THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING

ASPECTS OF ELECTRICAL COMMUNICATION

THE BELL SYSTEM

TECHNICAL JOURNAL

Demodulation of Wideband, Low-Power FM Signals*

By SIDNEY DARLINGTON

Some theoretical aspects of the demodulation of wideband, low-power FM signals are discussed. It is assumed that a band-limited, continuous, analog signal is supplied to the modulator and is recovered to a fidelity suitable for television, telephone, or carrier telephone. Much of the paper assumes that the baseband signal is sampled and clamped before it is applied to the frequency modulator. The combination has been called PAM-FM and is characterized by a piecewise constant transmitted frequency.

PAM-FM can be demodulated by spectrum analysis means not suitable for continuously varying frequencies. It is shown that a spectrum generator can be derived from the techniques of radar pulse compression, and is equivalent to an infinite set of correlators or matched filters plus means for scanning their terminals.

The spectrum analysis circuit forms are compared with demodulators using frequency detectors, with and without FM feedback, in regard to theoretical noise sensitivities. The theoretical sensitivities are quite similar for spectrum analysis and FMFB under conditions assumed. The comparisons disclose that frequency detectors (followed by filters) enjoy a disguised but efficient use of a differential phase coherence which is a characteristic of FM signals. A combination of spectrum analysis and frequency detection is described which has some of the theoretical advantages of both.

I. INTRODUCTION

This paper discusses some theoretical aspects of the demodulation of wideband, low-power frequency modulated signals. A wide trans-

---

mitted bandwidth permits a saving in power. Frequency modulation implies a constant power level, which makes peak power identical with average power. It is advantageous, for example, when the practical restrictions on peak power determine system power levels rather than restrictions on average power.

More specifically, the paper is concerned with FM systems subject to the following external requirements: A band-limited, continuous, analog signal is supplied to the input of a coder or modulator, which produces the transmitted signal. A demodulator reproduces the original baseband signal to a fidelity suitable for a television channel, a telephone channel, or a carrier system combining a number of telephone channels. For such purposes, for example, the average errors in the output must be more than 40 db below the baseband signal. It is assumed that a large FM index is used, to conserve signal power. These conditions are implicit in many of the conclusions. They will be referred to collectively as "the conditions assumed here."

Several different techniques and circuit forms are compared. The comparisons are concerned primarily, but not exclusively, with sensitivities to noise. Conventional FM receivers and circuits using FM feedback (FMFB) are included. However, more attention is paid to techniques which are closer to (but significantly different from) so-called frequency shift keying (FSK), a well-known method of data transmission.[1] Thus banks of correlators or matched filters appear in some of the proposed circuits, somewhat (but not exactly) as in FSK systems. Alternatively, the correlators or matched filters can be replaced by circuits resembling the pulse compressors of so-called Chirp radars,[2] and one (but not the only) purpose of the paper is to note how it can be done.

Circuits of different kinds are compared not only among themselves but also with theoretical bounds derived from general information theory. Thus the paper draws on four major disciplines within the general field of communication theory and practice, namely: conventional FM and FMFB, discrete data transmission, pulse compression radars, and information theory.

An expert in any one of the four disciplines may find some of the discussion quite familiar, and perhaps superfluous. However, it is unlikely that many readers will be thoroughly familiar with the pertinent parts of all the disciplines. Hence a somewhat tutorial approach has been adopted. However, some of the relations between disciplines and some of the circuit forms appear to be novel.

The purpose of the paper is to describe and compare the various

techniques and circuit forms in simple terms. Mathematical proofs are
outside the intended scope. Except in the Appendix, only the simplest
formulas are stated explicitly, and circuits are represented only by simple
block diagrams. A complete analysis is long, tedious, and mathematically
uninteresting; a good deal of it differs only in detail from established
applications to other problems. Some of the circuit forms have not
actually been built; the block diagrams can be filled in with circuit
details in many different ways, and best ways have not all been deter-
mined. The Appendix outlines very briefly some analytical and circuit
details, which may be needed for an appreciation of some of the conclu-
sions.

II. DEMODULATION BY SPECTRUM ANALYSIS

Much of the paper concerns systems in which the analog baseband
signal is sampled, as part of the initial modulation, but is *not* quantized.
Fig. 1 is a corresponding block diagram. Each sample is clamped during
the sample interval, and is supplied to a frequency modulator. Then
the transmitted frequency is constant over each sample interval, but
changes from interval to interval. Curve B of Fig. 2 illustrates the
variation in frequency with time. It differs from frequency shift keying
in the following way: The transmitted frequency may be anywhere in a
continuum of frequencies; it is not restricted to a finite number of dis-
crete frequencies. The distinction has important repercussions through-
out the paper.

If the sample interval is no greater than the Nyquist interval of the
baseband bandwidth, the sampling destroys no information (at least in
principle). It is assumed here that the sample interval equals the Ny-
quist interval.

Referring again to Fig. 1, the sequence of clamped samples at the
input of the frequency modulator may be called a pulse amplitude
modulation, or PAM representation of the original signal (with no gaps
between the pulses). The corresponding output of the frequency modu-
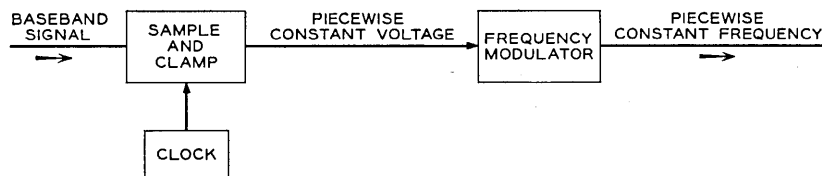lator has been called PAM-FM.[3] It is a known means of adapting time

Fig. 1 — Block diagram of a PAM-FM modulator.

division multiplex to frequency modulation.* (For multiplexing, signal samples are clamped for only fractions of sample intervals and are interleaved with the samples from other channels ahead of the frequency modulator.) We are concerned here with a quite different feature of PAM-FM. The piecewise constant transmitted frequency can be demodulated by means of circuitry which cannot handle the continuously varying frequency of the more usual FM signal.

It is assumed that the demodulator is synchronized to the constant frequency intervals, as received. Some synchronization means are suggested in the Appendix (Section A.9). Then either correlators or matched filters may be used to estimate the piecewise constant frequency, sample-by-sample. The block diagram in Fig. 3 illustrates the concept, without filling in circuit details. A set of correlators or filters, tuned to a sequence of closely spaced frequencies, furnishes a spectrum analysis of the signal plus noise received over each sample interval. The signal is estimated by finding the frequency at which the spectrum is largest.

The operation is complicated by the fact that the true frequency is anywhere in a continuum, and must be estimated to closer than 1 per cent of the bandwidth of the continuum. This implies something like 100 correlators or filters, or else means for interpolation which compare the outputs of adjacent units.

2.1 A Spectrum Generator

The set of correlators or filters furnishes an analog representation of the desired spectrum, in which positions along a sequence of output terminals correspond to discrete values of frequency. The techniques of radar pulse compression can be used to represent the same spectrum, with time as the analog of frequency, at a single output terminal. Externally, the circuit is equivalent to an infinity of correlators or filters, with scanning means to convert the spacially distributed outputs into a function of time.

The spectrum generation hinges on a sequence of two operations. Fig. 4 is a block diagram. The first operation beats the received signal with a varying-frequency local oscillator, to obtain the difference frequency. Fig. 5 illustrates the frequencies of the true signal, of the local oscillator, and of the signal at the output of the mixer. The true frequency is constant over each sample interval, as before. The oscillator frequency varies periodically, in synchronism with the signal samples. In particular, it varies linearly over each sample interval. Thus, at the

---

* For example, in telemetry systems.
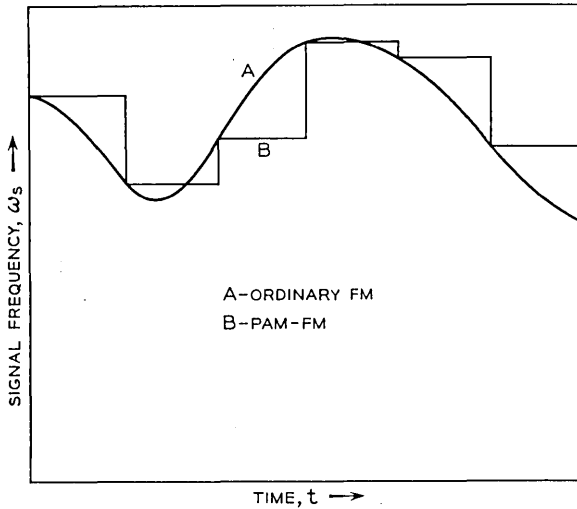
Fig. 2 — Instantaneous signal frequencies.

output of the mixer, the *variations* in frequency are the same over every sample interval, but the *average* varies from sample to sample.

The second operation transmits the modified signal through a pulse compressing (dispersive) line. The nominal delay (phase slope) varies linearly with frequency. Over any one sample interval, the instantaneous frequency varies linearly with time. Thus the nominal delay varies linearly with time. Fig. 6 illustrates the variations in delay with frequency and time.

The variations in delay are so scaled that the tail end of the signal sample just catches up with the head end. Then, on the basis of nominal delays, the entire signal sample emerges from the line in a single instant of time. Actually, of course, the nominal delay does not apply exactly to the time-varying instantaneous frequency. Thus the signal sample does not actually emerge from the line all at a single instant. However, under the conditions assumed it is squeezed into a small portion of the sample interval.
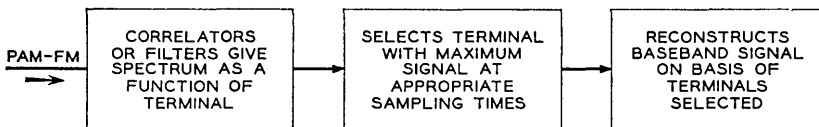
Fig. 3 — PAM-FM demodulation by correlators or filters.
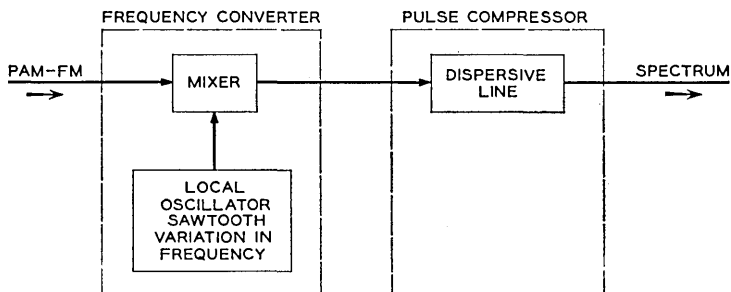
Fig. 4 — Spectrum generation.

The compression of a signal sample into a short pulse depends only on the variations in the instantaneous frequency, which are the same for each sample interval. On the other hand, the time of arrival at the output end of the line depends on the average frequency, which is the frequency of the true signal and varies from sample to sample. The baseband bandwidth and the FM index restrict the signal frequencies to a utilized RF bandwidth. With a suitable choice of circuit parameters, the corresponding variations in arrival time cover a little less than one sample interval. Then the true signal produces one pulse per sample interval, whose position in a (somewhat delayed) sample interval is a measure of the signal frequency. Fig. 7 illustrates the situation. In other terms, beating with a swept frequency and then pulse compressing converts PAM-FM into pulse position modulation, or PPM.*

It is now time to note specific formulas. For simplicity, let time t be zero at the center of a typical signal sample interval. Let the true signal, for that interval only, be

$$s(t) = \sqrt{2P_s} \cos (\omega_s t + \beta_s), \qquad -T/2 < t < T/2. \qquad (1)$$

Here T is the length of the sample interval, ω_s and β_s are the frequency and phase of the true signal, and P_s the signal power. Let the corresponding output of the mixer be

$$\hat{s}(t) = \sqrt{2P_s} \cos (\omega_s t + \beta_s - \tfrac{1}{2} q t^2), \qquad -T/2 < t < T/2 \qquad (2)$$

where q is an arbitrary constant. The instantaneous frequency is now $\omega_s - qt$, linear with respect to time. [Actual circuitry may introduce constant changes in amplitude, carrier frequency and phase angle, between (1) and (2), but these are trivial for present purposes.]

---

* In practice, the compressed pulse will have small side lobes, omitted in Fig. 7 for simplicity. See Fig. 8 below and also Section A.9 of Appendix.
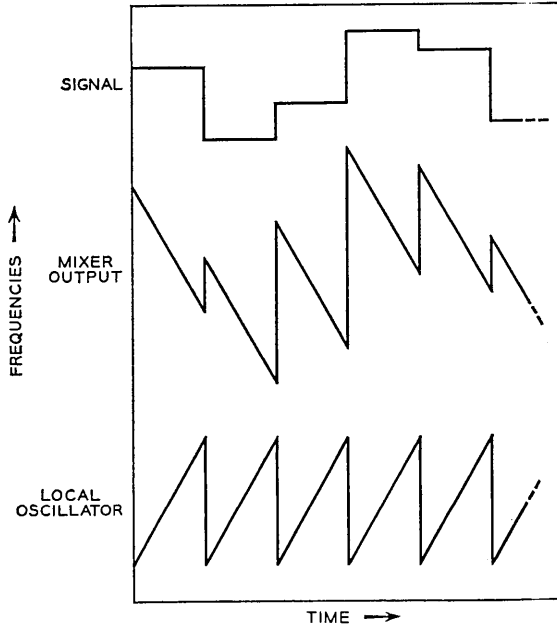
Fig. 5 — Frequency conversion.

The corresponding output of the pulse compression line is approximately

$$S(t) = \sqrt{2P_s}\, F\,(\omega_s - \omega_k)\, \cos\,(\omega_c t + \tfrac{1}{2}\, q t^2 + \beta_s - \beta_c)$$

$$\omega_k = \omega_c + qt \qquad\qquad\qquad\qquad (3)$$

$$F(\lambda) = \frac{\sin \lambda\, \dfrac{T}{2}}{\lambda}\,.$$

The expression assumes that ω_c is large compared with $|\,\omega_s - \omega_c\,|$ and $|\,\omega_k - \omega_c\,|$. For present purposes, ω_s and ω_k lie in the utilized RF band, and ω_c is the midband, or carrier, frequency. A derivation of (3) from (2) is outlined in the Appendix (Section A.1).

The processed signal $S(t)$ may be described as a high-frequency sinusoid multiplied by an envelope function. The frequency, $\omega_c + qt$, varies with time, but it is independent of the received signal. On the other hand, the phase angle is $\beta_s - \beta_c$, in which β_c is a property of the transmission line, but β_s is the phase angle of the unprocessed signal $s(t)$. The envelope is $\sqrt{2P_s}\, F(\omega_s - \omega_k)$. It is a function of time, but
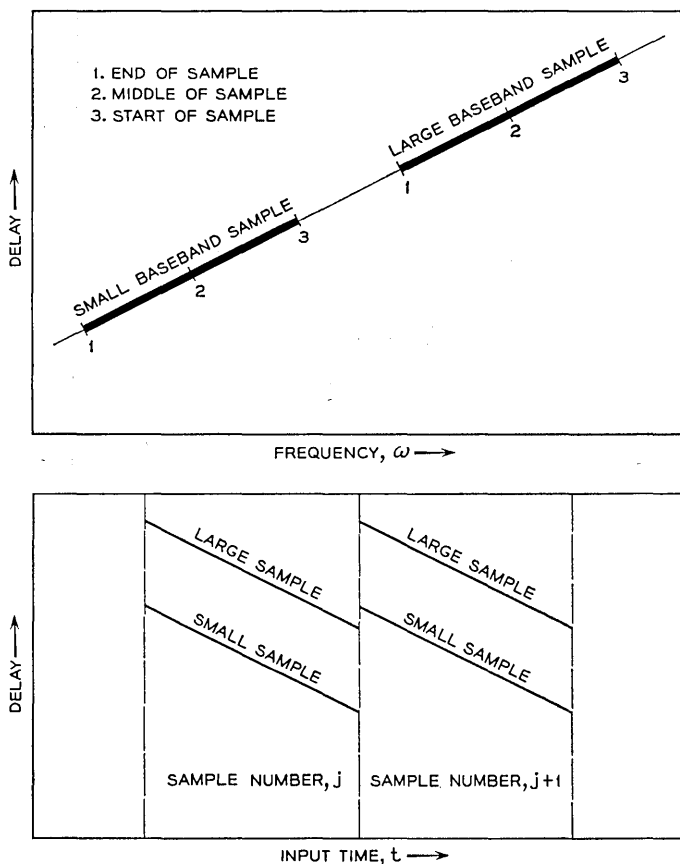
Fig. 6 — Delay vs frequency and time.

the time is an analog representation of the frequency variable ω_k. The signal frequency ω_s enters the envelope function as a parameter.

Fig. 8 is a qualitative plot of $F(\omega_s - \omega_k)$. The abscissae correspond simultaneously to time and ω_k. The largest F occurs at $\omega_k = \omega_s$. Thus the frequency ω_s may be determined by noting the time of the maximum F, and interpreting the time in terms of ω_k. The envelope F may be separated from the sinusoid by means of an envelope detector at the output of the line. Fig. 9(a) is a block diagram.

For some purposes, it is convenient to divide $S(t)$ into two components, as follows:

$$S(t) = \sqrt{2P_s}\, F(\omega_s - \omega_k)\, \cos \beta_s \cos \left(\omega_c t + \tfrac{1}{2}\, qt^2 - \beta_c\right)$$

$$- \sqrt{2P_s}\, F(\omega_s - \omega_k)\, \sin \beta_s \sin \left(\omega_c t + \tfrac{1}{2}\, qt^2 - \beta_c\right). \tag{4}$$
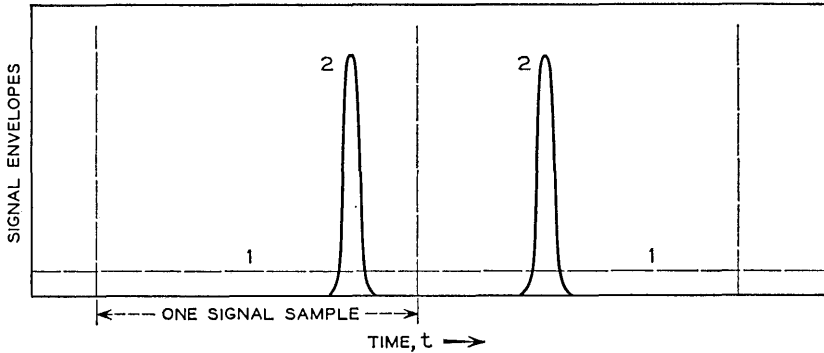
Fig. 7 — Signals at terminals of dispersive line: 1, input signal, before pulse compression; 2, output signal, after pulse compression.

The two sinusoids are independent of the signal $s(t)$. Physically, the two envelope functions can be resolved by means of phase detectors. Fig. 9(b) is a block diagram.

Consider the Fourier transform of a time function equal to $s(t)$ in the one sample interval, and zero elsewhere. More specifically, consider the transform at positive frequencies ω_k near ω_c. If the same approximations are made, as in the derivation of (3), the real and imaginary parts of the transform are the same as the two envelope functions in (4). The envelope function in (3) corresponds to the transform of the envelope of the original time function.

The same remarks apply a little more generally. Suppose the ampli-
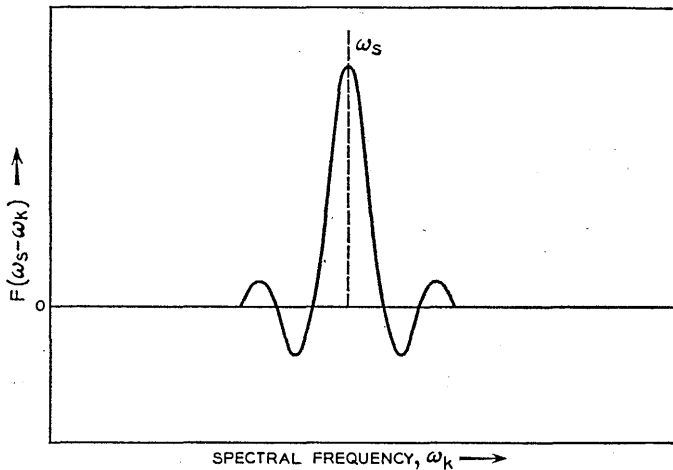
Fig. 8 — The function $F(\omega_s - \omega_k)$.
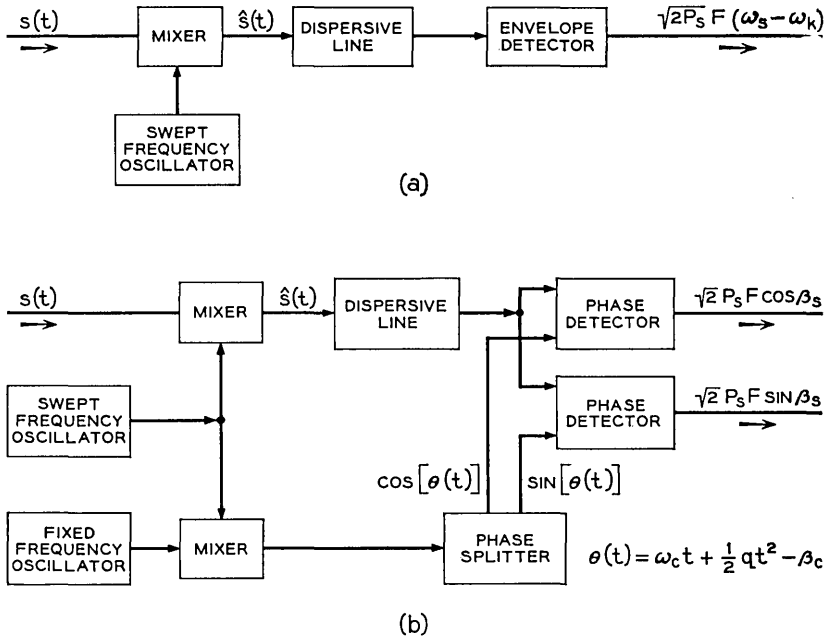
Fig. 9 — Detection of envelope and components: (a) envelope, (b) components.

tude of the received signal $s(t)$ is modified, as well as the frequency, before it reaches the pulse compressor. Then $\hat{s}(t)$ becomes

$$\hat{s}(t) = \sqrt{2P_s}\, A(t) \cos\left(\omega_2 t + \beta_2 - \tfrac{1}{2}\,qt^2\right), \qquad -T/2 < t < T/2. \quad (5)$$

Suppose the envelope $A(t)$ is symmetrical about the center of the sample interval. Then (3) and (4) apply except that $F(\lambda)$ is now the transform of a time function equal to the new envelope during the sample interval, and again zero elsewhere. For the analogous radar application, see Ref. 2.

