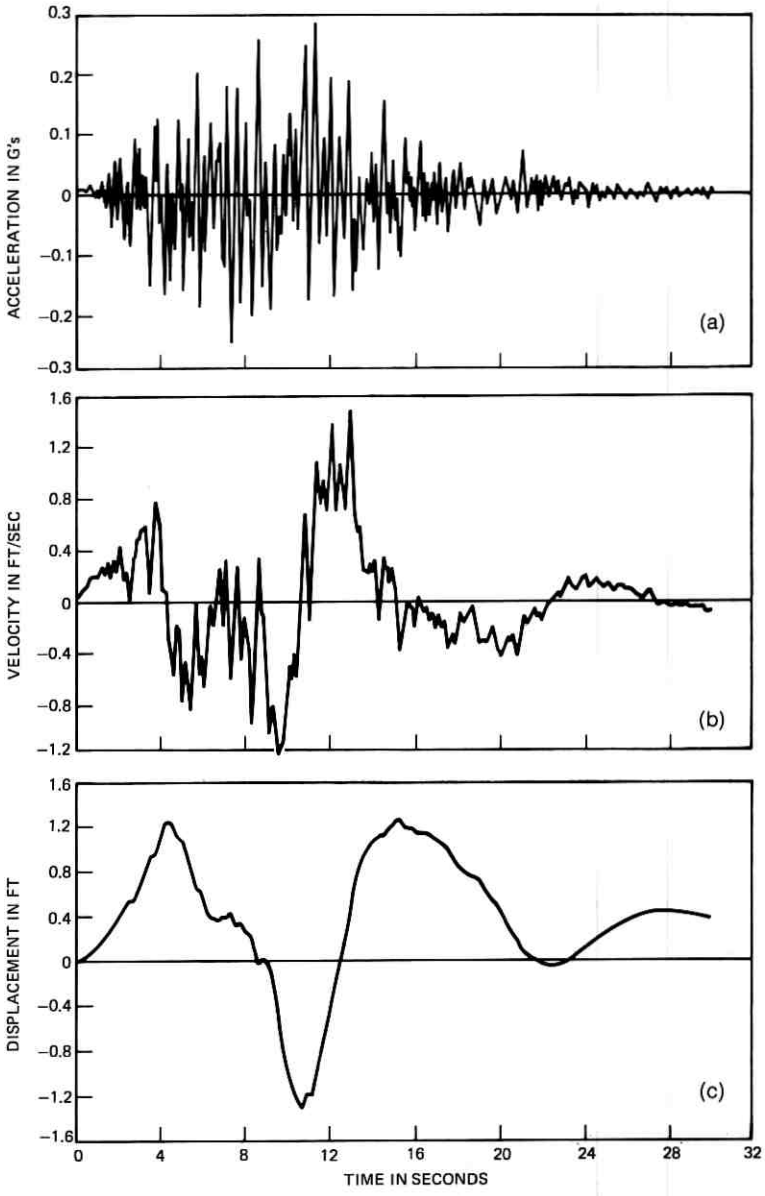


Fig. 9—Acceleration, velocity, and displacement time-histories of synthetic earthquake with $M = 6.3$.

dominant frequency is close to the average (15 radians/s), it is simpler to approximate other horizontal ground motions by scaling the previously mentioned waveforms with the factors shown in Table I which are the $SI_{0.2}$ values normalized to 1.0 for a magnitude 6.3 earthquake. This procedure will yield peak accelerations for different magnitudes of earthquakes that are approximately in agreement with Housner's data[23] (Fig. 7).

Figures 10a, 10b, and 10c are acceleration, velocity, and displacement response spectra of the accelerogram shown in Fig. 9a for damping ratios of 2 percent, 5 percent, and 10 percent. It is apparent that the acceleration, velocity, and displacement time functions of Figs. 9a, b and c together with these response spectra, are quantities associated with a particular sample member of the entire ensemble of earthquake ground motions. If a nondeterministic approach to structural analysis or design is used, the mean values and information about the variability of the ensemble response spectra are needed. The average response spectra and the corresponding standard deviations of the 25 synthesized earthquake samples are obtained by the Monte Carlo computation as described earlier and shown in Fig. 11. Note that these spectra represent only the elastic structural responses. For response analysis or design when inelastic behavior of structures is expected, the additional energy dissipation in the structural-foundation system should be taken into consideration.

It is understood that the vertical motion of an earthquake is generally less severe than, and is not of as much concern to structural designers as the horizontal motion. The accelerograms and response spectra of vertical motions may be taken as one-half to two-thirds the horizontal accelerograms and spectra shown in Figs. 9 and 10, respectively. This

TABLE I—SPECTRUM INTENSITY SCALING FACTORS

The following scaling factors are Spectrum Intensity Factors[23] ($SI_{0.2}$) normalized to 1 for an earthquake of Richter magnitude 6.3. Free-field horizontal ground motions and response spectra may be estimated for design purposes by multiplying the amplitude coordinates on Figs. 9, 10, and 11 by these scaling factors.

Richter Magnitude	Scaling Factor
5	0.6
5.7	0.8
6	0.9
6.3	1.0
6.7	1.1
8	1.5

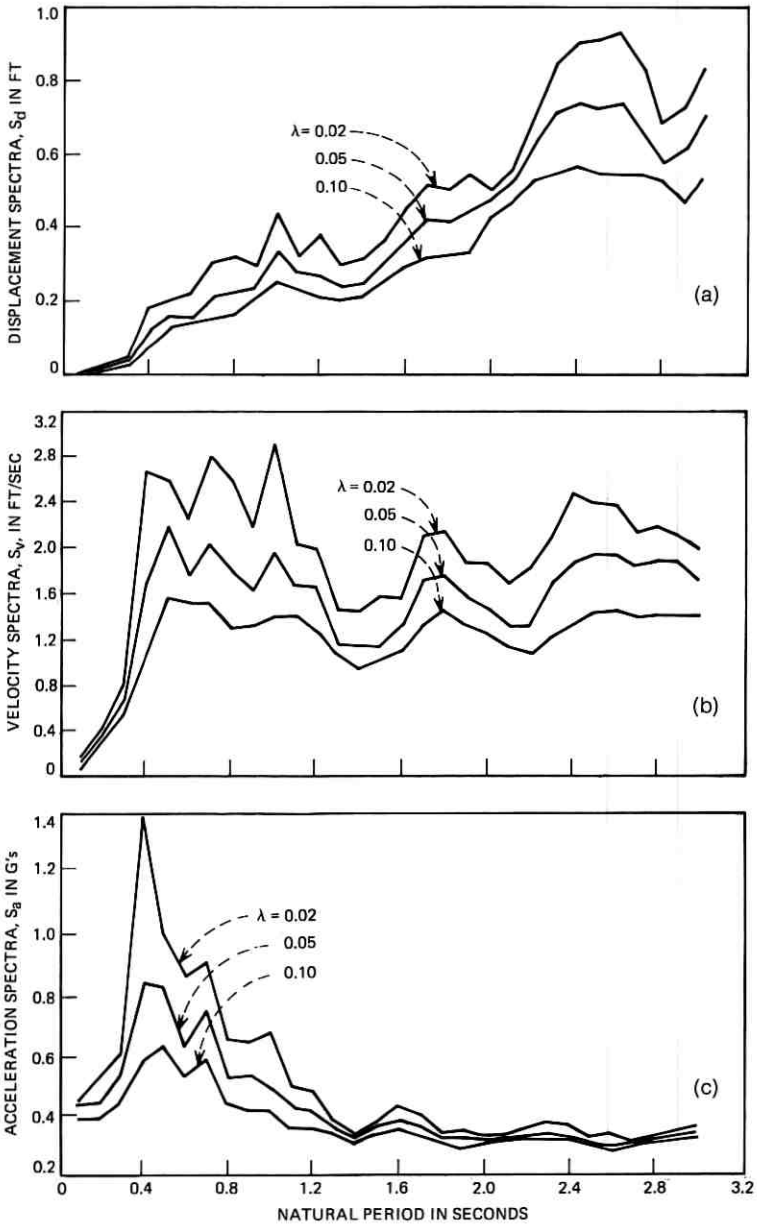


Fig. 10—Response spectra of earthquake accelerogram shown in Fig. 9.

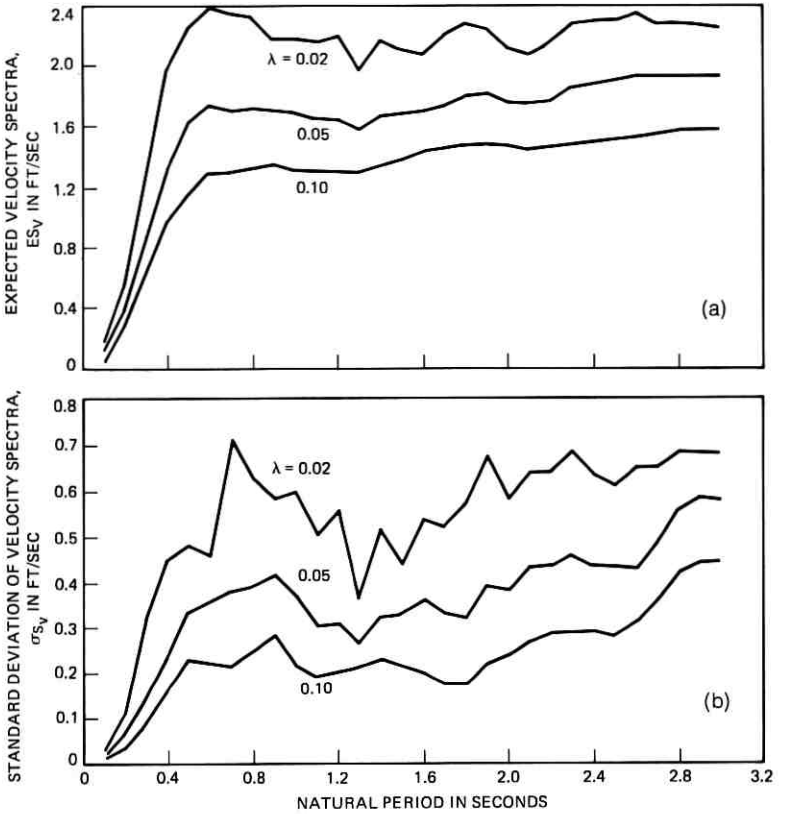


Fig. 11—Expected velocity spectra and standard deviation of an earthquake process with $M = 6.3$.

recommendation is made without the support of extensive theoretical analysis or data reduction as is done for the horizontal motion. However, it agrees with the limited data presently available, and it is consistent with the design practices adopted for other structures such as dams and nuclear power plants.[25]

3.3 *In-Building Motion Environments*

Telephone equipment installed within multistory buildings can generally be expected to encounter motion environments of greater intensity than equipment in single-story structures. Environments for the latter will be essentially free-field accelerograms, but multistory buildings amplify ground motions. It is not a trivial matter to specify

exactly how much amplification will take place, however, because the structural response of a building depends strongly on the properties of the structure and its founding soil. Soil properties affect not only the characteristics of the ground motions but also the fundamental frequency of the building. This in turn can drastically affect the way a building will vibrate and is generally referred to as soil-structure interaction.[26,27]

Typical fundamental frequencies for telephone buildings up to 20 stories tall and founded on various soils are also given in Fig. 8. If one examines building-frequency formulas found in the UBC, it will be noted that some nontelephone buildings have lower frequencies than those shown in Fig. 8. Telephone buildings are usually extremely well designed and constructed, and in spite of the massive equipment contained within them, their natural frequencies are generally higher than those of conventional office buildings.

It is important to understand that tall buildings, even the rigid ones constructed for telephone service, usually have fundamental frequencies that are below the dominant frequency contained in most earthquakes. This is fortunate because when the dominant frequency of an earthquake matches the fundamental frequency of a building, the acceleration responses of the structure can be much greater than in situations when these two frequencies do not match. This fact is illustrated in Fig. 12, which resulted from a study of the potential responses of a twenty-story building when subjected to earthquakes with various dominant frequencies.[27] It should be noted that the amplification of horizontal ground accelerations would clearly be highest when the earthquake's dominant frequency happens to match the structure's fundamental frequency and the building is founded on a stiff soil (V_s = 4000 ft/s). This hypothetical situation is unlikely to occur however; in fact, historical data that are shown on the frequency coordinate of Fig. 8 suggest that a frequency match should not be expected for telephone buildings that are over eight stories. A frequency match for buildings under eight stories will not result in amplifications that are as large as those given in Fig. 12. The estimated maximum ratios of in-building acceleration to horizontal ground motion that can be used for design purposes is shown in Fig. 13, which is a composite plot applicable to all telephone buildings up to 20 stories tall. These analytically-derived amplification factors may be regarded as upper-bound values that take into account the effects of the natural frequencies of the buildings, expected dominant frequencies of potential earthquakes, and all possible soil conditions on which buildings would conceivably be erected. Approximate in-building horizontal accelerograms can be artificially synthesized in a manner

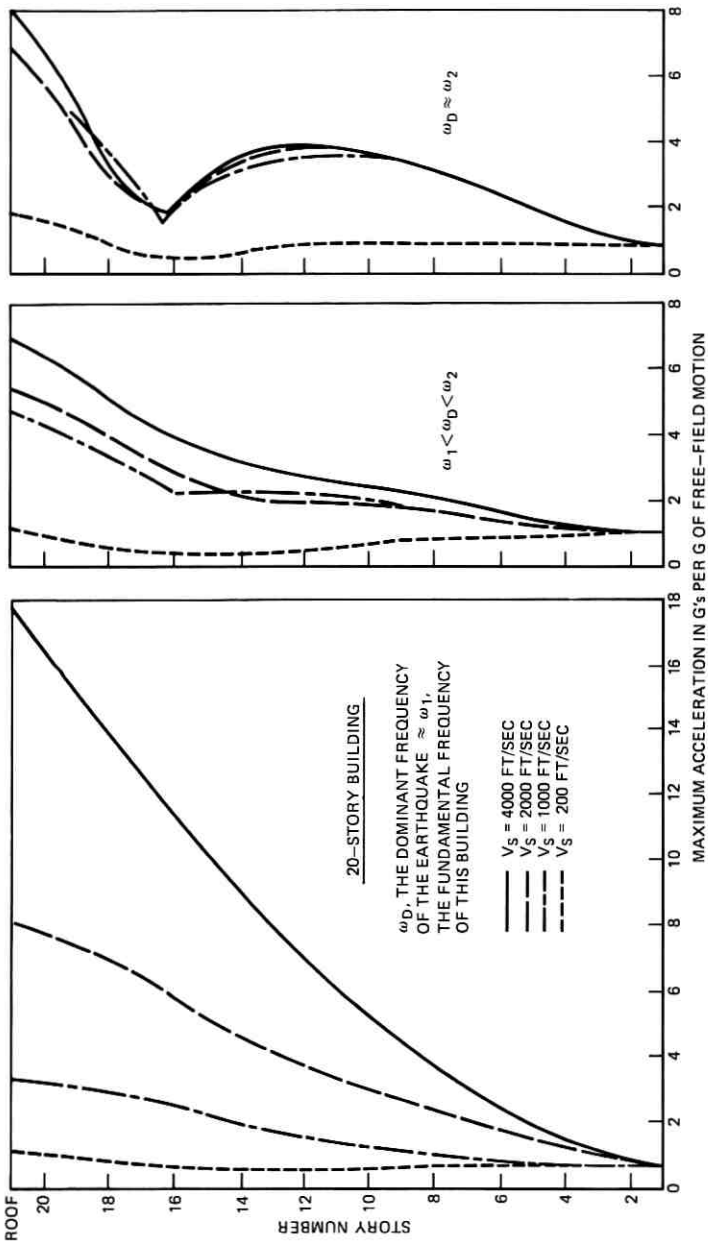


Fig. 12—Peak horizontal acceleration responses of 20-story buildings founded on various soils.

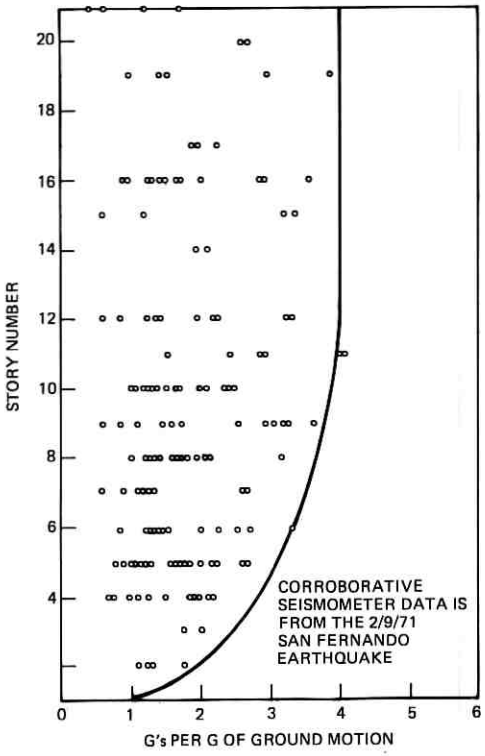


Fig. 13—Estimated practical upper bound for in-building horizontal accelerations during an earthquake.

similar to the generation of free-field motions, except that the value of $SI_{0.2}$ must be first multiplied by an amplification factor taken from Fig. 13. As an alternative to that procedure, the waveform shown in Fig. 9 and the corresponding response spectra in Figs. 10 and 11 may be multiplied by the appropriate scaling factor from Table I and also by the amplification factors given in Fig. 13. Note that these approximations are valid assuming there is no resonance between the building and the floor-mounted equipment.

IV. CONCLUDING REMARK

An extensive statistical analysis of seismicity data and earthquake motions has been made. Simple yet realistic stochastic models are used to describe the earthquake occurrence process and the random local

ground motion. Information useful for seismic design of the telephone system is generated according to the theory presented and through the use of an available strong earthquake catalog.

It should be pointed out that the main concern of this study is to provide information to assist in the physical design of earthquake-resistant structures by describing a realistic nationwide earthquake environment in meaningful engineering terms. The next step will be to establish appropriate structural design criteria. A simple approach for this would be to base such criteria on the most severe earthquake environment corresponding to the structure's expected service life (say 40 to 50 years) and knowledge of its stress, strain and deformation tolerance limits under dynamic loads. For important structures such as telephone central offices which house various sensitive and expensive electronic communications equipment, a more rigorous approach based on both the seismic risk and cost analyses is desirable. An optimal design strategy in terms of earthquake and structure parameters can be reached by achieving a balance in the total construction cost and the expected loss due to earthquake damage. The mathematical formulation and detailed analysis of such an optimum seismic design procedure are reported in a forthcoming paper in the B.S.T.J.[28]

Finally, it should be emphasized that the earthquake environments that are described are not intended, nor should they be construed, to be prophetic descriptions of future earthquakes. However, structures that are designed to adequately withstand these environments should consequently be expected to have a high probability of survival against earthquakes during their service life.

APPENDIX A

Approximate Solutions for the Peak Value Statistics of $x(t)$

Some approximate solutions can be obtained for the statistics of the peak amplitude of $x(t)$ in the following manner. Assume that for $\phi(t)$ in Fig. 2, the lengths of $t_1 - 0$ and $t_3 - t_2$ of the initial build-up and final decay are relatively short compared with the length $t_2 - t_1$ of the strong-motion phase of the accelerogram, and that the occurrence time $t^{(i)}$ of the extreme peak $z^{(i)}$ of a sample function $x^{(i)}(t)$ will always occur in the range $[t_1, t_2]$. These assumptions are justified from analyses of the behavior of structures under earthquake motion. Under these conditions and setting $\phi \equiv 1$, eq. (11) becomes

$$m_{|a|}(t) = \nu_n \exp\left(-\frac{a^2}{2\sigma_n^2}\right), \tag{12}$$

where $\nu_n = \sigma_{\dot{n}}/(\pi\sigma_n) = \omega_g/\pi$ is the expected number of zero crossings of the process $n(t)$ per unit time from both above and below. The problem is then reduced to a stationary one. Letting $t_0 = t_2 - t_1$ and using the logic that follows eq. (11) readily lead to the following approximate solutions:

$$F(z) = \exp\left(-\nu_n t_0 e^{-z^2/2\sigma_n^2}\right) \tag{13}$$

$$f(z) = \frac{\nu_n z t_0}{\sigma_n^2} \exp\left(-\frac{z^2}{2\sigma_n^2} - \nu_n t_0 e^{-z^2/2\sigma_n^2}\right), \tag{14}$$

$$\bar{z} \cong \left(A + \frac{\gamma}{A}\right)\sigma_n, \tag{15}$$

$$\overline{z^2} = (A^2 + 2\gamma\sigma_n^2), \tag{16}$$

and

$$\sigma_z^2 \cong \frac{\pi^2\sigma_n^2}{6A^2}, \tag{17}$$

where $A = (2 \ln \nu_n t_0)^{1/2}$ and γ is the Euler constant. If the ground acceleration is represented by a filtered white noise of constant power spectral density S, it can be shown that $\sigma_n = (S\pi/2\xi_g\omega_g^3)^{1/2}$ and that $\sigma_{\dot{n}} = \omega_g\sigma_n$, where ω_g and ξ_g are the natural frequency and damping constant of the simple ground filter. From eqs. (15) and (17) and the expression for σ_n, $\bar{z}$ and σ_z can be expressed in terms of t_0 and S.

APPENDIX B

Stationary Approximation for the Statistics of Response Process $u(t)$

The results for statistics of the response process $u(t)$ using the stationary approximation method are:

$$\sigma_u^2 = \frac{\pi\bar{S}_M B}{A_o(A_1 B - A_o A_3^2)}, \tag{18}$$

$$\sigma_{\dot{u}}^2 = \frac{\pi\bar{S}_M A_3}{A_1 B - A_o A_3^2}, \tag{19}$$

and

$$\nu_u = \frac{1}{\pi}\left(\frac{A_o A_3}{B}\right)^{\frac{1}{2}} \tag{20}$$

where $A_o = \omega_g^2\omega_o^2$, $A_1 = 2\omega_o\omega_g(\xi_o\omega_g + \xi_g\omega_o)$, $A_2 = \omega_o^2 + \omega_g^2 + 4\xi_o\xi_g\omega_o\omega_g$,

$A_3 = 2(\xi_o\omega_o + \xi_g\omega_g)$, $B = A_2A_3 = A_1$, and

$$\bar{S}_M = \frac{2\xi_g\omega_o^2\bar{z}_M^2A^2}{\pi(A^2 + \gamma)^2}. \tag{21}$$

The quantity $\bar{S}_M$ in eq. (21) is the expected uniform power spectral density of the white noise which will produce an earthquake process $x(t)$ of magnitude M with an expected peak ground acceleration $\bar{z}_M$ [see eq. (15)].

Now the response spectrum can be expressed in terms of $\bar{z}_M$. The relationship between $\bar{z}_M$ and M is shown in Fig. 3 based on the data from Housner.[23] The expected (pseudo-) velocity spectrum, defined as $S_v(\omega_o, \xi_o) = \omega_o \sup_t |u|$, is then given by

$$ES_v(\omega_o, \xi_o) = \omega_o K\sigma_u, \tag{22}$$

where $K = (2 \ln \nu_u t_o)^{1/2} + \gamma(2 \ln \nu_u t_o)^{-1/2}$. The standard deviation of the pseudovelocity spectrum, in analogy to eq. (17), is given by

$$\sigma_{s_v} = \omega_o\sigma \sup_t |u| = 1.28\omega_o\sigma_u/K. \tag{23}$$

REFERENCES

1. "Earthquake Engineering Research," Report to National Science Foundation, U. S. Dept. of Commerce Clearinghouse document, PB188636, 1969.
2. Fagel, L. W., Foss, J. W., and Liu, S. C., "The San Fernando Earthquake of February 9, 1971—Damage to Telephone Communications and Other Facilities," Unpublished work.
3. Uniform Building Code, Vol. I, International Conf. Building Officials, Pasadena, California, 1970.
4. DeCapua, N. J., Fagel, L. W., and Liu, S. C., "Earthquake Analysis of Telephone Building and ESS Equipment," Unpublished work.
5. "Earthquake History of the United States," Part I and Part II, U. S. Government Printing Office, 1961.
6. Algermissen, S. T., "Seismic Risk Studies in the United States," Proc. Fourth World Conference on Earthquake Engineering, Chile, *1* (January 1969), pp. 14–27.
7. Cox, D. R., and Lewis, P. A. W., *The Statistical Analysis of Series of Events*, London: Methuen, 1966.
8. Shlien, S., and Toksoj, M. N., "A Clustering Model for Earthquake Occurrences," Bull, Seism. Soc. Am., *60*, No. 6 (December 1970), pp. 1765–1787.
9. Vere-Jones, D., and Davies, R. B., "A Statistical Survey of Earthquakes in the Main Seismic Region of New Zealand, Part 2, Time Series Analysis," New Zealand J. Geol. Geophys, *9*, No. 3 (1966), pp. 251–284.
10. Knapoff, L., "The Statistics of Earthquakes in Southern California," Bull. Seismic Soc. Am. *54*, No. 6 (December 1964), pp. 1871–1873.
11. Isacks, B. L., Sykes, L. R., and Oliver, J., "Spatial and Temporal Clustering of Deep and Shallow Earthquakes in the Fiji-Tonga-Kernadec Region," Bull. Seism. Soc. Am. *57*, No. 5 (October 1967), pp. 935–958.
12. Ferraes, S. G., "Test of Poisson Process for Earthquakes in Mexico City," J. Geophys. Res. *72*, No. 14 (July 1967), pp. 3741–3742.
13. Page, R., "Aftershocks and Microaftershocks of the Great Alaskan Earthquake of 1964," Bull. Seism. Soc. Am., *58*, No. 3 (June 1968), pp. 1131–1168.

14. Aki, K., Hori, M., and Maturoto, H., "Aftershocks Observed at a Temporary Array Station on the Kenai Peninsula from May 19 to June 7, 1964," *The Prince William Sound, Alaskan Earthquake of 1964 and Aftershocks*, Vol. II, Parts B and C, 1969, pp. 131–156.
15. Kallberg, K. T., "Seismic Risk in Southern California," Research Rept. R69-31, Massachusetts Institute of Technology, Dept. of Civil Engineering, June 1969.
16. Vere-Jones, D., "Stochastic Models for Earthquake Occurrence," J. Royal Statist. Soc. B., *32*, No. 1 (1970) pp. 1–62.
17. Epstein, B., and Lomnitz, C., "A Model for the Occurrence of Large Earthquakes," *Nature, 211*, No. 5052 (August 1966), pp. 954–956.
18. Gumbel, E. J., "Statistical Theory of Extreme Values and Some Practical Applications," National Bureau of Standards, Appl. Math. Series 33, 1954.
19. Gutenberg, B. and Richter, C. F., "Earthquake Magnitude, Intensity, Energy and Acceleration," Bull. Seismological Soc. Amer. *46* (1956), pp. 105–143.
20. Liu, S. C., "Synthesis of Stochastic Representations of Ground Motions," B.S.T.J., *49*, No. 4 (April 1970) pp. 521–541.
21. Jennings, P. C., Housner, G. W., and Tsai, N. C., "Simulated Earthquake Motions," Tech. Rept. Earthquake Engineering Research Lab., California Inst. Tech., Pasadena, California, April 1968.
22. Liu, S. C., "On Multiplicative Processes," unpublished work.
23. Housner, G. W., "Intensity of Earthquake Ground Shocking Near the Causative Fault," Proc. Third World Conference on Earthquake Engineering, New Zealand, *1* (1965), pp. IV-94 to IV-111.
24. Liu, S. C., "Statistical Analysis and Stochastic Simulation of Ground-Motion Data," B.S.T.J., *47*, No. 10 (December 1968), pp. 2273–2298.
25. DiPol, C. V., "Seismic Design Criteria for Nuclear Power Plants," American Society of Civil Engineers National Meeting, New Orleans, La. February 1969, preprint paper no. 1791.
26. Liu, S. C., and Fagel, L. W., "Earthquake Interaction by Fast Fourier Transform," J. Eng. Mech. Div., ASCE, *97*, No. EM4 (August 1971), pp. 1223–1237.
27. Fagel, L. W., and Liu, S. C., "Earthquake Interaction for Multistory Buildings," J. Eng. Mech. Div., ASCE, *98*, No. EM4 (August 1972), pp. 929–945.
28. Liu, S. C., and Neghabat, F., "A Cost Optimization Model for Seismic Design of Structures," B.S.T.J., *51*, No. 10 (December 1972).

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 9, November, 1972
Printed in U.S.A.

A Study of Data Multiplexing Techniques and Delay Performance

By J. F. HAYES and D. N. SHERMAN

(Manuscript received May 25, 1972)

This paper is concerned with the study of data multiplexing techniques which provide local distribution for user populations whose source characteristics may be categorized as Inquiry/Response. The techniques studied are Polling, Random Access, and a Loop System. In each method a group of users is multiplexed onto the same line which is connected to a Central Processor. The Central Processor forms the interface between the local system and long haul facilities, but does no computation on arriving messages other than to route them appropriately. The advantage of such systems is flexibility of operation and economy through sharing of equipment at the Central Processor. We focus our attention on the average round-trip message delay and use this as a measure for comparisons of the three techniques.

Source models, in terms of calls per busy hour and number of bits per message for users as well as for computer responses, are developed which are appropriate in an Inquiry/Response context. Other factors taken into consideration are transmission rates, synchronization delay time, and allocation of available capacity to the transmission of overhead information, such as polling messages and acknowledgment traffic.

The results of the study are presented in graphic form, where the average delay due to traffic in the system is plotted as a function of the number of stations. Principal conclusions can be summarized as follows:

Polling—It is found that this system is sensitive to the synchronization delay which takes place each time a user station transmits to the central facility. For reasonable choices of system parameters as many as 100 stations can share the Central Processor without exceeding an average of 1-second round trip message delay.

Random Access—This system is not sensitive to synchronization delay. Over most of the range of load parameters, the Random Access system shows lower average delay than the other two systems.

Loop System—The study revealed that this system has average delay

performance comparable to the other two. The Loop system is not sensitive to synchronization delay.

I. INTRODUCTION

One of the major functions of data communication networks is providing means by which users can access a distant computer in an *interactive* manner. Time-sharing, inquiry-response and credit checking are examples of applications which have a common communications requirement. They require rapid set up times for access to their respective computers and, as a consequence of their interactiveness, often require high-capacity return channels for return traffic.

Providing access to a distant computer involves providing two major data transmission media, viz. long haul facilities and local distribution. (See Figs. 1 and 2.) While there are important problems associated with each part, we shall focus our attention on local distribution techniques. We are motivated in this direction by the relatively high cost of providing local distribution compared to the total cost of transmission.

The results presented in the sequel are most appropriate to the situation where users desiring access to distant computers are geographically clustered. In this instance, traffic from several users is concentrated at what are designated as User Stations. We shall consider multiplexing techniques whereby traffic from User Stations is conveyed to a central point which we call the Central Processor. The role of the Central Processor is to route or direct message flow to and from User Stations and computers.

This paper is devoted to an analysis of the roundtrip delay of three techniques which multiplex traffic onto common facilities. The analysis is based on models of user traffic and computer responses to user mes-

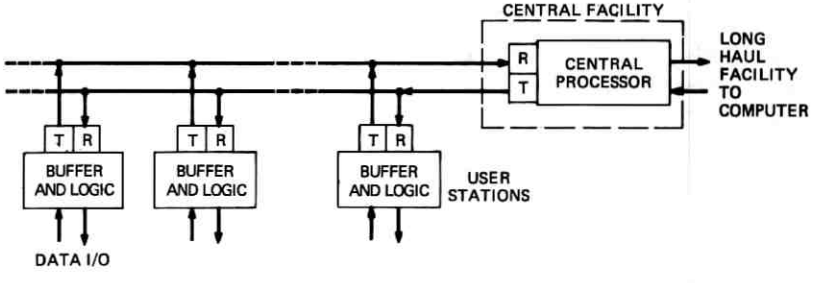


Fig. 1—Polling and random access systems.

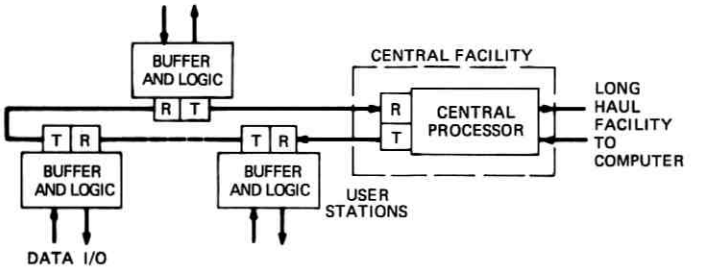


Fig. 2—Loop system.

sages. In the applications we consider, traffic from the user is bursty, i.e., short periods of activity followed by relatively long idle periods. Each message from a user elicits a response from the computer.

In the sequel we shall analyze the three alternatives for local distribution:

(*i*) Polling—(See Fig. 1) Stations are polled by a Central Processor and transmit over a common line. Return traffic is multiplexed over a second common line by the Central Processor.

(*ii*) Random Access—(See Fig. 1) Here User Stations transmit, at will on a common line, all messages that are generated. Positive acknowledgments are issued by the Central Processor upon error-free reception of the messages. Until an acknowledgment is received, the message is held in the station's buffer. Again, return messages travel over a second common line.

(*iii*) Loop System[1]—(See Fig. 2) User Stations share the same line. Traffic already on the line has priority, and newly generated messages are multiplexed into gaps in the line traffic.

In each of these techniques, traffic to and from different User Stations share common facilities. The basic difference among the techniques lies in the means of controlling traffic accessing the system through the User Station. In the Polling System, the Central Processor exercises tight control over entering messages. As we shall see, in order to do this, overhead is incurred which causes an increased delay. In contrast, entering traffic in the Random Access System is loosely controlled by the Central Processor and delay attributable to overhead is reduced considerably. In the Loop System, as in the Random Access System, messages are multiplexed on the line at the User Station without direct control by the Central Processor.

Our analysis of performance concentrates on buffering or queueing

delay. The results of this analysis will be used to calculate the roundtrip queueing delay encountered by a user-generated message. A message going to the computer from a particular station is competing for the line with messages from other stations. Consequently it encounters delay before it has sole access to shared facilities. Further, messages returning from the computer may be delayed in a queue at the Central Processor before being transmitted over the common line. We shall designate the sum of these two delays as the roundtrip delay. Of course, a message will encounter other delays such as propagation delay from the Central Processor to a distant computer and service time in the computer. However, these delays are independent of the local distribution system delays and can simply be added to the roundtrip delay that we calculate.

A common thread running through each of these systems is interactive queues. At each of the User Stations storage is assigned, if necessary, to queue up messages. However, since all the queues share the same server, the queues are not independent of one another. The exact treatment of interactive queues is mathematically difficult, therefore, in carrying out the analysis, certain approximations have been made.

The three systems described above are compared on the basis of average roundtrip delay. Since we have computed the moment generating functions of forward and return delay, higher order moments can be found easily enough. From these higher order moments one can calculate other measures of performance. In so doing, the question of correlation between forward and roundtrip delay must be considered. We did not concern ourselves with this correlation since it has no bearing on the average roundtrip delay.

In developing the models for the systems, we have attempted to take into account the constraints encountered in real communications systems. In the concluding section of the paper, examples using parameters of the Digital Data System[2] are presented.

II. GENERAL CONCLUSIONS

At the conclusion of the analytical investigations, we spend the final section of this paper in describing the delay performance of the systems. There also we compare, where possible, the advantages and disadvantages of various possible implementations. While the details appear later, it is worthwhile to give some general conclusions about the systems' behavior here.

For a given number of stations (N) and call rate (λ), the average

message delay in the Polling System depends strongly on how quickly receivers can synchronize (s) as well as the available channel capacity. For instance, one can imagine that in the Polling System, polls take place over a separate channel of capacity δ. Now let us suppose that two separate systems could be designed, the first having zero synchronization time and the other taking 10 ms for the Central Processor to synchronize the polls. We would find as the number of stations increased, that the 10 ms spent at each station ultimately degrades the performance substantially. This effect can be seen by comparing Figs. 3 and 4 where we have depicted the above situation. For a large number of stations ($\sim$ 100) the two lower right families of curves rotate counterclockwise, indicating increasing delay. Note that even if we poll with capacity δ = 100 bits/s (the upper right families of curves in Figs. 3 and 4), the delay is almost independent of the synchronization time and call rate. This is primarily due to the fact that we are forced to poll fewer stations (to maintain reasonable delay) and thus incur a smaller overhead induced delay.

If the aim of the system is to serve a small number of stations each of which has a high calling rate, the synchronization time is relatively unimportant. Thus it may be desirable to use a low speed line with a small synchronization time. In Fig. 5, we depict this situation and note that it is possible in this case to achieve similar performance for a small number of stations ($\leqq$ 25) as is achieved with high speed line (Fig. 3).

Since the Central Processor in the Random Access technique exercises loose control over traffic flow, there is a much smaller effect of overhead and correspondingly smaller delay (see Figs. 6 and 7).

In the applications we consider, the volume of return traffic is greater than the volume of forward traffic, therefore, the first station in a Loop system suffers the largest delay. In considering the Loop system we focus on this station. Analytical results obtained in the sequel indicate that the Loop System compares favorably with each of the previous techniques. In the Loop System, such quantities as synchronization time and retransmission delay do not arise. What is relevant in this system is the processing delay, which involves examining the addresses of packets passing through the station.

III. GLOSSARY OF IMPORTANT QUANTITIES

N—number of User Stations

λ—arrival rate of messages to the User Station (messages/s or calls/busy hour)

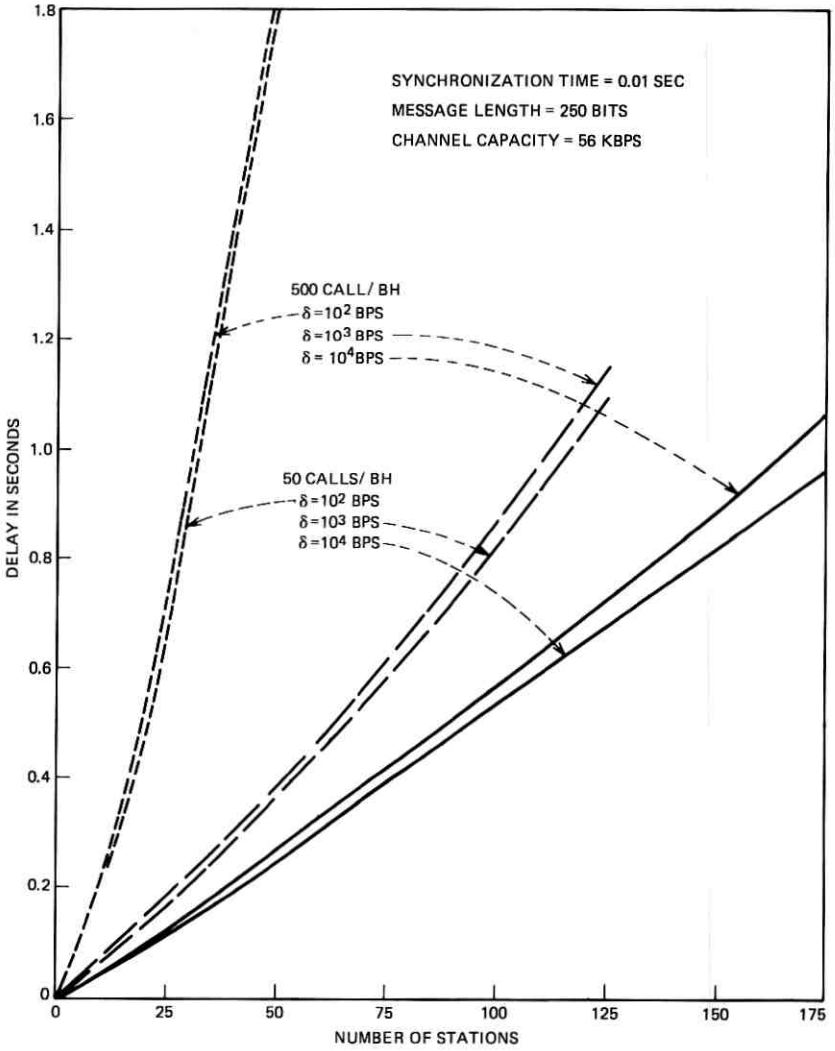


Fig. 3—Polling system delay in forward direction.

B_F—message length generated at User Stations (bits)
C_F—capacity of forward link (bits/s)
m—time duration of message while on forward link (s)
B_R—return message length (bits)
C_R—capacity of return link (bits/s)

M—time duration of message returning from the Central Processor (s)

w—minimum time between polls in Polling System and minimum retransmission time in Random Access System (s)

$\bar{t}_c$, $\overline{t_c^2}$—mean and second moment of the time required to poll and readout messages in Polling System

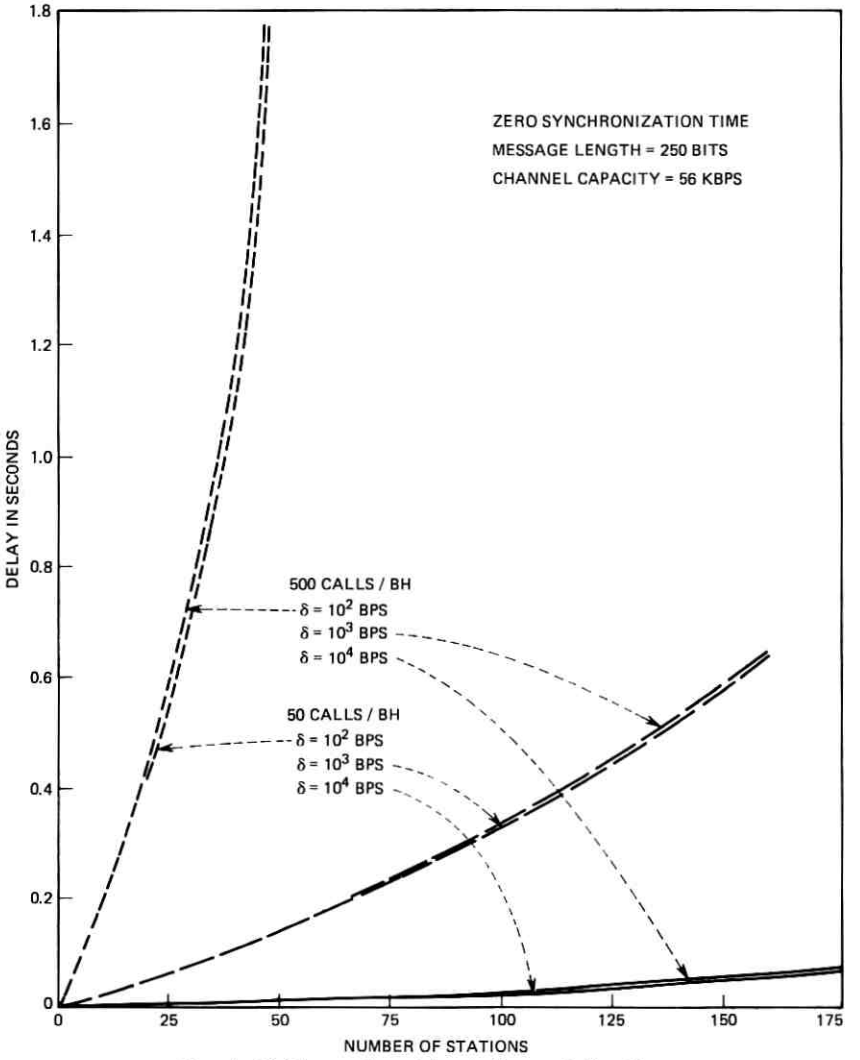


Fig. 4—Polling system delay in forward direction.

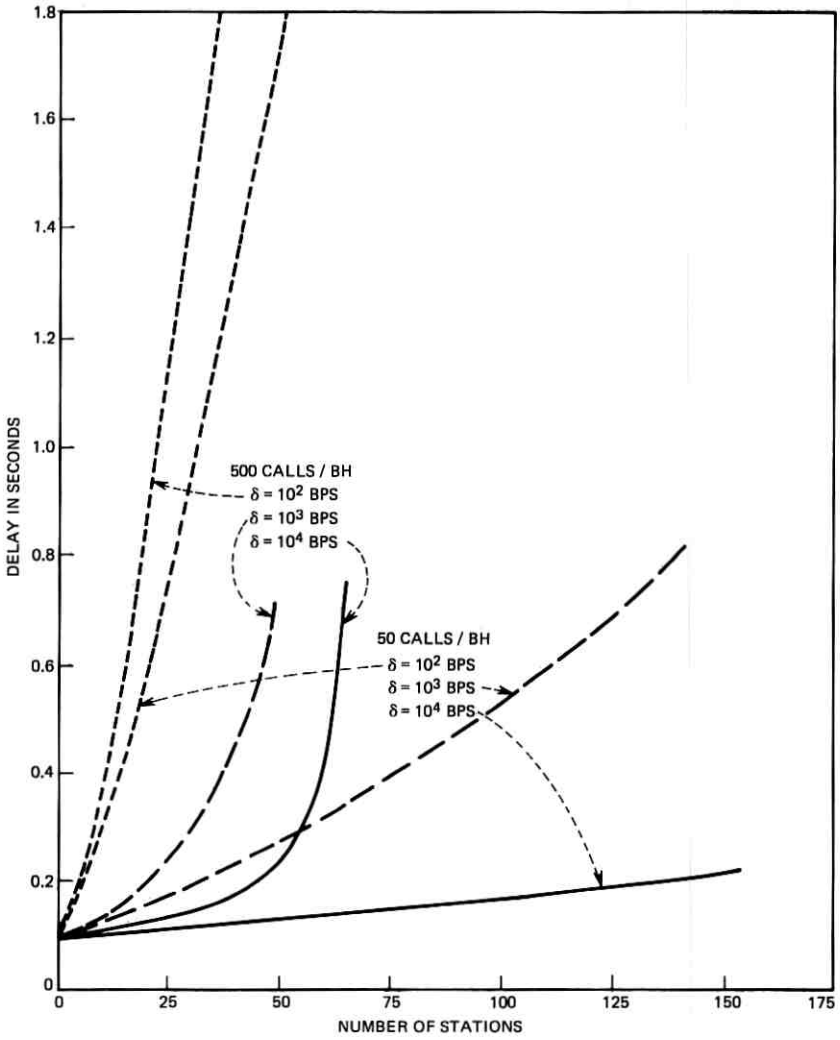


Fig. 5—Polling system delay in forward direction.

ρ—intensity of arriving traffic
d_{RA}—inbound message delay in Random Access System (s)
d_L—inbound message delay in Loop System (s)
d_P—inbound message delay in Polling System (s)
d_R—message delay on return link (s)
δ—capacity of return link allocated to polling messages in Polling

System and to positive acknowledgements in Random Access System (bits/s)

s—synchronization time (s).

IV. SOURCE MODELS

As indicated in the introduction we are primarily interested in the bursty, interactive user. Such users are encountered, for example, in credit checking and Inquiry-Response applications. The models that we shall develop are also applicable to time-sharing computer systems where users access a distant computer. Traffic characteristics of such users have been studied,[3,4] so that estimates of parameters characterizing the source (holding time, rate, etc.) are available.

We model user traffic as consisting of Poisson arrivals of short, fixed-length messages. The arrival rate is designated as λ either in calls per busy hour or in messages per second. Each message consists of B_F bits. If the line serving a User Station has a capacity of C_F bits per second then $m = B_F/C_F$ seconds are required to transmit each message.

In our analysis we shall assume that each User Station has unlimited buffer capacity so that buffer overflow never occurs. From the delay formulas that we develop, one can also derive buffer occupancy statistics, and thus imply the buffer size required for small overload probability.

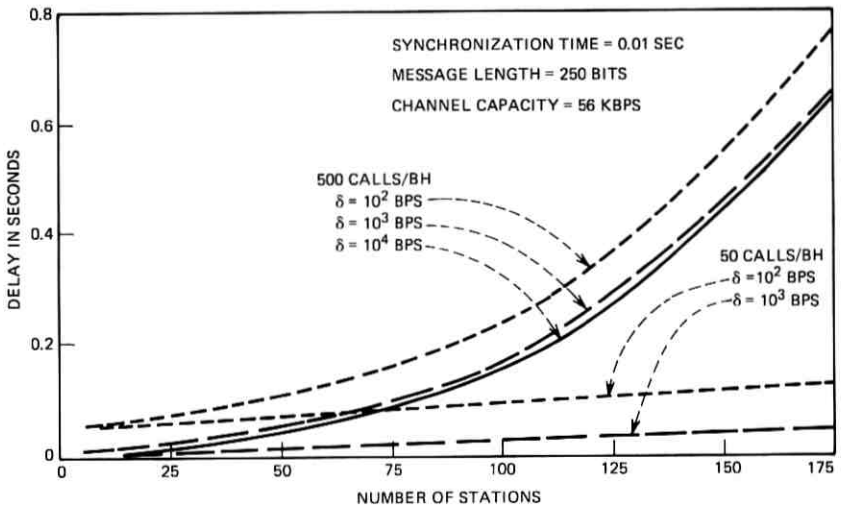


Fig. 6—Random access system, delay in forward direction, fixed-time retransmission.

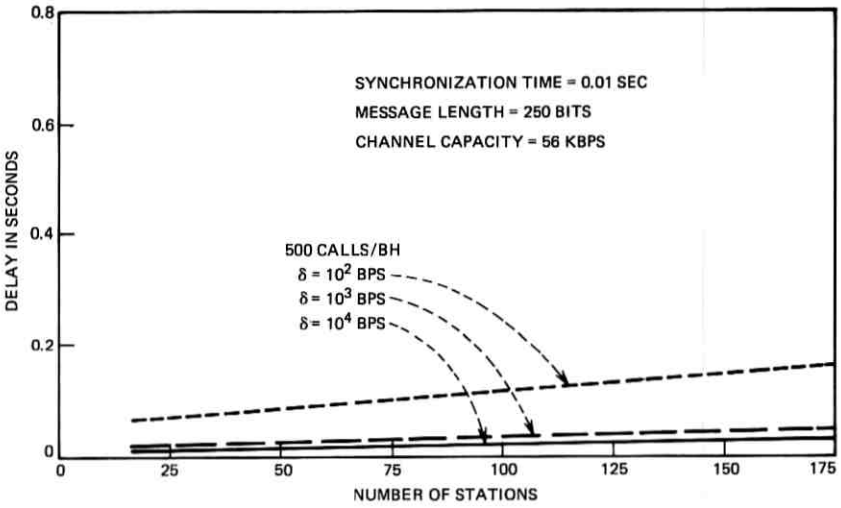


Fig. 7—Random access system, random retransmission, forward delay.

For the message arrival rates that we consider, the station buffers need not be very large in terms of message lengths to achieve adequately small overflow probability.

User messages arriving at the Central Processor are routed to a computer. Each message elicits a response from the computer. All of these return messages pass back again through the Central Processor. If messages arrive at the N user stations at average rate λ messages per second, then the average rate of return message flow through the Central Processor is $N\lambda$ messages per second. We assume that each return message has a fixed length of B_R bits, although our analysis can just as well be carried out for random message lengths.

A basic assumption in our analysis is that the return message flow is Poisson, i.e., messages arrive at the Central Processor at random. Several conditions serve to randomize the return message flow. It may well be that messages sharing the same local facilities are destined for different computers, so that the returns from different computers will tend to be uncorrelated. The computers will be serving many local communities of users, introducing a random delay between successive messages returning to the same Central Processor.

Under the assumption of a Poisson process of return messages, buffering must be provided at the Central Processor. Messages are read out of the buffer and transmitted to the User Station at a constant bit rate.

V. POLLING SYSTEM

The polling technique is not new, and the reader is referred to Martin[5,6] for discussions on the present approach and others. The type of polling under study is 'roll-call' polling in which the Central Processor works down a list of users and, in sequence, interrogates User Stations sharing the same line. Upon being polled, the User Station may transmit messages. Since the generation of messages is random with respect to the time of polling, messages must queue in a buffer located at the User Station until a polling message arrives. As indicated earlier, we focus our attention on the average delay encountered by a message.

The arrangement of lines and User Stations for the Polling System is shown in Fig. 1. User Stations are bridged across common lines. Notice that, in contrast with the star connection, only one receiver and one transmitter is required at the Central Processor. The Polling System can be implemented using a "daisy chain" configuration in which the line between two user stations loops back to the Central Processor.

Since polling messages are transmitted over a common line, User Stations must be capable of recognizing messages addressed to it. The same line is used in common by messages returning from the computer. Again the User Station must be able to recognize return messages addressed to it. As we shall see, there is an interplay between forward and return message delay, in as much as polling messages and return traffic share the same physical line.

The first component of the round-trip message delay that we consider is the delay at the User Station that a message encounters before it is put on the line. The calculation of this delay is complicated by the fact that the number of messages queueing at a particular station strongly depends upon the number of messages stored in all of the other stations. Since arrivals to any station are at a Poisson rate, the less frequently a station is polled, the more messages accumulate. Moreover, as messages accumulate, the polling cycle lengthens.

The polling problem has been analyzed by Leibowitz[7] by assuming a form of independence between queues.* The results of his analysis using this assumption compares well with the results of a rigorous analysis for a two-queue network. The nature of the independence assumption is such that the largest error occurs in a two-queue system.

* A rigorous solution of the polling problem has recently been found by Eisenberg.[8] However, in our application, which may involve a large number of stations, Leibowitz's approach leads to more tractable results.

Of primary importance in the calculation of delay in the forward direction is a parameter that we designate as the 'walk-time' between stations. In the situation where a station has no message to transmit, the Central Processor must still spend time interrogating it. Time is required to transmit a polling message. The speed with which polling messages can be transmitted is limited, since messages returning from the computer share the same line. Further, after polling a station, the Central Processor must wait for a possible response. This time involves, for example, establishing synchronization between the User Station and the Central Processor. We shall return to the walk-time after we have considered its effect on delay.

In calculating the delay of a message in this system, it is necessary to characterize the time between polls of a particular station. From the moment generating function derived in Appendix A we obtain the mean and the mean square values of the cycle time

$$\bar{t}_c = \frac{Nw}{1 - N\rho} \tag{1a}$$

$$\overline{t_c^2} = \frac{N(N-1)(w + \rho\bar{t}_c)^2 + N[w^2 + 2w\bar{t}_c\rho + \bar{t}_c m\rho]}{1 - N\rho^2}. \tag{1b}$$

In writing (1a) and (1b), we take advantage of the fact that the walk-time is a constant value, w, and the message length is a constant value $m = B_F/C_F$. Also we define $\rho = \lambda\bar{m}$.

Messages that arrive at a User Station are buffered, awaiting the arrival of the poll. When the poll arrives, messages are read out of the buffer on a first come first served basis. Thus a message must wait until the station is polled as well as until messages that have arrived before it are read out of the buffer. In Appendix A, the generating function of delay is derived. From this generating function it can be shown that the average delay is given by

$$\bar{d}_P = \frac{\overline{t_c^2}}{2\bar{t}_c}(1 + \rho) + m \tag{2}$$

where $\bar{t}_c$ and $\overline{t_c^2}$ are given by eqs. (1a and b).

Equations (1a) and (1b) show that the walk time, w, is an important parameter in the calculation of delay. We turn now to assigning values to it. Since the polling message must contain the address of the station being polled, it must be at least $\log_2 N$ bits long. Further, since it shares the line from the processor to the station with messages returning from the computer, not all of the return line capacity is available to transmit polling messages. We treat this situation by splitting the capacity of the return line. Suppose that of the C_R bit-per-second

capacity of the return line, δ bits per second, interleaved with the message bits, are allocated entirely for polling messages. The minimum time required to transmit a polling message is then

$$r = \{[\log_2 N] + 1\}/\delta \tag{3}$$

where $[x]$ is the largest integer less than x. The remaining $C_R - \delta$ bit-per-second capacity of the return line is allocated to return message traffic. This allocation may be effected by either FDM or TDM.

Other methods may be used to share the return line between polling and return messages. For example, messages may be transmitted over the same channel with priority given to polls or user traffic. The difficulty with this approach is that there may be a large difference in the lengths of the two kinds of messages. Thus, even though a polling message is short and has priority, it may wait a relatively long time until a transmission of a return message is completed.

Since User Stations continuously receive from the same point, the Central Processor, no time is required to establish synchronization for return or polling messages. This may not be true in the reverse direction since the Central Processor is receiving from a different station on each poll. Thus in the calculation of walk-time it may be necessary to allow time for the establishment of synchronization between the User Station and the Central Processor. Let this time be denoted s. The walk-time between stations is then the sum.

$$w = s + \{[\log_2 N] + 1\}/\delta. \tag{4}$$

The effect achieved by splitting return line capacity on return message delay can now be calculated. Return messages are B_R bits in duration and arrive at the central processor at a Poisson rate of $N\lambda$ messages per second. The time required to transmit each message is $B_R/(C_R - \delta)$ seconds. The delay of a message in the Central Processor may be found from the analysis of an M/G/1 queue.[9] It can be shown that the average return message delay is

$$\bar{d}_R = \frac{N\lambda[B_R/(C_R - \delta)]^2}{2[1 - B_R\lambda N/(C_R - \delta)]} + B_R/(C_R - \delta). \tag{5}$$

The average roundtrip delay is the sum of $\bar{d}_P$ and $\bar{d}_R$.

VI. RANDOM ACCESS SYSTEM

The Random Access system was suggested by the University of Hawaii's ALOHA System.[10] The configuration of this system is the same as shown for the Polling System in Fig. 1. As in the Polling System, messages arrive at each User Station at rate λ messages per second.

Messages are transmitted to the Central Processor on a first come first served basis. The length of a message in bits and the line rate are such that m seconds are required to transmit each constant length message. This constant length includes overhead bits required by the Random Access technique.

Messages are transmitted at random by each station in the system. There is a probability of messages from different stations overlapping. In order to detect message overlap, parity check bits are transmitted along with each message and error detection is performed upon reception. Only if a message is received error free is an acknowledgment transmitted to the User Station. Until a User Station recognizes a positive acknowledgment addressed to it, the message is held in a buffer. If, after a specified period of time, no acknowledgment is received, the message is retransmitted. This continues until an acknowledgment for a particular message is received.

Messages returning from the computer are buffered and sent to the User Stations in the same manner as the Polling System. As in the Polling System, messages must contain addresses since messages destined for different stations share the same line.

A basic relation governing traffic in the Random Access system has been derived by Abramson.[10,11] In this analysis, messages are assumed to enter the common line at Poisson rate. Suppose that a particular message is transmitted at time t_o. This message will encounter no interference if no other message is put on the line in the time interval $t_o \pm m$. We have then

$$Pr \text{ [message retransmission]} = 1 - \exp(-2Rm) \tag{6}$$

where R is the total transmission rate on the line, including retransmissions. The rate of retransmission on the line is $R(1 - e^{-2Rm})$.

Now messages arrive at all user stations at rate λN messages per section. If there is no continuous buildup of messages at User Stations, we must have the relationship.

$$R = N\lambda + R(1 - \exp(-2Rm)).$$

This reduces to

$$N\lambda m = Rm \exp(-2Rm). \tag{7}$$

An examination of eq. (7) discloses that the maximum information transmission rate is $N\lambda m = 1/2e \approx 0.18$. As one attempts to exceed this, the retransmission rate increases so fast that less information gets through.

The basic assumption in this analysis is that line traffic, including retransmissions, forms a Poisson process. Simulation studies[12] show that eq. (7) holds up to $N\lambda m \cong 0.1$. For higher loadings, this simulation shows a much higher retransmission rate and, presumably, the Poisson assumption does not hold.

A basic quantity in the calculation of delay is the time that the User Station waits before retransmission if no positive acknowledgment for a prior transmission is received. Let us consider how this 'timeout' period affects the system performance that a user might experience. If two messages do interfere, then each station must initiate a retransmission. If the times of each retransmission are the same, then interference will persist. Thus some strategy must be used to avoid persistent interference as well as to provide for a short 'timeout' delay. To more fully appreciate what is meant by 'short' we have analyzed two strategies for retransmission.

The first strategy assigns to each station a fixed or designated 'timeout' delay. This approach has the advantage that it completely avoids persistent interference. It has the disadvantage that some stations will have large (on the order of the number of stations times a fixed delay per station) delay.

The second strategy (called randomized retransmission) attempts to make use of the fact that it is generally two stations that interfere, and that it is not necessary to distinguish between all N stations in the system. This approach asks interfering stations to select a retransmission time by selecting it from a random sequence of retransmission times. If each station has a different sequence, then there is a small probability of persistent interference. The approach has the advantage of shortening the retransmission delay, but has the disadvantage that there is a non-zero probability of repeated interference.

In the next two subsections we discuss the delay analysis for these strategies. Detailed calculations may be found in Appendix B.

6.1 *Fixed Timeout Delay*

Consider first the Random Access System implemented with fixed timeout retransmission. Suppose that a message is transmitted from a User Station at time $t = 0$. Since we must allow for synchronization and for the transmission of an acknowledgment message, the message can be acknowledged no sooner than time $t = s + \{[\log_2 N] + 1\}/\delta + m$. If no acknowledgment is received at this time, station i waits for $2im$ seconds ($i = 1, 2, \cdots, N$) and retransmits. This is repeated until a positive acknowledgment is received. If we assume that interferences

on successive retransmissions are independent, we have from eq. (6)

$$Pr\,[\text{message clearance time} = w + m + kT_i]$$
$$= Pr\,[k \text{ retransmissions}] = [1 - \exp(-2Rm)]^k \exp(-2Rm) \quad (8)$$

where $w = \{[\log_2 N] + 1\}/\delta + s$ and $T_i = w + 2mi$.

Average message delay at each station as well as higher moments of delay may be calculated by calling upon results from the analysis of an M/G/1 queue.[9] Messages arrive at a Poisson rate of λ messages per second and each message has a geometrically distributed service time as given by (8). From the generating function derived in Appendix B, it can be shown that the average delay for a message at station i is

$$\bar{d}_{RA} = \frac{2[wP - 2Nm(1-P)] - \lambda[w^2P + m^2 + (2Nm)^2(1-P)]}{2\{1 - \lambda[wP + m - 2Nm(1-P)]\}} + m \quad (9)$$

where $P = \exp(2Rm)$.

The return message delay in the random access system is the same as in the Polling System. Fixed length messages of M bits return at rate $N\lambda$ messages per second, with $C_R - \delta$ bits per second capacity available for transmission. The resulting message delay is given by eq. (5). The roundtrip delay is the sum of $\bar{d}_{RA}$ [eq. (9)] and $\bar{d}_R$ [eq. (5)].

6.2 *Randomized Retransmission Delay*

We turn now to consider a random retransmission technique. After an initial transmission, the User Station waits a minimum of $s + ([\log_2 N] + 1)/\delta$ seconds. If no acknowledgment has been received, the message is retransmitted after a randomly-distributed time interval. If no acknowledgment is received after this second transmission, the process is repeated. In our calculations we have taken the random timeout interval to be exponentially distributed with mean $1/\alpha$.

The probability of interference on the initial transmission is given by eq. (6). On subsequent retransmissions the probability of interference depends on α. For example, if α is very large, then retransmission for the two interfering stations occurs almost immediately after

$$t = s + \frac{[\log_2 N] + 1 + m}{\delta}$$

and the probability of interference is high. As α decreases, the probability of interference decreases. We approximate the probability of inter-

ference on retransmission as

$$Pr\ [\text{retransmission interference}] = 1 - \exp\ [-2m(R + \alpha)]. \quad (10)$$

The average time between retransmissions is

$$\bar{l}_r = s + \{[\log_2 N] + 1\}/\delta + 1/\alpha + m. \quad (11)$$

In Appendix B, the generating function of the time required to clear a message from a station's buffer is derived. From this we can show that the mean and the mean square times to clear a message from a station buffer are respectively

$$\bar{b} = w \exp(-2Rm) + X(1/\alpha + 2w) \exp\{+4m(R + \alpha)\} + m \quad (12a)$$

$$\overline{b^2} = w^2 \exp(-2Rm) - \frac{2Xw^2}{\alpha \exp\{-4m(R + \alpha)\}} + \frac{2X(1 + 2w\alpha)(1 + w\alpha + w\alpha Y)}{\alpha^2 \exp\{-6m(R + \alpha)\}} + 2m\bar{b} + m^2 \quad (12b)$$

where

$$X = [1 - \exp(-2Rm)] \exp\{-2m(R + \lambda)\}$$

and

$$Y = 1 - \exp\{-2m(R + \lambda)\}.$$

From the theory of the M/G/1 queue, it can be shown that the average delay is

$$\bar{d}_{RA} = \bar{b} + \frac{\alpha\overline{b^2}}{2(1 - \alpha\bar{b})} + m \quad (13)$$

where $\bar{b}$ and $\overline{b^2}$ are given by eqs. (12a) and (12b).

Notice from (10) and (11) that decreasing α reduces the probability of interference while increasing the average retransmission delay. Thus there is an optimum value of α that balances these effects, yielding a minimum average delay.

VII. LOOP SYSTEM

The configuration of User Stations and Central Processor for the Loop System is shown in Fig. 2. As in the Polling and Random Access Systems, the line between User Stations can be looped back to the Central Processor. Traffic flow on the line is in terms of fixed size message slots. Messages arriving at a User Station are multiplexed

into these slots, one message to each slot. Traffic that is already on the line has priority; consequently a User Station must wait for an empty message slot. Traffic returning from the computer is addressed. At each station these addresses are examined. If a message is addressed to a particular station, it is taken off the line. Notice that in the Loop System, return and forward messages share the same line, consequently the total line length is half that required for Polling and Random Access Systems serving the same User Stations.

Messages arrive at each User Station at a Poisson rate of λ messages per second. These messages are multiplexed on a first come first served basis. If a message arrives when the buffer is empty, it still must wait until the line is free before it can be multiplexed.

Analytical and simulation results for this system have been presented in Refs. 13, 14, and 15. It was shown that the average message delay is approximated by the expression.

$$t_L = m(1 + \rho_L) + \frac{m\rho(1 + \rho_L)^2}{2[1 - \rho(1 + \rho_L)]} + \frac{\bar{l}^2}{\bar{l}} \frac{\rho_L}{2(1 + \rho_L)[1 - \rho(1 + \rho_L)]} \tag{14}$$

where ρ_L is the ratio of the average durations of line busy and idle periods. $\bar{l}$ and $\bar{l}^2$ are respectively the mean and the mean square values of the durations of the line busy periods. In deriving eq. (14) we make use of the fact that the message length is a constant value of m seconds.

A crucial quantity in the calculation of message delay is the line busy period. The characterization of the line busy period is simplified somewhat if we look at delay for what is usually the most critical station. In many applications the User Station receives more data than it transmits and as a consequence the traffic diminishes as one moves around the loop from the Central Processor.

We have modeled the traffic to the Central Processor as consisting of Poisson arrivals of fixed length messages. Messages arrive at rate $N\lambda$ messages per second and $M = B_R/C_R$ seconds are required to transmit each message. The busy period of the line out of the Central Processor is the busy period of an M/D/1 queue. It can be shown that[9]

$$\bar{l} = \frac{B_R/C_R}{1 - \rho_R} \tag{15a}$$

and

$$\bar{l}^2 = \frac{(B_R/C_R)^2}{(1 - \rho_R)^3} \tag{15b}$$

where

$$\rho_R = B_R \lambda N / C_B .$$

The average duration of a line idle period is $1/N\lambda$. We have then

$$\rho_L = \frac{B_R N \lambda / C_R}{1 - \rho_R} = \frac{\rho_R}{1 - \rho_R}. \tag{16}$$

In order to provide a valid comparison with the other two systems, we must consider the processing time of a message at each station. A message, generated at the first station after the computer, passes through all of the other stations on the loop. At each station the address of the message is examined, entailing a delay of $\{[\log_2 N] + 1\}/C_R$ where $[x]$ is the largest integer less than x. Notice that since each station is receiving continuously from the same adjacent station, there is no synchronization delay. The cumulative delay of a message in going from the first User Station to the Central Processor is then

$$\bar{d}_L = \bar{l}_L + N([\log_2 N] + 1)/C_R \tag{17}$$

where $\bar{l}_L$ is given by eq. (14).

The return message delay for the Loop System is similar to the return message delay in the previous two systems. The delay in seconds is given by eq. (5) with $\delta = 0$. Thus the roundtrip delay is the sum of $\bar{d}_L$ [eq. (17)] and $\bar{d}_R$ [eq. (5)].

VIII. EXAMPLES OF SYSTEM BEHAVIOR

In this section, results of computations using the equations for *average* delay derived in the foregoing are presented. In presenting these results we were faced with the difficulty of choosing sets of parameters that provide meaningful comparisons. There is such a wide latitude in the choice of parameters for each of the three systems that one could easily bury the reader in a mass of curves and tables. Therefore, we have limited ourselves to relatively few cases illustrating system behavior. It is not difficult to supplement the results we present here since the expressions we have derived for average delay are relatively easy to evaluate.

Calculations have been made using values of user-related parameters (λ, B_F and B_R) which are appropriate in an Inquiry-Response context. Values for those parameters related directly to implementation (C_F, C_R and s) are chosen with the Digital Data System[2] in mind.

The results of the computations are shown in the form of sets of

curves where average message delay is shown as a function of the number of user stations in the system. The rate at which messages arrive at User Stations, λ, is given two values, 50 calls per busy hour and 500 calls per busy hour, representing, respectively, light and heavy calling rates. In the application we have in mind, messages in the forward direction tend to be short. We have taken this message length to be $B_F = 250$ bits. Three values were used for the lengths of messages returning from the computer 2.5×10^3, 5×10^3, and 10^4 bits. We note that this latter value is approximately the number of bits required to fill a CRT display.

Since the systems we consider achieve economies by sharing, it is reasonable to put as many stations as possible in a system by choosing high line capacities C_F and C_R. In the Digital Data System, for example, 56 kbits per second are available transmit to local stations over short distances. Thus we take $C_R = C_F = 56 \times 10^3$ bits per second. This choice is tempered by the fact that, in order to achieve this high rate, synchronous operation is required. As we have noted in connection with the Polling and Random Access systems, the Central Processor receives data from different stations on each transmission. A low estimate for the time required to adjust synchronization from reception to reception at this speed is $s = 10$ ms. As we shall see, this value of s may lead to high values of delay for the Polling System. Therefore, for comparison we examine Polling Systems where transmission in the forward direction is asynchronous with $C_F = 2400$ bits per second and $s = 1/2400$ seconds. We have also examined fully synchronous operation where $s = 0$ and $C_F = 56 \times 10^3$ bits per second.