The operations which convert (1) into (3) and (4) are all linear operations on the signal. If $s(t)$ is generalized to a sum of many constant-frequency sinusoids, the spectrum corresponding to a single sample interval can be generated by summing the results of the operations on the individual sinusoids. Referring again to the block diagrams, in Fig. 9(a) the output is the amplitude of the transform, and in 9(b) the two outputs are the real and imaginary parts. We will use the collection of sinusoids as a representation of the signal plus noise, received during one sample interval.

Thus pulse compression techniques generate analog representations

of the transforms of signal samples. The transforms are generated as functions of time. The constant q determines the time-vs-frequency scale, and can be chosen so that the utilized RF band is scanned in less than one sample interval. The width of the peak in Fig. 8 is merely the familiar "spectrum line width" of a sinusoid of finite duration. The detailed shape (in particular the tails) can be modified to some extent through initial multiplication by an envelope function (or, alternatively, by a shaping circuit at the output of the dispersive line).

The same remarks apply to infinite sets of correlators or matched filters, except that the spectra are generated at specific instants as functions of position along arrays of output terminals. One result is: all three embodiments are equally sensitive to noise accompanying the received signal. A choice between the three must depend on practical compromises, limitations, etc., associated with the design of actual circuits. (For the external equivalence between correlators and matched filters, see, for example, Ref. 4.)

III. SENSITIVITIES TO NOISE

Demodulation by correlators, matched filters, or spectrum generators, as described in the previous section, will be referred to collectively as demodulation by spectrum analysis. This section compares the effects of noise in such circuits and in conventional FM receivers and FMFB. Between conventional FM and FMFB, some effects of noise are quite similar and some quite different. The two circuit forms will be referred to collectively as demodulation by frequency detection.

It is assumed that the noise is Gaussian and that it is added to the signal before it reaches the demodulator. It may be, for example, thermal noise associated with first stages of amplification in the receiver. In demodulation by spectrum analysis, the noise adds random processes to the spectra analyzed. These may be described as two independent Gaussian processes added to the envelope functions in (4). The independent variable in the random processes is the spectral frequency ω_k, which is also represented by time in the pulse compression embodiment. The processes are described in a little more detail in Section A.2.

It is convenient to normalize the error formulas in terms of parameters r, R, and T, defined as follows:

ω_b = baseband bandwidth (0 frequency to cutoff)

ω_r = full excursion of instantaneous signal frequency (maximum − minimum)

r = ω_r/ω_b = bandwidth expansion ratio $\hspace{2cm}$ (6)

P_s = signal power   (6) (cont.)

P_n = noise power in a frequency interval equal to one baseband

R^2 = P_s/P_n = "signal power to noise density ratio" at the input of
    the demodulator

T = π/ω_b = baseband Nyquist interval.

Under the conditions assumed here, the bandwidth expansion ratio, r, is fairly large — order of 10 or 20. Power thresholds (defined in the next section) set lower bounds on R, in the neighborhood of 14 to 16 db. In practical applications, practical compromises may require a somewhat larger R, and the bandwidth of the receiver must be a little greater than ω_r (whether demodulation is by spectrum analysis or frequency detection). The power spectrum of the noise is assumed to be uniform at the input of the demodulator, over the pertinent frequency interval.

3.1 *Two Different Effects of Noise*

For present purposes, one must examine two different effects of the noise, on the recovered baseband signal at the output of the demodulator. Under the conditions assumed here, the effects of the noise on the demodulated baseband signal are quite small most of the time. These may be called small noise errors, and their rms is one measure of circuit performance. On the other hand, during occasional brief intervals, peaks in the noise have a dominant effect and temporarily replace the true signal by a random false signal. This is commonly called blocking. It usually persists over intervals comparable with a baseband sample interval. The average number of blockings per second is the blocking frequency.

Fig. 10 illustrates the two effects in terms of probability densities. It is a qualitative (not quantitative) plot of the probability density of the error, due to noise, in the demodulated baseband signal at any one instant. The peak near zero is substantially Gaussian and corresponds to the small noise errors. The long tails are flat and correspond to the probability that blocking will replace the true signal by a random signal. The transitions between the Gaussian peak and the flat tails are not considered further here. They are very difficult to calculate and must be strongly dependent on design details.

The blocking frequency decreases very rapidly as the power ratio R increases. A related parameter is the power threshold. Thresholds of FM circuits (and also phase lock) have been defined in numerous ways for numerous purposes. The definition which best suits our present needs is the following: The power threshold is the signal power just sufficient
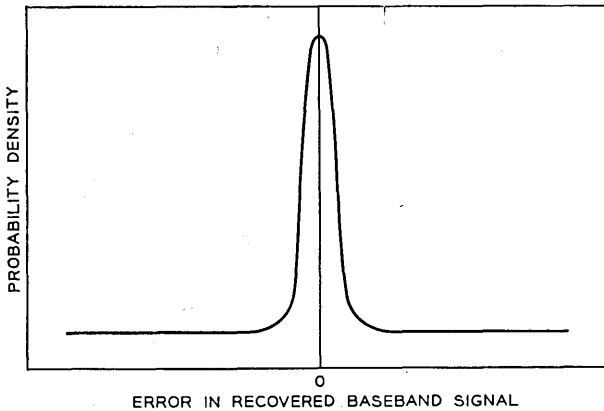
Fig. 10 — Qualitative form of error distribution.

to meet a specified limit on the portion of the samples which are blocked. It can be expressed in terms of the corresponding ratio, R, in db. Under the conditions assumed here, the specified limit on the blocking rate may be perhaps one in a thousand or one in ten thousand.

3.2 Small Noise Errors

Consider first the demodulation of individual signal samples by spectrum analysis. Both phase coherent and phase incoherent circuit forms are possible. More than one kind of phase coherence is of interest here. However, it will be simplest to start with the classical kind in which the phase of each constant frequency sample is independent of other samples and is determined uniquely by a rule known to the demodulator. This kind of phase coherence requires a degree of synchronization which may be impossible in practice. However, its theoretical properties bear on what follows.

Under the conditions assumed here, the corresponding small noise errors are approximately as follows:

For phase coherent demodulation:

$$\frac{\text{rms [small noise errors]}}{\text{max [true signal]}} = \frac{2\sqrt{3}}{\pi} \frac{1}{rR}. \tag{7a}$$

For phase incoherent demodulation:

$$\frac{\text{rms [small noise errors]}}{\text{max [true signal]}} = \frac{4\sqrt{3}}{\pi} \frac{1}{rR}. \tag{7b}$$

(The maximum true signal is here one half of a full signal excursion between equal + and − maxima.) Derivations are outlined in Section A.3.

According to (7), the small noise errors of phase coherent spectrum analysis are about 6 db smaller than those of phase incoherent spectrum analysis, assuming that the phases of signal samples are determined individually and uniquely by a suitable rule known to the demodulator. How do these compare with the small noise errors of demodulation by frequency detection?

Between conventional FM and FMFB, the small noise errors are approximately the same. More exactly, they are approximately the same functions of power level and bandwidths, which may themselves be quite different in practical applications of the two circuit forms. An approximate formula is

For demodulation by frequency detection:

$$\frac{\text{rms [small noise errors]}}{\text{max [true signal]}} = \frac{2}{\sqrt{3}} \frac{1}{rR}. \tag{8}$$

A well-known derivation is reviewed in Section A.4.

Superficially, conventional FM and FMFB appear to be phase incoherent. However, the (theoretical) small noise errors are almost the same as in the phase *coherent*, sample-by-sample spectrum analysis. They differ only by a voltage ratio $\pi/3$, or 0.40 db. This makes demodulation by frequency detection 5.62 db better, in regard to small noise errors, than the phase incoherent spectrum analysis.* It suggests that a more subtle form of phase coherence is at work, which perhaps can be realized also by a more subtle use of spectrum analysis.

Further evidence is as follows: Consider the usual description of noise reduction by conventional FM demodulation. (See again Section A.4.) The frequency detector, as such, produces a demodulated baseband signal plus a substantial amount of noise. However, when the FM index is large, most of the noise power is at frequencies above the baseband. Fig. 11 illustrates the usual form of the power spectrum. Then a filter which passes only the baseband eliminates most of the noise.

To approach the noise levels of phase coherent spectrum analysis, one must use an almost ideal baseband filter. But then the filter combines past outputs of the frequency detector over a "memory time" substantially longer than the baseband Nyquist interval. (Ideally it

---

* The 6-db difference has been noted before, with different interpretation, by, for example, Kotel'nikov.[5]
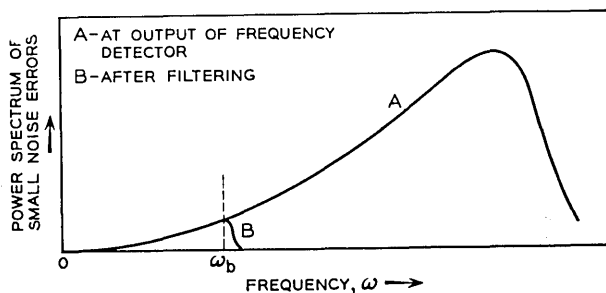
Fig. 11 — Small noise errors in frequency detection.

should be infinite.) Fig. 12(a) is a qualitative illustration of the appropriate weight function, or impulse response.

What happens if the filter is constrained to have a memory no longer than one baseband Nyquist interval? Suppose the true signal frequency is constant over that interval. Then the best weight function, within the constraint, is the parabola illustrated in Fig. 12(b).* The corresponding small noise errors turn out to be *exactly* as in phase *incoherent* spectrum analysis.

It is not at once clear how the longer memory of the ideal, unconstrained filter can reduce the (small noise) errors by anything like 5 or 6 db. The original baseband signals are substantially uncorrelated over intervals longer than one Nyquist interval. The effective correlation time of the noise process is even shorter. However, it is the *frequency* of the FM signal which has the correlation characteristics of the baseband. The *phase* is further characterized by the *continuity of phase rotations* required for a constant amplitude sinusoid of varying frequency. This may be regarded as a subtle kind of phase coherence which, in fact, is used effectively by the filter in demodulation by frequency detection.

The interpretation is clarified and supported by the following argument: Consider demodulation by spectrum analysis, and suppose the transmitted signal is generated by applying a piecewise constant control voltage to a frequency modulator. (See again Fig. 1.) Because the output of the modulator is a continuous sinusoid, the instantaneous phase rotation is continuous, even though its rate of change (which is the frequency) is discontinuous. The continuity of phase rotations, from sample to sample, has been called differential phase coherence.

* "Parabolic smoothing" is best for a finite interval, and a constant signal plus noise power proportional to ω^2. See, for example, Ref. 6.
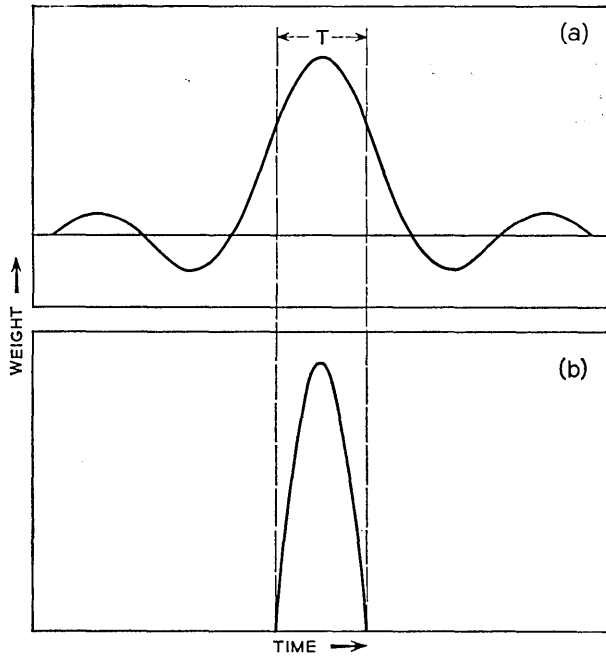
Fig. 12 — Filter weight functions: (a) ideal band-limiting filter; (b) optimum when constrained to one sample interval.

Fig. 13(a) illustrates the differentially coherent phase rotations. The slope of each straight line segment is the frequency during one signal sample interval. In contrast, if the transmitted signal is differentially phase incoherent, the phase rotations are discontinuous between samples, as in Fig. 13(b). This corresponds, for example, to forming a piecewise constant frequency signal by successive selections (or keying) from a set of phase incoherent oscillators.

Referring to Fig. 13(a), consider sample number k. The frequency can be estimated by an incoherent spectrum analysis of signal sample k by itself. [See again (7b) for the rms small noise errors.] Further information can be gleaned from spectrum analyses of samples $k - 1$ and $k + 1$. Specifically, estimates can be obtained from these samples of the phase rotations at the beginning and end of sample interval k. Only the *difference* between the two phase angles is actually needed, and hence the absolute phase reference required for the phase coherence of (7a) is no longer necessary.

The difference between the two estimated angles is the net phase rotation, modulo 2π, over sample interval k. Dividing by the duration
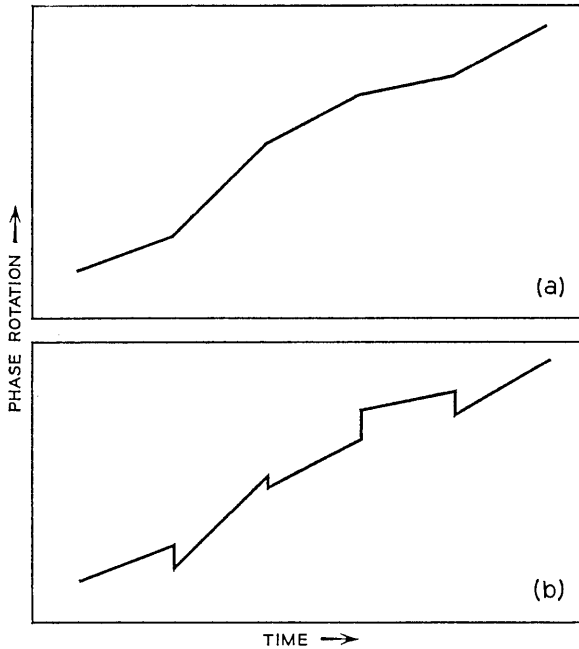
FIG. 13 — Phase rotations: (a) differential phase coherence; (b) differential phase incoherence.

T of the sample interval gives a second estimate of the frequency, but only to modulo $2\pi/T$. When the noise is small, as assumed, the first estimate is accurate enough to resolve the ambiguity. Then a weighted sum of the two estimates gives an improved estimate of the true signal frequency. (The small noise errors in the two estimates are substantially uncorrelated.) Small further improvements can be derived from frequency and phase estimates for additional sample intervals.

An optimum combination of phase and frequency measurements of all samples, $-\infty$ to $+\infty$, gives a 4.365-db theoretical improvement over sample-by-sample phase incoherent spectrum analysis. (The power ratio is $1 + \sqrt{3}$.) Of this, 3.979 db can be realized by using only samples $k-1$, k, $k+1$ to estimate the frequency of sample k. A derivation is described very briefly in Section A.5.

Why does one not realize the full 5.62 db apparent in conventional FM demodulation? It can be interpreted as a curious effect of the sampling of the original baseband signal, which is not part of the conventional FM system. The interpretation is supported by what follows.

Suppose the piecewise constant frequency is applied to the frequency

detector in an (idealized) conventional FM receiver, and that the noise level is low enough to justify the usual small noise approximations. The output is a piecewise constant true signal, like curve A of Fig. 14, plus noise. The noise can be reduced by sampling the output of a suitable filter, as suggested by Fig. 15. Can the ideal baseband filter be used, as for an unsampled signal?

Elementary information theory includes the following: If the samples were represented by a sequence of very short impulses, like curve B of Fig. 14, the ideal filter would be as effective as for the unsampled signal. However, because they are represented, in fact, by a piecewise constant signal, like curve A, the ideal filter has two shortcomings. It produces intersample interference. It responds to the wanted sample less efficiently than to an ideal impulse.

Suppose the filter is constrained to give no intersample interference, assuming each sample to be a constant signal over its entire sample interval. The best filter within the constraint gives 4.365 db improvement over incoherent sample-by-sample spectrum analysis, which is
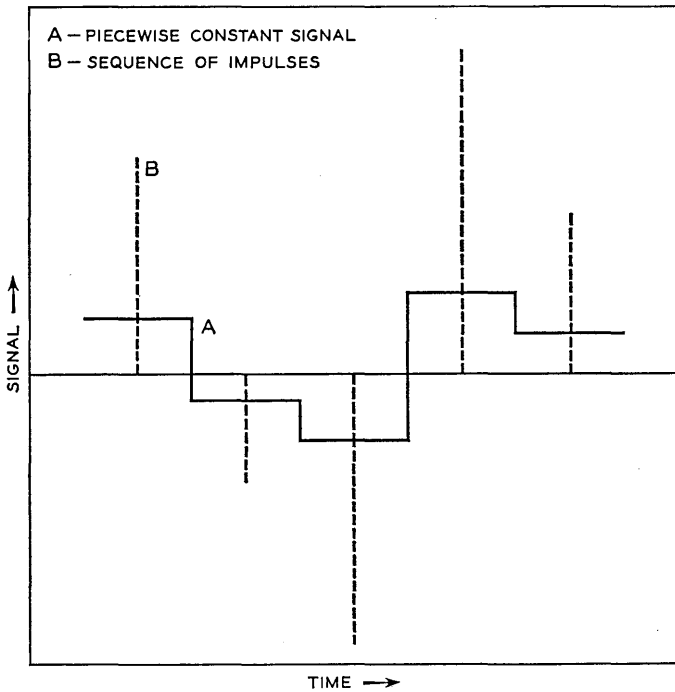
Fig. 14 — Filter inputs.

Fig. 15 — Filter and sampler after frequency detector.

exactly the same as the figure for multisample spectrum analysis using differential phase coherence. A derivation is outlined in Section A.6.

3.3 Thresholds

Consider first the thresholds of sample-by-sample spectrum analysis. Fig. 16(a) illustrates the spectrum of the usual signal-plus-noise sample. Fig. 16(b) illustrates the spectrum of the occasional sample which blocks. It assumes that the frequency of the spectral maximum is used as the estimate of the true frequency, as before. The blocking occurs when the spectrum of the noise sample has a peak, at a random frequency, which
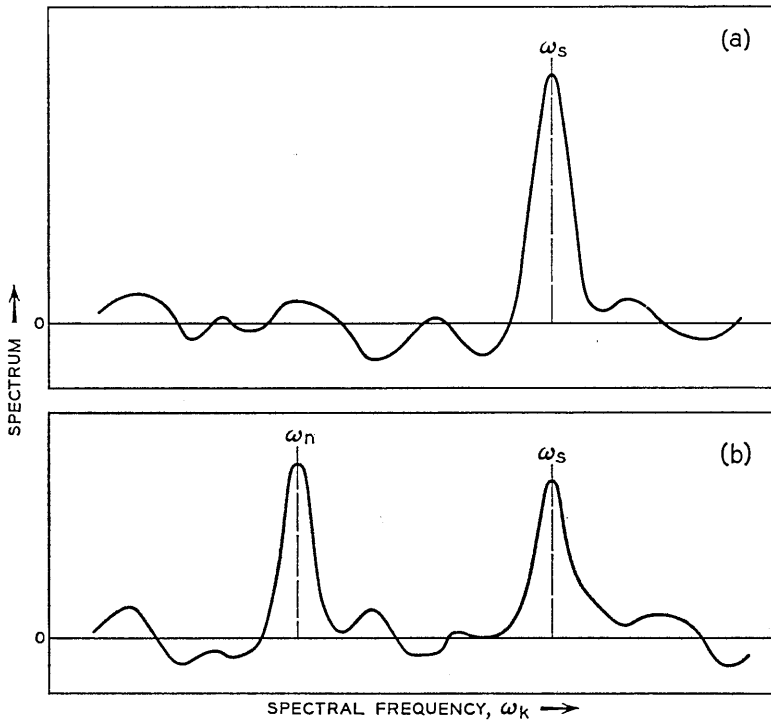
Fig. 16 — Spectrum of a single signal-plus-noise sample: (a) the usual sample; (b) the occasional sample which blocks.

exceeds the spectrum of signal-plus-noise at the true signal frequency. The remarks apply to both phase coherent and phase incoherent sample-by-sample spectrum analysis, provided the pertinent spectra are used for each.

The corresponding blocking probabilities are approximately as follows:

For phase coherent spectrum analysis:

$$P = \frac{r - 2}{2\sqrt{\pi}\,R} \exp\left(-R^2/4\right).$$ (9a)

For phase incoherent spectrum analysis:

$$P = \frac{r - 2}{4} \exp\left(-R^2/4\right).$$ (9b)

Here P is the probability that a typical sample is blocked, and blocking of different samples is uncorrelated.

Part of the derivation is the same as for the (gross) error rates of quantized frequency shift keying (FSK). However, there is an extra complication. In FSK, one is interested only in the spectrum at a finite set of discrete frequencies. The random process which is the noise spectrum is at most weakly correlated between the pertinent frequencies. Thus error rates have been approximated, for example, by assuming either zero correlation[7,8] or a manageably simple form of correlation.[9]

For our purposes, we must consider the spectra at all frequencies in a continuum, with the certainty that correlations are high across small frequency differences. An exact calculation would be extremely difficult. As an approximation, one can proceed as follows: Divide the pertinent frequency interval into, say η equal subintervals. Approximate the true spectrum in each subinterval by a constant. Assume that the constants for the η subintervals are independent random variables (over the ensemble of noise samples). Now one can estimate blocking probabilities as error rates in an η-frequency FSK system. Differences between (9) and equations in Refs. 7 and 8 reflect further approximations, appropriate under the conditions assumed here. They are described briefly in Section A.7, together with some further analytical details.