The parameter δ comes into play in the Polling and the Random Access system. Recall that δ is the portion of the return channel capacity allocated to the transmission of polling messages (Polling System) or positive acknowledgments (Random Access System). By varying this parameter, message delay in the forward direction is traded off against message delay in the reverse direction. For any particular system configuration, there is an optimum value for δ. However we have examined the effect of varying this parameter by choosing $\delta = 10^2$, 10^3 and 10^4 bits per second.

Sets of curves of delay in the forward direction for the Polling System are shown in Figs. 3, 4 and 5. In Figs. 7 and 8, return delay is shown as a function of the number of stations. As one might expect, traffic characteristics are important in judging the merits of implementations of the Polling System. For $\lambda = 50$ calls per busy hour, the 2,400 bps implementation (Fig. 5) yields lower delay than in the 56 kbps system

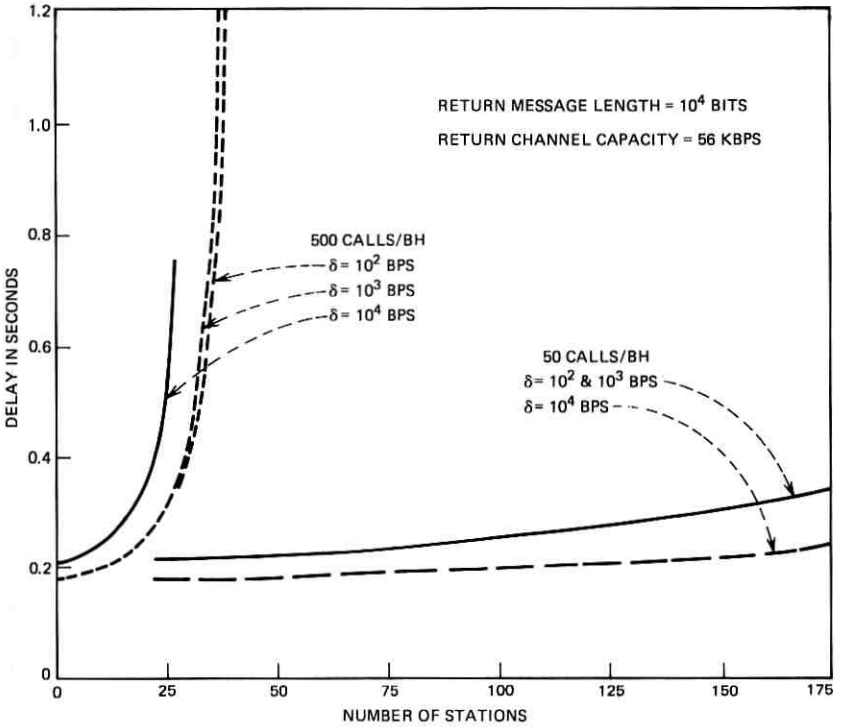


Fig. 8—Polling and random access system return delay.

with 0.01-second synchronization time (Fig. 3) in almost all cases. In fact, at this calling rate the fully synchronous 56 kbps system (Fig. 4) has delay which is lower than that of the 2,400 bps system by a relatively small amount. This advantage disappears under the heavy calling rate (500 calls per busy hour). The lower speed implementation is far more sensitive to calling rates.

Delay is sensitive to the parameter δ, which is that portion of the return channel allocated to polling messages. It is clear from that, for the traffic we consider, $\delta = 10^2$ bps is entirely too small. By increasing δ to 10^3 bps, there is a large decrease in forward delay and a small increase in return delay. It is not clear that further advantage is obtained if δ is increased to 10^4 bps. On Figs. 8 and 9, return delay may be very large for $\delta = 10^4$ bps and, for the same number of stations, low for $\delta = 10^3$ bps (e.g., $N = 130$ stations, $M = 2.5 \times 10^3/56 \times 10^3$ seconds on Fig. 9). For any given set of traffic characteristics there is an optimum value of δ which minimizes total delay.

The results of the computation of average forward delay for the Random Access System are shown on Figs. 6 and 7. The curves of average return delay are the same as for the Polling System and are shown on Figs. 8 and 9. Unlike the Polling System, the effect of s and δ does not accumulate with the number of stations, and consequently is not very sensitive to these parameters. The Random Access System is more sensitive to calling rate than the Polling System primarily because calling rate affects the probability of a message being retransmitted [see eqs. (6) and (7)]. In Fig. 6, the results shown are for the station with the longest fixed timeout interval ($i = N$) in eq. (11).

The fixed-time retransmission implementation of the Random Access System (see Fig. 6) compares very well with the Polling System. For example, for $s = 0.01$ seconds, $\delta = 10^3$ bits per second and $\lambda = 500$ calls per busy hour, the 100 station delay in the Polling System is nearly 0.9 seconds, (see Fig. 3) while the corresponding delay in the Random Access System is less than 0.2 seconds (see Fig. 6). The Random Access System with fixed-time retransmission also performs well in comparison with the $s = 0$ implementation of the Polling System.

Recall that in the Random Access random retransmission strategy, the timeout interval is exponentially distributed with mean $1/\alpha$. By a process of trial and error we have found that $\alpha = N\lambda$ yields a rough minimum of average delay for $N > 10$ stations. This value was used

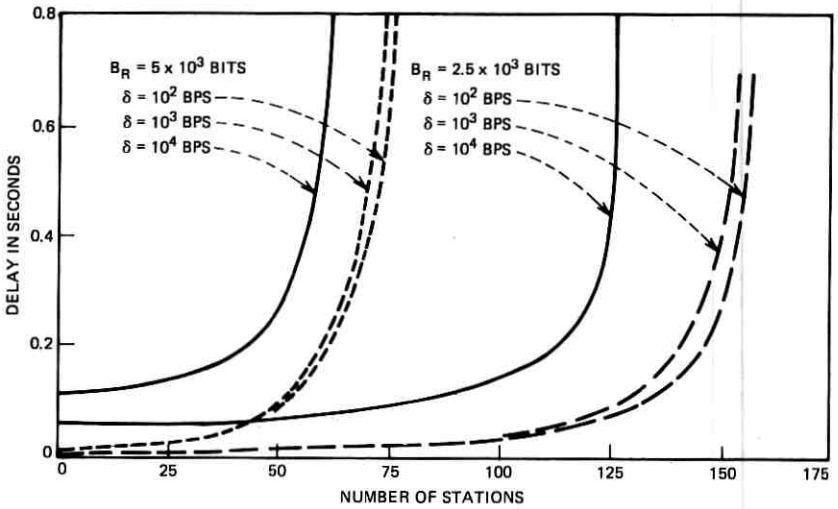


Fig. 9—Polling and random access system return delay; $\lambda = 500$ calls/BH.

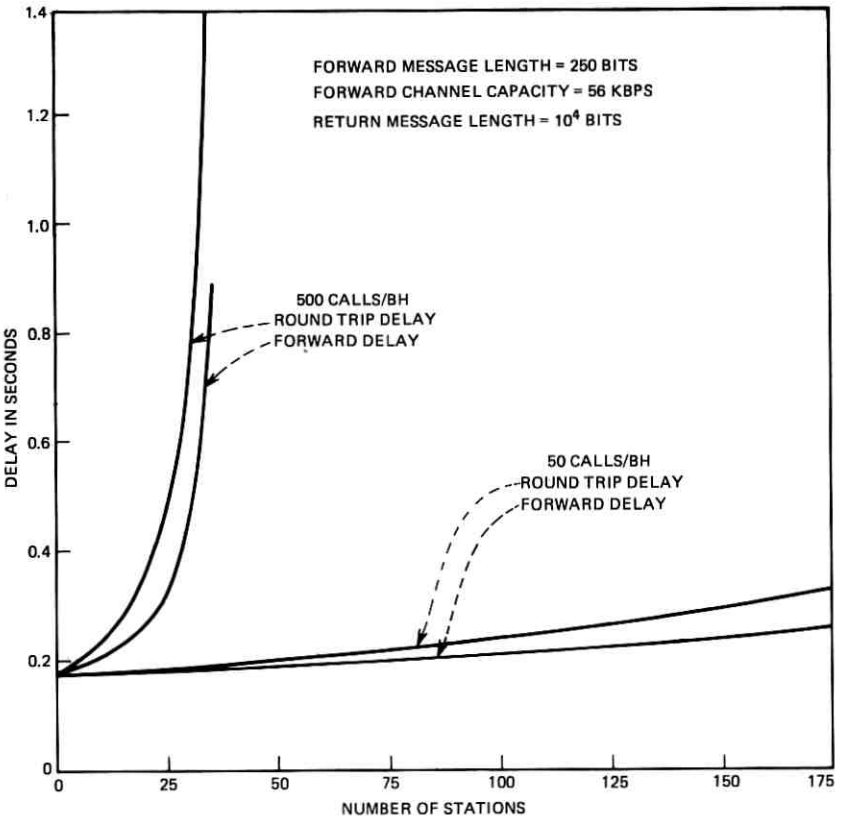


Fig. 10—Loop system forward and roundtrip delay.

to obtain the results shown in Fig. 7. As we see from Figs. 6 and 7, this realization of the Random Access System yields performance that is superior to fixed-time retransmission. Random retransmission compares favorably with the best implementation of the Polling System (see Fig. 4).

Our comparison may be biased somewhat in favor of the Random Access System since we have not taken into account overhead in that system. Recall that in order to detect errors, parity check bits along with information bits may be transmitted from the User Station. This will lengthen the message from the 250 bit message we have considered. A larger value of m will cause increased delay by increasing the probability of retransmission [see eqs. (6) and (7)] and by increasing the timeout interval [see eq. (11)]. However, we felt that it would not

be necessary to lengthen the message very much, and consequently delay would be approximately the values we have shown.

The results of average delay calculations for the Loop System are shown on Figs. 9 and 10. Forward and roundtrip delay for the station on the line immediately after the Central Processor are shown. Roundtrip delay for the Loop System is even more sensitive to return message duration than the Polling or Random Access Systems. This is because messages going from the user terminal to the Central Processor are blocked by return messages. However, in the stable region below the knee of the delay curve, the Loop System gives performance comparable to the best implementation of the Polling and Random Access Systems. For example, for $M = 2.5 \times 10^3/5.6 \times 10^4$ seconds, $\lambda = 500$ calls per busy hour, and $N = 100$ stations, the following average roundtrip delay estimates are obtained for each of the three systems: Loop System: 0.12 seconds (Fig. 11), Polling System: 0.16 seconds (Figs. 4 and 9), and Random Access System: 0.16 seconds (Figs. 7 and 9).

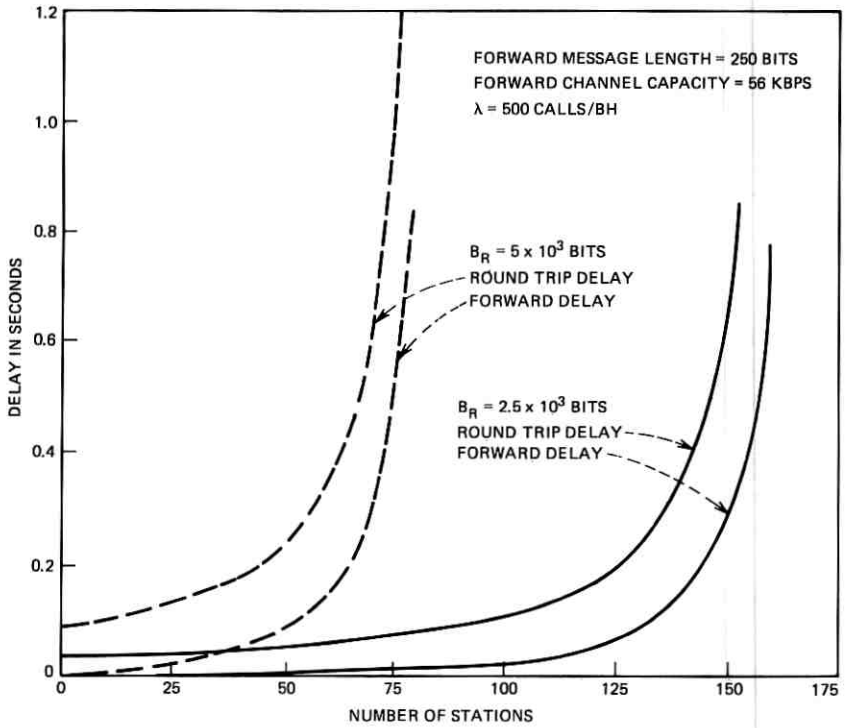


Fig. 11—Loop system forward and roundtrip delay.

IX. ACKNOWLEDGMENTS

We would like to thank R. J. Pilc for initial suggestions on the applicability of polling to local distribution, and M. Eisenberg for his assistance in developing the structure of the queueing model in the polling system.

APPENDIX A

In this appendix the moment generating function of forward delay in a Polling System is derived. Although other schedules may be analyzed, we concentrate on the situation where each station is polled once and only once in a cycle. Messages arrive for multiplexing at the User Stations at a Poisson rate of λ messages per second. In carrying out the derivation, we shall denote the generating function of the message length in seconds as $M(u)$. The generating function of the walk-time is denoted as $W(u)$. In the text, we shall apply our results to the case where message length and walk times are constants.

We consider first the "cycle time" of the polling sequence. This quantity is the time interval between polls of a particular station. As the polling sequence goes through a complete cycle, a random number of messages is encountered in each station's buffer. The analysis is simplified considerably if we assume that this number of messages is independent and identically distributed from station to station. Under this assumption, Leibowitz[7] shows that the moment generating function of the number of messages in each station's buffer at the time of polling is given by

$$P(x) = \{W(\lambda - \lambda x)P(\phi)\}^N \tag{18}$$

where $\phi = M(\lambda - \lambda x)$.

Since messages arrive at a Poisson rate, a relationship between the number of messages in the buffer and the cycle time can be derived.

$$Pr\ [n \text{ messages in a buffer at polling time cycle time } = \tau] = \exp\ (-\lambda\tau)(\lambda\tau)^n/n!.$$

Averaging over the cycle time we have

$$Pr\ [n \text{ messages in a buffer at polling time}] = \int_0^\infty \frac{\exp\ (-\lambda\tau)(\lambda\tau)^n}{n!} p_c(\tau)\, d\tau$$

where $p_c(\tau)$ is the probability density of the cycle time. By taking the Laplace-Stieltjes transform here it can be shown that,

$$P(x) = T_c(\lambda - \lambda x) \tag{19}$$

where $T_c(s)$ is the moment generating function of the cycle time. From (18) and (19) we have

$$T_c(u) = [W(u)T_c(\lambda - \lambda M(u))]^N. \tag{20}$$

Differentiating (20) and setting $u = 0$ yields the results on mean and mean square delay shown in eqs. (1a) and (1b) respectively.

Now a message arriving at a station must wait until the station is polled, and until all messages that have arrived before it have been multiplexed on the line. We first derive the generating function of the time the message must wait in the queue. We shall work under the assumption that messages arriving while prior messages are being multiplexed must wait until the next poll. Thus the queueing time of a customer can be written

$$d_P = \tau + \sum_{i=1}^{k} m_i \tag{21}$$

where τ is the time interval until the next poll and m_i , $i = 1, 2, \cdots, k$ are the lengths of k prior messages in the buffer. From (21) we can write the density of the delay d_P as

$$P(T) = \Pr[T < d_P \leqq T + dT] = \int_0^\infty d\tau \sum_{k=0}^{\infty} p_m^{(k)}(T - \tau)p(k, \tau) \tag{22}$$

where $p_m(s)$ is the probability density of the message length and $p(k, \tau)$ is the joint density of k and τ. It can be shown from elementary probabilistic arguments that

$$p(k, \tau) = \int_\tau^\infty dx \frac{p_c(x)}{\bar{t}_c} \frac{\lambda(x - \tau)^k}{k!} \exp\{-\lambda(x - \tau)\} \tag{23}$$

where $p_c(x)$ is the probability density of the cycle time. From (22) and (23) it can be shown that the moment generating function of the queueing delay is

$$T_{PQ}(u) = \frac{T_c\{\lambda - \lambda M(u)\} - T_c(u)}{\bar{t}_c[u - \lambda + \lambda M(u)]} \tag{24}$$

where $T_c(u)$ is the moment generating function of the cycle time. In order to find the total message delay, we must add the multiplexing time to the queueing delay. Since these quantities are independent random variables, the moment generating function of the message delay is

$$T_P(u) = T_{PQ}(u) \cdot M(u). \tag{25}$$

By differentiating $T_P(u)$ with respect to u and setting $u = 0$, the expression for delay given by eq. (2) of the text can be found.

APPENDIX B

In this appendix the generating function of message delay in the forward direction for the Random Access System is derived. Recall that after transmitting a message, a User Station waits for a positive acknowledgment from the Central Processor. The minimum time required to receive a positive acknowledgment is

$$w = \frac{[\log_2 N] + 1}{\delta} + s. \tag{26}$$

If no acknowledgment is received, the original message is retransmitted. We consider two transmission strategies; a fixed transmission time which is different for each station, and a random transmission time.

Under the fixed retransmission strategy, each station's transmission interval is different by at most $2m$ seconds, thus the retransmission interval for the ith station is $2mi$. This interval obviates the possibility of thc samc two stations repeatedly interfering with one another. If k retransmissions of a message are necessary to clear a message from the station's buffer, the total clearance time is

$$\tau = w + m + k(w + mi) \tag{27}$$

for the ith station.

If message flow on the line is Poisson, then the probability of two messages interfering with one another is

$$\Pr[\text{interference}] = 1 - \exp(-2Rm). \tag{28}$$

If we assume that subsequent retransmissions are independent trials, the total number of trials required to clear a message is geometrically distributed. We have

$$\Pr[k \text{ retransmissions}] = [1 - \exp(-2Rm)]^k \exp(-2Rm). \tag{29}$$

From (27) and (29) the moment generating function of the time required to clear a message is

$$S_{1i}(u) = E\{\exp(-u[w + m + kT_i])\} = \exp(-um)[\exp(uw + 2Rm) - \exp(2Rm - 2uim) + \exp(-2uim)]^{-1} \tag{30}$$

where

$$T_i = w + 2mi.$$

In the language of queueing theory, $S_{1i}(u)$ corresponds to the generating function of the service time of a customer (message). Messages arrive for multiplexing at a Poisson rate of λ messages per second. The message delay can be found from the theory of the M/G/1 queue. We have

$$D^i_{RA}(u) = \frac{(1 + \lambda S'_{1i}(0))uS_{1i}(u)}{u - \lambda + \lambda S_{1i}(u)}. \tag{31}$$

By substituting (30) into (31), differentiating with respect to u and setting $u = 0$, the expression for the mean value of delay given in eq. (9) can be found.

The derivation of the generating function of message delay for random retransmission proceeds along the similar lines. The retransmission interval is a geometrically-distributed random variable with mean value $1/\alpha$ for each station. If k retransmissions are required, then the time to clear a message is

$$\tau = (k + 1)w + m + \sum_{j=1}^{k} \xi_j \tag{32}$$

where ξ_j, $j = 1, 2, \cdots, k$ are exponentially distributed random variables. The distribution of k is slightly different than in the case of fixed time-out retransmission. If two stations have interfered, there is a non-zero probability that they will interfere on the subsequent retransmission. We account for this phenomenon by approximating the probability of retransmission by the expression

$$\Pr[\text{interference on retransmission}] = 1 - \exp(-2(R + \alpha)m). \tag{33}$$

The probability of no retransmissions is

$$\Pr[\text{no retransmission}] = \exp(-2Rm). \tag{34}$$

The probability of k retransmissions is

$$\Pr[k \text{ retransmissions}] = [1 - \exp(-2Rm)][1 - \exp(-2(R + \alpha)m)]^{k-1} \cdot \exp(-2m(R + \alpha)); \qquad k = 1, 2, \cdots. \tag{35}$$

From (31), (33) and (34) it can be shown that the generating function of message clearance is

$$S_{2i}(u) = E\left[\exp\left(-(k+1)w + m - \sum_{i=1}^{k}\xi_i\right)u\right]$$

$$= \exp(-um)\left[\exp(-2Rm - wu) + \frac{\lambda[1 - \exp(-2Rm)]\exp(-2m(R+\alpha))\exp(-2wu)}{\lambda + u - \lambda\exp(-wu)[1 - \exp(-2m(R+\alpha))]}\right]. \quad (36)$$

The generating function of message delay can be found from the theory of the M/G/1 queue. The generating function of message delay is the same as in (31) with $S_{1i}(u)$ replaced by $S_{2i}(u)$. The first two moments of the time needed to clear a message given by (12a) and (12b) are found from successive differentiations of (36). The formula for average delay given in (13) is well known.[9]

REFERENCES

1. Pierce, J. R., "Network for Block Switching of Data," IEEE Conv. Rec. (March 1971).
2. Muench, P. E., "Bell System Private Line Digital Data Service," IEEE Conv. Rec. (March 1971).
3. Jackson, P. E., and Stubbs, C. D., "A Study of Multiaccess Computer Communications," AFIPS, Conf. Proc., *34*, p. 491.
4. Jackson, P. E., and Fuchs, E., "Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models," Proc. ACM Conf., Pine Mountain, Georgia (October 1969).
5. Martin, J., *Teleprocessing Network Organization*, Englewood Cliffs: Prentice-Hall Inc., 1970.
6. Martin, J., *Design of Real-Time Computer Systems*, Englewood Cliffs: Prentice-Hall Inc., 1967.
7. Leibowitz, M. A., "An Approximate Method for Treating a Class of Multiqueue Problems," IBM J. Res. Development, *5*, No. 3, 1961, pp. 204–209.
8. Eisenberg, M., "Queues with Periodic Service and Changeover Time," to be published, J. Oper. Res.
9. Cox, D. R., and Smith, W. L., *Queues*, Methuen and Co. Ltd., 1961, pp. 52–59.
10. "The ALOHA System," University of Hawaii Technical Report, 1969, AFOSR 70-0416 TR, AD 701417.
11. Abramson, N., "The ALOHA-System-Another Alternative for Computer Communications," University of Hawaii Technical Report B70-1, April 1970, AD 707853.
12. Bortels, W. H., "Simulation of Interference of Packets in the ALOHA Time-Sharing System, University of Hawaii Technical Report B70-2, March 1970.
13. Hayes, J. F., and Sherman, D. N., "Traffic Considerations of A Ring Switched Data Transmission System," B.S.T.J. *50*, No. 9 (November 1971), pp. 2947–2978.
14. Anderson, R. R., Hayes, J. F., and Sherman, D. N., "Simulated Performance of a Ring Switched Data Network," *COM-20*, No. 3, IEEE Trans. on Comm. Tech. (June 1972).
15. Avi-Itzhak, B., "Heavy Traffic Characteristics of a Circular Data Network," B.S.T.J. *50*, No. 8 (October 1971), pp. 2521–2549.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 9, November, 1972
Printed in U.S.A.

On Fast Start-Up Data Communication Systems Using Pseudo-Random Training Sequences

By R. W. CHANG and E. Y. HO

(Manuscript received June 21, 1972)

This paper analyzes the start-up performance of automatic transversal equalizers when maximum-length pseudo-random sequences of short periods are selected as the training signals for fast start-up purposes. Single-sideband Nyquist systems are considered because they represent the limiting case of vestigial-sideband systems with small excess bandwidth. It is shown that the equalizer is capable of fast start-up except in some rare situations which can be avoided by using proper timing, phase, and equalizer initial settings. The results also show that the equalizer tap convergence rate is independent of the phase characteristic of the communication channel and of the choice of the pseudo-random sequences which have the same period.

The equalizer is set up in the training period by minimizing the mean-square error between the equalizer output and the transmitted pseudo-random sequence, which is different from the mean-square error for random data. Surprisingly, we have found that, even for pseudo-random sequences of very short periods, this start-up algorithm results in only a slight degradation in system performance. Accordingly, good system performance can be expected immediately after the system switches from the training mode to the data mode.

I. INTRODUCTION

Pseudo-random sequences have been used in the past as training signals for setting up automatic transversal equalizers during start-up periods.[1,2] For fast start-up, it is desirable to know how the equalizer settling time depends on the choice of the pseudo-random sequence, the channel characteristics, and the initial receiver conditions. These problems are examined in the first part of this paper for single-sideband Nyquist systems. We present basic theories from which the reader

can work out numerical examples of special interest. The results are compared with those obtained previously[3] for a different class of training signals (isolated test pulses). An important difference between these two cases is pointed out.

When pseudo-random sequences are used, it is most convenient to adjust the equalizer tap gains to minimize the mean-square error between the equalizer output and the transmitted pseudo-random sequence. It is not immediately clear how closely this simple algorithm optimizes the data set performance for transmission of random data (because an equalizer setting optimum for pseudo-random sequence transmission is not necessarily optimum for random data transmission, particularly when pseudo-random sequences with very short periods are used for fast start-up purpose). This problem is examined in Section IV and the analysis is illustrated by examples.

Section V summarizes the results of this paper. The reader mainly interested in the conclusions and their implications may read Section V next.

II. MATHEMATICAL MODEL AND FUNDAMENTALS

An amplitude modulation data communication system with a conventional tapped delay line transversal equalizer is depicted in Fig. 1. During data transmission, the transmitter transmits the information digits, $\{d_i\}$, sequentially at time instants $t = \cdots, t_1 - T, t_1, t_1 + T, \cdots$. The equalizer output is sampled sequentially at the symbol rate to recover the information digits. Let the ith equalizer output

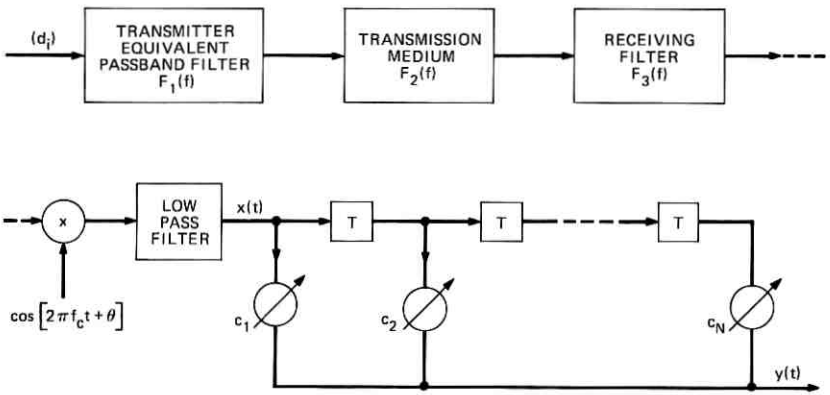


Fig. 1—Block diagram of an amplitude-modulation data communication system.

sample be y_i . We adopt the familiar mean-square error (MSE) criterion and adjust the gain controls of the equalizer to minimize the MSE between y_i and d_i . We assume $\{d_i\}$ is an ergodic process, hence the mean-square error can be written as

$$\begin{aligned} \mathcal{E} &= \lim_{\nu \to \infty} \frac{1}{\nu} \sum_{i=1}^{\nu} (y_i - d_i)^2 \\ &= \langle (y_i - d_i)^2 \rangle \end{aligned} \tag{1}$$

where $\langle x \rangle$ denotes the time average of x.

It can be seen from Fig. 1 that

$$y(t) = \sum_{k=1}^{N} C_k x(t - (k-1)T), \tag{2a}$$

and

$$x(t) = \sum_{i=-\infty}^{\infty} d_i h(t - iT), \tag{2b}$$

where $h(t)$ is the overall system (without equalizer) impulse response. For the sake of simplicity, we shall shift the time origin and use the abbreviations $y_i = y(iT)$, $x_i = x(iT)$, and $h_i = h(iT)$. Thus (1) can be written as

$$\mathcal{E} = \mathbf{C'AC} - 2\mathbf{C'V} + \langle d_i^2 \rangle, \tag{3}$$

where

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix}, \tag{4}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}, \tag{5}$$

$$\begin{aligned} a_{ij} &= \langle x_{-i+1} x_{-j+1} \rangle \\ &= \sum_m \sum_n \langle d_m \, d_n \rangle h_{-i+1-m} h_{-j+1-n} \end{aligned}$$

$$i = 1, 2, \cdots, N, \quad j = 1, 2, \cdots, N, \tag{6}$$

$$\mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}, \tag{7}$$

and

$$\begin{aligned} V_k &= \langle x_{i-k+1}\, d_i \rangle \\ &= \sum_m \langle d_i\, d_m \rangle h_{i-k+1-m}, \qquad k = 1, 2, \cdots, N. \end{aligned} \tag{8}$$

Let $\partial \mathcal{E}/\partial C_i$ be the partial derivative of $\mathcal{E}$ with respect to C_i, $i = 1$ to N, and let $\partial \mathcal{E}/\partial \mathbf{C}$ represent an $N \times 1$ column vector the ith element of which is $\partial \mathcal{E}/\partial C_i$. From (3), we obtain

$$\frac{\partial \mathcal{E}}{\partial \mathbf{C}} = 2\mathbf{AC} - 2\mathbf{V}. \tag{9}$$

The optimum value of $\mathbf{C}$ that minimizes the MSE $\mathcal{E}$ is denoted $\mathbf{C}_{opt}$. It is clear from (9) that

$$\mathbf{C}_{opt} = \mathbf{A}^{-1}\mathbf{V}. \tag{10}$$

Thus, the minimum value of $\mathcal{E}$ when $\mathbf{C} = \mathbf{C}_{opt}$ is

$$\mathcal{E}_{min} = \langle d_i^2 \rangle - \mathbf{V}'\mathbf{A}^{-1}\mathbf{V}. \tag{11}$$

Let $\mathbf{e}$ denote the difference between $\mathbf{C}$ and $\mathbf{C}_{opt}$; i.e.,

$$\mathbf{e} = \mathbf{C} - \mathbf{A}^{-1}\mathbf{V}. \tag{12}$$

Now consider the adjustment of the equalizer. As is well known, the equalizer can be adjusted in the training period by transmitting either a succession of isolated test pulses or a special class of pseudo-random sequence (pseudo-noise sequence[3]). The case of sending isolated test pulses during the training period has been considered by Chang.[4] In this paper, we examine the case of sending pseudo-noise sequences.†

In the training period, a binary pseudo-noise sequence is applied to the transmitter input. Since an adjustment is made at the end of each training sequence period and since we wish to make the largest number of adjustments during a fixed training time, we consider the shortest‡ possible pseudo-random sequence period (i.e., the case where

† The best known sequences of this type are the m-sequences (also known as maximal-length linear recurring sequences or maximum-length pseudo-random sequences).

‡ When the period of pseudo-random sequence is shorter than the length of the transversed equalizer, the analysis is difficult because A^{-1} may not exist.

the period of the pseudo-noise sequence is equal to the length of the transversal equalizer).

Let the pseudo-noise sequence be denoted by $\beta_1\beta_2\beta_3 \cdots \beta_N \beta_1\beta_2\beta_3 \cdots \beta_{k'}$, where $\beta_{k'}$ is the last bit. From (2b), the input to the equalizer can be written as

$$x(t) = \sum_{k=1}^{k'} \beta_k h(t - kT). \tag{2b}$$

Practically speaking, we may assume that $h(t)$ is time limited. Then it can be shown that for t larger than a certain value, say t_0, $x(t)$ will be a periodic function of period NT; i.e.,

$$x(t) = x(t + NT), \quad t_0 < t \quad \text{and} \quad t + NT \leqq k'T. \tag{13}$$

In the training period, the values of a_{ij}, V_k, $\mathcal{E}$, $\mathbf{C}_{\text{opt}}$, $\mathbf{A}$, and $\mathbf{V}$ are denoted by a_{ij}^*, V_k^*, $\mathcal{E}^*$, $\mathbf{C}_{\text{opt}}^*$, $\mathbf{A}^*$, and $\mathbf{V}^*$, respectively. From (6), we obtain

$$\begin{aligned} a_{ij}^* &= \sum_m \sum_n \langle \beta_m \beta_n \rangle h_{-i+1-m} h_{-j+1-n} \\ &= \sum_k \sum_m h_{-i+1-m} h_{-j+1-m+kN} \\ &\quad - \frac{1}{N} \sum_k \sum_{l=1}^{N-1} \sum_m h_{-i+1-m} h_{-j+1-m+kN+l} \quad \text{all} \quad i, j. \end{aligned} \tag{6}$$

From (8)

$$\begin{aligned} V_k^* &= \sum_m \langle \beta_m \beta_i \rangle h_{i-k+1-m} \\ &= \sum_j h_{-k+1+jN} - \frac{1}{N} \sum_j \sum_{l=1}^{N-1} h_{-k+1+jN+l}, \qquad k = 1, 2, \cdots, N \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j = \text{integers}. \end{aligned} \tag{8}$$

The partial derivative $\partial\mathcal{E}^*/\partial C_i$, $i = 1$ to N, can be computed from each block of N samples of $x(t)$, and the gain control C_i is adjusted by an amount proportional to $\partial\mathcal{E}^*/\partial C_i$ at the end of each block. For example, $\partial\mathcal{E}^*/\partial C_i$ is computed from x_{i+1} to x_{i+N} and C_i is adjusted after x_{i+N}. Then $\partial\mathcal{E}^*/\partial C_i$ is computed from x_{i+N+1} to x_{i+2N} and C_i is adjusted after x_{i+2N}. The optimum tap setting that minimizes the MSE, $\mathcal{E}^*$, is

$$\mathbf{C}_{\text{opt}}^* = (\mathbf{A}^*)^{-1}\mathbf{V}^*.$$

We now proceed to examine the convergence of $\mathcal{E}^*$.

III. SINGLE-SIDEBAND NYQUIST SYSTEMS

In this section, we consider single-sideband data communication systems which transmit at the Nyquist rate with $\sin x/x$ pulses (hereafter referred to as single-sideband Nyquist system). Such systems are considered because they represent the limiting case of sharp rolloff vestigial-sideband systems.

The transfer functions of the transmitting filter, transmission medium, and receiving filter (see Fig. 1) are $F_1(f)$, $F_2(f)$ and $F_3(f)$, respectively. The $F_i(f)$ are of the following form

$$F_i(f) = |F_i(f)| e^{J\beta_i(f)}, \quad i = 1, 2, 3 \tag{14}$$

where J is used to denote the imaginary number $\sqrt{-1}$.

In a single-sideband Nyquist system, $|F_1(f)|$ and $|F_3(f)|$ are specified by

$$\begin{aligned} |F_1(f)| &= 1 \qquad f_1 \leqq |f| \leqq f_2 \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{15}$$

and

$$\begin{aligned} |F_3(f)| &= 1 \qquad f_1 \leqq |f| \leqq f_2 \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{16}$$

In general, with lower single-sideband transmission, the carrier frequency, f_c, is set equal to f_2. Let $H(f)$ denote the Fourier transform of $h(t)$, which is the overall system impulse response at the equalizer input. It can be shown that

$$\begin{aligned} H(f) &= \tfrac{1}{2} |F_2(f - f_c)| \\ &\quad \cdot e^{J[\beta_1(f-f_c)+\beta_2(f-f_c)+\beta_3(f-f_c)-2\pi(f-f_c)t_1+\theta]}, \qquad 0 \leqq f \leqq f_2 - f_1 \\ &= \tfrac{1}{2} |F_2(f + f_c)| \\ &\quad \cdot e^{J[\beta_1(f+f_c)+\beta_2(f+f_c)+\beta_3(f+f_c)-2\pi(f+f_c)t_1-\theta]}, \qquad -(f_2 - f_1) \leqq f \leqq 0 \\ &= 0 \qquad \text{otherwise,} \end{aligned} \tag{17}$$

where θ represents demodulating carrier phase. The signaling interval is

$$T = \frac{1}{2(f_2 - f_1)}. \tag{18}$$

Since the time samples $h(iT)$, $i = \cdots 0, 1, 2, \cdots$ are taken at the Nyquist rate, we obtain from the sampling theorem, Parseval's theorem,

and (17)

$$g(i-j) \triangleq \sum_{m=-\infty}^{\infty} h_{-i+1-m}h_{-j+1-m}$$

$$= 2(f_2 - f_1)\int_{-\infty}^{\infty} h(t - iT + T)h(t - jT + T)\,dt$$

$$= (f_2 - f_1)\int_0^{f_2-f_1} [\cos 2\pi f(i-j)T][|\, F_2(f - f_c)\,|]^2\,df. \quad (19)$$

Substituting (19) into (6), we obtain

$$a_{ij}^* = \sum_k g(i - j + kN) - \frac{1}{N}\sum_k \sum_{l=1}^{N-1} g(i - j + kN + l) \text{ all } i, j. \quad (20)$$

It is clear from (19) and (20) that a_{ij}^* is independent of the demodulating carrier phase θ, the system timing t_1, and the phase characteristics $\beta_i(f)$ of the system. We also note that for a fixed N, a_{ij}^* is independent of the choice of the pseudo-noise sequence. [The pseudo-noise sequence $\beta_1, \beta_2, \cdots$ does not appear in (19) or (20).] Using the method in Ref. 4, it is concluded that the equalizer tap convergence rate is independent of the demodulating carrier phase θ, the system timing t_1, the phase characteristics of the system, and the choice of the pseudo-noise sequence for fixed N.

Note from (19) and (20) that a_{ij}^* depends on the amplitude characteristic $|\, F_2(f)\,|$ of the transmission medium. Since amplitude distortion is not severe in private line systems, in the following discussion we assume that

$$|\, F_2(f)\,| = 1, \quad f_1 \leqq f \leqq f_2\,. \quad (21)$$

Substituting (21) into (19), and neglecting a normalizing constant $(f_2 - f_1)^2$, we obtain

$$g(i-j) = 1, \qquad i = j$$
$$= 0, \qquad i \neq j. \quad (22)$$

Substituting (22) into (20) gives

$$a_{ij}^* = 1, \qquad i = j$$
$$= -\frac{1}{N} \qquad i \neq j. \quad (23)$$

Therefore, the eigenvalues of $\mathbf{A}^*$ are

$$\lambda_k = 1 + \frac{1}{N}, \qquad k = 1, 2, \cdots, N-1$$

$$\lambda_N = \frac{1}{N}. \tag{24}$$

The eigenvector $\mathbf{u}_N$ corresponding to λ_N is an $N \times 1$ vector whose elements are all unity. Since all but the last eigenvalue are equal, the equalizer can settle rapidly except in the case where the initial $\mathcal{E}^*$ contains a large component $(\mathbf{e}_0'\mathbf{u}_N)^2\lambda_N$, where $\mathbf{e}_0$ is the initial tap setting error vector[3]. Since λ_N is small and since $(\mathbf{e}_0'\mathbf{u}_N)^2$ cannot be exceptionally large with proper timing, phase, and initial equalizer settings†, it is very unlikely that $\mathcal{E}^*$ would contain a large component $(\mathbf{e}_0'\mathbf{u}_N)^2\lambda_N$. Therefore, the equalizer can settle rapidly in the training period.

IV. FURTHER ANALYSIS OF SYSTEM PERFORMANCE

At the end of the training period, the equalizer taps are set very nearly to $\mathbf{C}_{\text{opt}}^*$. The data set is then switched to the data transmission mode. Since statistics of the true data differ from those of the training pseudo-random sequence, the optimum tap settings, $\mathbf{C}_{\text{opt}}^*$, obtained for a training sequence cannot also be the optimum one for the true data. Thus, system performance degradation during the early stage of data transmission is expected, even if the data set is equipped with an adaptive equalizer. We now proceed to determine this degradation.

We assume zero-mean independent information digits (binary or multilevel). The signal level is normalized such that

$$\langle d_i^2 \rangle = 1. \tag{25}$$

The mean-square error, $\mathcal{E}$, can be obtained from (3),

$$\mathcal{E} = (\mathbf{C}_{\text{opt}}^*)'\mathbf{A}(\mathbf{C}_{\text{opt}}^*) - 2(\mathbf{C}_{\text{opt}}^*)'\mathbf{V} + 1, \tag{26}$$

where $\mathbf{A}$ and $\mathbf{V}$ are given by (5) and (7), respectively. From (6), (8), (19), and (25), we obtain

$$a_{ij} = g(i - j), \tag{27}$$

and

$$V_k = h_{-k+1}. \tag{28}$$

The a_{ij}^* and V_k^* can be rewritten as

$$a_{ij}^* = a_{ij} + \Delta_{ij}(N), \tag{29}$$

† For example, one may use the method described in Ref. 1.

and

$$V_k^* = V_k + \gamma_k(N), \tag{30}$$

where

$$\Delta_{ij}(N) = \sum_{k\neq 0} g(i-j+kN) - \frac{1}{N}\sum_{k}\sum_{l=1}^{N-1} g(i-j+kN+l), \tag{31}$$

and

$$\gamma_k(N) = \sum_{j\neq 0} h_{-k+1+jN} - \frac{1}{N}\sum_{j}\sum_{l=1}^{N-1} h_{-k+1+jN+l}\,. \tag{32}$$

From (26), we have

$$\begin{aligned}\mathcal{E} &= \mathbf{C}'_{\text{opt}}\mathbf{A}\mathbf{C}_{\text{opt}} - 2\mathbf{C}_{\text{opt}}\mathbf{V} + 1 + (\delta\mathbf{C})'\mathbf{A}(\delta\mathbf{C}) \\ &= \mathcal{E}_{\min} + (\delta\mathbf{C})'\mathbf{A}(\delta\mathbf{C}),\end{aligned} \tag{33}$$

where $\mathbf{C}_{\text{opt}}$ is the optimum tap setting for the true data and $\delta\mathbf{C}$ is the difference between $\mathbf{C}^*_{\text{opt}}$ and $\mathbf{C}_{\text{opt}}$,

$$\delta\mathbf{C} = \mathbf{C}^*_{\text{opt}} - \mathbf{C}_{\text{opt}}\,. \tag{34}$$

The last term in (33) represents the system performance degradation and is non-negative.

The mean-square error during the early stage of data transmission can now be determined from (33) and (34). As the period of the training PN sequence approaches infinity, $\lim_{N\to\infty}\Delta_{ij}(N)\to 0$, $\lim_{N\to\infty}\gamma_k(N)\to 0$, and $\lim_{N\to\infty}\delta\mathbf{c}\to\mathbf{0}$. Hence, $\mathbf{C}^*_{\text{opt}}$ approaches $\mathbf{C}_{\text{opt}}$ asymptotically as N increases.

We now assume some specific channel characteristics and apply these formulas to determine the initial performance degradation.

Example 1: A baseband channel with a flat amplitude characteristic and a typical quadratic delay characteristic[5] is assumed. The delay at the Nyquist frequency is taken to be $\beta_m T$ seconds. The phase characteristic is of the form

$$\beta_2(f) = 8\pi\beta_m T^3 f^3/3. \tag{35}$$

The system impulse response, $h(t)$, can be calculated from

$$h(t) = 2\int_0^{1/2T} \cos\left(2\pi ft + \beta(f)\right) df. \tag{36}$$

In this example, we consider a typical value $\beta_m = 2$. One hundred one samples of $h(t)$ (from $t_0 - 50T$ to $t_0 + 50T$) are taken with $T = 1$ and

$t_0 = 0.6T$. For a 7-tap equalizer, the minimum mean-square error attainable is 0.03347. The mean-square error obtained from (33) is 0.0393. The results for a 15-tap equalizer are 0.01484 and 0.01785, respectively. It is clear from these numbers that the performance degradation caused by using a PN sequence in the training period is negligible. This can be further illustrated by sketching the vector $\delta\mathbf{C}$ in (34). Since the amplitude characteristic is constant, we have

$$\mathbf{A} = \mathbf{I} \tag{37a}$$

and

$$\mathbf{A}^* = \mathbf{I} - \mathbf{\Delta}, \tag{37b}$$

where

$$\mathbf{\Delta} = \begin{bmatrix} 0 & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & 0 & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & 0 & \cdots & \frac{1}{N} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & 0 \end{bmatrix} \tag{37c}$$

The inverse of $\mathbf{A}^*$ is found to be

$$(\mathbf{A}^*)^{-1} = \mathbf{I} + \frac{N-1}{N+1}\mathbf{I} + \frac{N^2}{N+1}\mathbf{\Delta}. \tag{38}$$

From (10), (34), (37a), and (38), we have

$$\delta\mathbf{C} = \frac{N-1}{N+1}\mathbf{V} + \frac{2N}{N+1}\boldsymbol{\gamma} + \frac{N^2}{N+1}\{\mathbf{\Delta V} + \mathbf{\Delta}\boldsymbol{\gamma}\}. \tag{39}$$

Substituting (28), (32), and (37c) into (39), we obtain

$$\delta\mathbf{C} = \begin{bmatrix} \sum_{j \neq 0} h_{[(N-1)/2]+t_0+jN} \\ \vdots \\ \sum_{j \neq 0} h_{-[(N-1)/2]+t_0+jN} \end{bmatrix}. \tag{40}$$

In Fig. 2, the time samples h_k, $k = -\infty$ to ∞, of $h(t)$ are sketched.

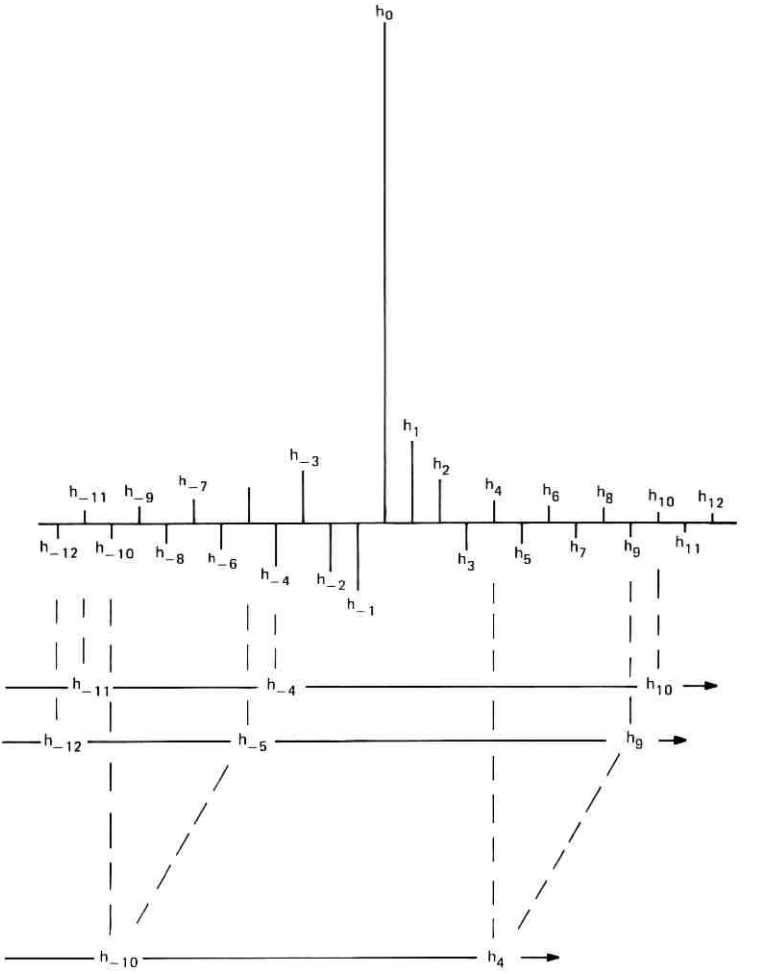


Fig. 2—Illustration of the elements of **δC**

For $N = 7$ (7-tap equalizer), the first element of **δC** is the sum of the infinite sequence $\cdots$, h_{-11} , h_{-4} , h_{10} , $\cdots$; the second element of **δC** is the sum of the infinite sequence $\cdots$, h_{-12} , h_{-5} , h_9 , $\cdots$; etc. It can be seen that the large time samples h_{-3} to h_3 are not included in these sums. This explains why **δC** should have small elements. By repeating the above for $N = 15$, one can easily see that the performance degradation due to the use of PN training sequence is small and that this degradation approaches zero as N increases.

Example 2: We continue example 1 but change β_m to its minimum value zero. The infinite sum $\sum_{i\neq 0} h_{-k+t_0+iN}$ can now be evaluated in closed form. By using the sampling theorem or the formulas for Psi (Digamma) functions,[6] it is obtained after some manipulations

$$\sum_{i\neq 0} h_{-k+t_0+jN} = (-1)^{-k+N}\frac{\sin \pi t_0}{\pi N} \cdot\left\{\frac{N}{t_0-k} - \frac{\pi}{2}\left(Ctg\,\frac{\pi(t_0-k)}{2N} + tg\,\frac{\pi(t_0-k)}{2N}\right)\right\} \tag{41}$$

where $0 \leqq t_0 \leqq 0.5T$ is assumed. The performance degradation can now be determined in closed form

$$(\delta\mathbf{C})'\mathbf{A}(\delta\mathbf{C}) = \sum_{k=-[N-1]/2}^{[N-1]/2} \frac{\sin^2 \pi t_0}{\pi^2 N^2} \cdot\left\{\frac{N}{t_0-k} - \frac{\pi}{2}\left(Ctg\,\frac{\pi(t_0-k)}{2N} + tg\,\frac{\pi(t_0-k)}{2N}\right)\right\}^2 \tag{42}$$

$\mathcal{E}_{\min}$ and $(\delta\mathbf{C})'\mathbf{A}(\delta\mathbf{C})$ are plotted in Fig. 3 for $N = 7$, 15, and 31 and $t_0 = 0.05$, 0.1, 0.15, 0.2 and 0.25. It can be seen that the value of $(\delta\mathbf{C})'\mathbf{A}(\delta\mathbf{C})$ is approximately an order of magnitude less than that of $\mathcal{E}_{\min}$. Also note that $(\delta\mathbf{C})'\mathbf{A}(\delta\mathbf{C})$ reduces almost by half when N is doubled. These results again show that the performance degradation caused by using PN training sequence is negligible.

V. CONCLUSIONS AND DISCUSSIONS

We have analyzed the start-up performance of a transversal equalizer for the case where a maximum-length pseudo-random sequence is used as the training signal to adjust the equalizer in the training period. The equalizer taps are adjusted by the gradient method to minimize the mean-square error, $\mathcal{E}^*$, between the equalizer output and the transmitted pseudo-random sequence. The pseudo-random sequence has been denoted $\beta_1, \beta_2, \cdots, \beta_N, \beta_1, \beta_2, \cdots$, where N is the period of the sequence. We have considered the case where N is equal to the number of taps of the equalizer. The following results are obtained:

(*i*) For a fixed N, the initial value of the mean-square error $\mathcal{E}^*$, the convergence rate of $\mathcal{E}^*$, and the minimum value of $\mathcal{E}^*$ are all independent of the specific values of the β_k's. Therefore, the same performance is obtained with any of the many pseudo-random sequences available. For example, a maximum-length pseudo-random sequence can be cyclic shifted to produce N

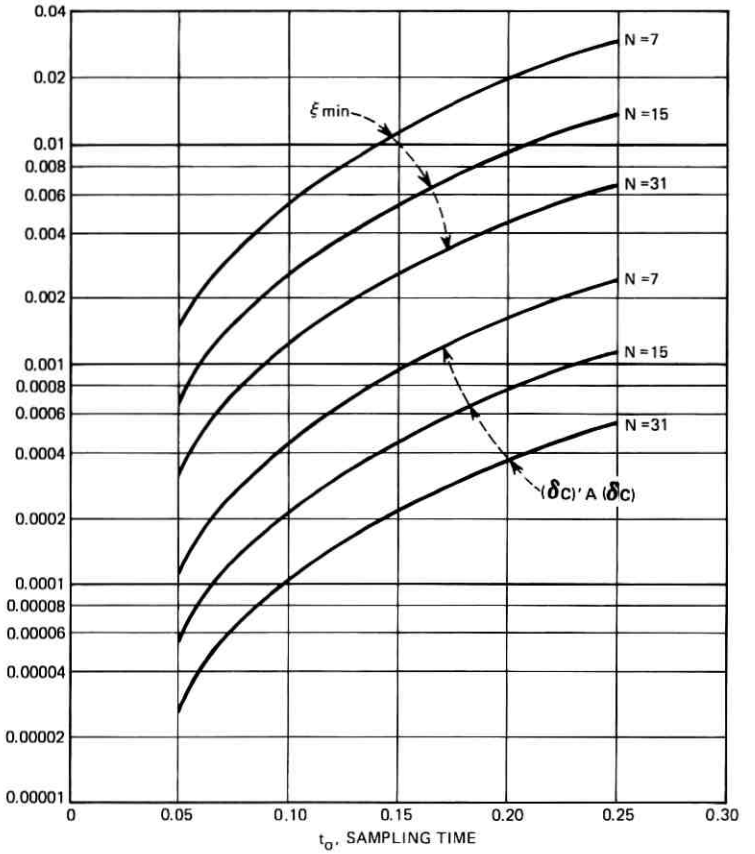


Fig. 3—Computational results of Example 2.

pseudo-random sequences. Any one of these N sequences can be used in the training period with the same result.

(*ii*) The initial value of $\mathcal{E}^*$ depends on the phase characteristic of the communication channel, and the timing and phase settings at the receiver. However, the convergence rate of $\mathcal{E}^*$ is independent of all these parameters. This result is similar to the one obtained previously[4] for the case where isolated test pulses are used as the training signal.

(*iii*) Unlike the isolated test pulse case, the eigenvalues of the correlation matrix here are not close together. For example, when the channel has a flat amplitude characteristic, the first $N - 1$ eigenvalues λ_1 to λ_{N-1} are equal to $1 + 1/N$, while the Nth

eigenvalue λ_N is equal to $1/N$. The mean-square error can be decomposed into N components, each associated with one of the N eigenvalues. The tap gain adjustment reduces rapidly the $N - 1$ components associated with λ_1 to λ_{N-1}, but the component associated with λ_N decreases very slowly. Therefore, as discussed at the end of Section III, care must be exercised in setting the timing, phase, and equalizer taps at the beginning of the training period so that the component associated with λ_N has a small initial value. Note that this precaution is not required when isolated test pulses are used, because in that case the eigenvalues are all close together and the components of the mean-square error all decrease rapidly.[4]

(*iv*) The analysis shows that the tap settings obtained with maximum-length pseudo-random sequences with very short periods are nearly optimum for random data transmission. More specifically, the equalizer taps are adjusted in the training period to minimize the mean-square error $\mathcal{E}^*$ between the equalizer output and the transmitted pseudo-random sequence. When such tap settings are used for actual data transmission, the mean-square error between the equalizer output and the transmitted random data can be written as

$$\mathcal{E} = \mathcal{E}_{\min} + \epsilon$$

where $\mathcal{E}_{\min}$ is the minimum attainable value of $\mathcal{E}$, and ϵ is nonnegative, because tap settings obtained with pseudo-random sequence do not necessarily minimize $\mathcal{E}$. Formulas for computing ϵ were developed in Section IV and illustrated by numerical examples. It can be seen from these formulas and Figs. 2 and 3 that ϵ decreases rapidly as N increases (N is the period of the pseudo-random sequence and also the number of equalizer taps). The computations show that ϵ decreases approximately by the factor $1/N$ (for example, ϵ reduces approximately by half when N is doubled). The computations also show that ϵ is about an order of magnitude less than $\mathcal{E}_{\min}$. (This is so for N as small as seven.) Therefore, tap settings obtained with pseudo-random sequences are nearly optimum for actual data transmission.

(*v*) As can be seen from the computations in Section IV, $\mathcal{E}_{\min}$ is rather large when N is small. For example, for a system with typical channel delay distortion (see example 1) and S/N = 30 dB, $\mathcal{E}_{\min}$ can be 15 dB above the thermal noise level when

$N = 7$, and 12 dB above the thermal noise level when $N = 15$. These large mean-square errors are due to the fact that for single-sideband Nyquist systems the overall system impulse response decays very slowly with time. Thus, for very sharp rolloff VSB systems (such as 4-percent rolloff), it is necessary to use a large number of equalizer taps (such as 31 or more).

REFERENCES

1. Lucky, R. W., and Rudin, H. R., "An Automatic Equalizer for General-Purpose Communication Channels," B.S.T.J., *46*, No. 9 (November 1967), pp. 2179–2208.
2. Mueller, K. H., Private Communication.
3. Golomb, S. W., *Digital Communications with Space Applications*, Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1964.
4. Chang, R. W., "A New Equalizer Structure for Fast Start-Up Digital Communication," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1969–2014.
5. Lucky, R. W., Salz, J., and Weldon, E. J., Jr., *Principles of Data Communication*, New York: McGraw-Hill, 1969, pp. 68–69.
6. Gradshteyn, I. S., and Ryzhik, I. M., *Table of Integrals Series and Products*, New York and London: Academic Press, 1965.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 9, November, 1972
Printed in U.S.A.

Stability Considerations in Nonlinear Feedback Structures as Applied to Active Networks

By M. BAUMWOLSPINER

(Manuscript received April 24, 1972)

Active filters have recently acquired widespread use in the realization of frequency-selective networks. Unlike their passive counterparts, active filters have the potential of oscillating.

Furthermore, it has been observed that the onset of oscillations in biquad active filters is dependent upon signal level. This led to the recognition that nonlinear stability theory would be necessary to comprehend this behavior.

This paper develops a technique to analyze the stability of networks containing linear and nonlinear elements interconnected in multifeedback structures. This is accomplished by extending the concept of the "Describing Function" to include networks containing nonlinearities with frequency-dependent linear feedback. The technique is then applied to explain qualitatively and quantitatively nonlinear effects in op-amps and their relation to the stability of frequency-selective networks containing them (e.g., the Multiple Amplifier Biquad, MAB, and the Single Amplifier Biquad, SAB). The technique is also applied to explain frequency shifts in amplitude-limited oscillators. The most valuable result of this analysis is the discovery of nonlinear feedback circuits which circumvent the conditional stability of high-frequency biquads. This has allowed us to obtain Q's of 50 at 100 kHz in a MAB employing 709 op amps. Similarly, a MAB employing 702 op amps was made to operate at 2 MHz with a Q of 10.

I. INTRODUCTION

In this paper, we shall deal primarily with a frequency-domain approach of analyzing networks containing linear and nonlinear elements interconnected as multifeedback structures. Particular applications will include the Single Amplifier Biquad[1] (SAB), the Multiple Amplifier Biquad[2] (MAB), and amplitude stabilized oscillators.

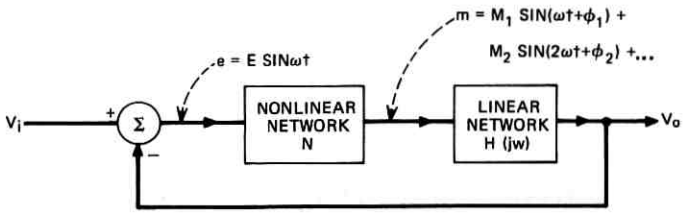


Fig. 1—A typical system with separable networks.

Our interest in this subject is motivated by conditional stability problems in biquads designed to operate above certain critical frequencies. Above these frequencies, these filters exhibit oscillatory modes when certain excitations are or were present. In the past, these conditional stability problems have been attributed to slew rate limiting in the operational amplifiers. However, we shall demonstrate later that although slew rate limiting may affect the stability, it is neither a necessary nor sufficient condition for conditional stability to occur in the MAB circuit.

We shall draw heavily on the Describing Function (DF) and the Dual Input Describing Function (DIDF) Techniques.[3-5] These techniques have the virtue of imparting a conceptual understanding of the problem. Techniques based on the Liapunov stability criterion such as the Popov criteria and circle criteria have been looked into and seem to offer overly restrictive sufficient conditions for stability in most of our reasonably high Q applications. In addition, they fail to extend easily to multiple nonlinear loops.

To facilitate the application of these techniques to the filter conditional stability problem, several interesting results concerning operational amplifiers will first be derived. Earlier nonlinear op-amp models have represented an op-amp by saturating elements intermingled with frequency-dependent networks in a *feedforward* configuration. It will be seen that this is insufficient to predict insertion phase measurements of an op-amp in its nonlinear region. However, the usual position of compensation elements in an op-amp is in a *feedback* path around the gain stages yielding higher effective (Miller effect) capacitances. As a consequence, the linear and nonlinear elements become interconnected in feedback structures which impart varying phase characteristics as a function of amplitude. This becomes highly significant when considering active resonant circuits. Most important, when high Q circuits are realized, any additional phase lag around a loop may be sufficient to increase the pole Q to the point of oscillation. Similar phase shifts may

occur in oscillators in which nonlinearities are purposely introduced to stabilize their amplitude. In oscillators these effects are manifested by a discrepancy between the linearly-computed and actual frequency of oscillation.

Finally, we will present circuits which produce either phase lag or lead as a function of amplitude. These circuits may be useful for nonlinear compensation of the aforementioned problems.

II. THE DESCRIBING FUNCTION TECHNIQUE

The describing function method is an outgrowth of the Harmonic Balance technique used by Krylov and Bogoliubov,[6] in nonlinear mechanics. The method reduces a nonlinear differential equation into a linear relationship by assuming a sinusoidal solution. The method is most useful when the system contains sufficient lowpass filtering to allow higher harmonics to be neglected. However, if the describing function technique is inadequate due to its neglect of higher order harmonics, the DIDF may help in solving the problem.

2.1 *Input-Output Concept of A Nonlinear Element*

We begin our analysis by reviewing some basic concepts of the DF technique. Consider the system of Fig. 1, where the linear and nonlinear parts are assumed separable. N is described in terms of its effect on a sinusoidal waveform. In particular, the describing function is defined as

$$\text{DF} = g(E) \triangleq \frac{\text{Fundamental of } m}{\text{Fundamental of } e} = \frac{M_1}{E} \angle \phi_1 . \tag{1}$$

In general, the DF will be a complex quantity. However, if N is a single-valued nonlinear function, its input-output characteristic will not enclose any area (see Fig. 2) and its DF will always be real as shown in Appendix A.

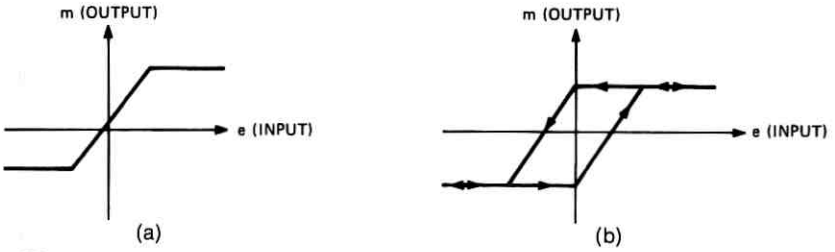


Fig. 2—(a) A single-valued nonlinearity (real DF). (b) A nonlinearity having a hysteresis loop that encloses a given area (complex DF).

On the other hand, the DIDF is defined for a two sine wave input, with one wave at a multiple frequency of the first, i.e.,

$$e = E \cos(\omega t + \phi) + E_1 \cos n\omega t.$$

The DIDF is the ratio of the desired frequency component in the output over the same frequency component at the input. The DIDF has been worked out by West, et al.,[7] for a saturating type of nonlinearity operating on a sinusoidal input in the presence of the third harmonic. This function can be used in determining the required perturbation of the DF when the third harmonic is not sufficiently low for the DF to be directly applicable.

With the above ideas applied to Fig. 1, we may extend linear stability criteria to obtain the stable and unstable limit points for this case. Assuming that there is sufficient lowpass filtering following N in Fig. 1, the input-output relationship of the system is:

$$\frac{v_0}{v_i} = \frac{g(E)H(j\omega)}{1 + g(E)H(j\omega)} \tag{2}$$

$$= \frac{H(j\omega)}{\frac{1}{g(E)} + H(j\omega)}. \tag{3}$$

We now apply the Nyquist criterion, but instead of taking the critical point as -1 and incorporating $g(E)$ into $H(j\omega)$, we take $-[1/g(E)]$ as the critical point.

Figure 3c shows the conventional Nyquist plot and the locus of $-[1/g(E)]$ with E as a parameter for the system the block diagram of which is given in Fig. 3a. The describing function of the nonlinearity,[8] as shown in Fig. 3b depicts the variation in amplitude (the phase shift being zero) of the fundamental of sin ωt operated on by a dead-zone nonlinearity. We note that if E is small, the critical point is not encircled, and the system will be stable. As E is increased, we reach point A, and the system becomes unstable. As a consequence, E increases till we reach point B where the system enters into a stable limit cycle. The limit cycle is stable since if E increases, the system becomes stable causing E to decrease. On the other hand, if E decreases, the system instability is such that E will increase. The intersection point B defines the amplitude (E) and frequency (ω) of oscillation. It is to be noted that in the above case the frequency of oscillation occurs at the intersection of $H(j\omega)$ with the real axis since N introduces no phase shift in this case. However, when N introduces phase shift, this will not be

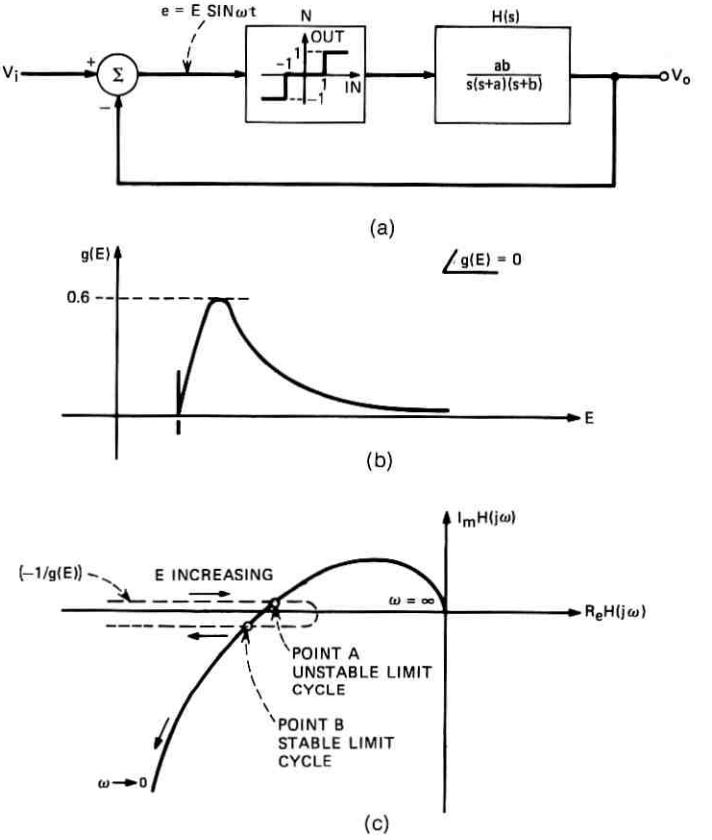


Fig. 3—(a) System with dead-zone relay. (b) DF of dead-zone relay. (c) Nyquist plot.

the case. Furthermore, even when N has a real DF and the second harmonic is not sufficiently filtered out, the DIDF predicts a phase shift of the fundamental through N.

2.2 *Input-Output Concept of A Feedback Structure with Nonlinear Elements in it*

With the above as background, we proceed further with the DF concept by determining the sinusoidal input-output relationship for a feedback system with a nonlinearity. We will do this by considering the specific example of Fig. 4, which resembles an op-amp with unity feedback. As before, we assume sufficient lowpass filtering to eliminate the effect of the second harmonic.

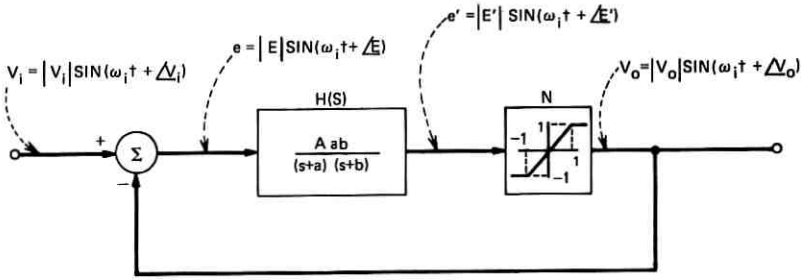


Fig. 4a—System with saturating nonlinearity.

It is clear from Fig. 4a that

$$V_i - V_0 = E \quad \text{or} \quad V_i = E + V_0 . \tag{4}$$

In addition, we know the relationship between E and V_0 from the describing function of Fig. 4b and the linear transfer function $H(S)$. Therefore, we can obtain the relationship between V_i and V_0. This is done most appropriately by considering the phasor diagram of Fig. 5b.

First, we obtain the transfer characteristics of the linear block, $H(j\omega)$ as shown in Fig. 5a. Incidentally, this is the Nyquist plot for the linear region of operation. Assuming momentarily that we hold $\omega = \omega_1$, $H(j\omega)$ and $g(E')$, determine the necessary magnitude and phase of E to give a particular V_0. From eq. 4, we then find the corresponding V_i by taking the vector sum as is shown for several V_0 in Fig. 5b. Observe that the boundary between the linear and nonlinear region is $(V_0^{(2)}, V_i^{(2)}, E^{(2)})$. In the linear region $\angle V_0 - \angle V_i$ is independent of amplitude, while in the nonlinear region $\angle V_0 - \angle V_i$ is a function of amplitude as shown in Fig. 5c. As a result of $H(j\omega)$, there will be a different curve of this type for each frequency considered. These curves together with similar ones for $|V_0|/|V_i|$ define the DF for the system

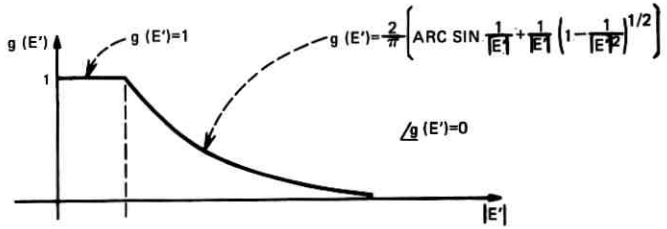


Fig. 4b—Describing function of saturating nonlinearity.

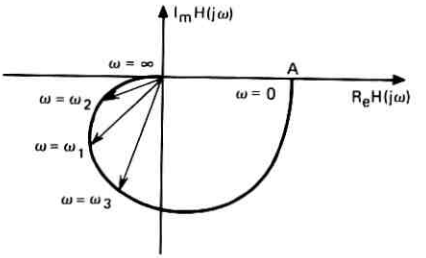


Fig. 5a—Nyquist plot of $H(j\omega)$.

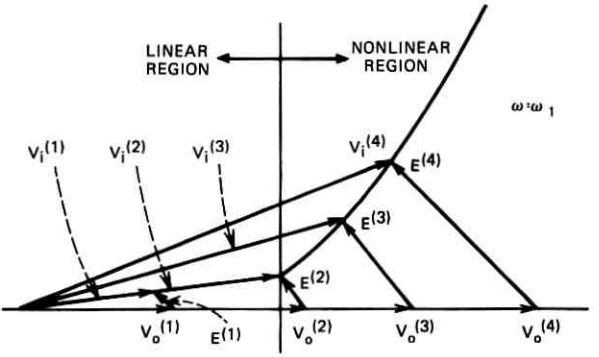


Fig. 5b—Vector diagram for determining V_i.