The approximation to the spectrum may be described further as follows: The covariance of the spectrum of the noise sample is approximated by perfect correlation over each subinterval and zero correlation between subintervals. The actual correlation across the (radian) frequency difference $\omega_2 - \omega_1$ is

$$\frac{\sin (\omega_2 - \omega_1) \dfrac{T}{2}}{(\omega_2 - \omega_1) \dfrac{T}{2}} \tag{10}$$

(see Section A.2). Equations (9) correspond to subintervals of width $2\omega_b$, which is the $| \omega_2 - \omega_1 |$ at the first zeros of the true covariance function.

We have defined the threshold as the signal power required to meet a specified limit on the blocking frequency. The corresponding power ratio R, used in (9), must give the single-sample blocking probability P which corresponds to the specified blocking frequency.

Under the conditions assumed here, P is very small, say 0.001 or 0.0001. Then the exponentials in (9) are very small, and small percentage changes in R produce much larger percentage changes in P. As a result, changes in the coefficients, multiplying the exponentials, can be compensated by much smaller changes in R. For example, a two-to-one change in a coefficient is offset by something like a $\frac{1}{2}$-db change in R. Two consequences are as follows: The threshold changes only slowly with the bandwidth expansion ratio r. The threshold is rather insensitive to the size of the frequency subintervals used in the approximation described above.

Numerical examples of thresholds will be tabulated in Section IV, together with small noise errors.

Slepian[10] has derived from general information theory some important upper and lower bounds on the thresholds (as here defined) of *quantized* systems, constrained to code baseband samples individually, for transmission over channels wider than the baseband. It is interesting to compare the thresholds (9) with Slepian's bounds, even though (9) refers to *unquantized* systems. Since the bounds depend on the number of quanta, one must first decide on the appropriate quantization.

Transmission and demodulation of a quantized signal, as such, involve no counterpart of the small noise errors in unquantized systems. However, when the original baseband signal is unquantized, transmission in quantized form implies quantization or round-off errors relative to the original signal. Then, in judging system quality, one can compare the quantization errors in a quantized system with the small noise errors in an unquantized system. Thus it is interesting to compare thresholds determined by (9) with Slepian's bounds for quantized systems such that the rms quantization errors match our rms small noise errors.

Our present purposes are served by a very rough comparison, using

graphical data in Slepian's paper. Under the conditions assumed here, the thresholds (9) are only very little above Slepian's lower bound. The differences are very roughly $\frac{1}{4}$ db for sample-by-sample phase coherent demodulation and one db for the phase incoherent form.

In principle, the thresholds can be reduced even a little further by combining phase and frequency estimates derived from more than one sample interval. We have seen that a second estimate of the frequency of sample k can be derived from the phases of samples $k - 1$ and $k + 1$. The same is true of the phase of sample k. This permits the phase coherent threshold to be approximated with only differential phase coherence. More complicated operations yield a further improvement. Referring again to Fig. 16(b), blocking occurs when a noise peak exceeds the signal peak, in the spectrum of the signal plus noise, and is chosen in its place. The additional phase information can be used to improve the choice between the two peaks. However, the 2π phase redundancy severely limits the improvement. For the conditions assumed here, a rough estimate is a ten-to-one reduction in the blocking frequency, or something like a one-db reduction in the threshold at the old rate (relative to phase incoherent spectrum analysis). A few further details are noted in Section A.8.

The improved threshold may be slightly below Slepian's lower bound. This is not improper, since it is obtained by violating Slepian's assumption of sample-by-sample coding and decoding.

Now consider the thresholds of conventional FM demodulators and FMFB. Fig. 17 compares simplified block diagrams of the two circuit

(a)

(b)
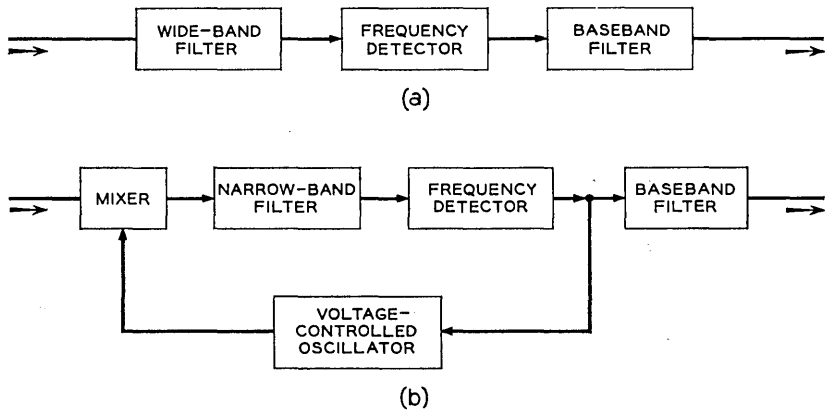
Fig. 17 — Demodulators using frequency detection: (a) conventional FM demodulator; (b) demodulator using FM feedback.

forms. The blocking phenomenon is a well-known characteristic of
these circuits. Under the conditions assumed here, the thresholds are
significantly lower (permit lower signal power) in FMFB circuits than
in conventional FM receivers. The advantage derives from the relative
bandwidths of the filters just ahead of the frequency detectors, and
thereby depends on a fairly large bandwidth expansion ratio, r (which
is here 10 or 20). This is, of course, the reason why FMFB is of current
interest, for example for satellite communication systems.[11]

Because of the nonlinear feedback loop, it is extremely difficult to
calculate for FMFB the quantitative thresholds required for specific
blocking rates. However, important parameters have been identified
and studied, for example by Enloe.[12] Good circuits have been built and
demonstrated for voice and television channels, with thresholds which
are not far above the theoretical lower bounds. Since the quantitative
blocking rates have not been determined, the margins above the bounds
are not known exactly.

3.4 *Comparisons with Other Methods*

At noise levels and blocking rates appropriate for television, telephone,
and carrier telephone, FMFB and spectrum analysis of PAM-FM have
lower theoretical thresholds than binary PCM. The binary symbols
are less sensitive to noise than, say, PAM-FM samples received at the
same rate. If this were the whole story, binary phase modulation would
have the smaller threshold by a power ratio of about two.* Actually,
of course, the symbol rate must be greater than the baseband sample
rate by a factor, say ρ, equal to the number of binary symbols per
sample. This, in itself, raises the power threshold by factor ρ. Thus, if
there are more than two symbols per sample, the theoretical threshold
for binary phase modulation is *larger*, by a power ratio of about $\rho/2$.

The threshold ratios are about the same if one compares the binary
PCM with the following FSK system: A set of, say, 10 discrete fre-
quencies is used, spaced orthogonally in the usual signal theory sense.
One frequency from the set is transmitted during each baseband sample
interval. But this system has only 10 quantum levels. To obtain, say,
100 quantum levels one must either transmit two symbol intervals per
sample (which raises the threshold 3 db), increase the channel bandwidth
by a factor of 10, or pack the frequencies much more closely than the
orthogonal spacing. With close spacing, errors of one quantum level

---

* Binary phase modulation requires less power than binary frequency modula-
tion. See, for example, Sunde.[13]

are more probable than larger errors, and there comes a point where they are more like the small noise errors of the analog systems.

In principle (but not likely in practice) thresholds can be *reduced* by using systems with *fewer* symbols or samples per second than the baseband sample frequency. For example, two baseband samples can be transmitted as a single analog sample provided the signal-to-noise ratio can be doubled (>80 db instead of >40 db). Transmission at the reduced sample rate yields a small reduction in threshold. It is paid for by an enormous increase in the channel bandwidth, which is required for the higher signal-to-noise ratio.

If more and more samples are combined, Shannon's fundamental channel capacity is undoubtedly approached. Turin[14] and Golay[15] have demonstrated that two closely related systems do, in fact, approach the theoretical capacity.*

Our formulas for demodulation by spectrum analysis assume that the true signal is estimated by finding the maximum point in the pertinent spectrum. The same is true of the analysis of FSK error rates in Refs. 7, 8 and 9. A well-known substitute for the determination of a maximum uses a circuit whose output is zero except when a signal-plus-noise (in this case the spectrum) exceeds a preset threshold. The threshold is set so that, most of the time, the peak due to the true signal and only that peak gets through.

Under the conditions assumed here, the threshold circuit form increases the theoretical power threshold by very roughly 3 db. More exactly, the blocking probability is dominated by an exponential factor $\exp(-R^2/8)$ as opposed to $\exp(-R^2/4)$ in equations (9).

IV. CONCLUSIONS

The techniques of radar pulse compression can be used to generate spectra of signal samples as analog functions of time. It can be done in real time in the sense that the spectrum of each signal sample is scanned in a time no greater than the sample interval. The spectra are the same as would be generated by infinite sets of correlators or matched filters. Spectrum generation of this sort may be useful for various purposes, particularly where the parameter ranges are suitable for the sort of hardware which has been developed for radar pulse compression.

Demodulation by frequency detection (with or without feedback) reduces the *small noise errors* by a disguised but efficient use of differen-

---

* The increase in channel bandwidth as Shannon's limit is approached is merely a property of these specific modulation schemes. In principle, it is necessary only to increase the length of the pieces of the signal which are coded as units.

tial phase coherence, which is a characteristic of FM signals. Demodulation by spectrum analysis can also take advantage of the differential phase coherence, although the pertinent operations are fairly complicated. The piecewise constant signal frequency, needed for the spectrum analysis, reduces the effectiveness by 1.24 db in the theoretical small noise errors (which can be offset by a 15 per cent increase in the FM index).

Under the conditions assumed, and for thresholds as defined here, the theoretical *power thresholds* of the spectrum analysis are very close to Slepian's lower bound. The power threshold of FMFB appears to be quite close, but just how close has not been determined.

Thus, under conditions appropriate for television, telephone, and carrier telephone systems, the theoretical noise sensitivities are very little different in FMFB and in PAM-FM with demodulation by spectrum analysis. Both techniques pose numerous practical problems, relating to, for example, stability requirements, switching time requirements, synchronization to signal samples, over-all complexity, nonlinearity in response to true signal, etc. FMFB has the advantage that it has already been used, although under somewhat special conditions.

Some theoretical thresholds and small noise errors are collected in Tables I and II, for various blocking probabilities P and bandwidth ratios r. They were calculated by (7) and (9) and refer to demodulation of PAM-FM by phase coherent and incoherent, sample-by-sample spectrum analysis. A few remarks on circuit problems are collected in Section A.9.

The noise figures obtainable with practical circuits are of course somewhat poorer. The degradations may be due to rather different practical compromises in circuits using spectrum analysis and in FMFB. Comparisons between practical noise figures may be different for different applications.

Under some conditions, a combination of spectrum analysis and frequency detection may be preferable to either alone. Fig. 18 is a block diagram of one out of many possible arranagements. A spectrum analyzer furnishes a first estimate of the frequency of a PAM-FM signal, using phase incoherent, sample-by-sample spectrum analysis. The estimated frequency variations are generated locally by a voltage-controlled oscillator. A mixer subtracts the oscillator frequency from the frequency of the received signal. (The block labeled "delay" allows for the operation time of the spectrum analysis.) Then the output of the mixer is very low index FM, corresponding to the errors in the first frequency estimate, plus noise.

TABLE I — THRESHOLDS AND SIGNAL-TO-NOISE RATIO
FOR PHASE COHERENT SPECTRUM ANALYSIS

Probability of Blocking P	Bandwidth Ratio $r = \omega_r/\omega_b$	Threshold Ratio P_s/P_n	$\left[\dfrac{\text{Max. Demod. Signal}}{\text{rms Small Errors}}\right]$*
		(db)	(db)
0.01	10	12.1	31.3
0.005	10	12.7	31.9
0.002	10	13.4	32.6
0.001	10	13.9	33.0
0.0005	10	14.3	33.5
0.0002	10	14.8	33.9
0.0001	10	15.2	34.5
0.01	20	12.8	38.0
0.005	20	13.4	38.5
0.002	20	14.0	39.1
0.001	20	14.4	39.5
0.0005	20	14.8	39.9
0.0002	20	15.3	40.4
0.0001	20	15.6	40.7
0.01	40	13.4	44.6
0.005	40	13.9	45.0
0.002	40	14.4	45.6
0.001	40	14.8	46.0
0.0005	40	15.2	46.3
0.0002	40	15.6	46.8
0.0001	40	15.9	47.1

* At threshold signal power.

Because of the low index, it is now appropriate to use a narrow-band filter (passing something over two baseband bandwidths) followed by a frequency detector and a low-pass filter. The sampled output of the filter furnishes a correction to the first frequency estimate. The theoretical threshold of the combination is the same as for phase incoherent spectrum analysis. The theoretical small noise errors are the same as for demodulation of PAM-FM by frequency detection. The theoretical improvement over the small noise errors of the first frequency estimate is 4.365 db.

If the spectrum analysis is accomplished by correlators or matched filters, a moderate number may be sufficient even though the over-all errors must be >40 db below the true signal. The error determination by frequency detection can correct for a fairly coarse quantization of the first estimate at the same time that it reduces the errors due to noise.

The over-all circuit may be described as open-loop tuning to the passband of the narrow-band filter, as opposed to closed-loop tuning in FMFB.

TABLE II — THRESHOLDS AND SIGNAL-TO-NOISE RATIOS
FOR PHASE INCOHERENT SPECTRUM ANALYSIS

Probability of Blocking P	Bandwidth Ratio $r = \omega_r/\omega_b$	Threshold Ratio P_s/P_n	$\left[\dfrac{\text{Max. Demod. Signal}}{\text{rms Small Errors}}\right]$*
		(db)	(db)
0.01	10	13.3	26.5
0.005	10	13.8	27.1
0.002	10	14.4	27.6
0.001	10	14.8	27.9
0.0005	10	15.2	28.4
0.0002	10	15.7	28.8
0.0001	10	16.0	29.3
0.01	20	13.9	33.1
0.005	20	14.3	33.5
0.002	20	14.9	34.0
0.001	20	15.3	34.4
0.0005	20	15.6	34.7
0.0002	20	16.0	35.1
0.0001	20	16.3	35.4
0.01	40	14.4	39.6
0.005	40	14.8	40.0
0.002	40	15.3	40.5
0.001	40	15.6	40.8
0.0055	40	16.0	41.1
0.0002	40	16.3	41.5
0.0001	40	16.6	41.8

* At threshold signal power.
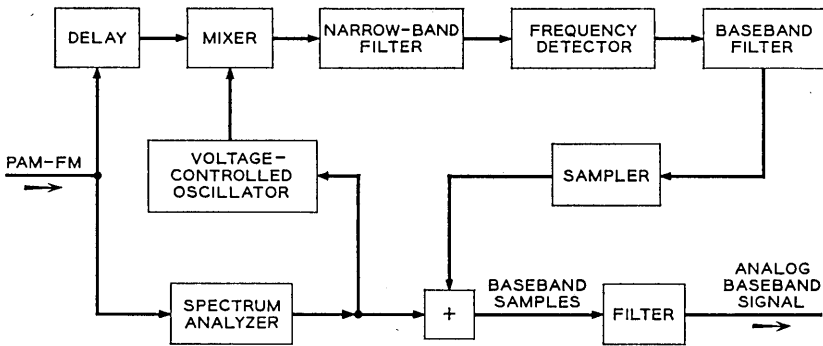
Fig. 18 — A combination of spectrum analysis and frequency detection.

APPENDIX

A.1 Spectrum Generation by Pulse Compression

   For a signal sample, modified by the local oscillator, assume:

$$\hat{s}(t) = \sqrt{2P_s}\, E(t) \cos \left(\omega_s t - \tfrac{1}{2}\, qt^2 + \beta_s \right)$$

$$E(t) = 0 \text{ outside of interval } -T/2 \leqq t \leqq +T/2$$

$$E(-t) = E(t).$$

For the impulse response of the pulse compression line, assume:

$$w(t) = \cos \left(\omega_c t + \tfrac{1}{2}\, qt^2 - \beta_c \right).$$

When $|\, \omega - \omega_c\,| \ll \omega_c$, the frequency function is

$$Y(i\omega) = (\quad) \exp \left[-i\, \frac{(\omega - \omega_c)^2}{2q} \right].$$

   The output of the line is $\hat{s}*w$. Integrate only over $E(t) \neq 0$:

$$S(t) = \sqrt{2P_s} \int_{\tau=-T/2}^{+T/2} E(\tau) \cos \left(\omega_s \tau - \tfrac{1}{2}\, q\tau^2 + \beta_s \right)$$

$$\cdot \cos \left[\omega_c(t - \tau) + \tfrac{1}{2}\, q(t - \tau)^2 - \beta_c \right] d\tau.$$

   Express the integrand as a sum of cosines. Neglect the high-frequency
term. Then:

$$S(t) = \sqrt{2P_s} \int_{\tau=-T/2}^{+T/2} \tfrac{1}{2} E(\tau) \cos \left[\omega t + \tfrac{1}{2}\, qt^2 + \beta_s - \beta_c \right.$$

$$\left. + (\omega_s - \omega_c - qt)\tau \right] d\tau.$$

Resolve into components per sin, cos $[(\omega_s - \omega_c - qt)\ \tau]$.

Recall that $E(\tau)$ is even. Then $E(\tau) \sin [(\omega_s - \omega_c - qt)\ \tau]$ is odd in τ.

$$S(t) = \sqrt{2P_s}\ F(\omega_s - \omega_c - qt) \cos (\omega t + \tfrac{1}{2}\ qt^2 + \beta_s - \beta_c)$$

$$F(\lambda) = \int_{\tau=-T/2}^{+T/2} \tfrac{1}{2}E(\tau) \cos (\lambda\tau)\ d\tau.$$

When $E(\tau) = 1,\ -T/2 \leqq \tau \leqq +T/2,\ F(\lambda) = \dfrac{\sin \lambda \dfrac{T}{2}}{\lambda}.$

A.2 Noise Contributions to Observed Spectrum

Following Rice,[16] but sacrificing some details of rigor to brevity, let the noise at the demodulator input be:

$$n(t) = \int_{\omega_1}^{\omega_2} x(\omega) \cos (\omega t + \beta_n)\ d\omega + \int_{\omega_1}^{\omega_2} y(\omega) \sin (\omega t + \beta_n)\ d\omega.$$

The interval ω_1 to ω_2 includes all signal frequencies ω_s.

Phase β_n = an arbitrary parameter in noise representation.

$x(\omega),\ y(\omega)$ = uncorrelated, zero average random variables, with uniform variances, and zero autocorrelations except across infinitesimal frequency intervals.

Let Ave denote an ensemble average, or expectation.

Let $w_1(\omega)$ and $w_2(\omega)$ be arbitrary, except for the pertinent conditions of integrability.

$$\text{Ave} \left\{ \int_{\omega_1}^{\omega_2} x(\omega)\ w_1(\omega)\ d\omega \int_{\omega_1}^{\omega_2} x(\omega)\ w_2(\omega)\ d\omega \right\} = \sigma^2 \int_{\omega_1}^{\omega_2} w_1(\omega)\ w_2(\omega)\ d\omega$$

$$\text{Ave} \left\{ \int_{\omega_1}^{\omega_2} y(\omega)\ w_1(\omega)\ d\omega \int_{\omega_1}^{\omega_2} y(\omega)\ w_2(\omega)\ d\omega \right\} = \sigma^2 \int_{\omega_1}^{\omega_2} w_1(\omega)\ w_2(\omega)\ d\omega$$

$$\text{Ave} \left\{ \int_{\omega_1}^{\omega_2} x(\omega)\ w_1(\omega)\ d\omega \int_{\omega_1}^{\omega_2} y(\omega)\ w_2(\omega)\ d\omega \right\} = 0.$$

P_b = noise power in one base bandwidth = $\omega_b \sigma^2$.

Let $N(\omega_k)$ = the noise part of the spectrum of one signal-plus-noise sample.

Apply Section A.1, with $\omega_s = \omega$ and $\omega_c + qt = \omega_k$, to integrands in $n(t)$.

$$N(t) = N_1(\omega_k) \cos (\omega_c t + \tfrac{1}{2} q^2 - \beta_c)$$

$$+ N_2(\omega_k) \sin (\omega_c t + \tfrac{1}{2} q^2 - \beta_c)$$

$$N_1(\omega_k) = \int_{\omega_1}^{\omega_2} x(\omega)\, F(\omega - \omega_k)\, d\omega,$$

$$N_2(\omega_k) = \int_{\omega_1}^{\omega_2} y(\omega)\, F(\omega - \omega_k)\, d\omega$$

$N_1(\omega_k), N_2(\omega_k)$ = independent Gaussian random processes in ω_k.

Appropriate choices of w_1, w_2 in the above expectation integrals give autocovariances of N_1, N_2.

$$\text{Ave } [N_\gamma(\omega_k)\, N_\gamma(\omega_j)] = \sigma^2 \int_{\omega_1}^{\omega_2} F(\omega - \omega_k)\, F(\omega - \omega_j)\, d\omega, \qquad \gamma = 1, 2.$$

Approximate the integration by integrating from $-\infty$ to $+\infty$.
Refer to Section A.1 and use $E(t) = 1$, $-T/2 \leqq t \leqq +T/2$.

$$\text{Ave } [N_\gamma(\omega_k)\, N_\gamma(\omega_j)] = \pi\sigma^2 \frac{\sin (\omega_k - \omega_j) \dfrac{T}{2}}{(\omega_k - \omega_j)}.$$

Let $\omega_j = \omega_k$, refer to (3), and recall that $R^2 = \dfrac{P_s}{P_b} = \dfrac{P_s}{\omega_b \sigma^2}$, $T = \dfrac{\pi}{\omega_b}$.

$$\frac{\text{Max of Signal Spectrum}}{\text{rms } N_\gamma(\omega_k)} = \sqrt{\frac{P_s}{P_b}} = R, \qquad \gamma = 1, 2.$$

A.3 Small Noise Errors in Sample-by-Sample Spectrum Analysis

Refer to Fig. 4, $S(t)$ of (3), and $N(\omega_k)$ of Section A.2.
Use $\omega_k = \omega_c + qt$ and $\beta_n = \beta_s$.

$$S(t) + N(t) = [\sqrt{2P_s}\, F(\omega_s - \omega_k) + N_1(\omega_k)]$$

$$\times \cos (\omega t + \tfrac{1}{2} qt^2 + \beta_s - \beta_c)$$

$$+ N_2(\omega_k) \sin (\omega t + \tfrac{1}{2} qt^2 + \beta_s - \beta_c).$$

Assume (for small noise errors only):

$$N_1^2, N_2^2 \ll 2P_s F^2(0), \qquad \omega_k - \omega_s = \epsilon, \qquad \epsilon^2 \ll \omega_b^2.$$

Phase Incoherent Spectrum Analysis. Neglecting N_2^2, the envelope is $\sqrt{2P_s}\, F(\omega_s - \omega_k) + N_1(\omega_k)$.
Form a power series in ϵ and solve for max with ϵ small.

$$\text{Ave } \epsilon^2 = \frac{\text{Ave} \left(\dfrac{\partial N_1}{\partial \omega_k}\right)^2}{2P_s \left(\dfrac{\partial^2 F(\epsilon)}{\partial \epsilon^2}\right)^2_{\epsilon=0}}.$$

Evaluate by (3) and Section A.2 to get (7b).