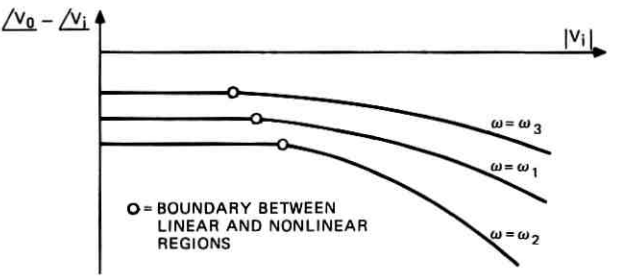


Fig. 5c—Phase shift vs input signal.

of Fig. 4a. This describing function differs slightly from those considered thus far in that it is a function of both amplitude and frequency.

A most important result of the input-output concept developed in this section is the capability of being able to analyze a rather complicated nonlinear network by breaking it up into its constituents. Each one is analyzed individually and then combined as shown in Fig. 6. After

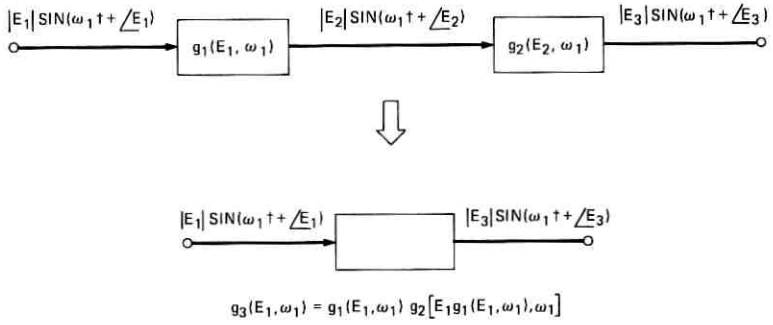


Fig. 6—Interconnection of DF's

reducing the network we apply a variation of the analysis employed for the example of Fig. 3. The difference lies in the fact that the DF is now a function of frequency and amplitude. As a consequence, we obtain many DF's, each for a particular frequency, of which the negative reciprocal must not intersect the Nyquist plot at that frequency for stable operation. Usually, as we will see, the critical frequencies correspond to those close to the critical point in the Nyquist plane.

A network which is well suited to this analysis technique is shown in Fig. 7. This is a typical Multiple Amplifier Biquad Filter using monolithic op-amps. Here, it is clear that if we employ the op-amp model shown in Fig. 7b, we have a multifeedback structure containing several nonlinearities. Deriving the DF of each closed loop op-amp and combining results, we can determine if the phase lag around the loop at certain amplitudes is sufficient to cause the biquad to oscillate. This is the subject of the next section.

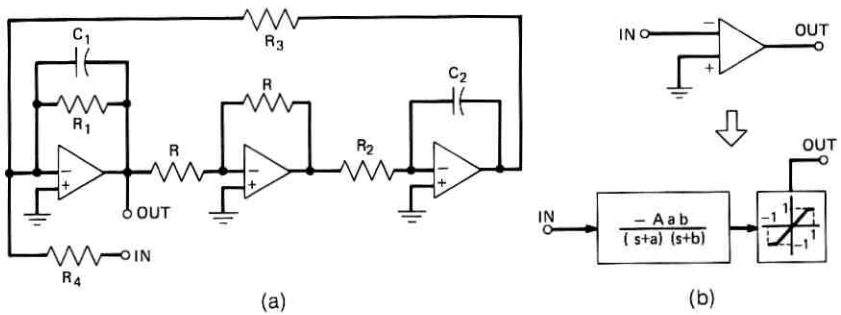


Fig. 7—(a) MAB circuit. (b) A possible model for the op-amp.

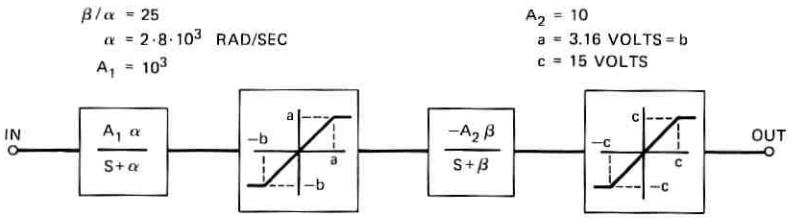


Fig. 8—Fairchild μA709 op-amp model.

III. APPLICATIONS

In this section we shall apply the techniques developed earlier to op-amps, active filters, and oscillators. Since active filters (e.g., MAB and SAB circuits) usually employ monolithic op-amps, we shall develop first an accurate model of the linear and nonlinear aspects of the op-amp. The op-amp model will generally depend on the type (e.g., 709, etc.) and compensation used. However, we shall demonstrate how to arrive at the model and present typical circuits.

3.1 *Operational Amplifiers*

3.1.1 *Open Loop Characteristics of Op-Amps*

We have already presented one possible model of a typical op-amp in Fig. 7b. It is possible to modify the model by placing an amplitude limiter at the input or splitting the linear transfer function by inserting an amplitude limiter. An excellent model, as viewed from pulse and slew-rate measurements on a Fairchild μA 709 op-amp is shown in Fig. 8. The DF for this model does not predict any extra phase shift as a function of amplitude. This follows from Appendix A, where it is shown that this type nonlinearity (no hysteresis) introduces no additional phase shift. However, if we consider the DIDF, we may indeed get extra phase lag as a function of amplitude. This results from the presence of harmonics, generated by the first nonlinearity, at the input of the second nonlinearity. The amount of phase lag may be estimated

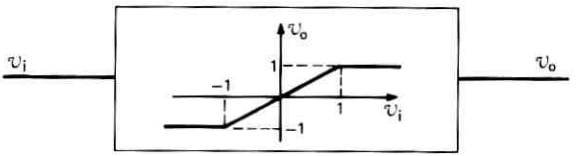
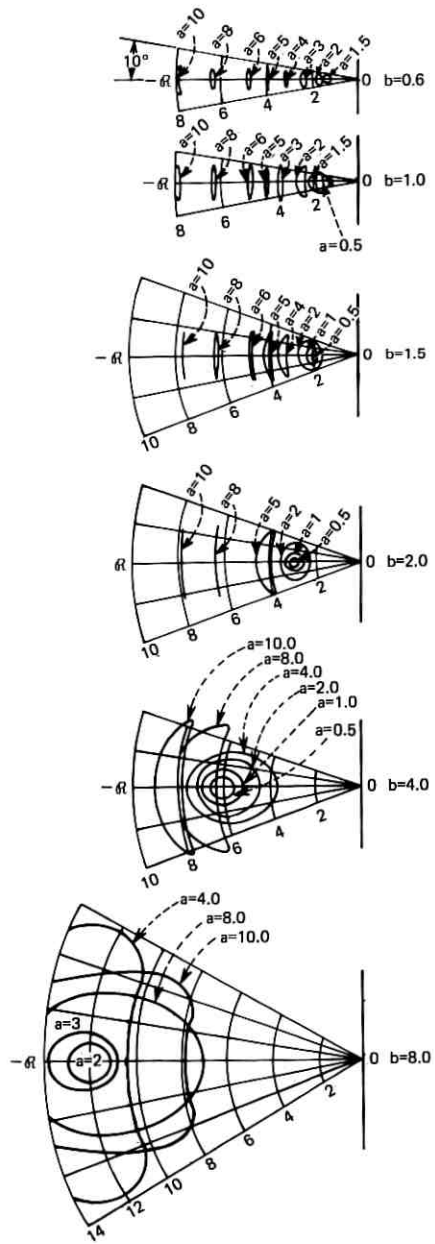
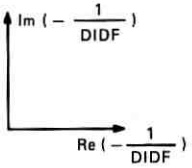


Fig. 9a—Symmetrical saturator.





a = AMPLITUDE OF FUNDAMENTAL

b = AMPLITUDE OF THIRD HARMONIC

RADIAL LINES ARE ALL 10° APART AND REPRESENT LINES OF CONSTANT DIDF PHASE; ANNULAR LINES ARE LINES OF CONSTANT DIDF AMPLITUDE.

$\nu_o = \nu_i \qquad -1 \leqslant \nu_i \leqslant 1$

$\nu_o = -1 \qquad \nu_i < -1$

$\nu_o = 1 \qquad \nu_i > 1$

Fig. 9b—Input-output relationships for fundamental in presence of third harmonic for saturating nonlinearity.

by using the set of DIDF curves for a symmetrical amplitude limiter[7] shown in Fig. 9. The curves for each value of a represent the DIDF as a function of the phase shift between the fundamental and third harmonic at the input of the saturator.

From these figures, it can be shown that a maximum fundamental phase lag of 11.1 degrees may be obtained when the op-amp of Fig. 8 is heavily overdriven (i.e., a square wave input to the second saturator). Yet, when the phase characteristics of a 709 op-amp were measured in the lab, no such effect was observed. The curves obtained are shown in Fig. 10. These curves display phase lead as a function of amplitude instead of phase lag. This result is extremely important, for it verifies the stability inherent in biquads at low frequencies as we shall see later. In summary, a better model of the op-amp is needed.

To determine this model, we first consider a simple circuit depicted in Fig. 11. This circuit is a transistorized amplifier which provides output limiting of the signal. Its applications include FM limiters and

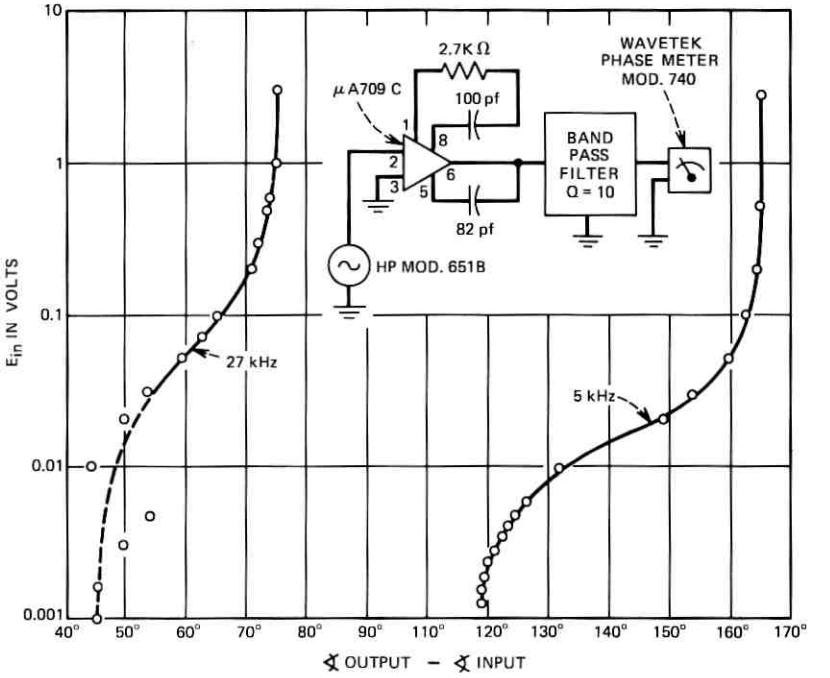


Fig. 10—Measured phase characteristic of an open loop 709 op-amp.

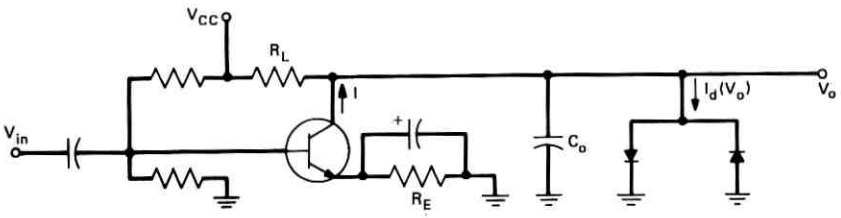


Fig. 11a—Transistor amplifier with output limiting.

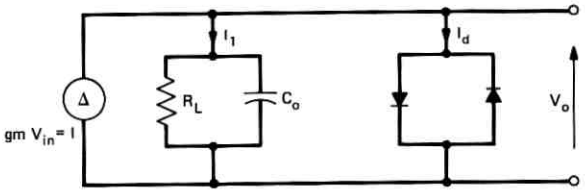


Fig. 11b—Model of Fig. 11a assuming transistor is in linear region.

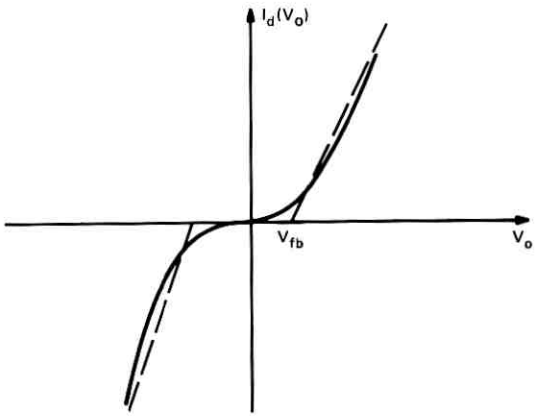


Fig. 11c—Characteristics of diode pair.

amplitude-stabilized oscillators. Defining Z as

$$Z = \frac{R_L}{SC_0R_L + 1} \tag{5}$$

it follows from the circuit that:

$$I = V_0/Z + I_d(V_0) \tag{6}$$

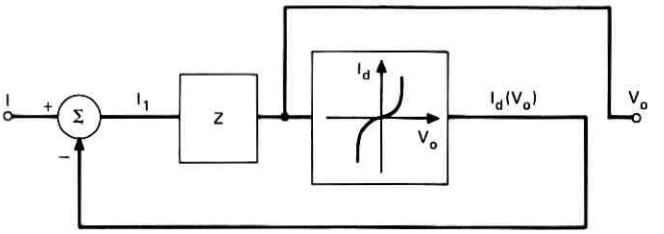


Fig. 12a—Block diagram of Fig. 11b.

or

$$V_0 = Z[I - I_d(V_0)]. \tag{7}$$

This equation is represented by the block diagram of Fig. 12a. Knowing the relationship between I_1 and I_d, we proceed to draw the phasor diagram of Fig. 12b. Initially at low levels, I_d is small compared to I_1 because the diodes are cut off. However, if I_1 is just short of turning the diodes on, a slight increase in I_1 will cause a large increase in I_d. These two situations are represented by $(I_d^{(1)}, I_1^{(1)})$ and $(I_d^{(2)}, I_1^{(2)})$ respectively, in Fig. 12b. Taking their vector sum, we obtain $I^{(1)}$ and $I^{(2)}$. Interestingly we note that the phase difference between I and I_d and hence V_{in} and V_o has decreased. By taking the small signal case as a reference, we conclude that the closed loop network exhibits phase lead with increasing amplitude.

A significant conclusion to be drawn from this example is that although the simple circuit appeared to contain just a clipper (which would indicate no amplitude-dependent phase shift) the feedback present alters the situation sufficiently to predict phase shift as a function of amplitude.

If the same form of analysis is carried out for the internal circuitry

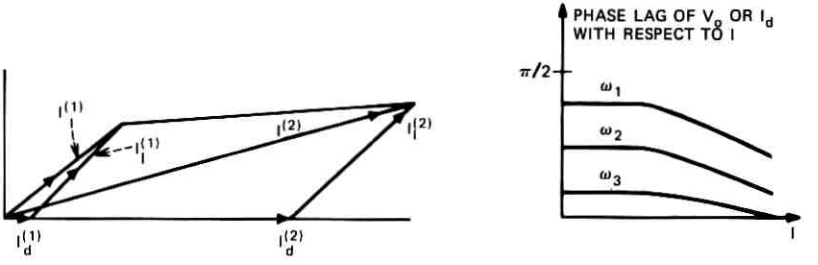


Fig. 12b—Input/output relationship.

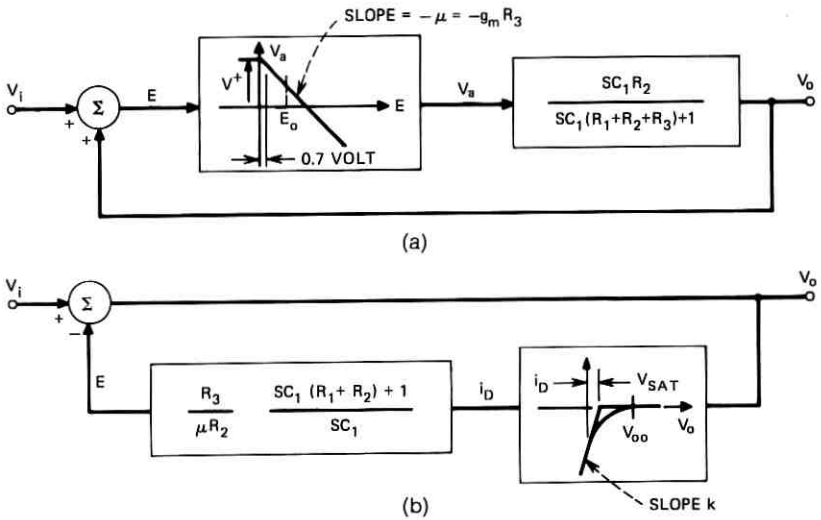


Fig. 13—709 op-amp input compensation circuity model. (a) Cutoff region. (b) Saturation region.

of the 709 op-amp, similar results are obtained. This has been done in Appendix B with particular attention to the circuitry involved in the input and output compensation. Most often, compensation in an op-amp is obtained by making use of the Miller effect to obtain low-frequency breakpoints with relatively small capacitors. Both the input and output compensation in a 709 op-amp take advantage of this effect. As a result, the transistor nonlinearities have RC feedback around them. The equivalent circuits of the input compensation circuitry in the cutoff and saturation regions, as derived in Appendix B, are shown in Fig. 13. In this figure we have left out some linear transfer function blocks at the input and output side, since these have no effect on the

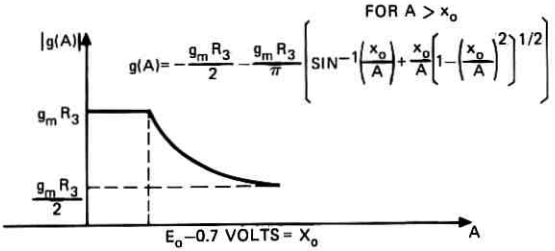


Fig. 14a—DF for nonlinear element in cutoff region model.

nonlinear characteristics. The following analysis will show that both of the circuits in Fig. 13 provide phase lead as amplitude increases.

Figure 14a shows the DF of the nonlinear element[8] in the cutoff region model under the assumption that the quiescent point is at $E = E_0$. Figure 14b depicts the phase relationships of the linear element in that model. With this information and Fig. 13a, we construct the phasor diagram of Fig. 14c. Specifically, since the nonlinear element is single valued and therefore contributes no amplitude sensitive phase shift, $-V_0$ will lead E according to the linear element and the inversion due to the nonlinear element. Entering the nonlinear region of the model, the error voltage E has to increase at a faster rate than V_0 to overcome the attenuation effect of the clipping element. From the relationship

$$V_i = (-V_0) + E \tag{8}$$

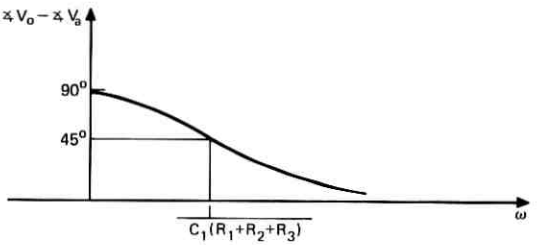


Fig. 14b—Phase characteristic for linear element in cutoff region model.

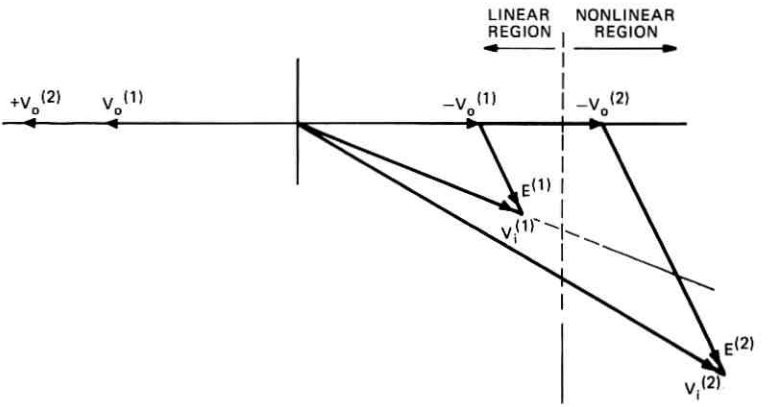


Fig. 14c—Phasor diagram for the closed loop in cutoff region model.

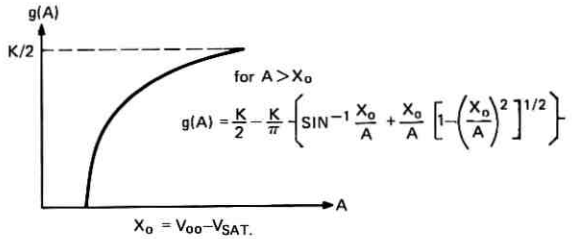


Fig. 14d—DF for nonlinear element in saturation region model.

we determine the third vector V_1. As a result, the angles between $(-V_0^{(1)}, V_i^{(1)})$ and $(-V_0^{(2)}, V_i^{(2)})$ do change with increasing drive. However, at high frequencies where E and $-V_0$ are initially close to being in phase, the increase of phase lead of V_0 relative to V_i is minimal. The frequency breakpoint below which the nonlinear phase lead effect may be expected is typically,

$$f \approx \frac{2}{2\pi(R_1 + R_2 + R_3)C_1} \approx 50 \text{ kHz}. \tag{9}$$

On the other hand, in the saturation region, the model of Fig. 13b applies. Here the DF for the nonlinear element is given in Fig. 14d under the assumption that the quiescent point is at $V_0 = V_{00}$. Furthermore, we have approximated the exponential nonlinearity of the diode in a piece-wise linear manner. It follows directly that the phase relationship of E with respect to V_0 is given by the phase characteristic of the lag network following the nonlinear element in Figure 13. Drawing the phasor relationships for the closed loop at a fixed frequency where the lag network has singificant phase lag yields Fig. 14f. Again, the phasor diagram reveals that the phase lead of V_0 with respect to V_i increases as amplitude increases. The frequencies over which this phase lead

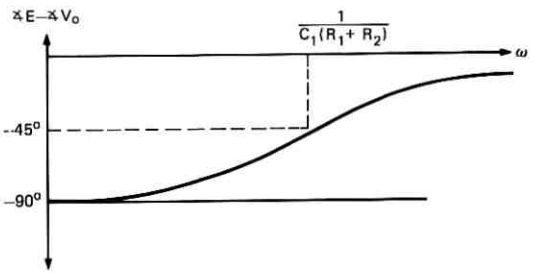


Fig. 14e—Phase characteristics for the linear element in saturation region model.

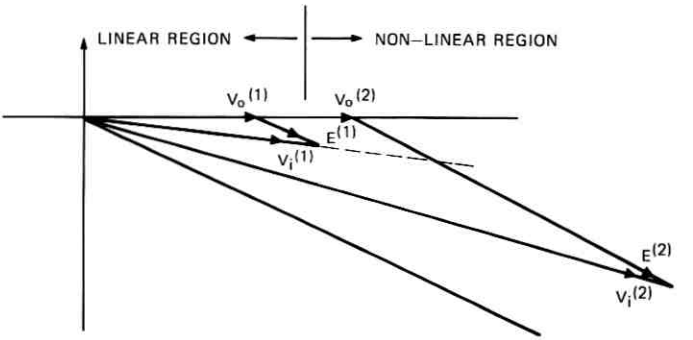


Fig. 14f—Phasor diagram for the closed loop in the saturation region model.

effect will be present is approximately the range of frequencies where the linear network provides phase lag. From Fig. 14e, this occurs for frequencies below:

$$f \approx \frac{2}{2\pi C_1(R_2 + R_1)} \approx 100\,\text{kHz}. \tag{10}$$

Above this frequency range, $E^{(1)}$ and $E^{(2)}$ will have approximately the same orientation as $V_0^{(1)}$ and $V_0^{(2)}$, forcing $V_i^{(1)}$ and $V_i^{(2)}$ to maintain their orientation roughly independent of amplitude.

The above analysis also applies to the output compensation circuitry provided that it is properly interpreted. The output compensation circuitry for a 709 op-amp and its nonlinear model are shown in Fig. 15. Here, unlike the input compensation, the RC network is fed back from an emitter follower circuit which simplifies the model and increases the

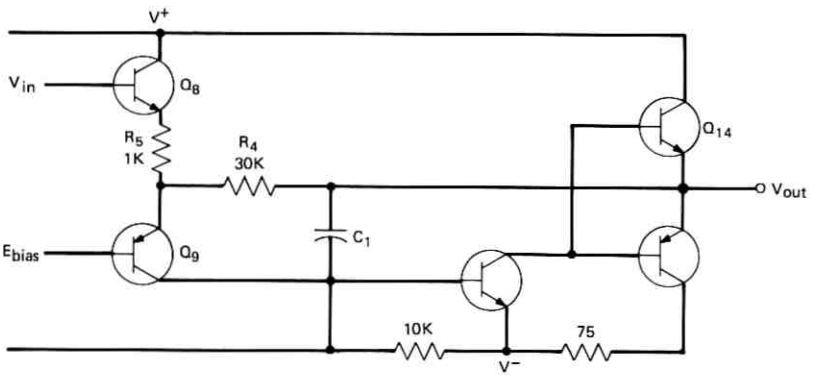


Fig. 15a—Output compensation of 709 op-amp.

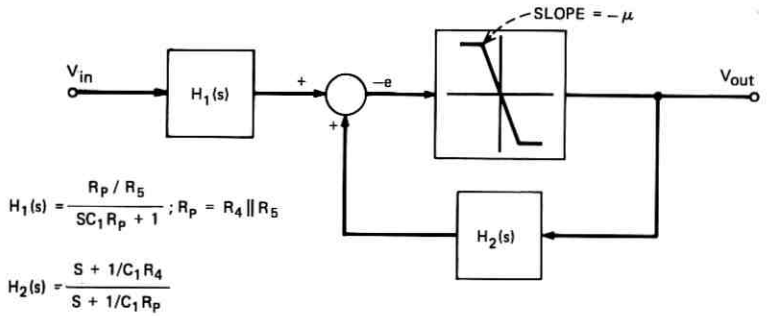


Fig. 15b—Nonlinear model of 709 op-amp.

breakpoint frequencies. In addition, the RC network is fed back to a lower impedance termination which also has the effect of increasing the breakpoint frequencies. In a typical 709 op-amp, this breakpoint frequency occurs at (the equivalent of eq. 9 with R_2 and R_3 being zero)

$$f \approx \frac{1}{2\pi C_1 R_4} \approx 1.5 \times 10^6 \text{ Hz}.$$

This frequency is significantly higher than those resulting from the input compensation (see eqs. 9 and 10). As a result, the output compensation will be quite irrelevant to the conditional stability of biquad filters in the 0 to 100 kHz range. Above 100 kHz, the amplifier will no longer be used in a biquad filter, since the open loop gain is less than 40 dB, which is insufficient for most precision applications.

We now have an op-amp model consistent with the experimental data of Fig. 10. Having discussed these open loop characteristics of an op-amp, we shall next use these concepts to derive the phase properties associated with closed-loop operational amplifiers.

3.1.2 *Closed-Loop Characteristics of Op-Amps*

We shall confine ourselves, in this section, to gain inverters, integrators, and leaky integrators. The circuit of Fig. 16a is the general case which includes all the above-mentioned configurations. The model of the closed loop op-amp shown in Fig. 16c is derived as follows. Assuming that the op-amp has infinite input impedance and letting

$$e_{in} = 0,$$

the voltage developed across the input of the op-amp will be

$$e = \frac{R_1}{R_1 + \left[\frac{1}{SC_1} \| R_a\right]} e_{out} = \frac{SR_1C_1 + R_1/R_a}{SR_1C_1 + [R_1/R_a + 1]} e_{out} . \quad (11)$$

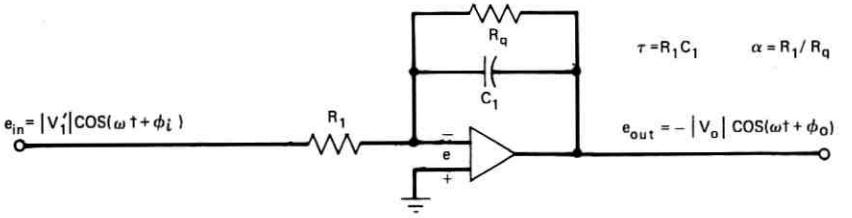


Fig. 16a—Closed loop op-amp.

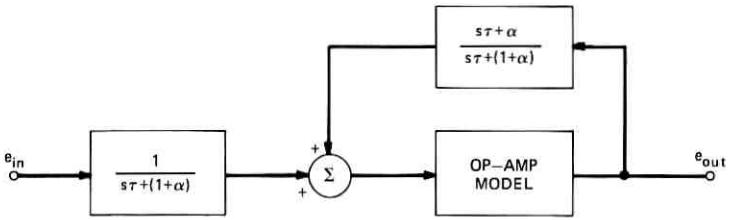


Fig. 16b—Model for Fig. 16a.

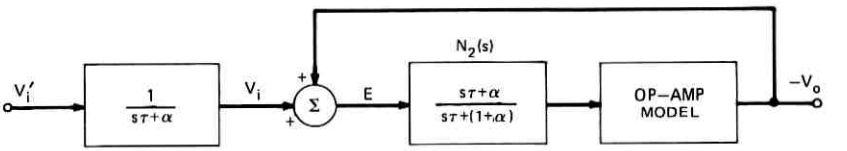


Fig. 16c—Model for Fig. 16a.

Likewise, letting

$$e_{out} = 0,$$

the input voltage of the op-amp will be

$$e = \frac{\frac{1}{SC_1} \,||\, R_q}{R_1 + \left[\frac{1}{SC_1} \,||\, R_q\right]} e_{in} = \frac{1}{SR_1C_1 + [R_1/R_q + 1]} e_{in} . \qquad (12)$$

The block diagram of the closed-loop op-amp, obtained through the superposition of eqs. 11 and 12, is shown in Fig. 16b. By the reduction method, we move the feedback network past the summing node and make the proper correction to the input network. This leaves us with the model of Fig. 16c. A significant advantage of this model is that the ideal transfer function has been separately realized by the network

between V_i' and V_i. Consequently, the feedback network following the ideal transfer function network gives a direct measure of the error introduced by the nonideal characteristics of the op-amp.

For the model of the op-amp itself, we may either use the networks of Fig. 13 in a cascade combination, or we may directly employ the DF curves shown in Fig. 10. The latter method is to be preferred since it is based on actual measurements and requires a minimum of computation. Fig. 17 is the phasor diagram for an inverter (i.e., Fig. 16 with $\tau = 0$ and $\alpha = 1$) at $f = 5$ kHz. In the linear region, we initially obtain a very small error voltage E and consequently V_i and V_0 are almost equal. This error voltage, although small, has a distinct phase lead of approximately 70 degrees with respect to the output voltage V_0. This phase shift is introcuded by the frequency characteristics of the op-amp and the network $N_2(s)$ in Fig. 16c. However, for the unity gain inverter, $N_2(j\omega)$ is a real quantity and, therefore, will not add additional phase. As we enter the nonlinear region, the magnitude and phase of the error voltage, E change as dictated by the describing function (DF) of the op-amp. The phase changes according to Fig. 10 and the amplitude according to the real DF of a saturation type of nonlinearity shown in Fig. 4b. Although the op-amp does not behave exactly as a saturation nonlinearity; the magnitude of the DF for the op-amp is well approximated by this DF, unlike the phase shift of the DF which for the saturation nonlinearity is zero.

Having obtained the phasor diagram of Figure 17, we note that increasing the input V_i predicts an increasing phase lag of the output $-V_0$ relative to the input V_i. This effect has been verified experimentally as shown by the curves in Fig. 18. At 27 kHz where the op-amp has

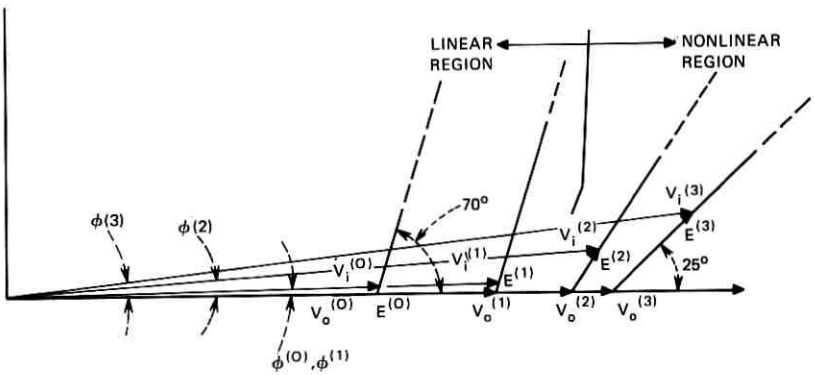


Fig. 17—Phasor diagram for unity gain inverter (using a 709 at 5kHz).

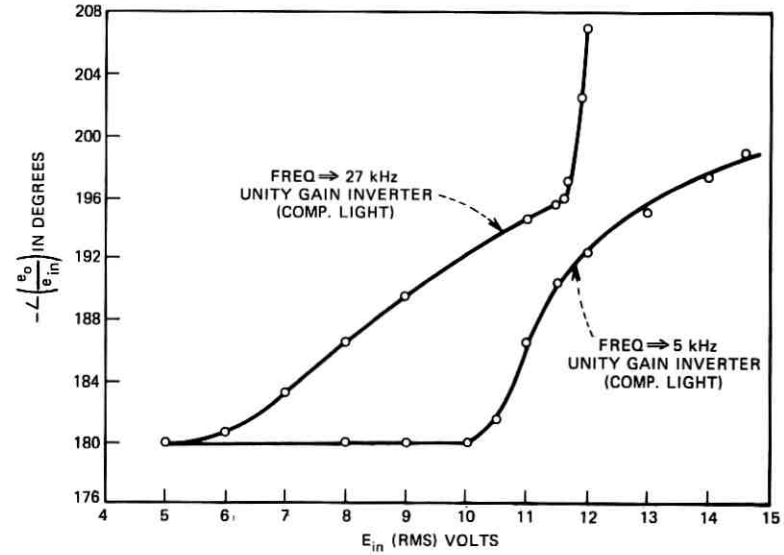


Fig. 18—Measured phase characteristics of a unity gain inverter employing a 709 op-amp.

reduced gain, the effective saturation nonlinearity comes in earlier producing the phase lag at a lower input level.