Phase Coherent Sample-by-Sample Spectrum Analysis. Refer to (1). Make the phase β_s a linear function of ω_s :

$$S(t) = \sqrt{2P_s} \cos [\omega_s t + (\omega_s - \omega_c)(T/2)].$$

Find the components of $S(t)$ and $N(t)$ in phase with a locally generated $\cos \{[\omega_c + q(T/2)]t + \frac{1}{2} qt^2 - \beta_c\}$.

Refer to (3). Let S_c be the component of S.

$$S_c(t) = \sqrt{2P_s} \, F(\omega_s - \omega_k) \cos [(\omega_s - \omega_k)(T/2)]$$

$$= \sqrt{2P_s} \, F[2(\omega_s - \omega_k].$$

It can be shown that the frequency variable is also doubled between covariances of $N_1(\omega_k)$ and its counterpart here. Hence noise is accounted for with $\frac{1}{2}$ the frequency errors ϵ.

If the frequency-dependent signal phase appears artificial, change the time scale to $\hat{t} = t + T/2$.

$$S(\hat{t}) = \sqrt{2P_s} \cos \left(\omega_s \hat{t} - \frac{T}{2} \omega_c\right), \qquad 0 \leqq \hat{t} \leqq T.$$

A.4 Small Noise Errors in Frequency Detection

The FM signal is now unsampled. For simplicity assume a constant signal frequency. Resolve the noise per signal phase.

$$s(t) + n(t) = [\sqrt{2P_s} + n_a(t)] \cos (\omega_s t + \beta_s)$$

$$+ n_b(t) \sin (\omega_s t + \beta_s)$$

$$s(t) + n(t) = \rho \cos [\omega_s t + \beta_s + \varphi(t)], \qquad \tan \varphi = \frac{n_b(t)}{\sqrt{2P_s} + n_a(t)}.$$

The unfiltered frequency error is $\dot{\varphi}$. Refer to Section A.2 to get:

When $n^2 \ll 2P_s$, $\text{Ave } \dot{\varphi}^2 = \dfrac{\text{Ave } \dot{n}_b^2}{2P_s} = \dfrac{\sigma^2}{2P_s} \displaystyle\int_{\omega_1}^{\omega_2} (\omega - \omega_s)^2 \, d\omega.$

The ideal baseband filter passes only $| \omega - \omega_s | \leqq \omega_b$.

$$\text{Ave (Filtered } \epsilon)^2 = \frac{\sigma^2}{2P_s} \int_{-\omega_b}^{+\omega_b} \lambda^2 d\lambda = \frac{\sigma^2 \omega_b^3}{3P_s} = \frac{\omega_b^2 P_b}{3P_s} = \frac{\omega_b^2}{3R^2}.$$

A.5 Small Noise Errors in Multisample Spectrum Analysis

Refer to (1) and Fig. 13(a). Let $(\omega_\sigma, \beta_\sigma) = \omega_s$ and the midsample phase β_s of sample σ. With differential phase coherence,

$$\beta_\sigma - \beta_{\sigma-1} = (\omega_\sigma + \omega_{\sigma-1})(T/2).$$

Let n_σ, m_σ = noise contributions to observed ω_σ, $(2/T) \beta_\sigma$.
Define $x_\sigma, y_\sigma, z_\sigma$ and note the relation to errors:

$$x_\sigma = \omega_\sigma + n_\sigma, \qquad y_\sigma = (2/T)\beta_\sigma + m_\sigma$$

$$z_\sigma = (x_\sigma + x_{\sigma-1}) - (y_\sigma - y_{\sigma-1}) = (n_\sigma + n_{\sigma-1}) - (m_\sigma - m_{\sigma-1}).$$

Let $\omega_\sigma + \epsilon$ = the following estimate of ω_σ :

$$\omega_\sigma + \epsilon = x_\sigma - \sum_{j=-\infty}^{+\infty} Q_j z_j.$$

Let σ_n^2 = Ave n_σ^2, σ_m^2 = Ave m_σ^2

$$\text{Ave } \epsilon^2 = \left[1 - 2(Q_\sigma + Q_{\sigma+1}) + \sum_j (Q_j + Q_{j+1})^2\right] \sigma_n^2$$

$$+ \left[\sum_j (Q_j - Q_{j+1})^2\right] \sigma_m^2.$$

Choose the Q_j's for min. Ave ϵ^2 by the calculus of variations.
Compare with Ave ϵ^2 for x_σ alone, which is σ_n^2.

$$\frac{\text{Min Ave } \epsilon^2 \text{ of sum}}{\text{Ave } \epsilon^2 \text{ of } x_\sigma \text{ alone}} = \frac{\sigma_m}{\sigma_n + \sigma_m}.$$

Further analysis like that of Sections A.2 and A.3 gives

$$\sigma_n^2 = 3\sigma_m^2$$

$$\frac{\text{Min Ave } \epsilon^2 \text{ of sum}}{\text{Ave } \epsilon^2 \text{ of } x_\sigma \text{ alone}} = \frac{1}{1 + \sqrt{3}} \text{ or } - 4.365 \text{ db.}$$

A.6 Small Noise Errors in Multisample Frequency Detection of PAM-FM

Refer to Section A.4 but assume only a piecewise constant signal frequency. Refer to Figs. 13(a) and (15).
Let $w(t)$ = filter weight factor, referred to the output sample time.
Assume $w(\pm \infty) = 0$.

$$\text{Filtered error} = \int_{-\infty}^{+\infty} w(t)\ \dot{\varphi}(t)dt = -\int_{-\infty}^{+\infty} \dot{w}(t)\ \varphi(t)dt.$$

Use $\varphi(t) = [n_b(t)/\sqrt{2P_s}]$ and a white noise approximation.

$$\text{Ave } \epsilon^2 = \frac{\sigma^2}{\sqrt{2P_s}} \int_{-\infty}^{+\infty} [\dot{w}(t)]^2 dt.$$

Find $w(t)$, which gives

(a) normalized response to constant frequency in sample σ,

(b) zero response to constant frequencies in samples other than σ,

(c) minimum Ave ϵ^2 within constraints a, b.

The calculus of variations makes $w(t)$ quadratic over each sample interval and continuous at the boundaries. Then Ave ϵ^2 is a quadratic sum of the boundary values. Minimizing the boundary values is like minimizing the coefficients Q_j in Section A.5 (with $\sigma_n^2 = 3\sigma_m^2$) and gives the same result.

A.7 Blocking Probability in Sample-by-Sample Spectrum Analysis

Refer to Sections 3.3 and A.3. Approximate $N_1(\omega_k)$, $N_2(\omega_k)$ by processes piecewise constant over η subintervals.

Approximate $\sqrt{2P_s}\ F(\omega_s - \omega_k)$ by $\sqrt{2P_s}\ F(0)$ over the subinterval s and zero elsewhere.

Let x_λ , $y_\lambda =$ the components of the signal-plus-noise spectrum, scaled (normalized) to unit variances. The probability densities are:

$$D_s = \frac{1}{2\pi} \exp\left[-\frac{(x_s - R)^2 + y_s^2}{2}\right],$$

$$D_\lambda = \frac{1}{2\pi} \exp\left(-\frac{x_\lambda^2 + y_\lambda^2}{2}\right), \qquad \lambda \neq s.$$

Phase Coherent Sample-by-Sample Spectrum Analysis. Rotation of the x_λ , x_s axes through $\pi/4$ gives quickly

$$P\{x_\lambda > x_s \mid \lambda \neq s\} = \frac{1}{2}\left[1 - \text{Erf}\left(\frac{R}{\sqrt{2}}\right)\right],$$

$$\text{Erf }(r) = \sqrt{\frac{2}{\pi}} \int_0^r \exp\left(-\frac{u^2}{2}\right) du.$$

This is the probability of a specific $x_\lambda > x_s$, out of $\eta - 1$ x_λ's, $\lambda \neq s$.

Under the conditions assumed here, the probability of any one or more is:

$$P \approx (\eta - 1)P\{x_\lambda > x_s\} = \frac{\eta - 1}{2}\left[1 - \text{Erf}\left(\frac{R}{\sqrt{2}}\right)\right]$$

$$\approx \frac{\eta - 1}{\sqrt{\pi}R} \exp\left(-\frac{R^2}{4}\right).$$

Per Section 3.3, use

$$\eta = \frac{r\omega_b}{2\omega_b} = \frac{r}{2}.$$

Phase Incoherent Sample-by-Sample Spectrum Analysis

$$P\{(x_\lambda^2 + y_\lambda^2) > (x_s^2 + y_s^2)\} = \frac{1}{2}\exp\left(-\frac{R^2}{4}\right).$$

Under the conditions assumed here, for one or more λ's, $\lambda \neq s$:

$$P \approx \frac{\eta - 1}{2}\exp\left(-\frac{R^2}{4}\right), \qquad \text{use } n = r/2 \text{ as before.}$$

The last approximation is here a simplification, not a necessity. For an exact formula (given D_s, D_λ as above) see Ref. 7 or 8.

A.8 Reduction of the Blocking Rate of Spectrum Analysis

Refer to Section A.5. Use x, y of A.5. For a second estimation of ω_s,

$$\omega_s = (1/T)(\beta_{s+1} - \beta_{s-1}) - \tfrac{1}{2}(\omega_{s+1} + \omega_{s-1})$$

$$\omega_s + \epsilon = \tfrac{1}{2}(y_{s+1} - y_{s-1}) - \tfrac{1}{2}(x_{s+1} + x_{s-1}) + (2\nu\pi/T)$$

$$\nu = \text{unknown integer due to phase ambiguities.}$$

Refer to Fig. 16(b). Find the integers ν for the best fits to frequencies of the two peaks in the signal-plus-noise spectrum.

With no weighting for the heights of peaks, the probability that the closest is the correct choice is of the order of 0.9 (under the system conditions assumed here).

The actual choice must use also the relative heights of the peaks.

Let $P_M(M_s, M_n) =$ the probability density of the maxima M_s, M_n at the peaks due to signal-plus-noise and noise only (respectively).

Let $P_\epsilon(\epsilon_s, \epsilon_n) =$ the probability density of the observed deviations ϵ_s, ϵ_n of the second ω_s from the location of the peaks, using best ν's.

Use subscripts 1, 2 for the M's and ϵ's before the identification of which peak is signal-plus-noise and which is noise only.

The best identification corresponds to the larger of

$$P_M(M_1, M_2)\, P_\epsilon(\epsilon_1, \epsilon_2) \quad \text{and} \quad P_M(M_2, M_1)\, P_\epsilon(\epsilon_2, \epsilon_1).$$

P_M gives a strong weighting except when M_2 is close to M_1.

But when $M_n > M_s$, the difference is usually small, and P_n only very rarely gives a strong weighting to a wrong choice.

$$\text{Let } u = M_s - M_n. \text{ Then } \frac{P(u)}{P(-u)} = e^{Ru}.$$

A complete calculation of the probability of a correct choice would require integration over a complicated portion of the 4-dimensional space of M_s, M_n, ϵ_s, ϵ_n.

A.9 Some Circuit Considerations

A few circuit considerations are described below in brief, purely qualitative, terms.

Synchronization of Spectrum Analysis to PAM-FM Samples. Assume the following: The spectra represent signal-plus-noise received during intervals locally selected by a precision oscillator or clock. The length T of the intervals is almost right, without synchronizing means. The problem is to synchronize the start time to the start times of the true signal samples.

Synchronizing signals might be obtained by any of several means. One uses a very narrow band transmission channel, to send synchronizing signals from the transmitter. Others derive synchronization error signals from the communication signal itself, which must fluctuate sufficiently to supply the necessary information. (When the true signal is constant from sample to sample, there is nothing to indicate the boundaries between samples.) An error in synchronization reduces the height of the peak in the signal spectrum (on the average). It also produces a discrepancy between values of ω_s obtained from the single sample spectrum and by the second method described in Section A.5. In principle at least, a synchronization error signal can be derived from either effect and can be averaged over many sample intervals to reduce the effects of noise on the synchronization.

Shape of the Signal Sample. In (3), the tails of the function F are neither small nor short. By Section A.1, they can be reduced by shaping the envelope $E(t)$ of the signal sample before forming its spectrum. A suitable filter in the output of the spectrum generator has the same effect. Since the best spectral maximum corresponds to the F of (3), a practical compromise is needed. The pulse shaping problem is an old one, but here intersample interference due to the tails is not the important problem, but rather the way the tails can increase the blocking probability (noise-plus-tails exceeding signal-plus-noise).

Channel Bandwidth. For both ordinary FM and PAM-FM the channel bandwidth must be a little wider than the full excursion, ω_r, of the instantaneous signal frequency ω_s. The so-called Carson's Rule calls for a channel width of $\omega_r + 2\omega_b$ for ordinary FM, and the appropriate rule for PAM-FM is at least not very different. FMFB and PAM-FM spectrum analysis can tolerate wider bands without significant changes in thresholds and small noise errors.

Transition Intervals. In idealized models of spectrum analysis, certain operations happen in zero time. In any actual circuits there will be nonzero switching times. Very roughly, if a fraction α of each sample interval is lost due to the switching times, the signal power must be increased by factor $1/(1 - \alpha)$. Thus 2 per cent lost time requires roughly 0.1 db more power. In a sense, switching times are spectrum analysis counterparts of feedback stability problems in FMFB, although the comparison is purely qualitative.

REFERENCES

1. Jordan, D. B., Greenberg, H., Eldredge, E. E., and Serniuk, W., Multiple Frequency Shift Teletype Systems, Proc. IRE, **43**, November, 1955, pp. 1647–1665.
2. Klauder, J. R., Price, A. C., Darlington, S., and Albersheim, W. J., The Theory and Design of Chirp Radars, B.S.T.J., **39**, July, 1960, p. 745.
3. Feldman, C. B., and Bennett, W. R., Bandwidth and Transmission Performance, B.S.T.J., **28**, July, 1949, p. 490.
4. Turin, G. L., An Introduction to Matched Filters, Trans. IRE-PGIT, IT-6, June, 1960, p. 311.
5. Kotel'nikov, V. A., *The Theory of Optimum Noise Immunity*, translated by R. A. Silverman, McGraw-Hill, 1959.
6. Blackman, R. B., Bode, H. W., and Shannon, C. E., Data Smoothing and Prediction in Fire-Control Systems, Summary Technical Report of Div. 7, NDRC **1**, Report Series No. 13, MGC 12/2, National Military Establishment Research and Development Board.
7. Turin, G. L., Probability of Error in NOMAC Systems, Lincoln Lab. Tech. Report No. 57, January, 1954.
8. Reiger, S., Error Rates in Data Transmission, Proc. IRE, **46**, May, 1958, pp. 919–920.
9. Nuttall, A. H., Error Probabilities for Equicorrelated M-ary Signals Under Phase-Coherent and Phase-Incoherent Reception, IRE Trans. on Info. Theory, IT-**8**, July, 1962, pp. 305–314.
10. Slepian, D., The Threshold Effect in Modulation Systems that Expand Bandwidth, IRE Trans. on Info. Theory, IT-**8**, No. 5, September, 1962, pp. 122–127.
11. B.S.T.J., issue devoted to the *Telstar* Experiment, **42**, July, 1963.
12. Enloe, L. H., Decreasing the Threshold in FM by Frequency Feedback, Proc. IRE, **50**, January, 1962, pp. 18–30.
13. Sunde, E. D., Ideal Binary Pulse Transmission by AM and FM, B.S.T.J., **38**, November, 1959, pp. 1357–1426.
14. Turin, G. L., The Asymptotic Behavior of Ideal M-ary Systems, Proc. IRE, **47**, January, 1959, pp. 93–94.
15. Golay, M. J. E., Letter to the Editor, Proc. IRE, **37**, September, 1949, p. 103.
16. Rice, S. O., Mathematical Analysis of Random Noise, B.S.T.J., **23**, July, 1944, p. 282; and **24**, January, 1945, p. 46.

Data Transmission over a Self-Contained Error Detection and Retransmission Channel

By F. E. FROEHLICH and R. R. ANDERSON

Error control of the detection and retransmission type requires an internal storage buffer when the data source cannot be stopped. With finite capacity there will be occasions when this internal buffer is overfilled. This paper investigates the relationships among the error statistics of the channel, the storage capacity of the buffer, the round-trip transmission delay and the bit rate from the source. It is shown that the process can be treated as a Markov chain. The solution algorithm is programmed for machine computation, and representative cases are solved numerically. For typical values selected from the telephone plant, it is found that buffer capacities of a few hundred bits would be adequate.

The technique described should be useful for solving other problems in queueing theory.

I. INTRODUCTION

Studies during the last few years have shown that in the transmission of digital data over telephone lines, high accuracy can be achieved when the message is encoded in an error detecting code. Correction can then be accomplished by a repeat transmission of the portion of the information containing the errors. These so-called "feedback" techniques have been shown to be very effective in controlling errors.[1,2,3,4]

For some sources of data it is inconvenient or impossible to have the source wait while previous data are being retransmitted. There are also cases where it is required that the output from the receiver be at a uniform rate. This memorandum describes a self-contained error detection and retransmission channel capable of accepting data from the source at a steady rate, or at any rate less than a specified maximum, and of delivering it to the sink at this same rate. The channel is "self-contained," meaning that the channel itself provides enough storage of in-

formation to permit the detection of errors and their correction by re-transmission without the data source and sink being aware that these processes are going on. The data source merely puts data into the trans-mission system at its own rate, and the data sink accepts highly reliable data from the system at the same rate. The relationships among system delay time, error probability, bit rate, and storage capacity are investi-gated.

The use of feedback error control with a data source which cannot be interrupted was briefly discussed by Reiffen, Schmidt, and Yudkin.[5] A. B. Fontaine has simulated such a system on a computer, using error data collected on private wire circuits.[6] Our analysis has indicated that shorter blocks could well have been used in the experimental simulation, which would have reduced the required storage capacity or increased the time to overflow.

II. THE DATA CHANNEL

A block diagram of the self-contained data channel is shown in Fig. 1. The transmitter consists of a buffer store, an encoder, a modulator, a reverse channel receiver and some logic. The transmission channel itself has a forward path and a reverse path, the latter carrying very little in-formation compared to the former. The receiver consists of a detector, a decoder, a buffer store, and a reverse channel transmitter plus logic.

The forward channel carries data (plus any necessary redundancy and starting codes); the reverse channel carries information indicating whether retransmission is required. Errors in the reverse channel will not appreciably affect the operation. The small amount of information required over this channel permits a high degree of redundancy. In ad-dition, a "fail-safe" code can be used, so that any undetected errors on the reverse channel result in unnecessary retransmissions (subsequently eliminated at the receiver) to ensure against loss of data.

To facilitate discussion, a specific model, chosen for its relative sim-plicity, is described. Modifications and improvements are apparent and will be briefly discussed. The method of operation is to accept data from the source continuously at a constant rate, R_S bits per second, which is less than the maximum rate, R_L, allowed by the data transmission sys-tem. The efficiency then, without considering the error-detecting code, is

$$E = R_S/R_L. \tag{1}$$

The data are transmitted at an effective rate of R_S until a retransmis-

sion is requested. After a retransmission request, data are sent at the
higher rate, R_L, until the system is returned to normal.

The change in rate could be made by switching the transmitting speed
of the data set. Another method to achieve the data rate change is con-
tinuous transmission at rate R_L with interspersed dummy or "fill-in"
bits as needed. The two methods are mathematically equivalent, and
we shall assume the latter for the discussion in this paper. Thus, in the
transmitting buffer the data are organized into blocks of N bits each
and sent to the encoder at a rate, R_L, faster than the maximum allow-
able input rate. In order to equalize the input and output rates of the
buffer, "fill-in" bits containing no information are inserted between the
blocks of message bits as shown in Fig. 2. The data then pass through
the encoder, where additional redundancy is added to allow for error
detection. At the encoder, one may either ignore the fill-in bits or en-
code them, but will probably use them to transmit additional useful
information. It is of course possible to place the error control encoder
before the buffer, but this increases the required buffer size without gain-
ing any apparent advantages. The signal is then modulated for transmis-
sion over the forward path.

Each block of information is retained in the transmitting buffer until
it is certain that there will be no retransmission request from the re-
ceiver. When a sufficient time interval has elapsed and no retransmis-
sion request is received, the block of data is erased from the transmitting
buffer. This time interval is taken to be T_D, the maximum round-trip
delay for which the system is designed. This includes the transmission
time in both directions plus any additional time for logical operations at
either end.

The system as described has a sort of natural block length, the number
of bits emitted by the source at rate R_S in time T_D

$$N = R_S T_D. \tag{2a}$$

With this block size, it is known that a retransmission request must apply
to the immediately preceding block of data bits.

It is shown later that shorter blocks have an advantage in reducing
the required buffer size, and hence we let

$$N = R_S T_D / k \tag{2b}$$

where k is an integer. For these shorter blocks, the system must assume
a maximum T_D or must include some provision for determining the ac-
tual round-trip delay time so that retransmission requests can be associ-
ated with the proper blocks of data.
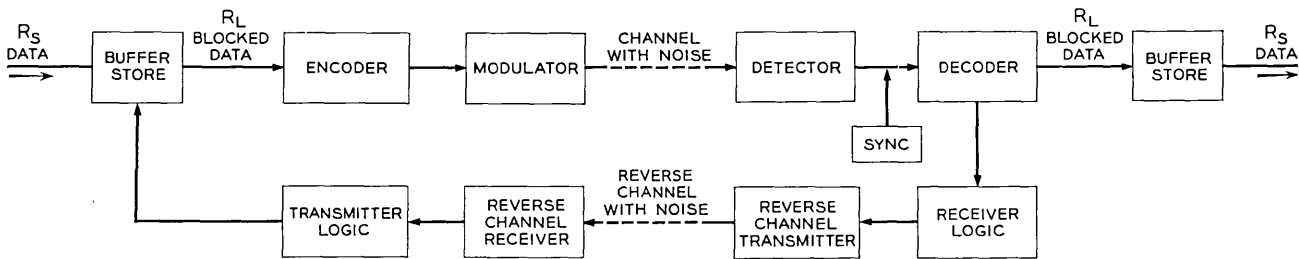
Fig. 1 — Complete self-contained error control channel.

SOURCE MESSAGE (RATE R_S)

| A | B | C | D | E | F | G |

$\leftarrow T_D \rightarrow$

$\leftarrow$ N BITS $\rightarrow$

EFFICIENCY, $E = R_S/R_L$

BLOCK SIZE, $N = R_S T_D$

TRANSMITTED MESSAGE (RATE R_L)

RETRANSMISSION REQUEST

| FILL | A | FILL | B | A' | B' | C | D | E | F | FILL | G |

$\leftarrow$ N BITS $\rightarrow$ $\leftarrow T_D \rightarrow$

$\leftarrow\rightarrow T_D$

$\leftarrow\rightarrow T_D$

$\leftarrow\rightarrow T_D$

$\leftarrow$ NORMAL MODE $\rightarrow$ $\leftarrow$ RETRANSMISSION MODE $\rightarrow$ $\leftarrow$ NORMAL MODE $\leftarrow$
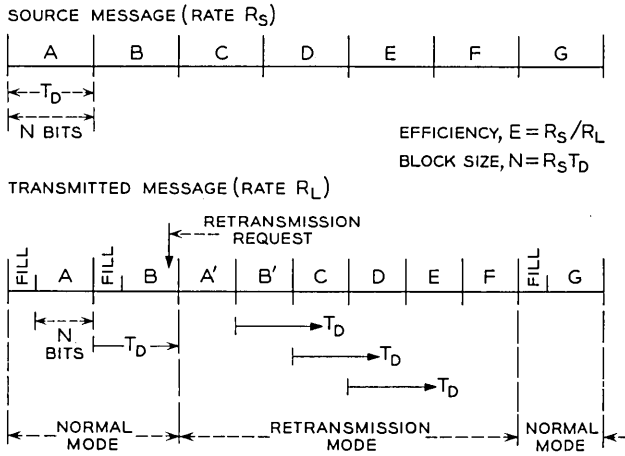
Fig. 2 — Example of time sequence at transmitter.

The number of bits, including both data and fill-in, from the buffer in the same time T_D/k is

$$N + M = R_L T_D/k. \tag{3}$$

In the receiver the demodulated signal is decoded and checked for errors. If no errors are found, the data block, with all redundancy removed, is placed in the receiving buffer. In case an error is detected in the received block of data, a retransmission request is sent to the transmitter via the reverse data channel, and no data are sent to the receiving buffer.

In the transmitter we impose the operating rule: in case a retransmission request is observed, the transmitter will complete the transmission of the current block of $N + M$ bits and then revert to the beginning of the block detected to be in error.† The transmitter then enters the retransmission mode and retransmits information starting with the block in error. During this period, the transmitting buffer continues to receive and store data from the source, thus increasing the quantity of information stored. In order to return the transmitting buffer to its normal state, the fill-in bits are now omitted between the transmitted blocks of data, so that bits will be removed faster than they arrive. This reduces the

---

† Another way to say this is that the transmitter takes no action on a retransmission request until the end of a full round-trip delay time, T_D, after sending the last bit of the block to be retransmitted. In this form the statement is also true when the transmitter is already in the retransmission mode. Note that in the latter case the time of decision is not necessarily at the end of a block.

information stored in the transmitting buffer and at the same time tends to refill the receiving buffer. The fill-in bits are omitted until both buffers have returned to their normal state.

The above sequence is illustrated in Fig. 2. Block A has been received in error. The retransmission request is noted by the transmitter before the completion of block B. At the conclusion of block B transmission, both A and B are retransmitted. Fill-in bits are now omitted until such time as the transmitting buffer returns to its normal state. This occurs after transmission of block F, if there are no additional retransmission requests.

We note immediately that, in case a number of nearby data blocks are found to be in error, the transmitting buffer may overflow. Similarly, the receiving buffer may empty out, so that for some time no information will be available to the data sink. The frequency of occurrence of these events depends, of course, on the error statistics of the channel, the storage capacity of the buffers, the round-trip transmission delay, the number of fill-in bits allowed between data blocks, and the size of the data block.

Questions to be answered about the self-contained data channel are: How often does the transmitting buffer store overflow and the receiving buffer empty completely? What delay is encountered by the information prior to delivery to the sink? What efficiency can this system achieve? What buffer store capacity is needed? In general, what are the relationships between buffer store size, block length, transmission efficiency, transmission delay, and average time between overflows, in any given message?

III. THE MARKOV PROCESS

In the following development, it will be assumed that retransmission requests are independent with probability P_r. For digital data transmission over telephone lines, individual bit errors are known to be not independent; however, for blocks which are long with respect to the bit error dependence, the retransmission requests will be nearly independent. There is some evidence that over voice telephone circuits at 1000–2000 bits per second the correlation among bit errors becomes so small after 10–15 bit intervals that the assumption of block error independence is acceptable.[1] An estimate of the probability of a retransmission is available, since the block error rate cannot be greater than the bit error rate times the block length.†

---

† Let λ be the bit error rate in B bits. Then λB is the number of bits in error. The number of blocks in error cannot be greater than λB. The total number of blocks is B/N so an upper limit of probability of block error is

We shall now devote ourselves to the question of the relationship between the storage capacity of the buffer and the average time between overflow of the buffer. It is evident that, since the number of data bits transmitted per unit time is not constant, an actual time calculation is inconvenient. We therefore quantize time into unequal units, such that the number of data bits transmitted per quantum is always the same.

The possible number of bits stored in the buffer form the states of a stochastic process. It will now be shown that, if these are considered only at certain moments of decision, the buffer states, y, form a finite Markov chain.

The only time a decision is made is exactly T_D seconds after the last bit of a block has been transmitted, and the decision consists of three parts:

(a) Which block shall be transmitted?

(b) Shall fill-in bits be transmitted following the data block?

(c) May the transmitting buffer erase a block of data?

The decision depends only on the state of the buffer and on whether a retransmission is requested; there are four cases:

(i) *Normal — The system is not in the retransmission mode, and retransmission is not required.* The buffer erases one block; the transmitter sends fill-in bits and then the next block in sequence from the source. By the time of the next decision, the buffer will have replaced the erased block with one block from the source. Thus, at the moment of the next decision, the total change in the buffer storage is zero. The time to the next decision is T_D/k.

(ii) *The system is not in the retransmission mode, but a retransmission is requested.* The buffer does not drop any bits. The transmitter backs up to the block at the beginning of the buffer in order to retransmit the block received in error. The transmitter shifts its mode and no fill-in bits are sent. The next decision will be made after one block has been completely transmitted plus T_D seconds, to allow time for another retransmission request to be received. During the retransmission time, EN bits come

$$\frac{\lambda B}{B/N} = \lambda N.$$

There may be multiple bit errors in a block, and some of the block errors may not be detected, so

$$P_r \leqq \lambda N. \tag{4a}$$

For the special case where bit errors are independent

$$P_r = 1 - (1 - \lambda)^N \doteqdot \lambda N. \tag{4b}$$

for λ much smaller than 1.

from the source, and during T_D, R_sT_D bits. The total increase in storage due to one retransmission is thus

$$I = R_sT_D + EN = R_sT_D(1 + E/k).\tag{5}$$

The time to the next decision is $T_D(1 + E/k)$.

(*iii*) *Off-Normal — The system has previously entered the retransmission mode and no additional retransmission is requested.* The buffer can drop the block which was received correctly. The transmitter continues with the block following the one just sent, without fill-in bits. The next decision will take place after the time required to transmit one block, in which time EN bits are added to the buffer. Since the buffer has dropped a full block, the amount of data in the buffer has decreased by

$$D = N(1 - E).\tag{6}$$

The time to the next decision is T_DE/k.

(*iv*) *The system is in the retransmission mode and another retransmission is required.* This is similar to case (*ii*), except that the transmitter shift is not required since it is already in the retransmission mode. The same number of bits will be discarded at the receiver, but, being already in the retransmission mode, none of these are fill-in bits, so the number of blocks to be retransmitted is greater by the ratio $(N + M)/N$. The transmitter remains in the retransmission mode and fill-in bits will not be sent. The increase in storage is given by (5), and the time to the next decision is $T_D(1 + E/k)$.

Let C be the total storage capacity of the transmitting buffer. When the source rate is constant, the transmitter can send the block as it is received. In this case, the smallest useful capacity, C_{min}, includes the one block to which the retransmission request applies, if received, plus the data which arrive from the source during the round-trip delay preceding the request

$$C_{min} = N + R_sT_D.\tag{7}$$

If the source rate may fluctuate and the start of transmission must be delayed, C must be larger. The worst case is that in which the source may intermittently stop so the transmitter must wait until the full block is received, in which case the minimum C is one block more. This additional block of storage to compensate for an intermittent source should probably not be charged to the error control system. The ability to provide this feature in a simple manner is, however, an advantage of the system.

There is another meaning for C_{min}. In the normal mode of operation

there must be just this many bits in storage at each time of decision. In setting up the Markov states below, we do not count this irreducible storage, but it is included in the final results for total storage capacity.

We have defined the state of the buffer, y, as the number of bits stored at any instant of decision. With a capacity of C bits, the range of this variable is

$$0 \leqq y \leqq C + 1. \tag{8}$$

The normal state is $y = 0$; overflow is $y = C + 1$.

We can now write down the transition probabilities, p_{ij}, of going from buffer state y_i to state y_j. Starting in the zero or normal state, the buffer stays in the normal state with probability $1 - P_r$ and increases by I with probability P_r

$$p_{0,0} = 1 - P_r, \qquad p_{0,I} = P_r. \tag{9a}$$

If the buffer is within D states of normal, at the next decision it will either return to normal or will increase by I

$$p_{y,0} = 1 - P_r, \qquad p_{y,y+I} = P_r \quad \text{for} \quad 0 \leqq y \leqq D \tag{9b}$$

If the buffer is more than D states from normal and more than I states from overflow, it will decrease by D or increase by I, but can neither return to normal nor overflow

$$p_{y,y-D} = 1 - P_r, \quad p_{y,y+I} = P_r \quad \text{for} \quad D < y \leqq C - I. \tag{9c}$$

If the buffer is within I states of overflow, the buffer will either decrease by D or go to overflow

$$p_{y,y-D} = 1 - P_r, \quad p_{y,C+1} = P_r \quad \text{for} \quad C - I < y < C + 1. \tag{9d}$$

In order to calculate the time to overflow, we force the buffer to stay in the overflow condition once it enters this state; i.e., the overflow state is made "absorbing"

$$p_{C+1,\,C+1} = 1. \tag{9e}$$

For all other transitions $p_{ij} = 0$. The transition matrix is

$$T = \{p_{ij}\}. \tag{9f}$$

In addition, we let the process start in the normal state with probability 1. The buffer state, in response to the retransmission signal, depends only on the buffer state at the previous moment of decision. This is the fundamental property for a process to be a Markov chain.

A schematic representation of the Markov chain described by equa-

tions (9) is given in Fig. 3. The over-all operation of the transmitter may be seen in Fig. 4, which shows the internal state diagram of a sequential machine which might be used to implement the transmitter. The states of the sequential machine are the same as the states of the Markov process, except that several of the latter may map into a single one of the former.

The arrow labels — A,B/C,D — are identified as follows. In all cases, a dash means the item is immaterial.

Transmitter inputs:

A — Has a retransmission request been received?
    0 — no   1 — yes

B — What is the state of the buffer?
    0 — empty (except for C_{min})
    I — partially filled
    1 — over-filled

Transmitter outputs:

C — May a block be dropped from storage?
    0 — no   1 — yes

D — Which block shall be sent next?
    D_{n-1} was the block which was just sent. D_n is the next block in sequence, and D_{n-m} is the mth block before.
    F_1 and F_2 are fill-in bits. Note that if $F_1 = F_2$, two states may be combined.

Fig. 4 also applies to the receiver, except for reinterpretation of the labels.

Receiver inputs:

A — Has an error been detected?
    0 — no   1 — yes

B — State of receiving buffer
    0 — full
    I — intermediate
    1 — empty

Receiver outputs:

C — Shall this block be sent to output store?
    0 — no   1 — yes

D — Shall a retransmission request be sent?
    These will all be 0 except the two labelled D_{n-m} and D_{n-m-1}, which will be 1.

For certain relations among the quantities involved, the matrix can be partitioned into several closed sets[7] of states, such that it is not possible to make the transition from a state in any one closed set to a state in
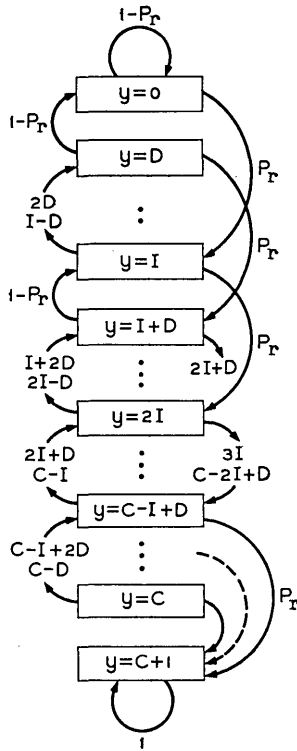
Fig. 3 — Markov state diagram.

any other such set. The states which cannot be entered from the normal state by any path may be removed from the matrix, thus reducing its size. This can be done by dividing out the greatest common factor in D, I, N, and C. A large number of the cases of interest are still included when this "normalizing factor" is made equal to D.

IV. CALCULATIONS

Following the method outlined in Kemeny and Snell,[8] we let Q be the transition matrix of all the transient states, i.e., matrix T, excluding the overflow state. Let J be the identity matrix. Then

$$G = (J - Q)^{-1} \tag{10}$$

exists and is called the fundamental matrix of the Markov process, with the following interpretation. Each element n_{ij} of G is the mean number
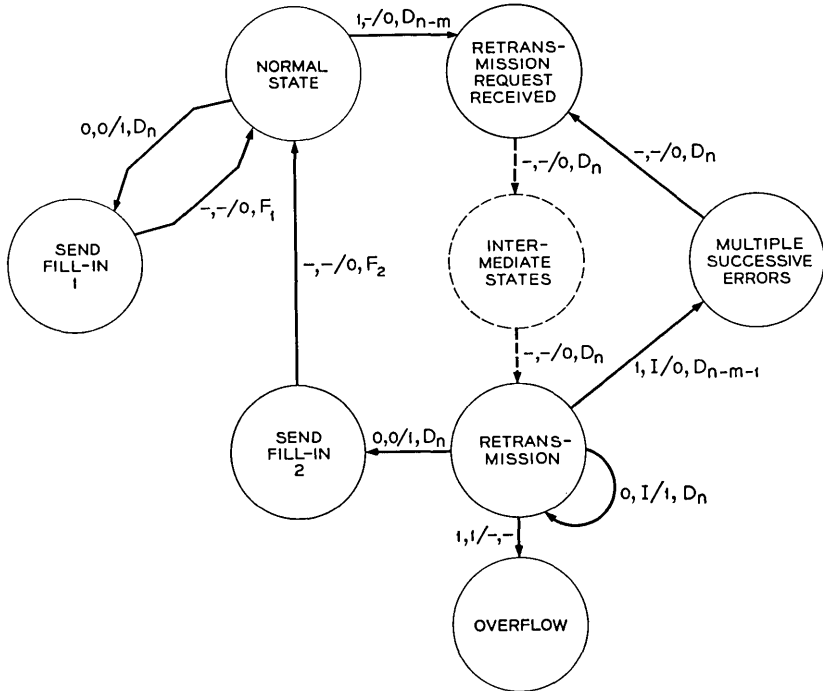
Fig. 4 — Diagram of internal states for transmitter.

of times the process is in state j, given that it started in state i. With $i =$ 0 for starting in the normal state, the row sum over j is the mean number of times the process is in any of the transient states, from which we can calculate the mean time to the first overflow. Thus, the average number of decisions before overflow is

$$\langle n \rangle = \sum_{j=0}^{c} n_{0j} . \tag{11}$$

Higher moments, in particular the second, can be found by additional operations on the fundamental matrix.[7]

A computer program was written to do the matrix arithmetic, and a few representative cases were solved numerically. The program computes the average number of blocks transmitted before overflow and the variance about this mean. The standard deviation is usually large, nearly equal to the mean. Typical examples are: when mean number of blocks before overflow was 23, standard deviation was 19; when mean was 949, standard deviation was 943; and when mean was 4795, standard

deviation was 4792. Thus the mean is a poor estimate of the actual time to overflow for any specific message, but is meaningful when a large number of transmissions are considered.

The calculations to this point have been in terms of the number of blocks, and we now convert back to actual time. Instead of a straight sum on n_{0j}, we multiply each term by the actual time taken.

There are four terms corresponding to the four cases described under the Markov process. The average time for each of the four cases is

$$(i) \ n_{00}(1 - P_r)T_D/k$$

$$(ii) \ n_{00}P_r(1 + E/k)T_D$$

$$(iii) \ \sum_{j=1}^{c} n_{0j}(1 - P_r)ET_D/k$$

$$(iv) \ \left[\sum_{j=1}^{c} n_{0j} - 1\right] P_r(1 + E/k)T_D \ .$$

The average time to overflow is the sum of these four:

$$\frac{t_{\text{ave}}}{T_D} = n_{00}\left(P_r + \frac{1 - P_r}{k} + \frac{EP_r}{k}\right) + \sum_{j=1}^{c} n_{0j}(P_r + E/k) \tag{12}$$
$$- P_r(1 + E/k).$$

V. RESULTS

As expected, the average time before the buffer overflows will increase when the buffer capacity is increased, and when the following variables are decreased: the bit rate, the round-trip delay, the probability of retransmission, the efficiency, and the block size. The number of variables can be reduced by measuring time in units of T_D, the round-trip delay, and bits in units of R_LT_D, the number of bits from the buffer in time T_D. Since the block error rate depends on the length of the block, the probability of retransmission is modified by the block length. The variables of the system, all of which are now dimensionless, become

$$C^* = C/R_LT_D$$
$$N^* = N/R_LT_D$$
$$E$$
$$P^* = P_rR_LT_D/N$$
$$t^* = t/T_D$$

A number of curves are plotted to show the expected time between overflows as a function of the probability of retransmission. For each curve, the size of the buffer, the block size, and the efficiency are held constant. When the expected time, t/T_D, is greater than about 100 (corresponding to several seconds of transmission for reasonable values of T_D), the curves are nearly linear on log-log paper, and only this portion is plotted.

To use the curves, it is assumed that the transmission line parameters, R_L and T_D, are known. In order to facilitate interpretation of the curves, some reasonable specific values have been assigned to these parameters and the corresponding values of time, buffer size, and block length have been calculated. The assignments are as follows: Let R_L be 2000 bits per second; this could be a 2400 bps data set with an $83\frac{1}{3}$ per cent efficient error-detecting code. Let T_D be 120 ms. Then $R_L T_D = 240$ bits, the total number of bits sent in one round-trip delay time. Some other parameters are given in Table I.

Fig. 5 shows the time gained by increasing the capacity of the buffer store. For this set of curves the efficiency is 0.5 and the block length is $0.5 R_L T_D$; that is, the block is as long as the maximum round-trip delay. When the efficiency is increased to 0.75 and 0.9, with the same block length $(0.5 R_L T_D)$, the results are as shown in Figs. 6 and 7, respectively. The storage capacity required to provide a specified time to overflow at a given probability of retransmission increases markedly with efficiency. The same effect is shown in Fig. 8, where the capacity is held constant for several efficiencies. The source bit rate at $E = 0.75$ is 50 per cent greater than at $E = 0.5$, and at $E = 0.9$ the bit rate is up by 80 per cent. The cost of this increased bit rate is either the extra buffer storage or the reduced time between overflows. Some of the data from Figs. 5–8 are

TABLE I — OPERATING PARAMETERS
(Given that $R_L = 2000$ b/s and $T_D = 120$ msec)

E	R_S (bits/sec)	N (bits)	I (bits)	C_{min} (bits)
0.5	1000	20	130	140
		120	180	240
0.75	1500	20	195	200
		120	270	300
		180	315	360
0.9	1800	20	234	236
		120	324	336
		216	410	432

TIME TO OVERFLOW IN HOURS
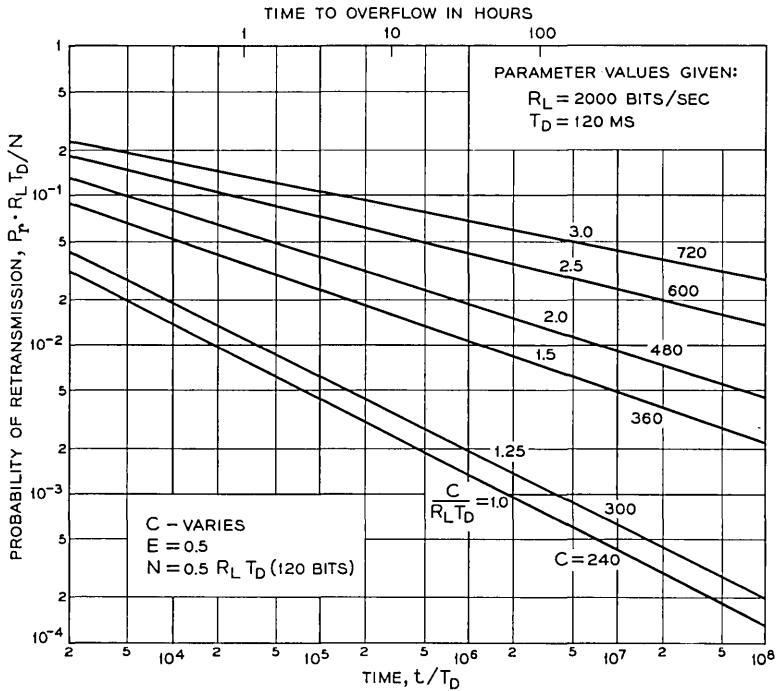
Fig. 5 — Effect of buffer size: $E = 0.5$.

shown in Table II, using the arbitrary assignments $R_L = 2000$ b/s, $T_D = 120$ ms, and $N = 120$ bits.