For the integrator, the same analysis applies but with the critical difference that $N_2(j\omega)$ in Fig. 16c is no longer a real quantity as in the case of the inverter. The phasor diagram for this case is shown in Fig. 19. We have chosen an integrator with $\alpha = 0$ and τ, the reciprocal of angular frequency, $2\pi \cdot 5$ kHz. As a result, $N_2(j\omega)$ will provide a 45-degree lead at 5 kHz, which will have the effect of moving the phase of the error voltage, E, 45 degrees closer to the vector V_0 than in the

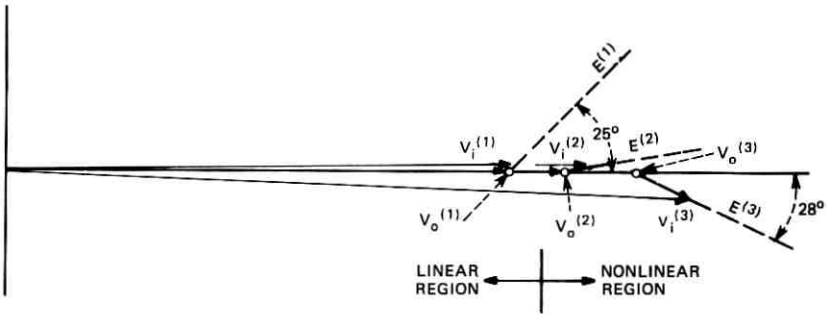


Fig. 19—Phasor diagram for integrator at 5kHz, $\tau = 1/2\pi 5$kHz, $\alpha = 0$.

case of the inverter. However, when the system is overdriven, E swings around to the lagging side of V_0, which consequently causes V_0 to lead V_i. As a result, in the integrator at this frequency, increasing the drive level produces a phase lead effect. This has been verified experimentally and the results are shown in Fig. 20.

3.2 *Active Filters*

3.2.1 *The Multiple Amplifier Biquad Filter*

The above discussion has provided a basic feel for what happens when the op-amp is employed in a simple feedback loop. We shall next discuss the performance of networks which use these inverters or integrators as building blocks.

The MAB[2] circuit was already shown in Fig. 7a in its bandpass configuration. The configuration generates a pair of poles with arbitrary Q and frequency location. To generate a pair of zeroes, outputs taken at different points are summed in a separate summing amplifier. The zeroes can also be generated by feeding the input to more than one op-amp as is done in the Multiple Input Biquad Filter (MIB). However, their basic operational frequency limitation manifests itself by a conditional stability problem in its pole forming loop. This problem is more acute when high Q's are desired, for then the resonator loop operates very close to the critical point of its Nyquist plot. This is shown in Fig. 21, where it is seen that the phase shift around the loop at the pole frequency gets very close to 180 degrees as the Q is increased.

As a consequence, a small amount of additional phase lag in any of the amplifier blocks (i.e., the inverter and the leaky and nonleaky integrators) making up the main loop can make the system unstable.

With the information already acquired, as to the characteristics of

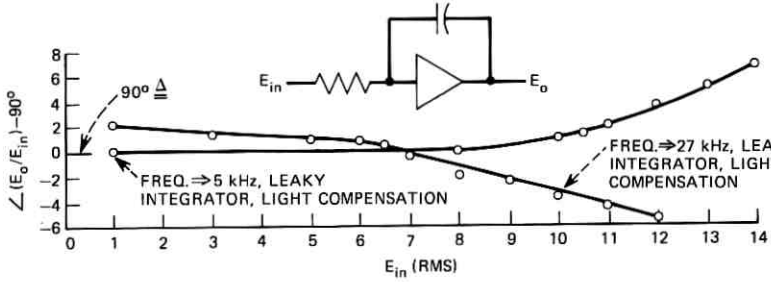


Fig. 20—Measured phase characteristics for integrator employing a 709 op-amp (the gain is unity).

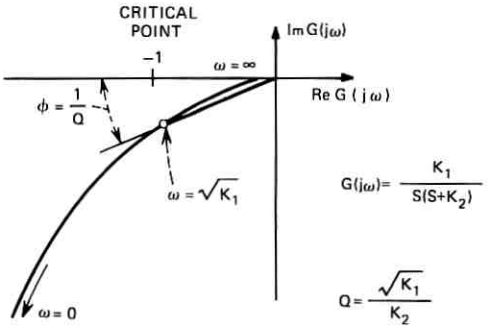


Fig. 21—Nyquist plot for main loop in MAB.

the individual amplifier blocks, the task of determining whether a given filter will be stable can be investigated. By adding up the phase shift curves of the individual amplifiers, as determined from Figs. 18 and 20, we are able to determine if the outer loop is stable with increasing drive level. As an example, we shall consider a MAB circuit which has a pole frequency at 5 kHz with a Q of 20. The design is such that

$$\omega = \frac{1}{\sqrt{R_3R_2C_1C_2}} = \frac{1}{R_3C_1} = \frac{1}{R_2C_2} = 2\pi \cdot 5 \cdot 10^3$$

and

$$Q = \frac{R_1}{R_3} = 20.$$

In this case, the integrators will have approximately unity gain at the resonant frequency of the pole. Consequently, all the three amplifiers enter the nonlinear region at roughly the same level. From the given Q we determine the phase lag needed to bring the main loop into oscillation at the pole frequency. For the reasonably high Q of 20, this is given by:

$$\phi_n = \frac{1}{Q} = 0.05 \text{ rad} \cong 2.9°.$$

Referring back to the phase characteristics of the inverter in Fig. 18, we note that in the inverter, phase lag is introduced as the input level is increased. However, this is counterbalanced by the phase lead introduced by the integrator (see Fig. 20).

It is to be noted that for very high input levels where the phase lag of the inverter becomes dominant, the circuit will, nevertheless, be stable. This results from the sharp drop in the magnitude of the DF

when the amplifiers are heavily saturated. The drop in the magnitude of the DF has the effect of shifting the critical point toward the left in the Nyquist plot of Fig. 21.

If the same circuit is scaled up to 27 kHz, the outer loop will undoubtedly be conditionally stable. In this case, both the inverter and integrator provide phase lag (Figs. 18 and 20) and at the level which produces 2.9-degree phase lag, the circuit will become unstable. As the signal level increases, due to the instability, the circuit will reach a stable limit cycle where the magnitude of the DF has decreased sufficiently to just touch the Nyquist curve.

The method of analysis employed thus far has its greatest advantage in that it is able to suggest many ways of getting around the stability problem in MAB filters. These methods center in either improving the manner in which the magnitude of the DF drops (possibly by making this occur earlier so that it becomes dominant) or improving the phase characteristics of the DF. One method which has been employed successfully is the use of a diode network across the leaky integrator. This has the effect of improving both of the DF characteristics, magnitude and phase. This circuit and the model for it are shown in Fig. 22. The model in Fig. 22c is derived very simply by taking

$$i = i' \text{ and } 0 \text{ volt at op-amp input.}$$

We then have,

$$i = \frac{-e_{in}}{R_1} = i_d(e_0) + e_0 SC_1 + \frac{e_0}{R_Q} = i'$$

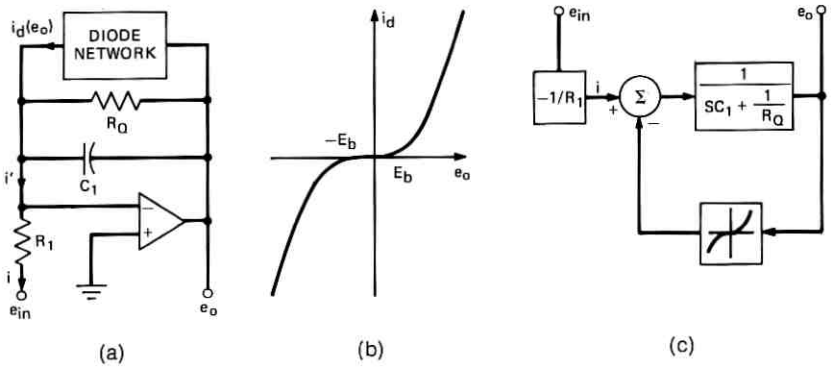


Fig. 22—(a) Diode network across integrator. (b) I/O characteristics of the diode network. (c) Model for (a).

which if rearranged, we get the equation describing the model

$$e_0 = \left[\frac{-e_{in}}{R_1} - i_d(e_0)\right] \frac{1}{SC_1 + \frac{1}{R_Q}}.$$

Drawing a phasor diagram for this circuit reveals that the magnitude of the DF drops and simultaneously phase lead is introduced as signal level increases. It is worthwhile to point out that limiting directly at the output of the op-amp has the opposite effect (as shown earlier in the example of Fig. 4); therefore, limiting cannot be blindly applied. This technique has been employed in a MAB using 709's to obtain a Q of 50 at 100 kHz. Above this frequency, the gain of a 709 op-amp is no longer sufficient to give a precision filter. Similarly, a MAB employing 702 op-amps was made to operate at 2 MHz with a Q of 10.

This method of nonlinear compensation has the disadvantage of limiting the dynamic range of the filter. Therefore, techniques which compensate the op-amps internally are more desirable. It has been shown in Section II how internal compensation affects the DF of closed loop amplifiers. In turn, these circuits may be rearranged to produce better characteristics. We will not dwell on this subject in great detail; however, we shall discuss the connection with slew rate. It is well known that slew rate limiting in operational amplifiers is caused by some form of either voltage or current limiting. Consequently, as a result of this limiting, we may expect that the DF and, therefore, conditional stability will be affected. However, this may not necessarily be the case for it may happen that the limiting element causing slew rate is not the dominant factor in determining the phase characteristic of the op-amp. This is most vividly illustrated by a 709 op-amp with input and output compensation. It has been experimentally observed that the input compensation affects the frequency range of conditional stability while the output compensation has a negligible effect. On the other hand, slew rate limiting is mostly affected by the output compensation. The dominance of the input compensation in controlling conditional stability is consistent with our derivation of the DF of an open loop 709 op-amp.[9]

3.2.2 *The Single Amplifier Biquad Filter*

The single amplifier biquad[1] circuit is shown in Fig. 23a. In Figs. 23b and c, we have the feedback structure, as seen by the op-amp, and its Nyquist plot. Here, once again, we deal with a single op-amp in a closed loop configuration and our interest lies in determining its stability.

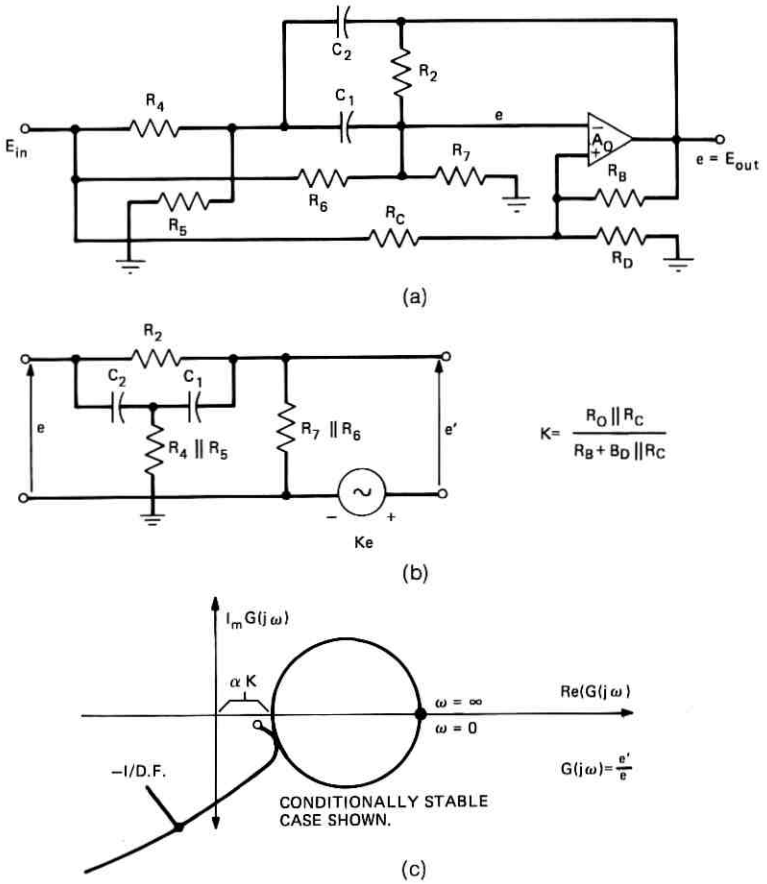


Fig. 23—(a) SAB Circuit. (b) Feedback structure as seen by op-amp. (c) Nyquist plot of feedback structure assuming $R_7 \| R_6 \gg R_2$.

From the open loop DF curves of the op-amp, we can plot the locus of the critical points as $-1/DF$. As shown in Fig. 23c, the circuit is conditionally stable as a result of more than 90 degrees of phase lag in the amplifier. One manner in which stability can be achieved is by appropriately including a diode network in the feedback structure such as to move the circle in Fig. 23c to the right when the drive level exceeds a predetermined threshold. This network can either be placed across R_2 or R_D. Another alternative, which is advantageous from a dynamic-range perspective, is to design the op-amp or its compensation such

that proper phase lead is introduced sharply as the level is increased. Modified forms of the compensation circuitry discussed in Section 3.1 can be used to achieve this effect.

3.3 *Amplitude Stabilized Oscillators*

Sine wave oscillators are usually designed in the linear domain by placing the poles slightly to the right of the $j\omega$-axis in the S-plane. If the network remained in its linear region, the network output would increase exponentially without bound. Invariably, therefore, nonlinearities are introduced to produce a stable limit cycle. Fig. 24 illustrates a typical circuit which makes use of a saturation type of nonlinearity. Since the DF of a saturating nonlinearity is real, the circuit will oscillate at the frequency of the tuned circuit, independent of the nonlinearity.

Many times, especially at high frequencies, nonlinearities with an effective real DF are difficult to obtain. In these cases, the computation of the DF (as a function of amplitude and frequency) is required to determine both the amplitude and frequency of the limit cycle. To illustrate this point, we refer to the oscillator of Fig. 25. (This circuit is being currently employed in a 20-MHz subcable oscillator. In this application, reliability necessitates an accurate determination of the frequency of oscillation.) A first look at the circuit may lead us to the conclusion that it behaves in the same manner as the oscillator shown in Fig. 24a. The pair of diodes would behave as the saturating nonlinearity while the frequency selectivity would be provided by the tank circuit. However, a second look reveals that this circuit is quite similar to that shown in Fig. 11. It can also be seen, recalling Fig. 12,

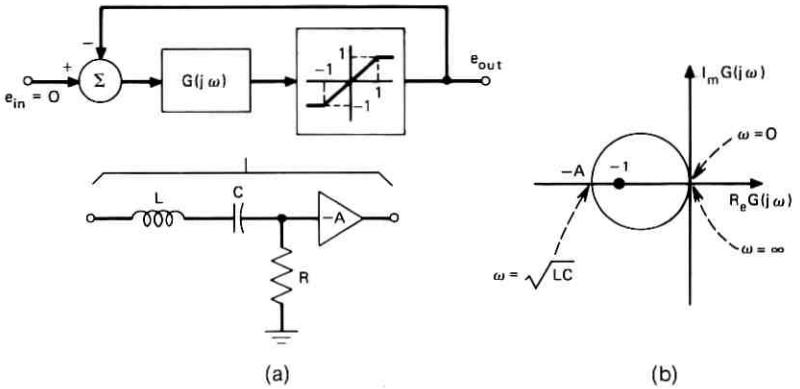


Fig. 24—(a) Amplitude-stabilized oscillator. (b) Nyquist plot.

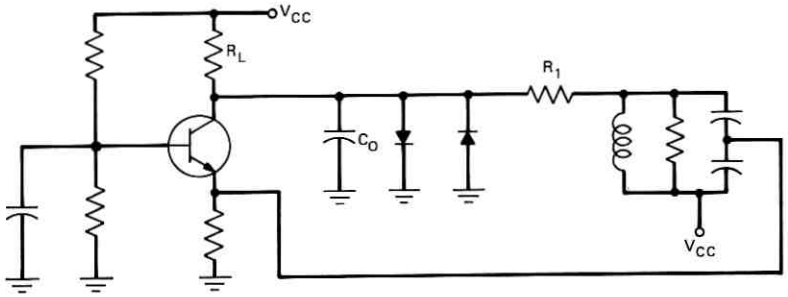


Fig. 25—Amplitude-stabilized oscillator by the use of diodes.

that the pair of diodes is effectively in a feedback configuration with frequency-sensitive elements, which produces a complex DF. Therefore, we conclude that this circuit will have an oscillation frequency dependent on the nonlinearity.

IV. CONCLUSION

A technique which analyzes the stability of frequency-selective feedback structures containing nonlinearities has been presented. This technique is an adaptation of the describing function technique to include multiple feedback structures. This is accomplished by extending the definition of the conventional describing function to include networks containing nonlinearities around frequency-dependent linear elements. By doing so, the describing function of typical operational amplifiers in open and closed loop configurations have been derived. These, then, serve as a means of predicting the conditional stability criteria in frequency-selective feedback networks such as the Multiple Amplifier Biquad and the Single Amplifier Biquad. A valuable result of this analysis is the discovery of circuits which circumvent in one way or another the conditional stability of high-frequency biquads. This result has had a highly beneficial impact on high-frequency biquads since it relieves the conditional stability problem and shifts emphasis to the maximum open-loop gain available at high frequencies. This open-loop gain determines the precision which high Q circuits can achieve at these high frequencies.

V. ACKNOWLEDGMENTS

Many of the techniques described in the memorandum came about through the special efforts of R. L. Ukeiley. He has spent long hours

in gathering important data relating to the maximum operating frequencies for MAB's and open and closed loop op-amp measurements. As a matter of fact, it was he who first observed the lack of correlation between slew rate and conditional stability. I am also indebted to J. J. Friend for his many suggestions and to R. C. Drechsler for his patient efforts in guiding this project.

APPENDIX A

A Property of the Imaginary Part of A Describing Function

Theorem: The imaginary part of the describing function, $G(A)$, associated with a multiple valued input-output function, $f(a)$, is given by

$$\text{Im } \{G(A)\} = \frac{-S_A}{\pi A^2}$$

where $S_A \triangleq$ area enclosed by $f(a)$ for $|a| < A$; positive when path taken by nonlinearity is in a counterclockwise direction.

Proof:

$$\text{Im } \{G(A)\} = \frac{1}{A\pi} \int_0^{2\pi} f(A \sin \omega t) \cos \omega t \, d(\omega t). \tag{13}$$

Let,

$$\mu = A \sin \omega t.$$

Then,

$$d\mu = A \cos \omega t \, d(\omega t).$$

It follows by substitution into eq. 13 that,

$$\text{Im } \{G(A)\} = \frac{1}{A^2\pi} \oint_{|\mu|<A} f(\mu) \, d\mu = -\frac{S_A}{\pi A^2} \qquad \text{Q.E.D.}$$

APPENDIX B

709 Op-Amp Models

The schematic of the input circuitry of a μ709 op-amp is shown in Fig. 26. The transistors Q_4 and Q_6 make up the second stage of amplification. Q_3, Q_5, and Q_7 decode the differential output of Q_1 and Q_2 providing a single-ended signal at the base of Q_4. The input compensation, a series RC network, is placed across base and collector of the Darlington pair, Q_4 and Q_6. This provides (together with the

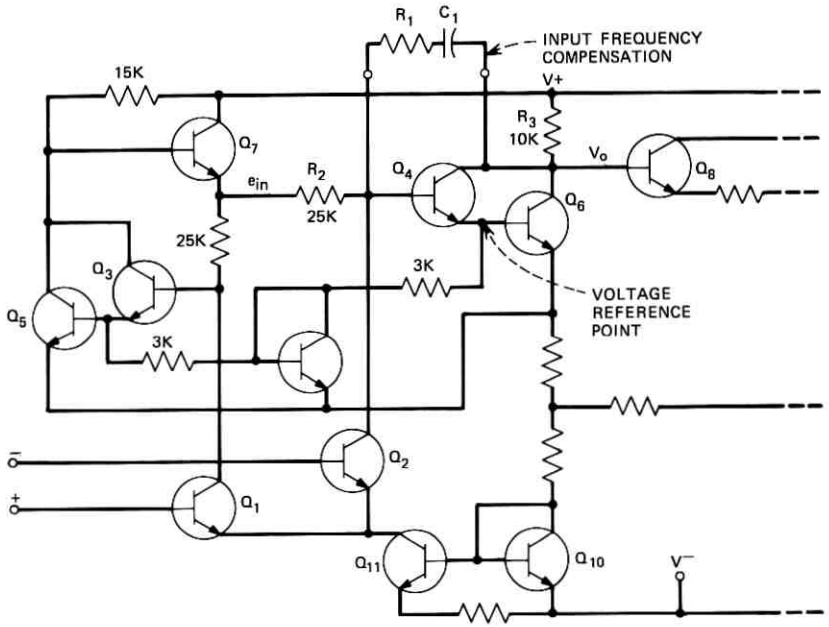


Fig. 26—Input circuitry of μA 709 op-amp.[10]

output compensation) the necessary frequency rolloff to stabilize the amplifier in the linear domain.

Fig. 27 depicts the approximate circuit model of the second-stage amplifier derived from the Ebers–Moll model.[11] We have assumed that the input impedance of the Darlington stage is much larger than R_2 and the load impedance presented by the emitter follower, Q_8, is negligible compared to R_3. The cutoff region is represented by the characteristics of V_a, shown in Fig. 27b. Saturation is introduced in the model by the diode D. In Fig. 27a the current I is given by

$$I = \frac{e_{in} - V_0}{R_2 + R_1 + \dfrac{1}{SC_1}} = \frac{V_0 - V_a(V_{be})}{R_3} + i_D(V_0)$$

Rewriting this equation, gives

$$V_0 = \frac{Z_2}{Z_1} e_{in} + V_a(V_{be}) \frac{Z_2}{R_3} - i_D(V_0) Z_2 \tag{14}$$

where,

$$Z_1 \triangleq R_1 + R_2 + \frac{1}{SC_1}$$

$$Z_2 = \frac{Z_1 R_3}{Z_1 + R_3}.$$

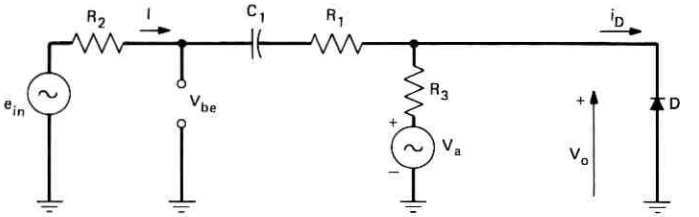


Fig. 27a—Circuit model for the second-stage amplifier.

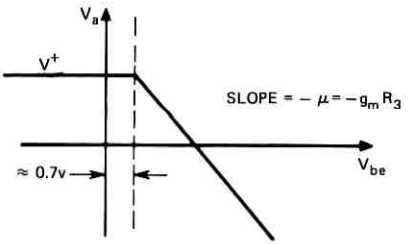


Fig. 27b—Nonlinear elements of the second stage amplifier corresponding to the cutoff region.

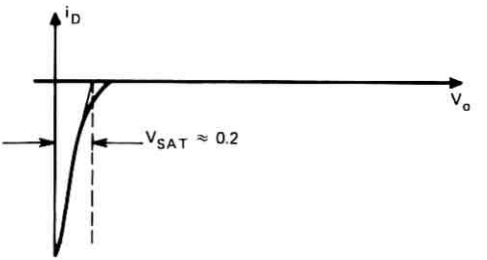


Fig. 27c—Nonlinear elements of the second stage amplifier corresponding to the saturation region.

We also obtain from Fig. 27a

$$V_{be} = e_{in} - \left(\frac{e_{in} - V_0}{Z_1}\right)R_2$$

$$= e_{in}\frac{R_2}{Z_3} + V_0\frac{R_2}{Z_1} \tag{15}$$

where,

$$Z_3 = \frac{Z_1 R_2}{Z_1 - R_2}.$$

Figure 28 is a block-diagram representation of eqs. 14 and 15. For simplicity, we distinguish between the cutoff and saturation regions. The result of this is shown in Fig. 29. In both models, the forward path containing Z_2/Z_1 in Fig. 28 is neglected since the other path through V_a has a much larger gain. We have also assumed, with some loss of generality, that the biasing is such that a sufficiently high drive level producing cutoff (saturation) does not cause saturation (cutoff) during some other portion of the cycle. However, in general, the problem can be solved by computing the DF for the inner loops involving i_D and V_a and employing these results to solve for the DF for the complete system.

The model of Fig. 29a represents the second-stage amplifier when driven into cutoff. It is obtained directly from Fig. 28 by making a slight rearrangement at the output in addition to the two previously

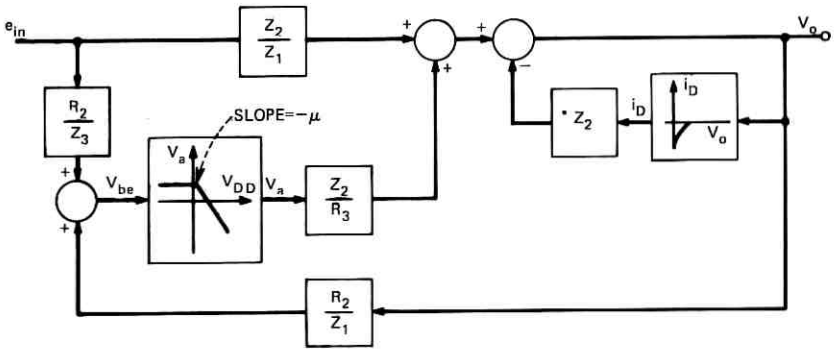


Fig. 28—Block diagram representation of eqs. 14 and 15.

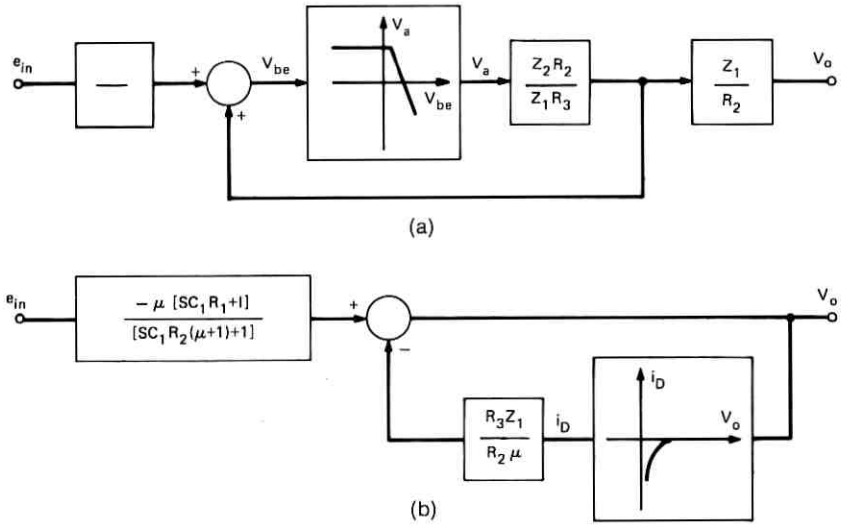


Fig. 29—(a) Model for the cutoff region. (b) Model for the saturation region.

mentioned assumptions. The network parameters are given by

$$\frac{Z_2R_2}{Z_1R_3} = \frac{SC_1R_2}{SC_1(R_1 + R_2 + R_3) + 1}$$

$$\frac{R_2}{Z_3} = \frac{SC_1R_1 + 1}{SC_1(R_1 + R_2) + 1}$$

$$\frac{Z_1}{R_2} = \frac{SC_1(R_1 + R_2) + 1}{SC_1R_2}.$$

The model shown in Fig. 29a depicts the saturation region. It is obtained from Fig. 28 by restricting V_a vs V_{be} to operate in the linear region. Consequently, this nonlinear element (V_a vs V_{be}) in Fig. 28 becomes a linear gain element. The process of reduction is shown in Fig. 30. Figure 30c, after factoring out the feed-forward path and assuming a large μ, yields the saturation model of Fig. 29b. The linear element in the feedback loop in terms of fundamental parameters is:

$$\frac{R_3Z_1}{R_2\mu} = \frac{R_3}{R_2\mu}\left[\frac{SC_1(R_1 + R_2) + 1}{SC_1}\right].$$

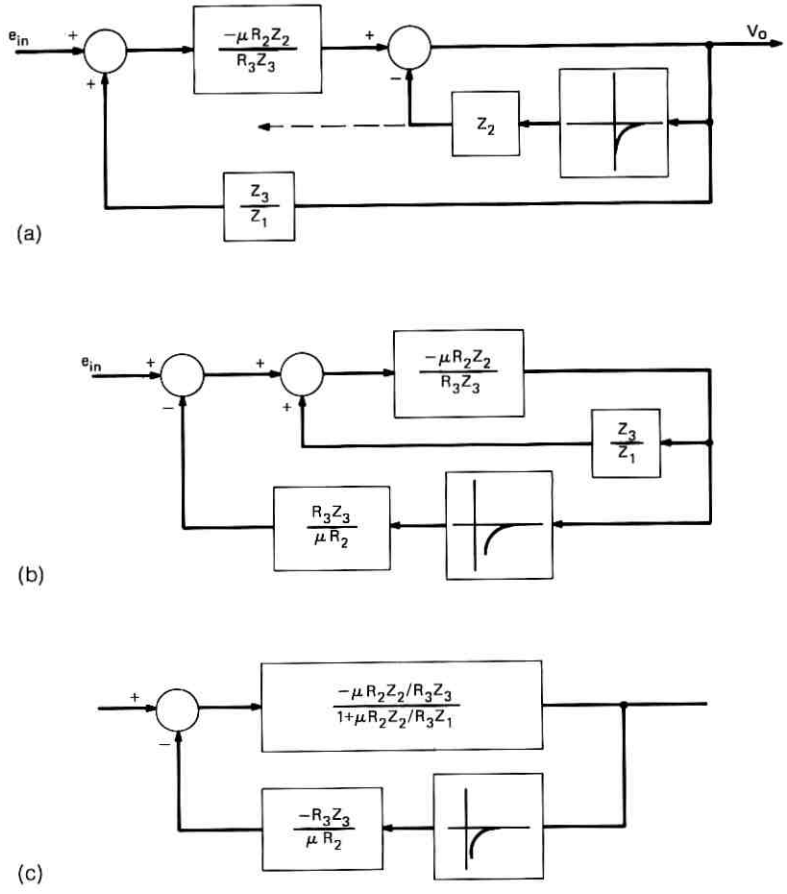


Fig. 30—Reduction of Fig. 28.

REFERENCES

1. Friend, J. J., "A Single Operational Amplifier Biquadratic Filter Section," 1970 IEEE International Symposium on Circuit Theory Digest, December 14–16, 1970, pp. 179–180.
2. Thomas, Lee C., "The Biquad: Part I—Some Practical Design Considerations" and "The Biquad: Part II—A Multipurpose Active Filtering System," IEEE Trans. Circuit Theory, *CT-18* (May 1971), pp. 350–361.
3. Goldfarb, L. C., "On Some Nonlinear Phenomena in Regulatory Systems," Automatika i Telemekanika, *8*, No. 5 (September–October 1947), pp. 349–383.
4. Gibson, J. E., *Nonlinear Automatic Control*, New York: McGraw-Hill, 1963, pp. 203–235, 343–438.
5. Siljak, D., *Nonlinear Systems*, New York: John Wiley, 1969, Chapters 3, 4, 5, 6, 7 and App. F and G.

6. Krylov, N., and Bogoliubov, N., *New Methods in Nonlinear Mechanics*, Moscow: State Publishing House, 1934.
7. West, J. C., Douce, J. L., and Livesly, R. K., "The Dual Input Describing Function and Its Use in the Analysis of Nonlinear Feedback Systems," Proc. IEEE, *103*, Part B, 1956, pp. 463–474.
8. Op. Cit., Siljak, D., pp. 472, 481, 498–510.
9. Ukeiley, R. L., private communication relating to unpublished data concerning frequency and slew rate limitations.
10. Giles, J. N., *Fairchild Semiconductor IC Applications Handbook*, Fairchild Semiconductor, 1967, pp. 57–59.
11. Angelo, Jr., E. J., *Electronics: BJT's FET's and Microcircuits*, New York: McGraw-Hill, 1969, pp. 245–250.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 9, November, 1972
Printed in U.S.A.

Waveguide Breakdown Effects at High Average Power and Long Pulse Length

By A. S. ACAMPORA and P. T. SPROUL

(Manuscript received March 17, 1972)

Analytical predictions of the power handling capabilities of waveguide systems generally have not considered the effects of high average power and long pulse length. It has experimentally been noted, however, that a substantial reduction in power handling capability below expected levels does occur as average power and pulse length are increased. This reduction can be attributed to the presence of loose particulate matter which is heated by the average power, causing localized rarefication of the dielectric gas fill as well as the expected voltage enhancement.

An unstable arcing situation is shown to exist when the arc duration exceeds some critical time. Typical pulse lengths in use today exceed this critical time and may result in continuous arcing. The use of control circuitry to terminate each particle-induced arc prevents continuous arcing and deletes the particle, and is therefore essential to stable operation at long pulse lengths and high average power.

I. INTRODUCTION

An understanding of microwave breakdown in gases has existed for many years[1,2] and has been successfully applied to the design of many high-power microwave systems operating at low duty cycles with pulse lengths on the order of several microseconds. However, recent testing of S-band WR284 microwave hardware filled with sulfur hexafluoride (SF_6) and operated at 500- to 1500-kilowatt average power, and 10- to 150-megawatt peak power with 50- to 150-microsecond rectangular pulses showed performance far below expectation. A series of tests and supporting analytical work were undertaken to explain this behavior. These have shown that the empirically observed reduction in peak power handling capability at high levels of average power is caused by the presence of loose particulate matter, microscopic in size and lossy at microwave frequencies. The experimental and analytical proof of the mechanisms responsible for this effect is described. A procedure is

presented employing short duration discharges to break up these particles, thereby increasing power handling capability.

Waveguide cavities and resonant ring structures of various lengths were tested because the empirical peak breakdown power level was found to decrease with resonator length. It is shown that this effect is related to the arc energy absorption which is dependent upon resonator length. A quantitative understanding of this phenomenon is developed and is applied to confirm the experimental results.

II. BREAKDOWN MECHANISMS

The parameters which affect the peak power handling capability of a microwave system are:

(*i*) the maximum electric field strength appearing within the system
(*ii*) the nature of the dielectric gas employed
(*iii*) the molecular density (or, equivalently, the absolute pressure and temperature) of the dielectric gas
(*iv*) the microwave frequency
(*v*) the breakdown volume.

Free electrons which always exist within the gas because of cosmic radiation and other random phenomena are accelerated by the electric field present and suffer collisions with neutral gas molecules. If the kinetic energy imparted to an electron by the field exceeds the ionization potential of the gas, the possibility exists of producing an additional electron upon collision with a neutral gas molecule. If the kinetic energy of the accelerated electron is less than the ionization potential, it may become attached to a neutral gas molecule upon collision. Diffusion produces a net flow of electrons from regions of high electronic density to regions of lower density. Some recombination of electrons and positive ions also occurs.

Breakdown occurs when the rate of electron production via ionization exceeds the combined rate of electron loss through the processes of attachment, diffusion, and recombination. For most high power transmission systems, the gas pressure is in the region of one to two atmospheres, and recombination rates are negligibly small. Also, since the dimensions of the breakdown volume are generally much greater than the characteristic diffusion length of the gas, electron loss via diffusion is also negligible. Hence, breakdown occurs when the rate of ionization exceeds the rate of attachment.

Both ionization and attachment rates are functions of the mean

electron energy[3] which, in the high pressure regime where collision frequency is far larger than the microwave frequency, is dependent on the ratio of the peak electric field strength (E) to the gas molecular density (N). Assuming the density is defined by the ideal gas law, the field strength at breakdown is given by

$$E = C\frac{\rho}{KT} \tag{1}$$

where

ρ = absolute gas pressure
K = Boltzman's constant
T = absolute temperature
C = constant, dependent upon dielectric gas

For SF_6, C = 3.69×10^{-19} volt-meter^2/molecule.