In all the above cases, the block lengths have been the same, $0.5\,R_L T_D$ (120 bits). Only when the efficiency is 0.5 does this represent the so-called "natural" block, i.e., the number of bits from the source in one round-trip delay time; at the increased bit rates of the higher efficiencies, the natural block length is also increased. The effect of increasing the block length in one case is shown in Fig. 9, which can be compared to Fig. 6. The required capacity for a given time to overflow has increased markedly. We therefore investigate the effects of shorter blocks.

Fig. 10 illustrates the case where each natural block is divided into three shorter blocks. A decision is made at the end of each arrow, and the fourth block back is either dropped from the buffer or is retransmitted. For example, when a retransmission is received while sending B_3, both A_1 and A_2 have been dropped and A_3 is the next block to be sent. With sufficiently inexpensive logic in the terminals, improved per-
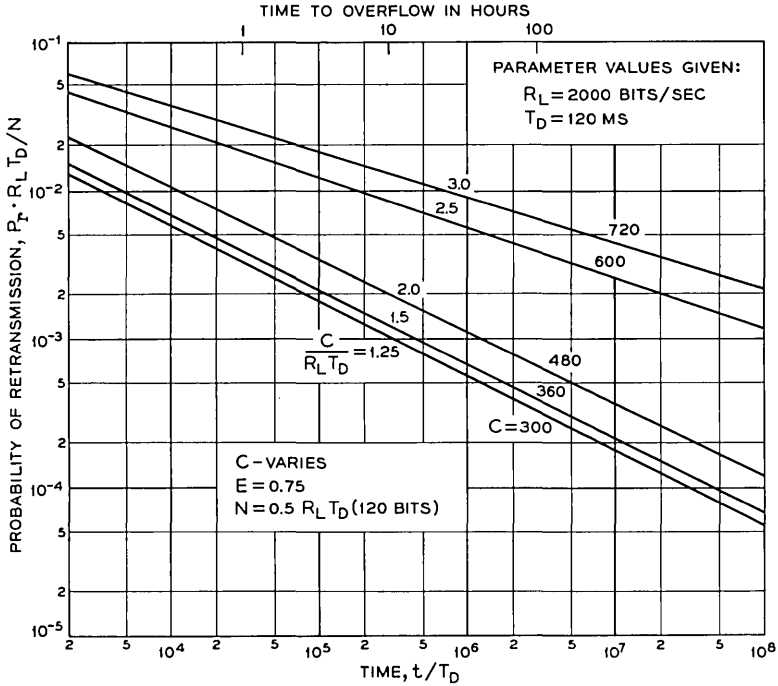
Fig. 6 — Effect of buffer size: $E = 0.75$.

formance is possible on short loops by using the actual value of T_D. In the example, we might have already dropped A_3 and therefore start the retransmission with B_1.

In Fig. 11 we show the effect of decreasing the block size, at constant capacity and efficiency. Similar results for a larger capacity and efficiencies of 0.5 and 0.75 are shown in Fig. 12.

It is somewhat difficult to visualize all of these effects when plotted separately. We attempt to summarize some of the results in Fig. 13. For these curves the normalized retransmission probability, P^*, is held constant, and buffer storage capacity is held to the minimum usable value, as given by (7); that is, the capacity is the natural block length plus the actual block length, and therefore decreases with either the block size or the efficiency. Both the latter are allowed to vary and we show the effect on the time to overflow.

There is little effect from changing the block size — except on the buffer capacity. One would therefore choose the smallest practical block.
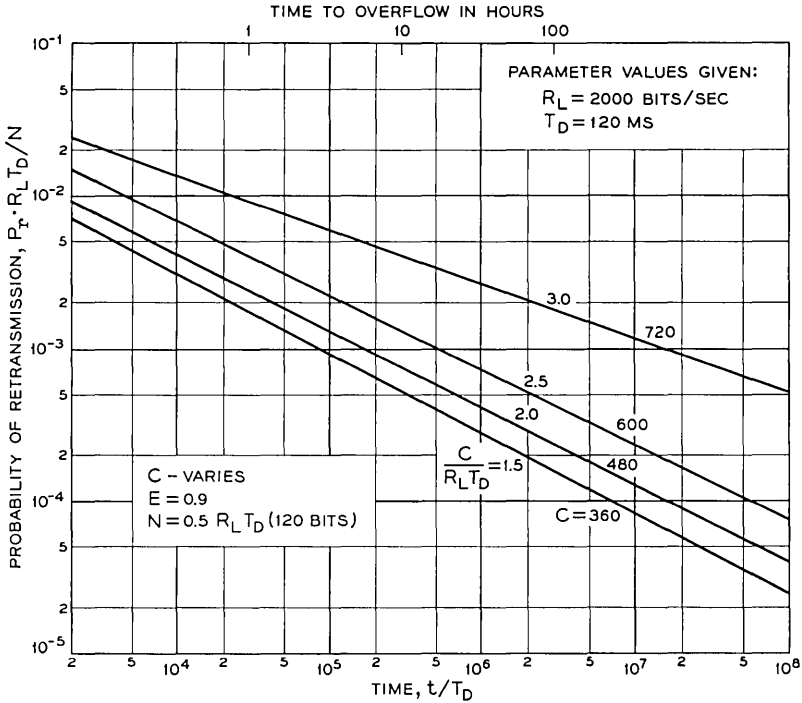
Fig. 7 — Effect of buffer size: $E = 0.90$.

However, as the efficiency is increased, the required buffer capacity is increased, although not rapidly, and the time between overflows decreases. As shown earlier (Figs. 5–7, 9) it is possible to regain this loss in time to overflow by modest increases in buffer capacity over the minimum used here. Since the increased efficiency increases the maximum source rate, this is certainly the direction to go, up to the point where the increased rate is worth less than the cost of the additional storage required.

VI. DELAY

For smooth flow to the sink the receiving buffer must have the same capacity as the transmitting buffer, and will normally be kept full. Thus the receiving buffer will introduce a delay in the message of

$$\tau = C/R_s .\qquad(14)$$

This is in addition to the delay of $T_D/2$ from the transmission line.
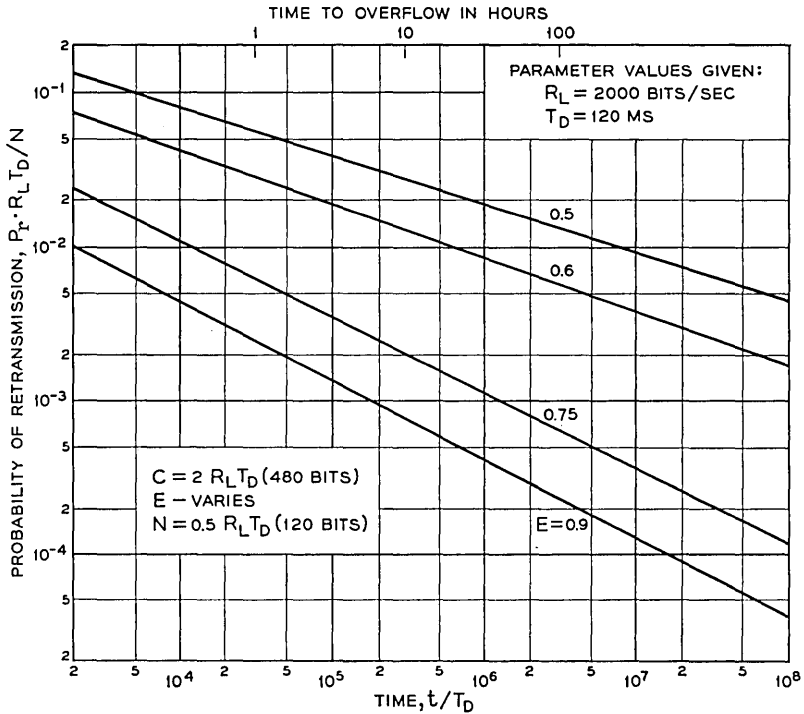
Fig. 8 — Effect of efficiency: buffer capacity fixed.

TABLE II — MEAN TIME TO OVERFLOW

(Given that $R_L = 2000$ b/s, $T_D = 120$ msec, $N = 120$ bits)

$E = R_S/R_L$	C (bits)	Ave. Time to Overflow (Hours)	
		$P^* = 0.01$	$P^* = 0.001$
0.5	C_{min} (240)	0.67	66.6
0.75	C_{min} (300)	0.12	11.2
0.9	C_{min} (336)	0.03	2.90
0.5	360	44.4	>1 year
0.75	360	0.15	14.9
0.9	360	0.04	3.14
0.5	480	245.3	>1 year
0.75	480	0.42	44.1
0.9	480	0.06	5.32
0.5	600	>1 year	>1 year
0.75	600	6.29	>1 year
0.9	600	0.15	17.93

TIME TO OVERFLOW IN HOURS
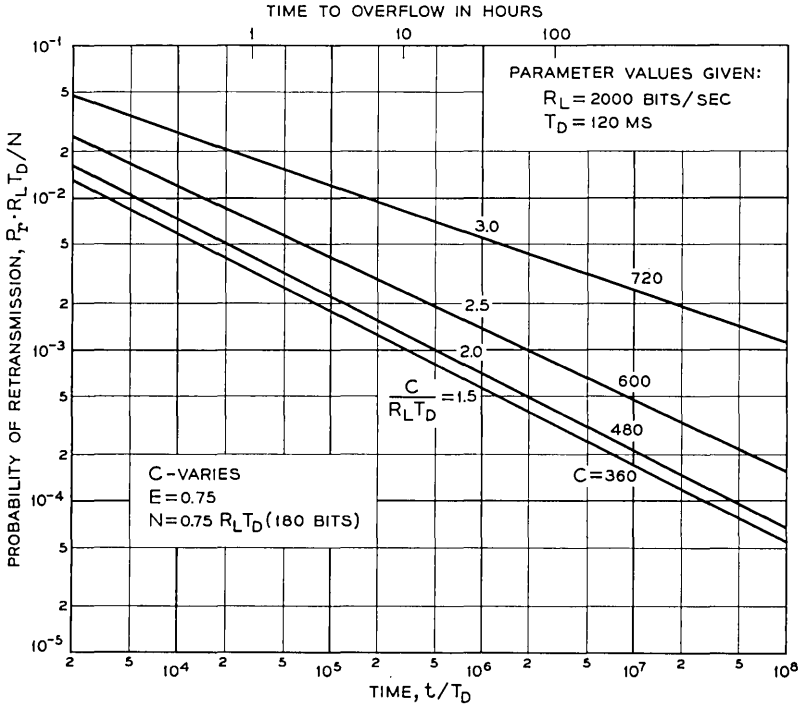
Fig. 9 — Effect of buffer size: longer block, $E = 0.75$.

SOURCE MESSAGE (RATE R_S)

| A_1 | A_2 | A_3 | B_1 | B_2 | B_3 | C_1 | C_2 | C_3 |

N BITS

EFFICIENCY, $E = R_S/R_L$

BLOCK SIZE, $N = R_S T_D/k$, FOR $k=3$

RETRANSMISSION REQUEST

TRANSMITTED MESSAGE (RATE R_L)

| A_1 | A_2 | A_3 | B_1 | B_2 | B_3 |

T_D

T_D

T_D

T_D

T_D
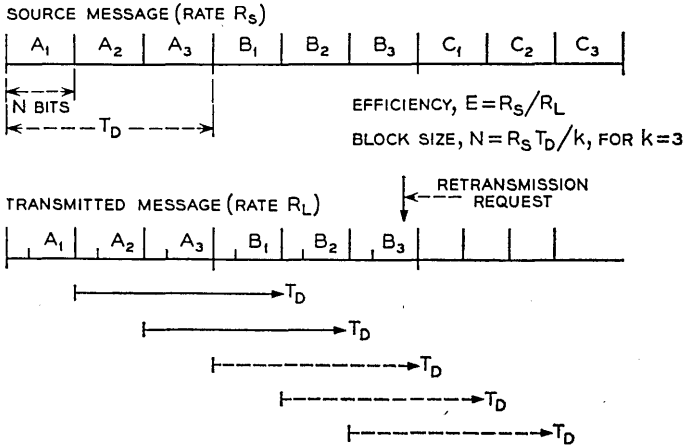
Fig. 10 — Example of time sequence with shorter blocks.
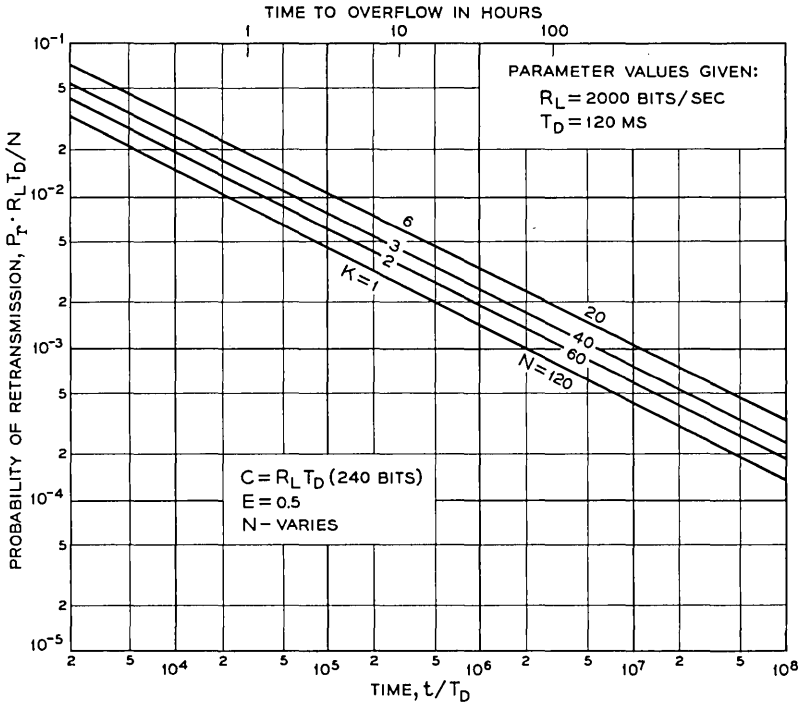
Fig. 11 — Effect of block size: fixed buffer capacity, $E = 0.5$.

There are other choices for operating the receiving buffer which will decrease the delay at the expense of irregularity of flow to the sink, which may be tolerable in many cases. If there were no receiving buffer at all, the delay would be zero except when retransmissions were required. When retransmissions are required, however, there would be additional delay until the block is received correctly, up to a maximum given by (14). The flow to the sink would not be smooth; each block would be delivered at rate R_L, followed by an interval when no data are being delivered. Various compromises between these extremes are possible. For example, buffer capacity of a single block would permit data to be delivered to the sink at the source rate with no interruptions until a retransmission is requested. Then the sink must alternately wait and accept data at the higher line rate until the process returns to normal. The delay is variable, the minimum being

$$\tau = N/R_s \tag{15}$$

with the maximum again given by (14).
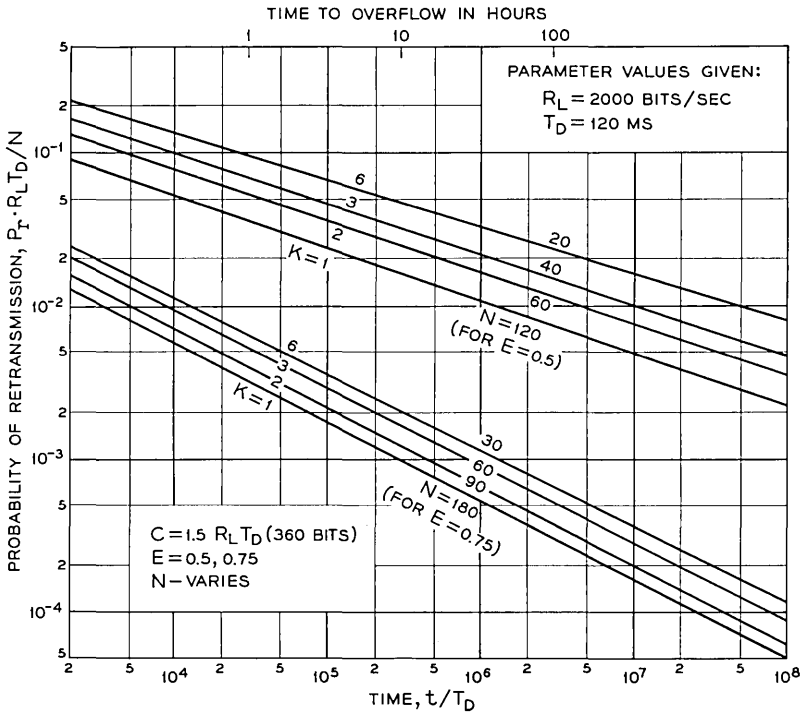
Fig. 12 — Effect of block size: fixed buffer capacity, $E = 0.5$ and 0.75.

This is one case where it has been possible to develop a calculable relationship between the message delay involved in error control and the resulting error rate.

VII. OTHER MODIFICATIONS

The system may be designed to take any of several actions when an overflow of the buffer occurs. The source and sink may be stopped, requiring manual resetting; they may be temporarily halted for a time sufficient for the system to clear; or, without stopping the source, the uncorrected data block may be delivered to the data sink, with or without an indication that the particular block contains errors.

One desirable modification would be to act sooner on receipt of the retransmission request. The transmitter would not continue to the end of the current block, but would immediately back up to the beginning of the block in error. This procedure could be quantized by using blocks a fraction of N in length. As indicated above, this procedure would require
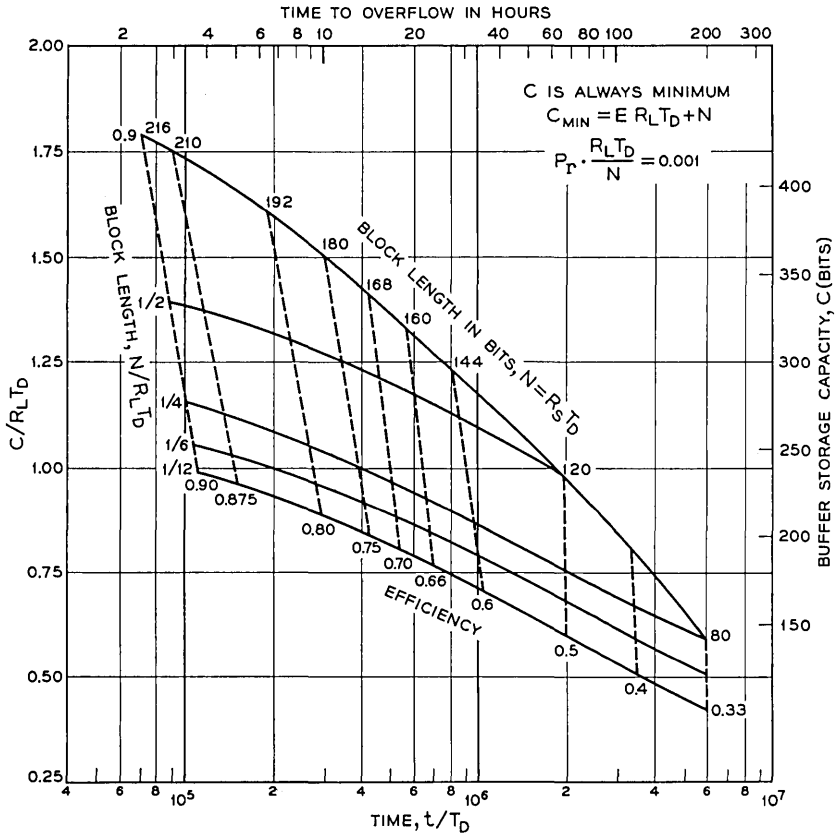
Fig. 13 — Capacity and time to overflow as functions of efficiency and block length.

either a knowledge of the actual round-trip delay or inclusion, in the retransmission requests and the retransmission, of an indication of the exact block (or fraction) involved. Another modification which would improve performance on shorter loops would be to make a preliminary measurement of the round-trip delay and adjust the operation accordingly. This could be done automatically.

Earlier, we mentioned the problem of an irregular input sequence and indicated that one additional block of storage was necessary. If this block is not counted, the performance level will be as given for a regular source, except for the possible gain arising from the probability of the intermittent source being stopped during the time when retransmissions are re-

quired. The output will be delayed an additional time corresponding to one block of data, but will be smoothed considerably — the rate will be constant except when waiting for the source.

It has been assumed that, once an error is detected, all subsequent received data are ignored until that block has been retransmitted and received correctly. With more complicated bookkeeping it would be possible to save some of these blocks, reducing the amount of retransmission required. On the other hand, since errors do occur in bursts on many transmission channels, the immediately succeeding block would have a higher-than-average error rate, and so might not be worth saving.

VIII. CONCLUSIONS

It has been shown that it is possible to calculate the performance of a self-contained error-control system by treating the system as a Markov process when the system consists of (a) an error-detection code, (b) provision for requesting and accomplishing retransmissions as necessary, and (c) buffer storage to allow smooth, uninterrupted flow from the source to the sink. Failure occurs when a sufficient number of retransmissions are requested in a short enough time that the total information to be stored exceeds the capacity of the buffer.

Whenever an overflow is about to occur, we could ignore the retransmission request and deliver the block as-is, in which case it appears to the sink as an error. It seems reasonable to require that this type of error should have about the same frequency of occurrence as undetected errors. For voice channels using reasonably simple codes, we might assume an undetected error rate of 10^{-8} or about one error per day.[1,3] We might also require the efficiency to be about that of the error detecting code.

With these criteria, it appears clear that one should not try to work with minimal storage, because of the relatively short time to overflow. Neither should one try to push the efficiency very high, or the required capacity grows out of bounds. A reasonable compromise for voice channels would be a buffer capacity somewhat less than 1000 bits.

We get a slightly different answer if we consider instrumentation. It is likely to be economically infeasible to build a buffer of this size with individual bit storage devices, especially since serial access is adequate. However, with bulk storage such as a circulating delay line or a magnetic tape loop, moderate increase in buffer size is not costly, and several thousand bits would be available about as cheaply as a few hundred. This would permit buffer efficiencies close to unity.

Results for any other specific cases can be easily calculated with this

computer program. It is apparent that a number of modifications in the model are possible and would serve to reduce the required storage. The transition matrix would merely have to be changed to correspond to the new model; the matrix arithmetic would be the same.

The details of the chosen model and the examples were taken from a specific data transmission problem. The techniques, both the model and the method of solution, are applicable to a wider variety of problems where buffering is a consideration.

We should like to acknowledge the assistance of H. O. Burton in consultation on the mathematics of the Markov process. We appreciate the continued encouragement of G. W. Gilman, who suggested the use of feedback error control with a data source which cannot be interrupted.

REFERENCES

1. Bennett, W. R., and Froehlich, F. E., Some Results on the Effectiveness of Error-Control Procedures in Digital Data Transmission, I.R.E., Trans. Comm. Syst., **CS-9**, March, 1961, pp. 58–65.
2. Schwartz, L. S., Some Recent Developments in Digital Feedback Communication Systems, I.R.E. Trans. Comm. Syst., **CS-9**, March, 1961, pp. 51–57.
3. Cowell, W. R., and Burton, H. O., Computer Simulation of the Use of Group Codes with Retransmission on a Gilbert Burst Channel, Trans. A.I.E.E. (Comm. & Elect.), No. 58, January, 1962, pp. 577–585.
4. Brown, A. B., and Meyers, S. T., Evaluation of Some Error Correcting Methods Applicable to Digital Data Transmission, 1958 I.R.E. Natl. Conv. Record, Pt. 4, March, 1958, pp. 37–55.
5. Reiffen, B., Schmidt, W. G., and Yudkin, H. L., The Design of an Error-Free Data Transmission System for Telephone Circuits, Trans. A.I.E.E. (Comm. & Elect.), No. 55, July, 1961, pp. 224–231.
6. Fontaine, A. B., Queueing Characteristics of a Telephone Data Transmission System with Feedback, A.I.E.E. Conference Paper 62-1441, Fall General Meeting, October 9, 1962.
7. Feller, W., *An Introduction to Probability Theory and its Applications*, Vol. I, 2nd ed., John Wiley and Sons, New York, 1957.
8. Kemeny, J. G., and Snell, J. L., *Finite Markov Chains*, D. Van Nostrand Co., New York, 1960.

Intermodulation Distortion in Analog FM Troposcatter Systems

By E. D. SUNDE

In broadband transmission over troposcatter paths, selective fading will be encountered with resultant transmission impairments, depending on the modulation method. An analysis has been made in a companion paper of such selective fading, based on an idealized model of troposcatter paths. It indicated that selective fading will be accompanied by phase nonlinearity which in a first approximation can be regarded as quadratic over a narrow band. A probability distribution for such quadratic phase distortion was derived. On the premise of quadratic phase distortion, the error probability owing to selective fading was determined for digital transmission by various methods of carrier modulation.

The same idealized model and basic premise of quadratic phase distortion is used here to determine intermodulation distortion in FM for a signal with the statistical properties of random noise. An approximate expression for intermodulation noise owing to specified quadratic phase distortion has been derived, applying for any method of frequency pre-emphasis in FM. In turn, median intermodulation noise as well as the probability distribution of intermodulation noise has been determined, as related to certain basic system parameters.

A comparison is made of predicted with measured intermodulation noise in four troposcatter systems with lengths from 185 to 440 miles. The results indicate that phase nonlinearity owing to selective fading can be approximated by quadratic phase distortion, or linear delay distortion, over an appreciable part of the transmission band ordinarily considered for troposcatter systems, with a probability distribution that can be determined from certain basic parameters of troposcatter links, such as the length and antenna beam angles. However, to predict intermodulation distortion on any system, further experimental data than are now available are required on beam broadening by scatter.

The present random multipath FM distortion theory is shown to afford

a significant improvement over an equivalent single-echo theory that has been applied on an empirical basis to troposcatter systems.

INTRODUCTION

An analysis has been made elsewhere[1] of error probabilities in high-speed digital transmission over idealized troposcatter paths, considering both random noise and intersymbol interference owing to pulse distortion caused by selective fading. The above analysis indicated that a principal cause of intersymbol interference is a quadratic component of phase distortion, or linear delay distortion. On the same basic premise an evaluation is made herein of intermodulation noise in analog transmission by frequency modulation, as now used for transmission of voice channels in frequency division multiplex. Expressions and curves are given of intermodulation noise in an idealized troposcatter channel for a signal with the properties of random noise, as related to certain basic system parameters and comparisons are made with the results of measurements on four troposcatter systems.[2,3]

In random multipath transmission the received wave can be considered the sum of a plurality of echoes, arriving over the various paths with varying amplitudes and different delays. Although this view is conceptually simple, it does not facilitate analysis of the statistical properties of the received signal and of signal distortion. In the combination of a number of time functions, such as echoes, the analysis is greatly facilitated by the use of Fourier transformation to determine the corresponding spectra. The latter can in turn be combined directly with appropriate attention to phase relations to obtain the resultant wave. For this reason it is preferable from the standpoint of analysis to regard the received wave as a multiplicity of sine wave components, rather than signal wave echoes, arriving over the plurality of transmission paths with varying amplitudes and phases. This is the method ordinarily used in the analysis of the statistical properties of narrow-band random noise, which has properties that with appropriate translation of the basic parameters are also applicable to random multipath transmission. It is the method underlying both the previous determination of error probabilities in digital transmission owing to noise and selective fading, and the present analysis of intermodulation noise in FM.

In certain radio systems the received wave can be considered the sum of a principal signal wave and a weaker echo, and comprehensive theoretical analyses have been published of intermodulation noise in FM owing to such echo distortion,[4,5,6] together with the results of simulative tests.[7] For these reasons this two-path model has been adopted as a

coarse simile to multipath transmission in some interpretations of the result of measurements of intermodulation noise in troposcatter systems.[3] The limitations of this simile are recognized in the latter publication,[3] in which it is suggested that a more refined analysis is desirable. The idealized multipath model used in the analysis of troposcatter digital transmission affords a significant improvement, though it has certain predictable limitations, as shown herein.

I. TRANSMITTANCE PROPERTIES OF TROPOSCATTER LINKS

In tropospheric transmission beyond the horizon the received wave can be considered the sum of a large number of components of varying amplitudes resulting from a multiplicity of reflections within the common volume of the antennas. Owing to variations in the structure of the common volume, caused largely by winds, there will be relatively slow changes in the many reflections and thus in the amplitudes of the component waves. When a steady-state sine wave is transmitted, the received wave will thus exhibit random variations in its envelope and phase, known as fading.

In addition to such transmittance variations with time at a particular frequency, there will be transmittance variations with frequency at any given instant, as illustrated in Fig. 1. At a given instant the amplitude and phase characteristics of the transmission path may be as indicated in Fig. 1(a) and at a later instant as in Fig. 1(b).

Let $u = \omega - \omega_0$ represent the radian frequency relative to a reference frequency ω_0. When the transmission vs frequency characteristic of a troposcatter channel varies slowly with time t, it can be represented by
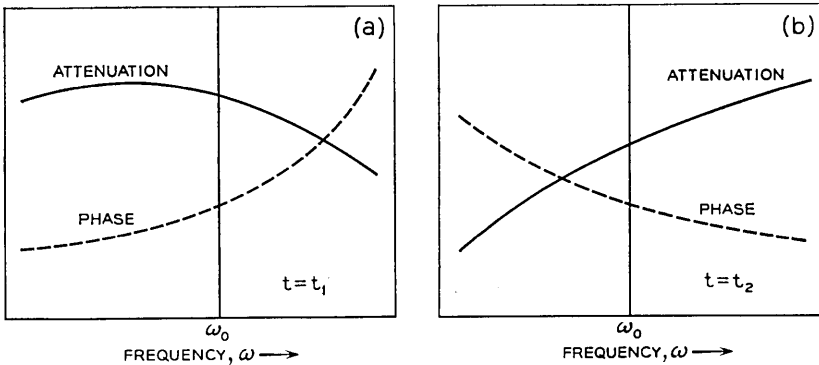

Fig. 1 — Illustrative variations in attenuation and phase characteristics with frequency at two instants t_1 and t_2.

$$T(u,t) \; = \; A(u,t)e^{-i\varphi(u,t)} \tag{1}$$

where

$A(u,t) \;=\;$ amplitude characteristic as a function of t for a fixed
     u, or as a function of u for a fixed time t

$\varphi(u,t) \;=\;$ phase characteristic.

If $u \,=\, u_0$ is fixed, both $A(u_0\,,\,t)$ and $\varphi(u_0\,,\,t)$ are random variables of the time t, as are the time derivatives $A'(u_0\,,\,t)$, $A''(u_0\,,\,t)$, $\varphi'(u_0\,,\,t)$, $\varphi''(u_0\,,\,t)$. The probability distributions of $A(u_0\,,\,t)$ and $\varphi(u_0\,,\,t)$ can be determined on the premise that they are the sum of a large number of randomly phased components. This results in a Rayleigh probability distribution of $A(u_0\,,\,t)$, in conformance with observations of rapid fading. To determine the probability distributions of A', A'', φ' and φ'', statistical information is required regarding the rapidity of fades. This ordinarily takes the form of the time autocorrelation functions of $A(t)$, or the related power spectrum of changes in transmittance amplitude. Such power spectra can be characterized by a certain equivalent fading bandwidth.

If the time is assumed fixed at $t \,=\, t_0$, then $A(u,t_0)$ and $\varphi(u,t_0)$ will have certain random fluctuations with the frequency u that can be characterized by probability distributions. This also applies to $\dot{A}(u,t_0)$, $\ddot{A}(u,t_0)$, $\dot{\varphi}(u,t_0)$, and $\ddot{\varphi}(u,t_0)$, where the dots indicate differentiation with respect to frequency u. The probability distributions of $\ddot{A}$, $\dot{\varphi}$, and $\dot{A}$ and $\ddot{\varphi}$ depend on the frequency autocorrelation functions, or the corresponding power spectra of variations with frequency. The latter depend on differences in transmission time over the various paths, and can be related to the maximum departure Δ from the mean transmission delay.

The amplitude and phase characteristics as a function of u at any time t_0 can in general be represented by a power series as

$$A(u,t_0) \; = \; a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \cdots \tag{2}$$

$$\varphi(u,t_0) \; = \; b_0 + b_1 u + b_2 u^2 + b_3 u^3 + \cdots . \tag{3}$$

Certain basic relations have been developed by Carson and Fry[7] and by van der Pohl,[8] for transmission impairments in FM resulting from attenuation and phase distortion. With the aid of these relations it can be shown that intermodulation noise is caused principally by phase distortion rather than by amplitude distortion. Moreover, it can be shown that the principal contributor is quadratic phase distortion represented by $b_2 u^2$, which corresponds to linear delay distortion $2b_2 u$.

II. PROBABILITY DISTRIBUTION OF QUADRATIC PHASE DISTORTION

From (3) it follows that

$$\ddot{\varphi}(u,t_0) \;=\; 2b_2 + 6b_3 u + \cdots. \tag{4}$$

For $u = 0$, i.e., at the reference or carrier frequency, the probability distribution of b_2 is the same as that of $\ddot{\varphi}(0,t)$. The latter probability distribution has been determined elsewhere[1] on the approximate premise of a linear variation in transmission delay, with maximum departures $\pm\Delta$ from the mean delay. In Fig. 2 is shown the probability that $\ddot{\varphi}$, or $2b_2$, exceeds $\Delta^2/3$ by a factor k. For example, there is a probability $p = 0.5$ that $\ddot{\varphi}$ exceeds $\Delta^2/3$ by a factor $k \approx 1.2$, and a probability $p = 0.1$ that $\ddot{\varphi}$ exceeds $\Delta^2/3$ by a factor $k \approx 19$.

Thus in general

$$\ddot{\varphi}_p \;=\; 2b_2(p) \;=\; k_p \Delta^2/3 \tag{5}$$

where $k_p\Delta^2/3$ is the value of $\ddot{\varphi}$, or $2b_2$ with a probability p of being exceeded.

Alternatively, the value of b_2 with a probability p of being exceeded is

$$b_2(p) \;=\; \frac{k_p}{6}\,\Delta^2. \tag{6}$$

Thus

$$b_2(0.5) \;\approx\; \frac{1.2}{6}\,\Delta^2 = 0.2\Delta^2 \tag{7}$$

$$b_2(0.1) \;\approx\; \frac{19}{6}\,\Delta^2 = 3.2\Delta^2 \tag{8}$$

$$b_2(0.01) \;\approx\; \frac{400}{6}\,\Delta^2 = 67\Delta^2. \tag{9}$$

Thus, when Δ is known, together with intermodulation noise for quadratic phase distortion, it is possible to determine the median value of average intermodulation noise, or the value exceeded with any other specified probability p.

III. INTERMODULATION NOISE FROM QUADRATIC PHASE DISTORTION

In a first-order evaluation of intermodulation noise, only the quadratic term $b_2 u^2$ in (3) would be considered, since it will be the principal contributor. The ratio of nonlinear distortion power to average signal power at the frequency ω will depend on the signal properties and on the pre-
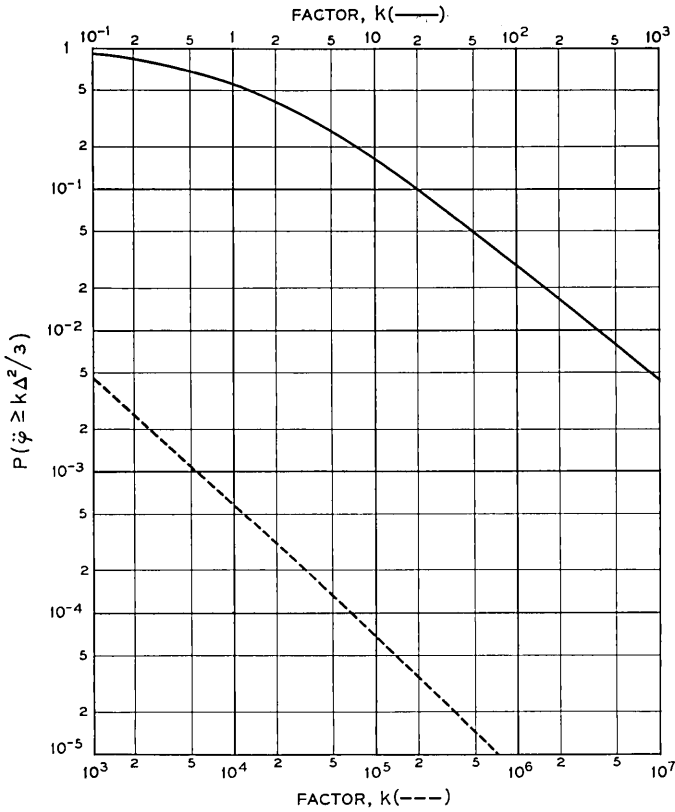
Fig. 2 — Probability that $\ddot{\varphi}$ or $2b_2$ exceeds $\Delta^2/3$ by a factor k.

emphasis used in frequency modulation. It will be assumed that the original message wave has a flat power spectrum of radian bandwidth $\Omega = 2\pi B$ and the statistical properties of random noise, and furthermore that the message wave is passed through a transmitting filter with a power transfer characteristic

$$t(\omega) = 1 + c(\omega/\Omega)^2$$
$$= 1 + c(f/B)^2. \tag{10}$$

At the receiving end a complementary filter is used to restore the message wave.

As discussed in the Appendix, exact determination of intermodulation noise from quadratic phase distortion presents formidable difficulties, except on the premise of slight phase distortion, which is not generally applicable to troposcatter systems. However, it is possible to obtain an

approximate solution without the above limitation. The following relation is derived in the Appendix for the ratio $\rho(f)$ of intermodulation noise to average signal power at the frequency $f = \omega/2\pi$

$$\rho(f) = \frac{B^2}{D^2} G(c,a) H(\gamma) \tag{11}$$

where c is defined by (10)

$$a = f/B = \omega/\Omega$$
$$B = \text{bandwidth of baseband signal} = \Omega/2\pi$$
$$D = \text{rms frequency deviation} = \underline{\Omega}/2\pi$$

and

$$\gamma = b_2\underline{\Omega}^2 = (2\pi)^2 b_2 D^2. \tag{12}$$

The function $G(c,a)$ depends on the pre-emphasis and is given by expression (108) in the Appendix, which is

$$G(c,a) = \frac{3a^2}{(1 + ca^2)(3 + c)} F(c,a)$$

$$F(c,a) = 2 - a + \frac{2c + c^2a^2}{3} [1 + (1 - a)^3] \tag{13}$$

$$- \frac{c^2a}{2} [1 - (1 - a)^4] + \frac{c^5}{5} [1 + (1 - a)^5]$$

This function is shown in Fig. 3 for pure FM and PM and for $c = 16$. The particular case of $c = 16$ and $a = 1$ will be considered further in the following, and for this case

$$G(16,1) = 0.192.$$

The function $H(\gamma)$ is shown in Fig. 4 and represents an approximation, as discussed in the Appendix. It will be noted that this function departs from proportionality with γ^2 for $\gamma \geqq 0.5$, reaches a certain maximum value and then diminishes.

IV. INTERMODULATION NOISE IN TROPOSCATTER PATHS

In accordance with (6), the value of b_2 with a probability p of being exceeded is $b_2(p) = k_p\Delta^2/6$. The corresponding value of γ is given by (12) as

$$\gamma_p = \frac{k_p\Delta^2}{6} (2\pi)^2 D^2 \tag{14}$$

$$= 6.6k_p(\Delta D)^2.$$
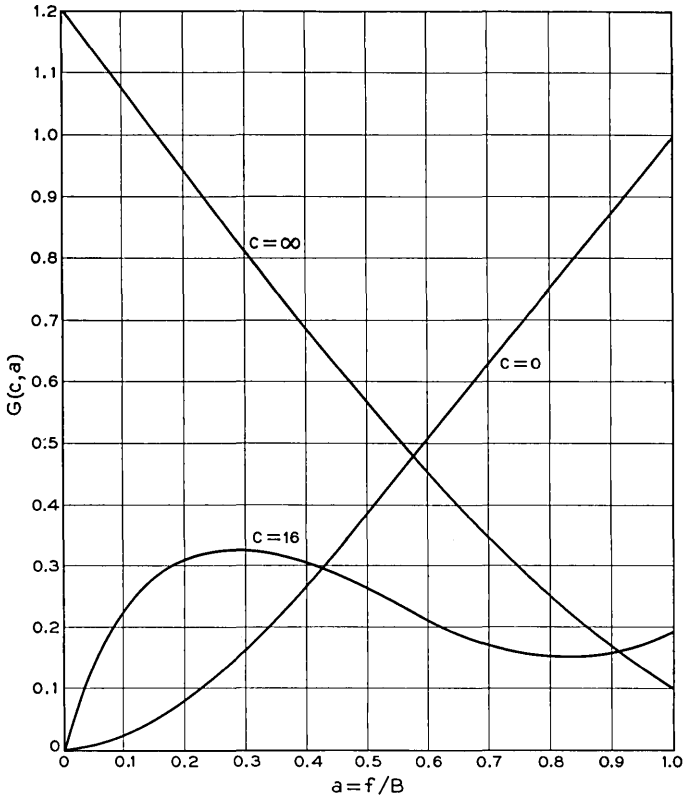
Fig. 3 — Function $G(c,a)$ for pure FM $(c = 0)$, pure PM $(c = \infty)$, and for pre-emphasized FM with $c = 16$.

Thus

$$\gamma_{0.5} \approx 8 \ (\Delta D)^2 \tag{15}$$

$$\gamma_{0.1} \approx 125 \ (\Delta D)^2 \tag{16}$$

$$\gamma_{0.01} = 2600 \ (\Delta D^2). \tag{17}$$

The corresponding ratios $\rho(f)$ at $f = B$ with a probability p of being exceeded

$$\rho_p(B) = 0.192 \left(\frac{B}{D}\right)^2 H(\gamma_p) \tag{18}$$

$$\rho_{0.5}(B) = 0.192 \left(\frac{B}{D}\right)^2 H(8\Delta^2 D^2) \tag{19}$$

$$\rho_{0.1}(B) = 0.192 \left(\frac{B}{D}\right)^2 H(125\Delta^2 D^2) \tag{20}$$

$$\rho_{0.01}(B) = 0.192 \left(\frac{B}{D}\right)^2 H(2600\Delta D^2). \tag{21}$$

V. DIFFERENTIAL TRANSMISSION DELAY Δ

Exact determination of the equivalent maximum departure from the mean transmission delay requires consideration of the antenna beam patterns as affected by scattering. On the approximate basis of equivalent antenna beam angles α, it follows from the geometry indicated in Fig. 5 that

$$\Delta \approx \frac{L}{v} \frac{\alpha + \beta}{2} \left(\theta + \frac{\alpha + \beta}{2}\right) \tag{22}$$

where $\beta \leqq \alpha$, v is the velocity of propagation in free space, L is the length of the link, and

$$\theta = \frac{L}{2R} = \frac{L}{2R_0 K} \tag{23}$$

where R_0 is the radius of the earth and the factor K is ordinarily taken as $4/3$.

The equivalent antenna beam angle α from midbeam to the 3-db loss point depends on the free-space beam angle α_0 and on the effect of scatter, which is related in a complex manner to α_0 and the length L, or alternatively θ. Narrow-beam antennas as now used in actual systems are loosely defined by $\alpha_0 \leqq 2\theta/3$. For these, $\alpha \approx \alpha_0$ on shorter links, while on longer links $\alpha > \alpha_0$ owing to beam-broadening by scatter. Analytical determination of α for longer links appears difficult, and only limited experimental data are available at present. For broad-beam antennas, $\alpha_0 \gg 2\theta/3$ and beam-broadening by scatter is in theory inappreciable.