For irregularities in the waveguide such as those which occur at flanges, in hybrids, etc., the local voltage gradient is increased by a field enhancement factor (β). Substituting into (1) the waveguide field strength required for breakdown becomes

$$E = \frac{C\rho}{\beta KT}. \tag{2}$$

The electric field strength generated in a rectangular waveguide transmitting a power P can be shown to be

$$E^2 = \frac{4Z_oP}{d_1d_2} \tag{3}$$

where d_1 and d_2 are the waveguide cross-section dimensions and Z_o is the wave impedance. Therefore, the breakdown power threshold can be derived from (2) and (3) as

$$P = \left(\frac{d_1d_2}{4Z_o}\right)\left(\frac{C^2}{\beta^2K^2T^2}\right)\rho^2. \tag{4}$$

For constant temperature, this is the familiar relationship between peak power breakdown and the square of the absolute pressure, known to be valid in the high-pressure region where electron loss by diffusion is negligible.

III. EFFECTS OF HIGH AVERAGE POWER

The relationship discussed above has been understood and effectively applied for many years. However, in the experimental work described

below, a significant reduction in peak breakdown power was observed as the average power increased. It is noted that, although eq. (4) is not explicitly dependent upon average power, temperature does appear. Hence, it was postulated that loose particulate matter was reducing the breakdown power level because a loose particle is not thermally attached to the waveguide wall; its temperature therefore increases as the average power increases. This causes a reduction in the molecular density of the gas at the location of the particle, thereby reducing the field strength needed to cause breakdown. The presence of a particle also produces localized field enhancement. These two localized effects combine to lower the breakdown power.

The electric field enhancement factor and the equations for temperature rise at various levels of average power are derived in Appendix A for spherical conducting particles. This derivation neglects intrapulse heating since, within the range of pulse lengths investigated, it is negligible compared to average power heating. Using these results and eq. (4), peak power breakdown levels can be calculated for various average power levels and particle sizes. Figure 1 shows the results of such calculations for WR-284 waveguide filled with SF_6 at 25 psig containing spherical copper particles from 4 to 120 microns in diameter. Also appearing on Fig. 1 are experimental points obtained by introducing copper spheres of known diameter into a traveling wave resonator[4] constructed of high-conductivity copper to minimize losses. Correlation between theory and experiment is seen to be good.

It is observed from Fig. 1 that, within the range of particle sizes investigated:

(*i*) Peak breakdown power levels at very low levels of average power are independent of particle size and are below that of ideal waveguide because of field enhancement only.

(*ii*) As the average power increases, the presence of heated particles causes further reduction in power handling capability due to localized reduction in gas density. Larger particles cause a greater reduction in power handling capability because they are heated to higher temperatures.

The analysis from which Fig. 1 was derived is applicable when the particle dimensions are small compared with the waveguide dimensions, and large compared with the characteristic diffusion length of the gas. It should not be inferred from Fig. 1 that particle-induced arcing cannot occur above 100 megawatts, since particles with dimensions smaller than the diffusion length will cause breakdown to occur above this level. An analysis of particle behavior in this region was not considered.

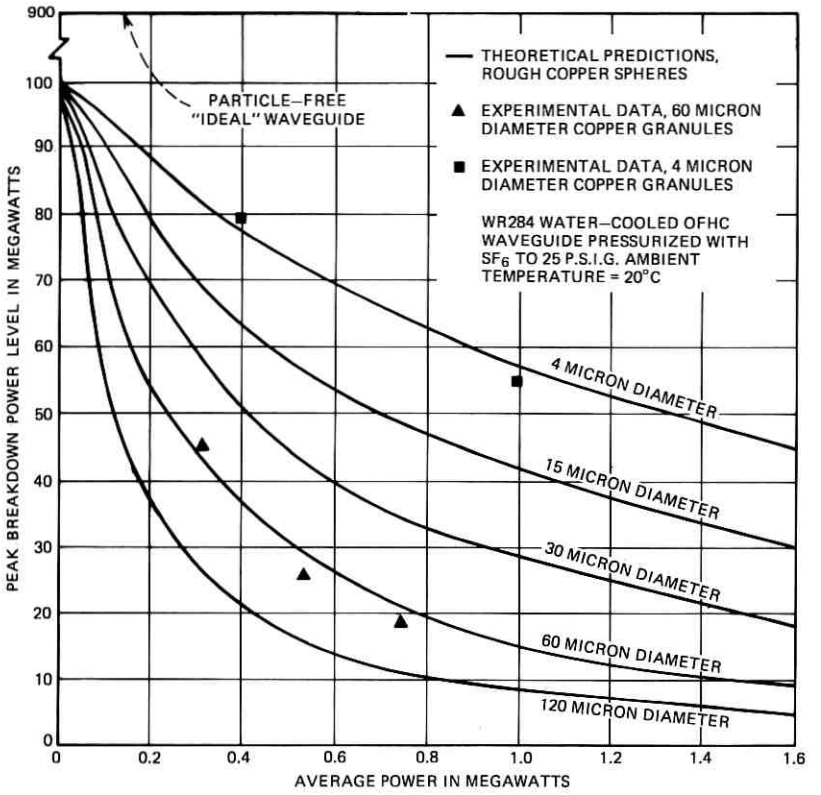


Fig. 1—Effect of particles on breakdown.

IV. EXPERIMENTAL WORK

The equipment used to obtain the experimental data is shown in schematic form in Fig. 2. A resonant ring was used to generate the high power levels required to produce breakdown with the RF power source available. A photograph of the equipment is shown in Fig. 3. View ports were provided to permit visual observation of arcing. Thermocouples were inserted in flanges and attached to the waveguide in various places to monitor temperature rise versus average power. Forward and reverse power was monitored by power meters and by oscilloscopes. A counter was used to record the number of arcs.

Because of the high average power levels, the waveguide was water-cooled. Two cooling channels were soldered on the top and bottom walls through which water was circulated at a rate of 7 gallons per minute. To permit assembly of the flange bolts, the cooling channels were

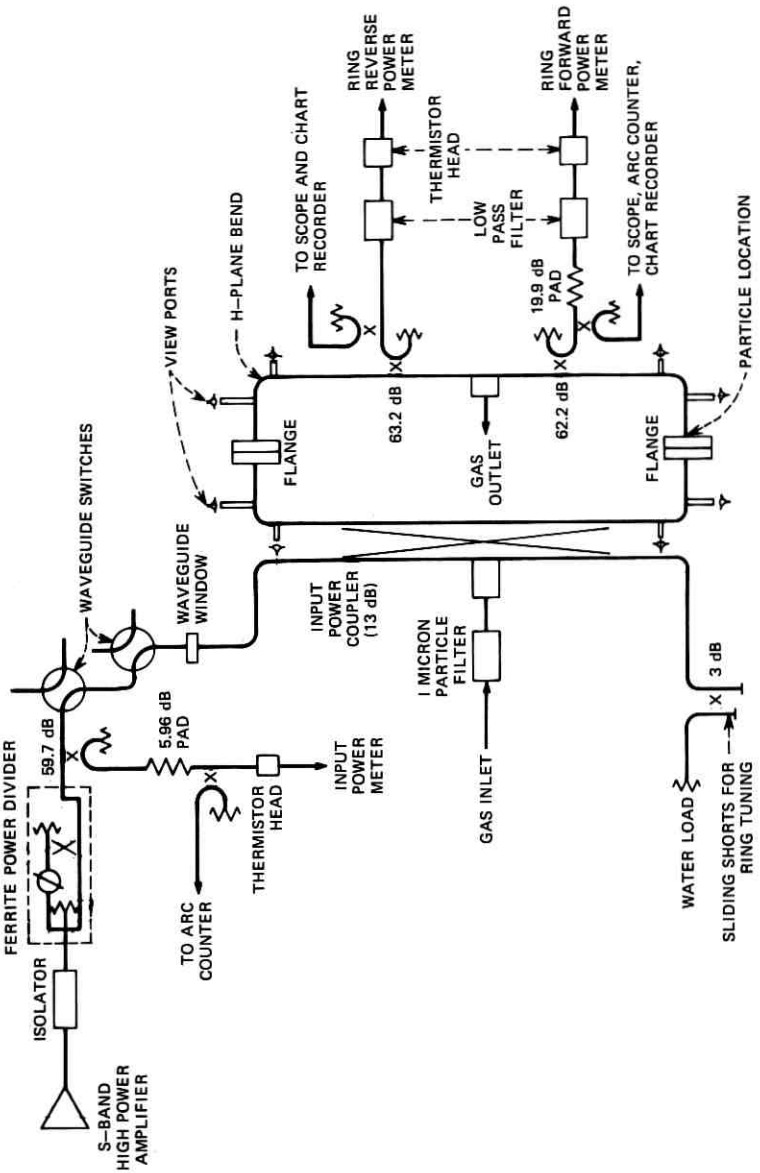


Fig. 2—Resonant ring experimental setup.

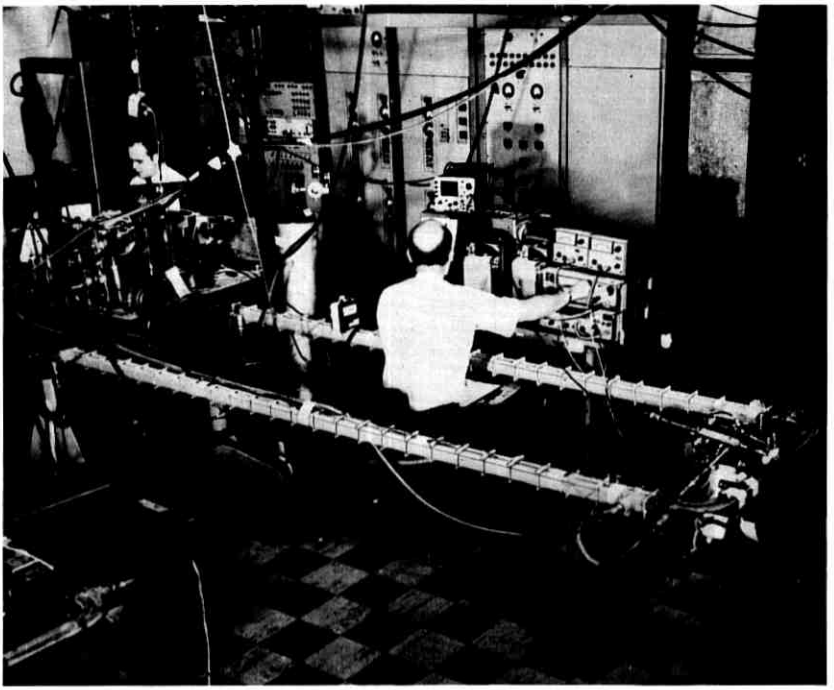


Fig. 3—Resonant ring.

terminated 2-1/2 inches from the flange. Because of this, the flange temperature was found to rise at a rate of 65°C/megawatt of average power. This factor has been included in the particle temperature calculations of Appendix A. The flange used was adapted from a vacuum-tight design, the principle feature of which was the use of a stainless steel knife edge and a soft copper gasket to provide the RF and gas seal. Figure 4 shows a model of the flange and cooled waveguide.

V. RF PROCESSING

Early experience with the setup described above showed arcing at substantially lower peak power levels than expected. By use of the viewports, it was observed that many arcs were being initiated at the power coupler and monitoring coupler slots. These were attributed to high gradients at sharp corners. By allowing arcing to continue, the test setup was "aged" to progressively higher and higher peak and average powers. When arcing no longer occurred at these points, it was observed

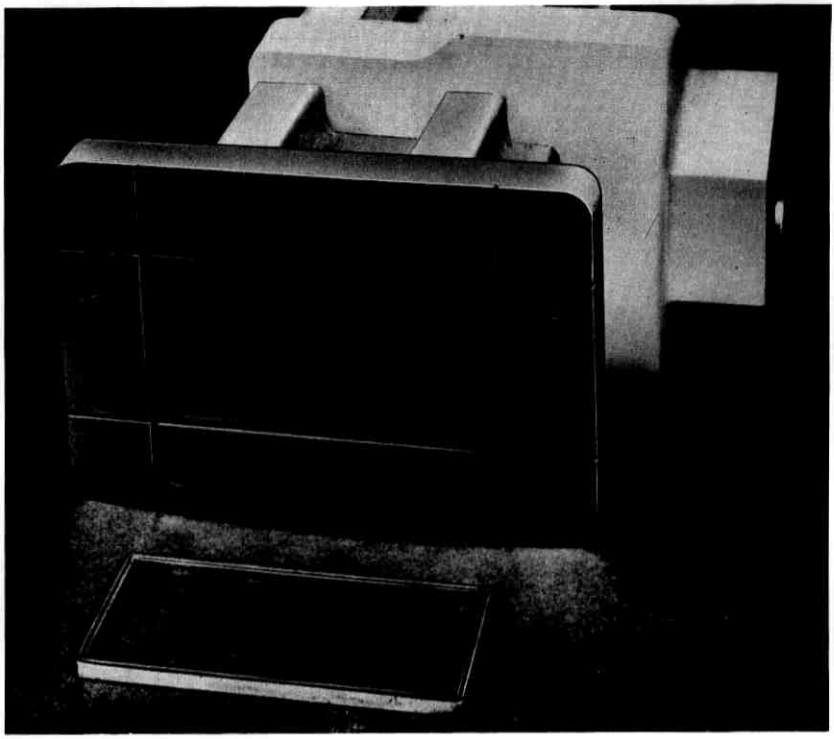


Fig. 4—Flange.

to occur at randomly located places, becoming more intense at higher levels of average power. This observation was a factor in generating the hypothesis that particulate matter was inducing breakdown. The average power limitation imposed by the particulate matter was experimentally found to be lessened by allowing the arcing to continue until stopped, a procedure which came to be known as "processing."

As shown in Fig. 1, when a high average power is present, very small particles can cause arcing at levels well below the inherent peak power handling capability of the system. Processing is accomplished by applying power to the assembled system, increasing the power levels until breakdown is induced by the particles and then waiting until arcing subsides. Processing arcs change the physical configuration of the initiating particles by (*i*) breaking them into smaller particles, (*ii*) removing them from the high field region via breakdown pressure waves, and/or (*iii*) vaporization. Figure 1 shows that, as the particles

become smaller, a higher peak power can be handled before arcing occurs. Therefore, by raising the power level in discrete steps and remaining at each level for a period of time sufficient to allow processing to occur, the power-handling capability of the microwave system increases to a higher value. This value can be maintained as long as the system is intact, that is, not opened. Disassembly always introduces particulate matter and some reprocessing is required to achieve the former stable power handling level.

Figure 5 shows an experimental processing sequence. The data was taken in a 10-foot ring set up as shown in Fig. 2. Parameters for this experiment were:

Pulse length—120 microseconds
Dielectric gas—SF_6 at 25 psig
Ambient temperature—28°C.

The charts appearing in Fig. 5 depict the number of arcs which occurred in two-minute intervals at the indicated power levels. In all tests, duty cycle was varied by holding the pulse length fixed and changing the pulse repetition rate. Arcing began during Run 1 at a peak power level of 12.6 megawatts with a 1.5 percent duty cycle. It is noted that (*i*) each time the power levels were raised to higher values, breakdown occurred, and (*ii*) at each power level, the number of arcs which occurred per two-minute interval decreased with time. At and above the peak power level of 66.7 megawatts, the average power was held constant at 1 megawatt. Performance became unstable at the peak power level of 139 megawatts.

Run 2 was then executed. It is noted that the processing arcs which occurred during Run 1 improved system performance to the extent that during Run 2, no discharges occurred below the power levels of 66.7-megawatts peak, 1-megawatt average. Also, fewer arcs occurred at the higher power levels than during Run 1. Performance became unstable at a power level of 151 megawatts peak. The results of this experiment demonstrate that processing a system to a high power level raises the power level at which subsequent discharges occur and reduces their severity.

Ideally, processing should be performed at power levels exceeding system requirements so that operation at the design power levels would be virtually arc-free. Such a procedure cannot be applied to a typical system which normally operates at the maximum available power level. Processing can still be applied in such a situation, however, if it is performed with the dielectric strength of the gaseous fill somehow reduced. This can be accomplished either by introducing a low dielectric

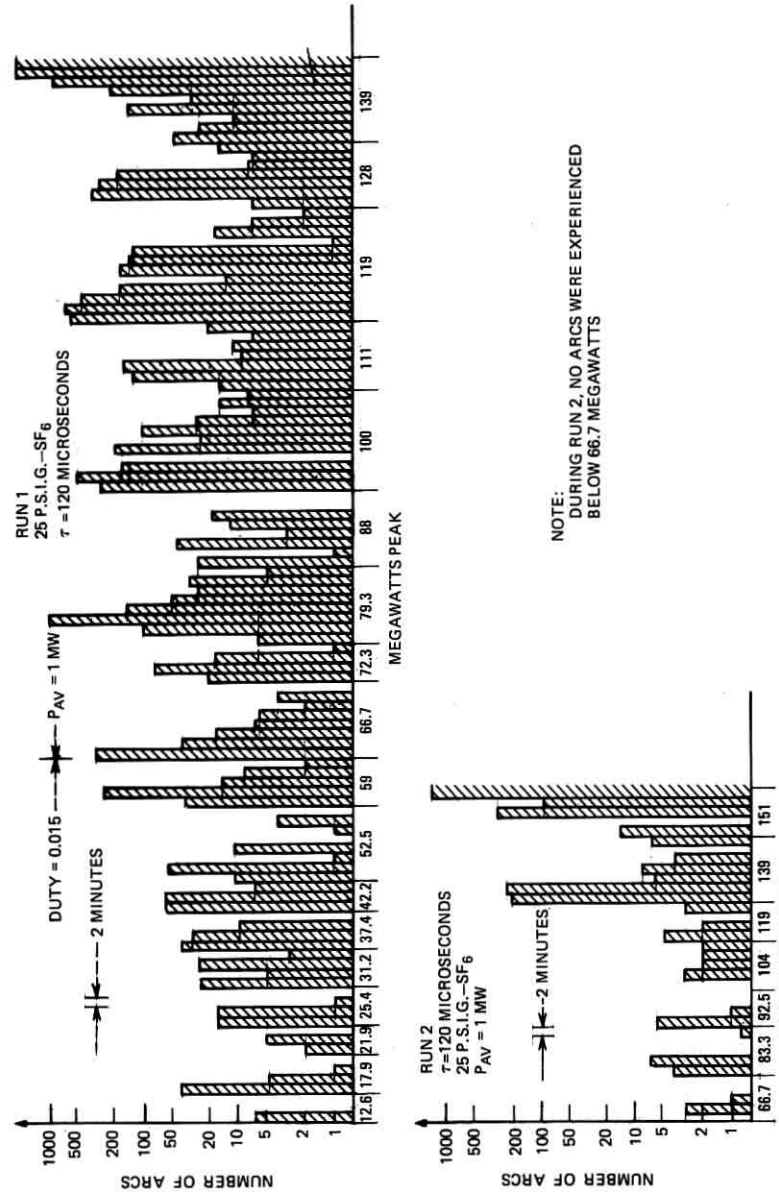


Fig. 5—Waveguide processing.

strength gas before processing, or by processing with the pressure of the normal dielectric gas reduced. Processing can then occur to a greater extent with the available power and will result in stable performance when the system is operated with normal dielectric strength.

Dry nitrogen was first investigated as a processing agent because of its low dielectric strength. It was found that after the initiation of a processing arc, nitrogen does not recover its dielectric strength sufficiently fast to prevent breakdown on all successive pulses. Hence, processing cannot be satisfactorily performed using this gas.

Processing with low pressure SF_6 was next investigated. A 32-foot resonant ring was selected for this purpose. Results are shown in Fig. 6. The ring was processed to power levels of 20 megawatts peak, 700 kilowatts average while filled with SF_6 at a pressure of 5 psig. After increasing the gas pressure to 25 psig, the peak power level was raised with the average power held constant; breakdown did not again occur until a peak power level of 35 megawatts had been achieved. Hence, processing with 5 psig of SF_6 provided almost 3 dB of margin in power handling capability when the system was repressurized to 25 psig. This data was successfully repeated after disassembly and reassembly of the structure. Hence, reduced pressure SF_6 can be employed in the processing of a practical system to assure some power handling margin.

VI. PROCESSING CONSIDERATIONS AND LIMITATIONS

The energy dissipated by processing arcs must be kept sufficiently low to prevent the temperature rise of the waveguide surface in the vicinity of the arc from exceeding the melting point of the waveguide material. If melting of the waveguide wall occurs, the turbulence of the arc can eject small masses of molten metal into the waveguide interior, preventing processing since the additional particles induce continuing breakdown. Hence, the arc rate, instead of subsiding with time, increases. Such a situation is referred to as massive breakdown and is evident in Fig. 5, Run 2 at the 151-megawatt level.

It will be shown that the power level at which massive breakdown occurs is inversely proportional to the square root of the time duration of a single arc. To analyze the effects of arc duration upon waveguide processing, it is necessary to determine (*i*) the formative time of a processing arc, and (*ii*) the percentage of incident power absorbed by a growing arc as a function of time. The processes contributing to the growth of an arc which begins at the surface of a small particle have not, to the authors' knowledge, been investigated in any depth. In our

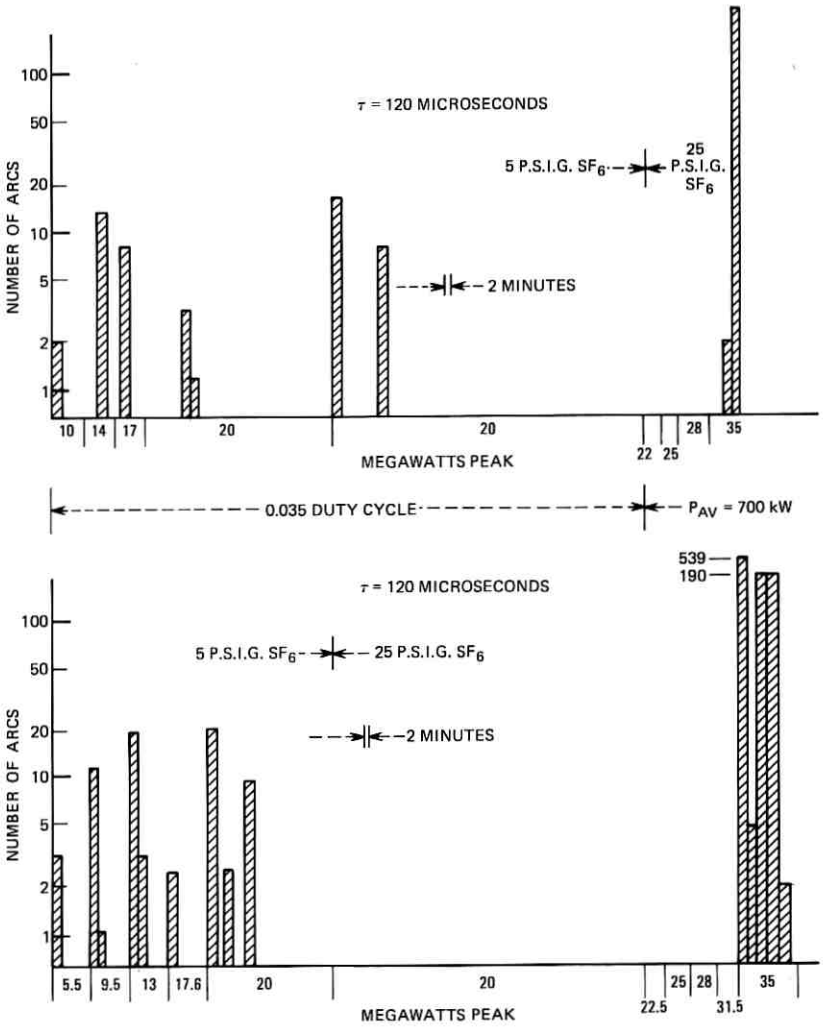


Fig. 6—Low-pressure processing.

investigation, no rigorous analytical determination of the desired two quantities was attempted; rather, both were determined experimentally. The experimental procedure is contained in Appendix B. The waveforms observed suggest that the power absorbed by a processing arc when the incident power is constant can be approximated by the form:

$$P_A = \alpha[1 - e^{-t/T}]P_o \tag{5}$$

where

P_o = incident power
T = formative time constant of the arc, independent of power level
α = steady-state absorption percentage.

For WR-284 waveguide pressurized to 25 psig with SF_6, it was found that $T = 2$ microseconds and $\alpha = 1.5$ percent.

The rise in temperature of the waveguide wall in the vicinity of a processing discharge can be determined by assuming that the thermal conductivities and heat capacities of the dielectric gas and the waveguide material are such that power dissipated within an arc is transferred instantaneously to the waveguide walls. Consider the semi-infinite block of metal shown in Fig. 7. Let $P_A(t)$ be the power per unit area at the metal surface $x = 0$. It can be shown[5] that the temperature rise within the metal for $t > 0$ is given by:

$$\Delta T = \frac{\gamma}{2} \int_0^t \frac{P_A(u)}{\sqrt{t-u}} e^{-[x^2/4(t-u)]}\, du \tag{6}$$

where u is a dummy variable of integration and γ is a constant dependent upon the thermal conductivity and heat capacity of copper. At the

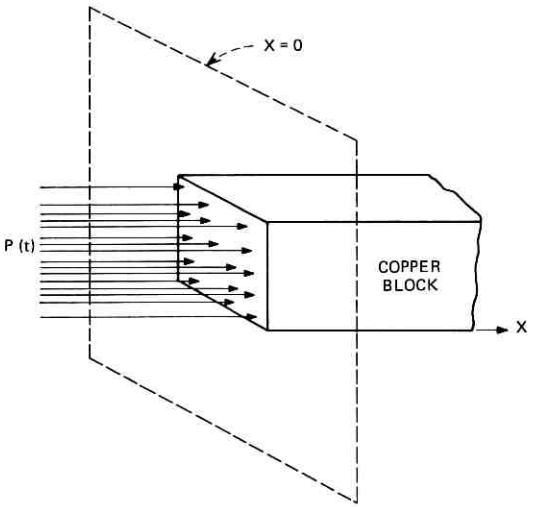


Fig. 7—Model employed to calculate ΔT.

surface of the metal, $x = 0$, this integral becomes:

$$\Delta T = \frac{\gamma}{2} \int_0^t \frac{P_A(u)}{\sqrt{t-u}}\, du. \tag{7}$$

The power absorbed by an arc is given by eq. (5). Consequently, by substituting (5) in (7), the temperature rise of the waveguide surface at time t after the arc is initiated is found to be:

$$\Delta T = \frac{\gamma}{2} \int_0^t \frac{\alpha[1 - e^{-t/T}]P_o/A}{\sqrt{t-u}}\, du \tag{8}$$

where A is the cross-sectional area of the arc.

For our situation, the arc duration, t, exceeds the arc formative time constant, T, so that:

$$\Delta T(o, t) \cong \frac{\alpha\gamma}{2} \frac{P_o}{A} \int_0^t \frac{1}{\sqrt{t-u}}\, du \tag{9a}$$

or

$$\Delta T \cong \frac{\gamma\alpha}{A} P_o \sqrt{t}. \tag{9b}$$

Therefore, the temperature rise experienced at the waveguide wall varies as the square root of arc duration, t. To prevent the waveguide wall temperature from exceeding the melting point of the waveguide material, the arc duration, t, must be kept small.

To test the validity of the above theory of processing limitation resulting from particle generation, processing was performed within WR-284 resonant structures of various lengths, pressurized with SF_6 at 25 psig. The duration of an arc which occurs within a resonant structure is determined by the time required to dissipate all of the energy which was stored within the resonator prior to breakdown. Since stored energy increases with resonator length, arc duration also increases with resonator length.

Whereas the incident power in a terminated line is a constant, the power incident upon an arc in a resonant structure decreases with time as the stored energy is consumed by the arc. An expression for this incident power as a function of time $P_o(t)$ is derived in Appendix B (eq. 60). The power absorbed by the arc is therefore:

$$P_A(t) = \alpha P_{F_o}(1 - e^{-t/T}) \exp\left\{-\frac{v_o\alpha}{l}[t - T(1 - e^{-t/T})]\right\} \tag{10}$$

where

P_{F_e} = resonator equivalent power level prior to breakdown
v_g = group velocity of electromagnetic wave within the resonator
l = resonator length.

Substituting this value of absorbed power into equation (6) yields:

$$\Delta T = \frac{\gamma}{2A} \int_0^t \frac{\alpha(1 - e^{-u/T}) \exp \{-(v_g/l)\alpha[u - T(1 - e^{-u/T})]\} P_{F_e}}{\sqrt{t - u}}. \quad (11)$$

Experimental verification of these relationships was accomplished on four resonant structures. The first two were waveguide cavities with input and output irises designed for optimum power multiplication. The other two were resonant ring structures. All structures were pressurized with SF_6 at 25 psig and processed as described above to the point of massive breakdown. The results are as shown in Table I.

Resonators 1 and 3 were operated at a constant 0.5-percent duty cycle. Resonators 2 and 4 were operated with the average power held constant at 1 megawatt. The massive breakdown power reached is noted. The last column is the calculated value of the relative wall temperature reached at massive breakdown using eq. 11. The calculations were performed numerically on a digital computer and are plotted in Fig. 8.

These calculated maximum values are relatively close, in spite of the fact that the structures are radically different in dimensions. The calculated wall temperature rise of the short resonator deviates most. This is probably because the power level at which breakdown occurred was relatively close to the theoretical breakdown limit of the waveguide which is 900 megawatts. These experimental data imply that massive breakdown occurs when the rise in waveguide surface temperature above ambient operating temperature exceeds a critical value of approximately 850 γ/A.

A microscopic examination of some of the parts used in these experiments confirmed the analysis. Arc marks were found to be present upon

TABLE I—BREAKDOWN POWER LEVEL VERSUS LENGTH

No.	Resonator Length (Meters)	Massive Breakdown Power Level (Megawatts)	ΔT Maximum
1	0.135	650	675 γ/A
2	3.05	150	875 γ/A
3	3.66	150	895 γ/A
4	9.65	70	960 γ/A

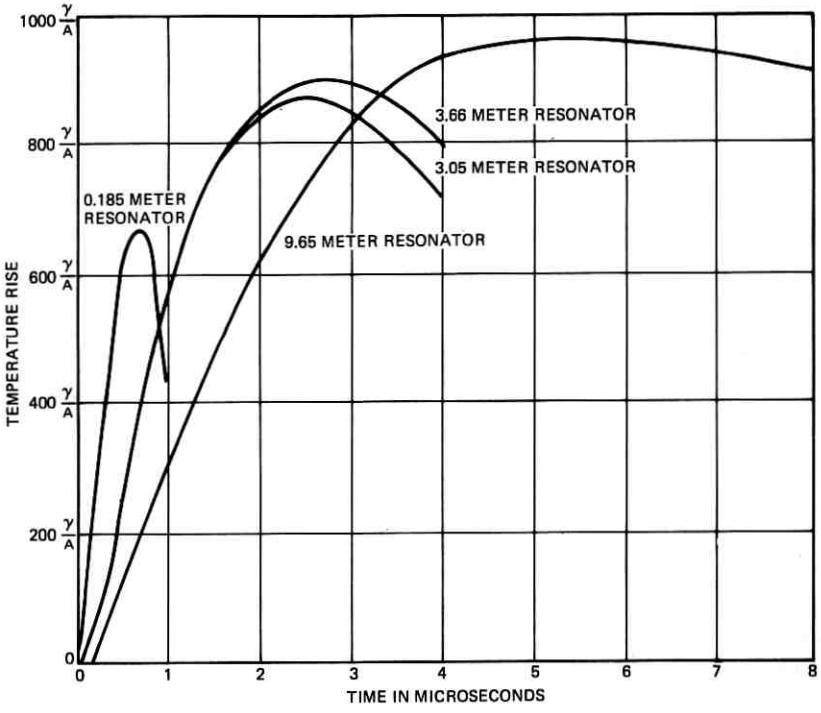


Fig. 8—Waveguide surface temperature rise during massive breakdown arcs.

the waveguide flange gaskets which were removed from the 10-foot and 38-foot resonant rings after massive breakdown. One such arc mark, magnified approximately 200 times is shown in Fig. 9. To achieve proper lighting, it was necessary to tilt the gasket under the microscope. Hence, only a thin horizontal strip is in focus. It is noted that the arc mark consists of an area of discoloration approximately 6 millimeters in diameter, and that a small hemispherical crater approximately 30 microns in diameter is located near the center of the discoloration. The area of discoloration is believed to have been caused by excessive heating, and the crater formed after the waveguide wall had melted. Referring to Fig. 1, it is obvious that the size of the conducting particle removed from the crater is capable of initiating additional discharges. The size of the particle generated is consistent with values reported for particles formed by dc discharges of roughly the same energy as experienced at massive breakdown in the resonators.[6]

Processing considerations in a practical system based upon the above

theory of particle generation and massive breakdown can be illustrated by the following example. Suppose a microwave system is to be designed to transmit 20 megawatts of peak power with 60-microsecond pulses, and that after 10 microseconds of a pulse has elapsed, an arc is initiated by a particle of suitable size. The arc duration is therefore 50 microseconds, and eq. (9b) predicts a temperature rise of:

$$\Delta T = \frac{\gamma\alpha}{A} \times 20 \times 10^6 \sqrt{50 \times 10^{-6}} = 141 \times 10^3 \frac{\alpha\gamma}{A}.$$

For SF_6, $\alpha = 1.5 \times 10^{-2} \Rightarrow \Delta T = 2100\,\gamma/A$. Hence, melting occurs, and massive breakdown will ensue. If, however, the initiation of the arc was detected (such as by comparing forward power levels at the amplifier and at the load) and the RF drive was blanked within, say, 4 microseconds after arc initiation, then:

$$\Delta T = \frac{\gamma\alpha}{A} \times 20 \times 10^6 \sqrt{4 \times 10^{-6}} = 40 \times 10^3 \alpha\gamma/A.$$

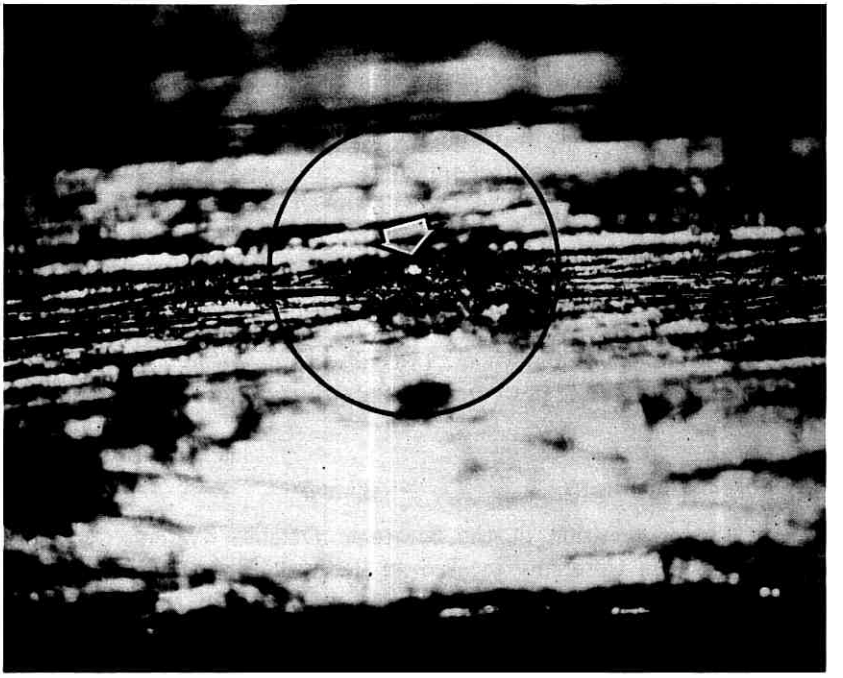


Fig. 9—Arc mark.

For $\alpha = 1.5 \times 10^{-2}$, $\Delta T = 600\,\gamma/A$. Thus, by employing this "early-off" protection, the critical temperature is not exceeded, and processing can be carried to completion in a high average power system.

VII. CONCLUSION

The theoretical and experimental investigation of the effects of high average power levels and long pulse lengths on waveguide breakdown has shown that:

(*i*) Loose particulate matter (e.g. thermally detached from the waveguide wall) is heated to high temperatures by high levels of average power. The resulting rarification of the dielectric gas in the vicinity of the heated particle, coupled with field enhancement at the particle boundary, causes breakdown to occur at values of peak power below that expected in the absence of particles.

(*ii*) Waveguide "processing" with RF discharges of short duration can be used to eliminate undesirable foreign matter and increase the power handling capability of a microwave system.

(*iii*) To achieve processing, arcs must terminate as soon as possible after they occur by blanking the remainder of the RF pulse. This prevents excessive energy dissipation in the arc which can cause waveguide surface melting.

(*iv*) The ultimate peak power level to which a system can be processed varies inversely with the square root of the allowed arc duration.

(*v*) Processing requires a dielectric gas which rapidly recovers its dielectric strength. Sulphur hexafluoride is ideal and perhaps the only practical choice. With its use, processing at reduced pressure can achieve some breakdown margin in the usual case where processing power is limited to the normal operating power.

VIII. ACKNOWLEDGMENTS

With the cooperation of the Strategic Defense Systems Division of the Raytheon Company located at Bedford, Massachusetts, much assistance was given and most of the data was taken by the Equipment Division of the Raytheon Company located at Wayland, Massachusetts. Miss Patricia Loth of the Hazeltine Corporation, Long Island, New York also contributed to the planning and conduct of the experiments.

APPENDIX A

Field Enhancement and Heating of Conducting Particles

Consider a spherical conducting particle located within a waveguide as shown in Fig. 10a. For simplicity, the particle is assumed to be centrally located and completely detached from any of the waveguide walls. When electromagnetic waves are propagated through the waveguide, the boundary conditions imposed by the particle cause the field to be distorted. Since the radius of the particle is much less than a wavelength, the quasi-static approximation can be applied to obtain a good estimate of the fields existing in the vicinity of the particle. The models shown in Figs. 10b and 10c will be used to obtain the quasi-static fields. The coordinate systems used in Figs. 10b and c are unrelated to each other or to that of Fig. 10a. The zero-order longitudinal magnetic field can be neglected, since the particle is assumed to be centrally located (H_z is zero at the center of the waveguide transverse section for the mode of interest, namely the TE_{10}).

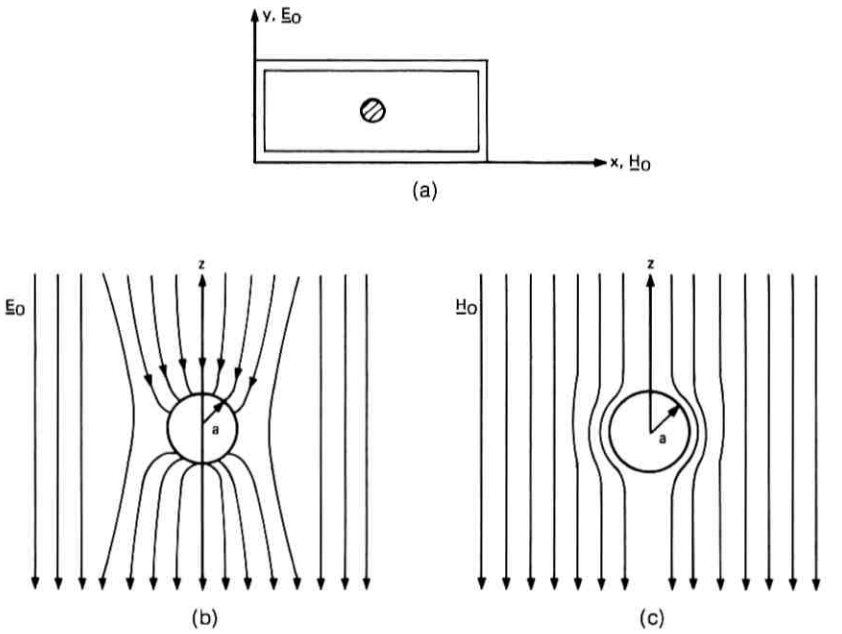


Fig. 10—Particle model: (a) Particle location. (b) E field model. (c) H field model.

The quasi-static approximation consists of expressing the total electromagnetic field in the form:

$$\mathbf{E} = \mathbf{e}_0 + \omega\mathbf{e}_1 + \omega^2\mathbf{e}_2 + \cdots \tag{12}$$

$$\mathbf{H} = \mathbf{h}_0 + \omega\mathbf{h}_1 + \omega^2\mathbf{h}_2 + \cdots \tag{13}$$

where

$$\nabla \times \mathbf{e}_0 = 0 \tag{14}$$

$$\nabla \times \mathbf{h}_0 = 0 \tag{15}$$

$$\omega\nabla \times \mathbf{h}_1 = +\epsilon_0 \frac{\partial \mathbf{e}_0}{\partial t}$$

$$\omega\nabla \times \mathbf{e}_1 = -\mu_0 \frac{\partial \mathbf{h}_0}{\partial t}$$

$$\vdots$$

$$\omega\nabla \times \mathbf{h}_n = +\epsilon_0 \frac{\partial \mathbf{e}_{n-1}}{\partial t} \tag{16}$$

$$\omega\nabla \times \mathbf{e}_n = -\mu_0 \frac{\partial \mathbf{h}_{n-1}}{\partial t}. \tag{17}$$

The approximation consists of truncating (12) and (13) after a sufficient number of terms. The zeroth order fields $\mathbf{e}_0$ and $\mathbf{h}_0$ are found from eqs. (14), (15) and the unperturbed fields E_0 and H_0.

The magnitudes of the higher-order correction fields are related to the magnitudes of the zeroth order fields by the factor $(a/\lambda)^n$, where λ is the wavelength, a is a characteristic dimension of the perturbation, and n is the index of the correction field being considered. Since $a \ll \lambda$, it follows that the first and all higher-order correction fields can be neglected, and the fields in the vicinity of the particle are closely approximated by the zeroth-order terms. These are found by solving LaPlace's equation in spherical coordinates for the electric and magnetic scalar potentials, taking the gradients of these functions, and applying the boundary conditions $\mathbf{e}_0 \mid_{r\to\infty} = -E_0\mathbf{i}_z$; $e_{0\theta} \mid_{r=a} = 0$, $\mathbf{h}_0 \mid_{r\to\infty} = H_0\mathbf{i}_z$; $h_{0r} \mid_{r=a} = 0$. The results are:

$$\mathbf{e}_0 = -E_0\left[1 + 2\left(\frac{a}{r}\right)^3\right]\cos\theta\mathbf{i}_r + E_0\left[1 - \left(\frac{a}{r}\right)^3\right]\sin\theta\mathbf{i}_\theta \tag{18}$$

$$\mathbf{h}_0 = H_0\left[1 - \left(\frac{a}{r}\right)^3\right]\cos\theta\mathbf{i}_r - H_0\left[2 + \left(\frac{a}{r}\right)^3\right]\sin\theta\mathbf{i}_\theta . \tag{19}$$

Now, from eq. (18), it follows that $\mathbf{e}_0$ assumes its maximum value at $\theta = 0$, $r = a$. This value is:

$$| \mathbf{e}_0 |_{\max} = 3E_0 . \tag{20}$$

Hence, the field enhancement factor, β, for spherical conducting particles is 3.

From eq. (19),

$$\mathbf{h}_0 |_{r=a} = -3H_0 \sin \theta \mathbf{i}_\theta \tag{21}$$

$$| I_s | = | \mathbf{h}_0(a) | = 3H_0 \sin \theta \tag{22}$$

where I_s is the surface current density induced in the particle by the field. If the microwave frequency is high enough such that the skin depth of penetration into the particle is much less than the particle's radius, the surface current density may be assumed to flow uniformly within one skin depth, δ. Hence, the current density J becomes:

$$\mathbf{J} = \frac{I_s}{\delta}. \tag{23}$$

The power dissipated per unit volume within one skin depth of the particle is therefore:

$$P_V = \frac{1}{2} \frac{| J |^2}{\sigma} = \frac{1}{2} \frac{| I_s |^2}{\sigma \delta^2} \tag{24}$$

where σ is the conductivity of the particle. The power dissipated per unit surface area of the particle therefore becomes:

$$P_A = \delta P_V = \frac{| I_s |^2}{2\sigma\delta}. \tag{25}$$

But

$$\delta = \sqrt{2/\omega\mu_0\sigma} \tag{26}$$

$$\therefore \quad P_A = \frac{| I_s |^2}{2} \sqrt{\frac{\omega\mu_0}{2\sigma}} \tag{27}$$

or

$$P_A = \frac{9H_0^2 \sin^2 \theta}{2} \sqrt{\frac{\omega\mu_0}{2\sigma}}. \tag{28}$$

The total dissipated power becomes:

$$P_{\text{diss}} = \int P_A \, dS \tag{29}$$

$$= \int_0^\pi \int_0^{2\pi} \frac{9H_0^2}{2} \sqrt{\frac{\omega\mu_0}{2\sigma}} a^2 \sin^3 \theta \, d\theta \, d\phi \tag{30}$$

or

$$P_{\text{diss}} = \frac{36H_0^2a^2}{3}\sqrt{\frac{\omega\mu_0}{2\sigma}}. \tag{31}$$

But the propagating power level, P, is related to H_0 by the expression

$$P = \frac{d_1d_2Z_0}{4}H_0^2 \tag{32}$$

where Z_0 is the TE_{10} wave impedance.

$$\therefore \qquad P_{\text{diss}} = \frac{144\pi a^2\sqrt{\frac{\omega\mu_0}{\sigma}}}{3Z_0d_1d_2}P. \tag{33}$$

Since

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}\frac{\omega/c}{\sqrt{(\omega/c)^2 - (\pi/d_1)^2}}, \tag{34}$$

$$P_{\text{diss}} = \frac{144\pi a^2}{3d_1d_2}\sqrt{\frac{\omega\epsilon_0}{2\sigma}}\left[1 - \left(\frac{c\pi}{d_1\omega}\right)^2\right]P \tag{35}$$

where c is the free-space velocity of light.

To calculate the temperature rise of the particle resulting from this dissipated power, let $\mathbf{n}$ be the heat flux density flowing from the particle:

$$\oint_{\substack{\text{particle}\\ \text{surface}}} \mathbf{n}\cdot d\mathbf{A} = P_{\text{diss}}. \tag{36}$$

Now, assuming heat flow by conduction through the gas,

$$\mathbf{n} = -\xi\nabla T \tag{37}$$

where ξ is the thermal conductivity of the gas. In the steady-state, conservation of energy implies that:

$$\nabla\cdot\mathbf{n} = 0 \tag{38}$$

$$\therefore \qquad -\nabla\cdot[\xi\nabla T] = 0. \tag{39}$$

For SF_6,[7]

$$\xi \cong \xi_o T \tag{40}$$

where $\xi_o = 4.3 \times 10^{-5}$ watts/meter-°K^2. Hence,

$$\xi_o\nabla\cdot[T\nabla T] = 0. \tag{41}$$

From circular symmetry, eq. (41) becomes

$$\frac{\xi_o}{r^2}\frac{d}{dr}\left[r^2 T \frac{dT}{dr}\right] = 0. \tag{42}$$

The solution of this differential equation is:

$$T = \sqrt{c_2 - \frac{c_1}{\xi_o r}} \tag{43}$$

where c_1, c_2 are constants determined by applying the boundary conditions

$$T\,|_{r\to\infty} = T_o\,, \qquad \int_{\substack{\text{particle}\\ \text{surface}}} \mathbf{n}\cdot d\mathbf{A} = P_{\text{diss}}\,.$$

The result is:

$$T = \sqrt{T_o^2 + \frac{P_{\text{diss}}}{2\pi\xi_0 r}}. \tag{44}$$

The absolute temperature of the particle, T_p, is therefore

$$T_p = T\,|_{r=a} = \sqrt{T_o^2 + \frac{P_{\text{diss}}}{2\pi\xi_0 a}} \tag{45}$$

where P_{diss} is given by eq. (35). It is observed that P_{diss} is dependent upon the electrical conductivity, σ, of the particle.

Since the electrical conductivity of a conductor varies inversely with the absolute temperature,

$$\sigma = \sigma_o \frac{T_p}{T_A} \tag{46}$$

where T_A is the ambient temperature and is assumed to be 293°K and σ_o is the conductivity at ambient temperature. For copper, the bulk conductivity is equal to 5.5×10^7 meters/ohm. Because microwave current densities flow within a narrow skin depth, ripples in the surface of the conductor cause a decrease in the effective conductivity near the surface.[8] At S-band frequencies, surfaces with rms variations on the order of several skin depths (such as rough spheres) have an effective conductivity of approximately 62 percent of the bulk conductivity. Hence, a value of 3.44×10^7 meters/ohm will be assumed for σ_o when calculating the temperatures of the copper spheres.

Substituting eqs. (35) and (46) into (45) yields:

$$T_p = \sqrt{T_o^2 + \frac{72aP_{av}\sqrt{\frac{T_p\omega\epsilon_o}{2}\left[1 - \left(\frac{c\pi}{d_1\omega}\right)\right]}}{3d_1d_2\xi_o\sqrt{\sigma_o T_A}}}. \tag{47}$$

Letting

$$B = \frac{72\sqrt{\frac{\omega\epsilon_o}{2}\left(1 - \frac{c\pi}{d_1\omega}\right)^2}}{3d_1d_2\xi_o\sqrt{\sigma_o T_A}} \tag{48}$$

yields:

$$T_p^2 = T_o^2 + BaP_{av}\sqrt{T_p}\,. \tag{49}$$

Experimentally, $d_1 = 2.84$ inches, $d_2 = 1.34$ inches.

$$\therefore \qquad B = 4.68 \times 10^2 \text{ in MKS units}$$

Finally, T_o is equal to the temperature of the waveguide in the region where the particle is located. Experimentally, the particles were introduced at the waveguide flanges, and the flanges were found to experience a temperature rise of 65°C/megawatt of average power above ambient. Hence, to be consistent with experiment,

$$T_o = T_A + 65 \times 10^{-6}P_{av} \tag{50}$$

where T_A is the ambient room temperature (293°K). Hence,

$$T_p^2 = (293 + 65 \times 10^{-6}P_{av})^2 + 4.68 \times 10^2 aP_{av}\sqrt{T_p}. \tag{51}$$

Values of T_p which satisfy (51) appear in Table II for various values of P_{av} and particle radius a.

APPENDIX B

Arc Formative Time and Power Absorption

The formative time of a processing arc is defined as the interval from arc initiation to complete interruption of power transmitted. Photographs of traveling-wave resonator waveshapes taken during the occurrence of processing arcs indicate that processing-arc formative times are independent of peak power level over the range of investigation (11 to 80 megawatts peak).

The formative time was measured within a terminated WR-284 system pressurized with SF_6 at 25 psig. Transmitted power pulses were

TABLE II—PARTICLE TEMPERATURE T_p AS A FUNCTION OF SIZE AND POWER LEVEL

Particle Radius (Microns)	Average Power (Watts)				
	10^5	2×10^5	4×10^5	8×10^5	1.6×10^6
60	388°K	478°K	648°K	960°K	1425°K
30	346°K	395°K	491°K	664°K	964°K
15	323°K	350°K	405°K	508°K	696°K
7.5	310°K	327°K	365°K	426°K	550°K
2	300°K	312°K	332°K	368°K	436°K

monitored on a storage oscilloscope at the termination. The formative time was found from the rate of decay of the transmitted power pulse whenever a discharge occurred at any point between the final power amplifier and the termination. A terminated (non-resonating) structure was employed for this measurement to divorce the arc formative time from the natural decay of a resonator caused by the additional arc loading. It was observed that the rate of decay of power transmitted past an arc was independent of the initiating power level and that the transmitter power waveshape decayed exponentially with a time constant of 2 microseconds. Hence, in WR-284 waveguide pressurized to 25 psig with SF_6, the power absorbed by an arc can be written as:

$$P_A = \alpha P_o[1 - e^{-t/T}] \tag{52}$$

where

t = time
P_o = incident power
α = steady-state percentage of incident power absorbed by arc
T = 2 microseconds.

The value of α was determined by use of a traveling wave resonator. When an arc is struck within a resonant structure, it begins to absorb the energy that was stored in the resonator and the fields decay to zero. The rate of decay of the fields is dependent upon the percentage of incident power absorbed by the arc. This relationship is derived as follows.

Let $P_F(t)$ be the instantaneous power level within the resonator; then, when an arc is initiated,

$$P_F(t + \tau) = [1 - a(t)]P_F(t) \tag{53}$$

where

τ = transit time through the resonator ($= l/v_g$ where l is the resonator length and v_g the group velocity)

$a(t)$ = percentage of incident power absorbed by arc as a function of time.

Since τ is small,

$$P_F(t + \tau) \cong P_F(t) + \tau \frac{dP_f}{dt}. \tag{54}$$

Substituting eqs. (52) and (54) into (53) yields

$$\tau \frac{dP_F}{dt} = -\alpha[1 - e^{-t/T}]P_F \tag{55}$$

$$\Rightarrow P_F = C \exp \{-\alpha/\tau[t + Te^{-t/T}]\}. \tag{56}$$

Now,

$$P_F(t = 0) = P_{F_0} \tag{57}$$

$$\Rightarrow P_F = P_{F_0} \exp \{-\alpha/\tau[t - T(1 - e^{-t/T})]\}. \tag{58}$$

But

$$e^{-t/T} \cong 1 - \frac{t}{T} + \frac{t^2}{2T^2} \tag{59}$$

$$\therefore \quad P_F \cong P_{F_0} \exp \{-\alpha/\tau[t - T(1 - 1 + t/T - t^2/2T)]\} \tag{60}$$

$$\cong P_{F_0} \exp \{-\alpha/\tau[t - t + t^2/2T]\} \tag{61}$$

$$\Rightarrow P_F(t) \cong P_{F_0} \exp \{-\alpha t^2/2T\tau\}. \tag{62}$$

Figure 11 shows the typical decay of resonator forward power during the occurrence of an arc. By fitting eq. (62) to this curve, the value of α can be found since the ring length l, the group velocity v_g, and the formative time constant T are known. The value of α was experimentally found to be

$$\alpha = 0.015. \tag{63}$$

Hence, in the steady-state, an arc struck in WR-284 waveguide filled with 25 psig of SF_6 absorbs 1.5 percent of the power which is incident upon it and reflects the remainder.

It is noted that the above method for determining the percentage of power absorbed by an arc provides greater accuracy than could be obtained by direct measurement of the power levels incident upon and reflected from the arc since the latter involves the subtraction of two nearly equal quantities.

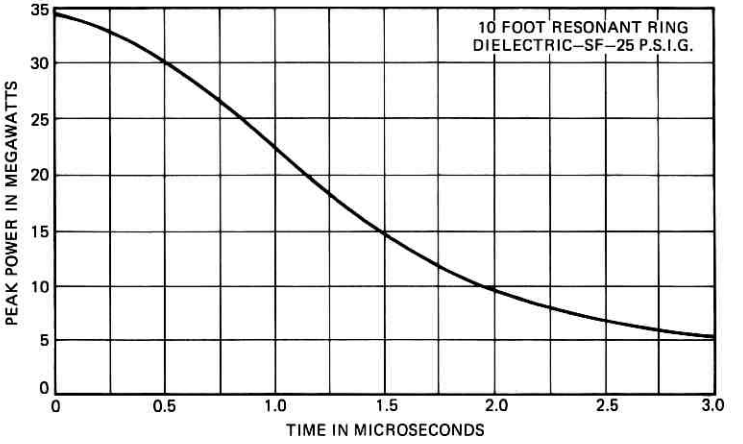


Fig. 11—Resonant ring power decay.

REFERENCES

1. MacDonald, A. D., *Microwave Breakdown in Gases*, New York: John Wiley & Sons, 1966.
2. Gould, L., and Roberts, L. W., "Breakdown of Air at Microwave Frequencies," J. App. Phy., *27*, No. 10 (October, 1956).
3. Brown, S. C., *Basic Data of Plasma Physics*, Cambridge, Mass.: Press, 1966.
4. Miller, S. J., "The Traveling Wave Resonator and High Power Microwave Testing," Microwave J., September 1960.
5. Duff, G. F. D., and Naylor, D., *Differential Equations of Applied Mathematics*, New York: John Wiley & Sons, 1966.
6. Namitokov, K. K., AD 696-315, Foreign Technology Div., Wright-Patterson AFB, Ohio, "Aggregate State, Composition and Structure of the Products 0—ETC (U)," (September 1969), FTO-MT-24-81-69.
7. Anderson, T. Reeves, A. L., Mears, W. H., and Orfeo, S. R., "The Use of Sulfur Hexafluoride in Waveguides," AIEE Conference Paper, January 1957.
8. *The Microwave Engineer's Handbook and Buyer's Guide*, Horizon House, 1965.

Contributors to This Issue

A. S. Acampora, B.S., 1968, M.S., 1970, Polytechnic Institute of Brooklyn; Bell Laboratories, 1968—. Mr. Acampora has been concerned with the design of high-power microwave systems and components for use in modern-phased array radars, and has also conducted data processing and reliability studies as they apply to military radar systems. He is presently investigating the role of collective effects in gas laser operation. Member, Eta Kappa Nu.

Milton Baumwolspiner, B.S.E.E., 1969, Polytechnic Institute of Brooklyn; M.S.E.E., 1970, Columbia University; Bell Laboratories, 1969—. Mr. Baumwolspiner has been engaged in the development of active, and more recently, digital filters. Presently, he is a member of the Logic Circuits Task Force which is involved in the development of all the services needed to provide custom MSI integrated circuits with universal gate arrays.

Robert W. Chang, B.S.E.E., 1955, National Taiwan University; M.S.E.E., 1960, North Carolina State University; Ph.D., 1965, Purdue University; Bendix Corporation, 1960–1963; Bell Laboratories, 1965—. Mr. Chang has worked on a variety of problems in data transmission and communication system theory. Member, Phi Kappa Phi, Eta Kappa Nu, Sigma Xi, IEEE.

William J. DeBonte, S.B. (Physics), 1965, Massachusetts Institute of Technology; M.S. (Physics), 1966, and Ph.D. (Physics), 1970, University of Pennsylvania; Bell Laboratories 1970—. Since coming to Bell Laboratories, Mr. DeBonte has been engaged in studies of models of domain structures applicable to magnetic bubble technology. Member, American Physical Society and Sigma Xi.

L. W. Fagel, B.S. (Mechanical Eng.), 1959, Stevens Institute of Technology; M.S., 1961, New York University; Bell Laboratories, 1959—. Mr. Fagel has done research in applied mechanics and structural

dynamics, and design and development work in conjunction with providing protection for Bell System building and antenna structures from nuclear explosion and earthquake excitations.

JEREMIAH F. HAYES, B.E.E., 1956, Manhattan College; M.S., 1961, New York University; Ph.D., 1966, University of California, Berkeley; Faculty, Purdue University, 1966–1969; Bell Laboratories, 1969—. Mr. Hayes is currently working on the modeling of computer communications networks. Member, IEEE, Sigma Xi, Eta Kappa Nu.

E. Y. Ho, B.S.E.E., 1964, National Taiwan University; Ph.D., 1969, University of Pennsylvania; Bell Laboratories, 1969—. Mr. Ho has been engaged in developing and analyzing automatic equalizers for data transmission systems. Member, IEEE.

S. C. LIU, B.S. (C.E.), 1960, National Taiwan University; M.S., 1964, and Ph.D., 1967, University of California at Berkeley; Bell Laboratories 1967—. Mr. Liu has done research in structural dynamics, random vibrations, and earthquake engineering. Recently he has been concerned with structural optimization problems. Member, American Society of Civil Engineers, Seismological Society of America.

DAVID N. SHERMAN, B.S.E.E., 1963, Polytechnic Institute of Brooklyn; M.S.E.E., 1964, and Ph.D., 1968, University of Southern California; Bell Laboratories, 1969—. Mr. Sherman's interests have been in the areas of data insertion in speech and the analysis of computer communications networks. He is currently involved in the design of FSK data modems. Member, Eta Kappa Nu and Tau Beta Pi.

P. T. SPROUL, B.S., 1937, E.E., 1955, Iowa State University; Bell Laboratories, 1937—. Mr. Sproul has been concerned with telephone and television transmission, radar relay systems and radar designs. He was responsible for design of the microwave portions of a modern-phased array radar including the high-power transmitter when the work was performed. He is currently in charge of digital design and software/digital tests of the same radar equipment. Senior Member IEEE, Member Eta Kappa Nu.

B.S.T.J. BRIEF

A Table Look Up Approach to Loop Switching

By L. H. BRANDENBURG and B. GOPINATH

(Manuscript received July 24, 1972)

In this paper we consider some questions of implementation of a scheme described in Section V of Ref. 1 for addressing message blocks in the Pierce loop system.[2] The scheme consists of using a stream of binary digits (0 and 1) as the address of the destination loop of a message such that the scalar product* of the address with a stream of binary digits of equal length stored at a loop gives the distance between the loop and the destination. The message is routed along a path that minimizes distance between source and destination. The binary streams used in this scheme can be obtained by factoring the distance matrix D of the graph representing the connection of loops into two binary-valued matrices P and Q such that

$$D = PQ^t,$$

where the superscript "t" stands for matrix transposition. The kth row of P represents the address to be prefixed to a message destined for loop k. The kth row of Q represents a binary sequence to be stored in loop k. The scheme discussed in Section V of Ref. 1 provides a particular way of factoring D. For completeness, we give a description of that factorization as follows.

Let s be the diameter of a graph on n vertices and let D be its distance matrix. (Note: $s = \max_{ij} D_{ij}$). We consider a PQ^t factorization of D with P and Q matrices each of order $n \times ns$. The ith row of P, P_i, is zero everywhere except in positions $(s(i-1)+1)$ to si where it is one. The ith row of Q, Q_i, is constructed as follows: for every j, there are exactly D_{ji} 1's in positions $(s(j-1)+1)$ to sj with the rest of the entries of Q_i equal to zero.

* The scalar product of two binary streams of equal length is the number of places they both have a "one".

$$P = \begin{array}{c} \begin{array}{cccccccc} & s & & 2s & & 3s & & ns \\ & \downarrow & & \downarrow & & & & \end{array} \\ \begin{bmatrix} 1\cdots1 & 0\cdots0 & 0\cdots0 & \cdots & 0\cdots0 \\ 0\cdots0 & 1\cdots1 & 0\cdots0 & & \\ & & & & \\ 0\cdots0 & 0\cdots0 & 0\cdots0 & \cdots & 1\cdots1 \end{bmatrix} \end{array}$$

$$Q = \begin{array}{c} \begin{array}{cccccccccc} & s & \overbrace{\quad} & & 2s & \overbrace{\quad} & & 3s & & \overbrace{\quad} \\ & \downarrow & D_{12} & & \downarrow & D_{13} & & \downarrow & & D_{1n} \end{array} \\ \begin{bmatrix} 0\cdots0 & 1\cdots1 & \cdots & 0 & 1\cdots1 & \cdots & 0 & \cdots & 1\cdots1 & \cdots 0 \\ 1\cdots10 & 0 & \cdots & 0 & 1\cdots1 & \cdots & 0 & & 1\cdots1 & \cdots 0 \\ & & & & & & & & & \\ \underbrace{1\cdots10}_{D_{n1}} & \underbrace{1\cdots1}_{D_{n2}} & \cdots & 0 & \underbrace{1\cdots1}_{D_{n3}} & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \end{array}$$

Obviously the length of addresses in this scheme is ns. The P matrix defined above, rows of which are addresses to be prefixed onto messages, is the same for all graphs with diameter s. The Q matrix, rows of which are stored in the loops in the network, contains the information that identifies a particular graph.

As will be discussed further, a simple mechanization of this scheme reveals that it is essentially an obvious table look up scheme. In a table look up scheme, each loop has stored the distance between it and every other loop in the networks, and each address essentially provides a signal to look up or read out a particular distance that is stored.

The above PQ^t factorization can be transformed into a table look up scheme in the following way. The integer i ($\leqq n$) uniquely specifies the ith row of P; thus, a message address need only consist of the binary representation of i, which, of course, requires at most $[\log_2 n]$* bits. Each loop, instead of storing simply a row of matrix Q, stores a binary array consisting of n rows and $[\log_2 s]$ columns. The ith row of this array contains the binary representation of the distance between the loop in question and the ith loop. The binary array can be implemented by a read only memory the ith row of which is accessed by the binary sequence representing integer i. Note that in the original PQ^t form of this scheme, both matrices P and Q were dependent on s, the diameter of the graph. In the table look up mechanization, s appears

* $[x]$ means "the least integer greater than x".

only in the dimension of the stored memory. (The authors had originally considered a table look up mechanization for minimum distance routing but had concentrated their early efforts on the more mathematically fruitful schemes discussed in Refs. 1 and 3. It is interesting to note at this point that the table look up mechanization has evolved in a natural way from a special case of one of those schemes.)

The table look up scheme meets the following important criteria:

(*i*) The simplicity of constructing addresses is important in the case of large graphs with no special structure. Even if one were to find minimum length addresses similar to the ones described in Sections I, II, and III of Ref. 1, we suspect that the construction of these addresses would be very complicated. In the present case, each address is obtained trivially as the binary representation of an integer.

(*ii*) Since present plans for length of message blocks envisage lengths of perhaps a few thousand bits, the length of the message address is an important parameter in any large loop system. The present method guarantees minimum length addresses, $[\log_2 n]$, provided that the graph is not constrained to have a particular structure.

(*iii*) The scheme can be adapted by prefixing some control digits to do alternate routing.

(*iv*) It is simple to determine the necessary changes in the stored memory required by updating (adding loops) or modifying the network, since such changes can be identified by inspection of the distance matrix.

(*v*) The speed of operation of this scheme can be very fast compared to the speed of the message.

(*vi*) If the graph is hierarchical in the sense of Pierce,[2] except that interconnections among loops at the local level are allowed, then routing can be accomplished by using the Pierce scheme at higher levels, and the present scheme at the local level.

At first it might appear that the present scheme requires longer address lengths than the Pierce addressing scheme[2] for the case of strictly hierarchical graphs. For most cases, we show this is not so. A hierarchical graph of k levels can be represented as in Fig. 1.

N_i represents the maximum of the number of branches connecting each vertex at the $(i-1)$st level, to vertices at the ith level. Pierce's scheme obviously requires addresses of length $\sum_{i=1}^{k} [\log N_i]$ bits. The present scheme requires at most

$$m = [\log (1 + N_1 + N_1N_2 + \cdots + N_1N_2 \cdots N_k)] \text{ bits.}$$

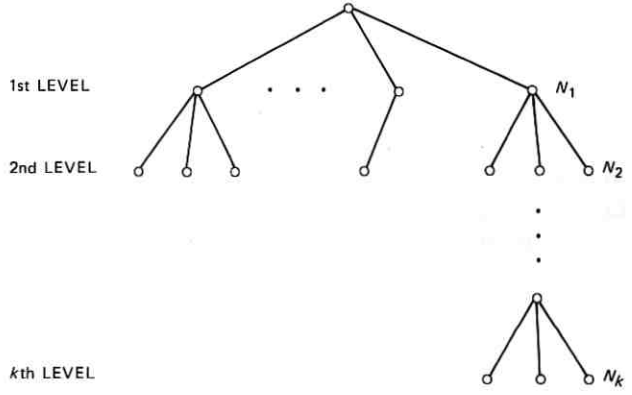


Fig. 1—A hierarchical graph of k levels.

But

$$m = \left[\log N_1 N_2 \cdots N_k \left(1 + \frac{1}{N_k} + \frac{1}{N_k N_{k-1}} + \cdots + \frac{1}{N_k N_{k_1} \cdots N_1}\right)\right].$$

If, at each level, there is at least one vertex with two or more branches descending to the next level (this will always be the case if the levels "fan out") then $N_i \geqq 2$ for each i. Then

$$m \leqq \left[\log N_1 N_2 \cdots N_k \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^k}\right)\right]$$

$$\leqq [\log 2N_1 N_2 \cdots N_k] \leqq [\log N_1 N_2 \cdots N_k] + 1.$$

Therefore, if each N_i is a power of 2, then Pierce's scheme gives addresses at most one bit shorter than those of the present scheme. However, if even one of the N_i's is not a power of 2, then the present method has address lengths at most equal to those of Pierce's scheme. In fact, even if each N_i is a power of 2, the Pierce scheme has the one bit advantage if and only if every vertex at the ith level has exactly N_{i+1} branches connected to the $(i + 1)$st level, for all levels.

In order to compare the present scheme with other schemes described in Ref. 1, we assume that the important parameters are of the following orders:

s = diameter of the graph ≈ 6
n = number of loops $\approx 10^3$
length of message block $\approx 4 \times 10^3$ bits.

Then for the present scheme, the address length $\approx$ 10 bits, or 1/4 percent of the message block length. The memory requirement at each loop is n $\log_2 s$ or $\approx 3 \times 10^3$ bits. For these numbers, the presently available memories, either core or the more advanced semiconductor memories, could be used to realize very high speed routing of messages.

For the other schemes in Refs. 1 and 3, the address length was conjectured not to exceed $2(n - 1)$ bits $\approx 2 \times 10^3$ bits, or 50 percent of the message block length. The memory requirement is of the same order or about 2/3 of that of the present scheme.

With regard to time of computing distance at each junction, the table look up scheme using semiconductor memory might require on the order of 100 nanoseconds; on the other hand, the schemes discussed in Refs. 1 and 3, based on inner product or Hamming distance calculations, might require on the order of 5,000 nanoseconds.

These numbers, together with the six points listed previously, provide evidence of the practical advantage of the table look up scheme over the other schemes in Ref. 1. We have, of course, omitted discussion of such important questions as reliability and the implementation of updating the stored memory in each loop due to changes in the network. However, these questions are of equal importance in the present scheme and the schemes in Ref. 1, since the basic routing rule is the same (minimum distance) and the sizes of the stored memories are comparable as shown above.

The authors are indebted to H. S. McDonald for helpful discussions concerning some of the practicalities of implementing a loop switching system.

REFERENCES

1. Brandenburg, L. H., Gopinath, B., and Kurshan, R. P., "On the Addressing Problem of Loop Switching," B.S.T.J., *51*, No. 7 (September 1972), pp. 1445–1470.
2. Pierce, J. R., "Network for Block Switching of Data," B.S.T.J., *51*, No. 6 (July–August 1972), pp. 1133–1145.
3. Graham, R. L., and Pollak, H. O., "On the Addressing Problem for Loop Switching," B.S.T.J., *50*, No. 8 (October 1971), pp. 2495–2519.