By way of numerical example, let $L = 170$ miles and $K = 4/3$, in which case $\theta = 0.016$ radian. With $\alpha_0 = 0.004$ radian $\ll 2\theta/3$ it is permissible to take $\alpha = \alpha_0$. With $\beta = \alpha = \alpha_0$, (22) gives $\Delta = 0.08 \times 10^{-6}$ second.

The differential delay Δ in general varies with time and for narrow-beam antennas can be considered the sum of two components

$$\Delta(t) = \Delta_0 + \Delta_1(t) \tag{24}$$

where Δ_0 is a fixed component obtained from (22) by taking $\alpha = \alpha_0$, the free-space beam angles. The variable component $\Delta_1(t)$ depends on

$$\gamma = b_2 \, \Omega^2 = b_2 (2\pi D)^2$$
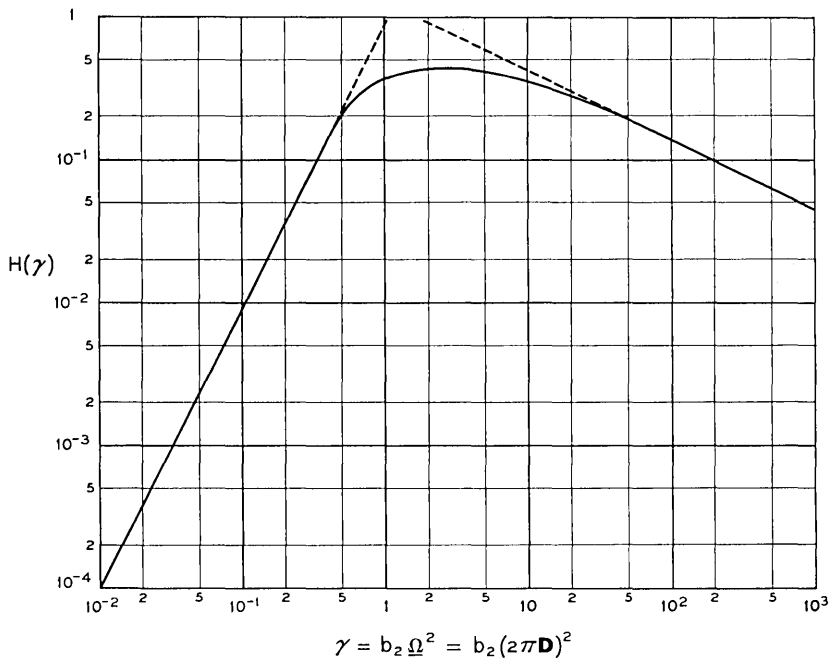
Fig. 4 — Function $H(\gamma)$. The parameter γ is the phase distortion in radians at a frequency corresponding to the rms frequency deviation $\Omega = 2\pi D$ radians/second.
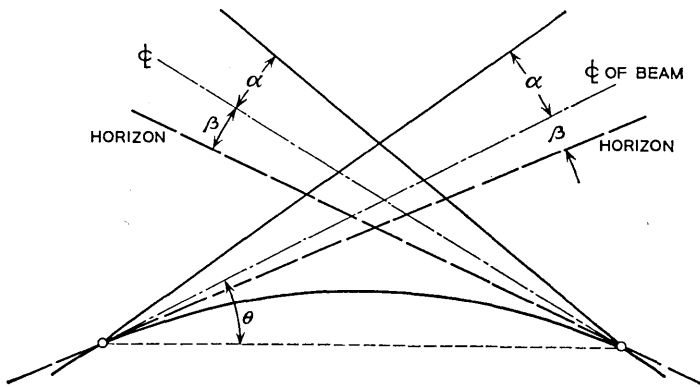

Fig. 5 — Definition of antenna beam angles α, take-off angle β and chord angle θ to midbeam. With different angles at the two ends, the mean angles are used in expressions for Δ.

scatter variation with time, as does path loss, and will have a certain correlation with path loss variations. Owing to the fixed component Δ_0, a weaker correlation exists between $\Delta(t)$ and path loss variations.

Because of the dependence of Δ on path loss, the ratio ρ_p of intermodulation noise to average signal power will depend somewhat on path loss However, for a given path loss ρ_p is independent of the average transmitter power and thus of the average signal power at the receiver.

VI. LIMITATIONS ON FIRST-ORDER DISTORTION THEORY

The above first-order approximation applies for sufficiently narrow signal bandwidths at the detector input such that terms in (3) of higher order than u^2 can be neglected. Results given by Rice for random variables (Section 3.4 of Ref. 10) indicate there is no correlation between $\dot\varphi$ and $\ddot\varphi$, so that distortion owing to the term $b_3 u^3$ will combine on a power addition basis with distortion resulting from $b_2 u^2$. Moreover, there is a negative correlation factor between $\dot\varphi$ and $\ddot\varphi$, so that on the average b_4 is negative whenever b_2 is positive, and conversely. Hence distortion produced by $b_4 u^4$ will on the average subtract directly on an amplitude basis from that resulting from $b_2 u^2$. In the range where the function $H(\gamma)$ increases linearly with γ^2, intermodulation noise owing to the term $b_2 u^2$ increases as $b_2^2 (\Delta D)^4$. In the same range, intermodulation noise from the term $b_4 u^4$ will vary as $b_4^2 (\Delta D)^8$ and may hence have a significant effect for adequately large values of ΔD even though b_4 be much smaller than b_2. As shown later, comparisons of measured intermodulation noise with predictions based on the above first-order theory indicate the increasing importance of the term $b_4 u^4$ in reducing intermodulation noise as ΔD is increased.

VII. TWO-PATH VS MULTIPATH DISTORTION THEORY

The above first-order distortion theory is a mathematically derived approximation that in principle yields valid results with appropriate limitations on signal bandwidth and frequency deviation, and which retains the multipath feature that is essential to this end. By contrast, the two-path or single-echo simile mentioned in the introduction has no such basis but has been adopted principally because of the convenience of available theoretical analysis.[4,5,6] A second reason is that single-echo distortion theory yields results that in some respects are quite similar to those obtained with multipath transmission, as shown below.

It is noteworthy that, by proper choice of echo amplitude and delay, results similar to those for median quadratic phase distortion can be

obtained. This is illustrated in Fig. 6, which shows the median ratio $\rho(B)$ obtained from (19) as a function of D for $B = 1$ mc/sec with $\Delta = 0.1$ and 0.5 microsecond. In the same figures are shown the ratios $\rho(B)$ obtained on the premise that the received wave consists of a main signal and an echo of equal amplitude delayed by 0.07 and 0.4 microsecond. The ratio $\rho(B)$ for the latter condition is obtained from a chart given in Fig. 9 of Ref. 3, applying for FM with virtually the same pre-
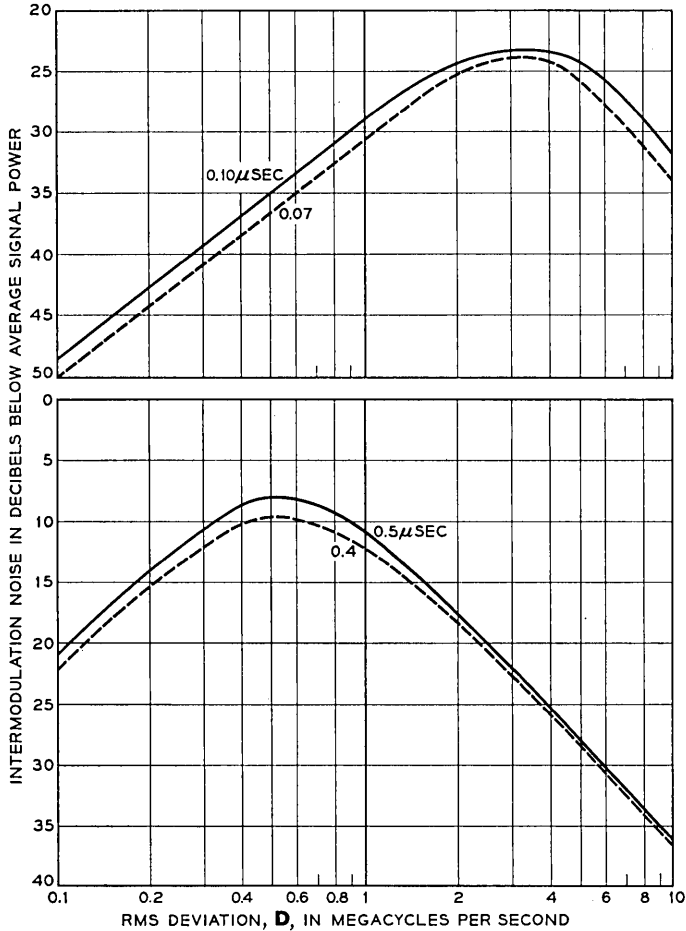

Fig. 6 — Comparison of intermodulation noise from single-echo distortion and quadratic phase distortion at $B = 1$ mc/sec: (solid lines) median intermodulation noise from quadratic phase distortion for indicated departures Δ from mean delay; (dashed lines) intermodulation noise from echo of same amplitude as signal with delays Δ, as indicated.

emphasis as assumed herein and given by (10). The above charts are based on echo distortion theory applying for echoes that are much weaker than the signal, but this premise is ignored here in extending the theoretical results to a fictitious echo of the same amplitude as the signal. In this connection it may be noted that simulative tests[9] indicate that intermodulation noise is nearly proportional to echo amplitude, even when the latter equals the signal amplitude. With both quadratic phase distortion and single-echo distortion, intermodulation noise is virtually proportional to the second power of signal bandwidth. Hence, the relative comparisons in Fig. 6 could also apply for other bandwidths than $B = 1$ mc/sec.

The above comparisons indicate that in applying equivalent single-echo FM distortion theory to multipath transmission as in troposcatter systems, with physically tenable echo delays, certain dilemmas will be encountered. The theory could be extended beyond its validity to fictitious echoes of the same amplitude as the signal, to obtain virtually the same median intermodulation noise as for quadratic phase distortion. This would exclude the possibility of greater intermodulation noise than the median value, since the greater echo is by definition the main signal. The other procedure would be to assume an echo that is smaller than the main signal, which is physically more acceptable and does not violate the basic premise underlying echo distortion theory. In this case intermodulation noise predicted on the basis of echo distortion theory would, at least in certain cases, be much smaller than actually observed and could not be made to conform with observations, unless the echo amplitude is increased to the same amplitude as the signal.

Thus, if the ratio of echo amplitude to signal amplitude is r, intermodulation noise power based on single-echo theory will be less than for multipath transmission by a factor r^2. Hence it becomes necessary to introduce a factor $1/r^2$ to make single-echo theory applicable to multipath transmission. In Ref. 3, this factor has been determined empirically from measurements to be discussed later, and is given as 9 db.

VIII. OBSERVED MEDIAN INTERMODULATION NOISE

Measurements have been made on four troposcatter links of the median value of intermodulation noise at the frequency $f = B$. The modulating wave in these tests had a flat power spectrum, and pre-emphasis was used that closely corresponded to $c = 16$ in (10).

The basic parameters of the systems on which the measurements were made are given in Ref. 3 and are summarized in Table I. In this table α_0 is the free-space antenna beam angle from midbeam to the 3-db

TABLE I — BASIC PARAMETERS OF TROPOSCATTER TEST
SYSTEMS IN CARIBBEAN (A) AND IN ARCTIC (B,C,D)

System	A	B	C	D
Length, miles	185	194	340	440
Radio frequency, mc	725	900	900	800
Antenna/diameter, ft	60, 60	30, 60	120, 120	120, 120
α_0 (radian)	0.0115	0.017	0.0058	0.0058
θ (radian)	0.015	0.016	0.031	0.034
Δ_0 (microsecond)	0.12	0.21	0.185	0.255

loss point, which may not conform with the angle α in (22) when scatter is considered. The values of K and θ are taken from Ref. 3, and θ differs slightly from that obtained from (23) owing to differences in antenna elevations. The take-off angle β is virtually zero and has been neglected. The value Δ_0 of Δ given in the table was calculated with $\alpha = \alpha_0$, rather than the actual beam angle with scatter. Systems A, B, C and D correspond to paths 1, 2, 4 and 3 in Ref. 3.

In Figs. 7 and 8 are shown the ratios ρ_i (B) expressed in db as a function of the rms frequency deviation D for different bandwidths B of the baseband signals.

IX. COMPARISON OF THEORETICAL WITH OBSERVED MEDIAN VALUES

In the same Figs. 7 and 8 are shown median values of intermodulation noise obtained from (19) for each case, based on values Δ_m of Δ that afford the best average approximation to the measurements. The latter values are somewhat greater than Δ_0, as indicated in Table II.

A ratio Δ_m/Δ_0 or $\alpha_m/\alpha_0 > 1$ is to be expected owing to beam-broadening by scatter, and the above ratios appear reasonable in the light of present knowledge. Thus, if the actual angles α were known so that Δ could be determined, it appears plausible that satisfactory conformance with observed intermodulation noise would be obtained.

As noted in Section V, Δ includes a component $\Delta_1(t)$ that varies with time depending on scatter conditions and which is correlated with path loss fluctuations. The ratio ρ thus depends on path loss as affected by scatter and has a certain correlation with path loss variation, as shown elsewhere.[3] Hence, if measurements had been made under different path loss conditions, the derived values Δ_m would have been somewhat different.

From Figs. 7 and 8 it will be noted that with the above choice of $\Delta = \Delta_m$ it is possible to obtain better agreement between predicted and observed intermodulation noise for small bandwidths B of the baseband

Fig. 7 — Comparison of measured and calculated median intermodulation noise: (dashed curves) measured median intermodulation noise in top channels at indicated frequencies in kc; (solid curves) calculated median intermodulation noise for idealized model with the following values of the equivalent maximum deviation Δ from the mean transmission delay: system A, $\Delta_m = 0.12$ microsecond ($\Delta_0 = 0.12$); system B, $\Delta_m = 0.25$ microsecond ($\Delta_0 = 0.12$).

signal and small deviations D than for large bandwidths and frequency deviations. This probably resides in the circumstance that the phase distortion terms of higher order than $b_2 u^2$ have been neglected in the above first-order theory, as discussed in Section VI.

The measured median ratios given in Figs. 7 and 8 are plotted in Fig. 9 against the ratios predicted by first-order theory. It will be noted that measured intermodulation noise is less than predicted for signal-to-interference ratios less than about 30 db, owing to reduction in intermodulation noise by phase distortion of higher order than $b_2 u^2$ that has been neglected in first-order theory. The results in Fig. 9 permit an approximate empirical correction to first-order theory.

As discussed in Section VII, with single-echo distortion theory virtually the same median intermodulation noise is obtained as with the above first-order theory, provided the echo is equal in amplitude to the
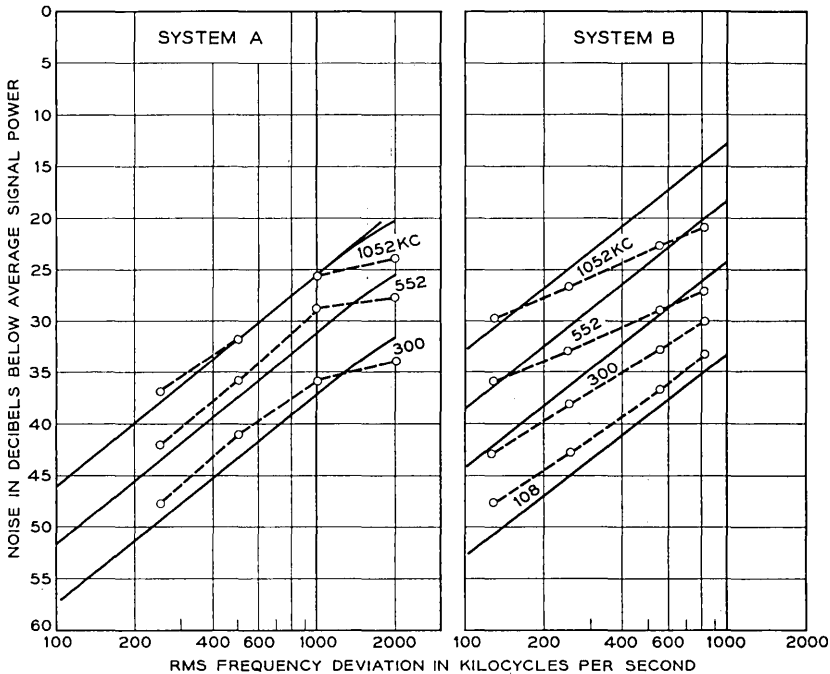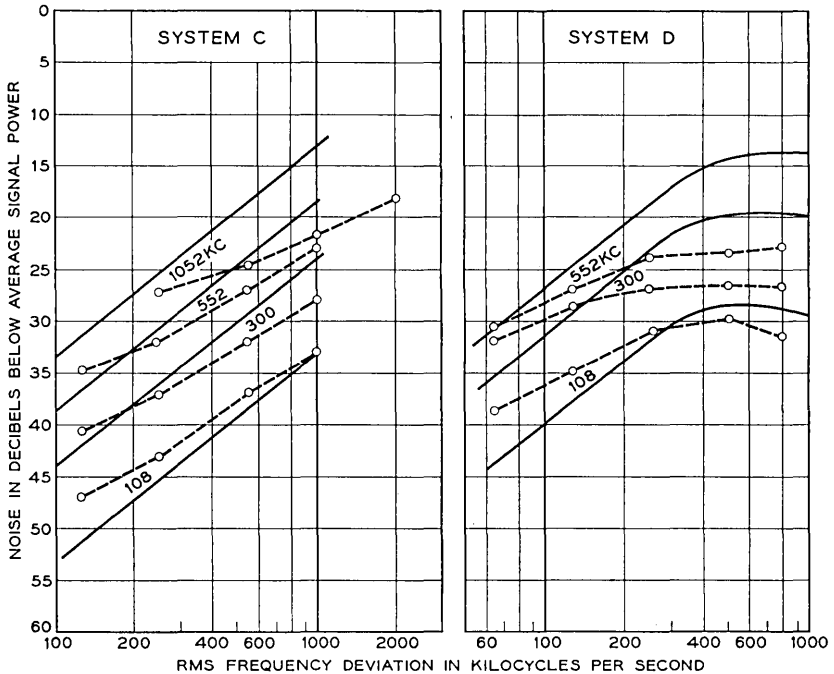
Fig. 8 — Comparison of measured and calculated median intermodulation noise: (dashed curves) measured median intermodulation noise in top channels at indicated frequencies in kc; (solid curves) calculated median intermodulation noise for idealized model with the following values of the equivalent maximum deviation Δ from the mean transmission delay: system C, $\Delta_m = 0.25$ microsecond ($\Delta_0 = 0.185$); system D, $\Delta_m = 0.55$ microsecond ($\Delta_0 = 0.255$).

mean signal. For smaller echoes, predicted intermodulation noise must be less. This conforms with results presented in Figs. 12 and 14 of Ref. 3, which show that intermodulation noise predicted from single-echo theory is significantly smaller than observed. To obtain a satisfactory average relation between predictions and observations, the predicted values must be increased by 9 db, as in Fig. 15 of Ref. 3

TABLE II — RATIO Δ_m/Δ_0

System	A	B	C	D
Length, miles	185	194	340	440
Δ_0 , microsecond	0.12	0.21	0.185	0.255
Δ_m , microsecond	0.12	0.25	0.25	0.55
Δ_m/Δ_0	1.0	1.2	1.35	2.15
α_m/α_0	1.0	1.1	1.35	2.15
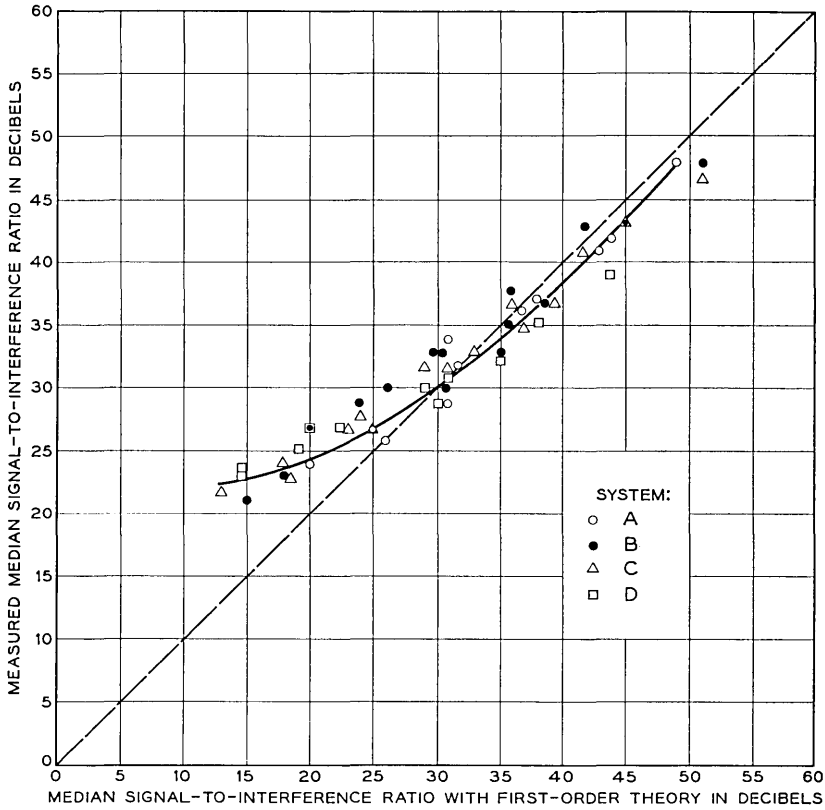
Fig. 9 — Comparison of measured median signal-to-interference ratios with median values based on first-order approximation with best choice of differential transmission delay Δ.

X. PROBABILITY DISTRIBUTION OF INTERMODULATION NOISE

From (18) it is apparent that the probability distribution of ρ is directly related to that of $H(\gamma_p)$. This function is shown in Fig. 10 as related to $(\Delta D)^2$ for $p = 0.5$, 0.1 and 0.01. It should be recognized that this function as given herein is approximate, and that the errors are likely to be greater for small values of p than for median intermodulation noise as considered previously.

From the curves in Fig. 10 it is possible to obtain approximate curves of the probability distribution of intermodulation noise, applying for various values of $\Delta^2 D^2$ as shown in Fig. 11. These curves show that the probability distributions vary markedly with the above parameter, in conformance with a few probability distributions derived from observa-
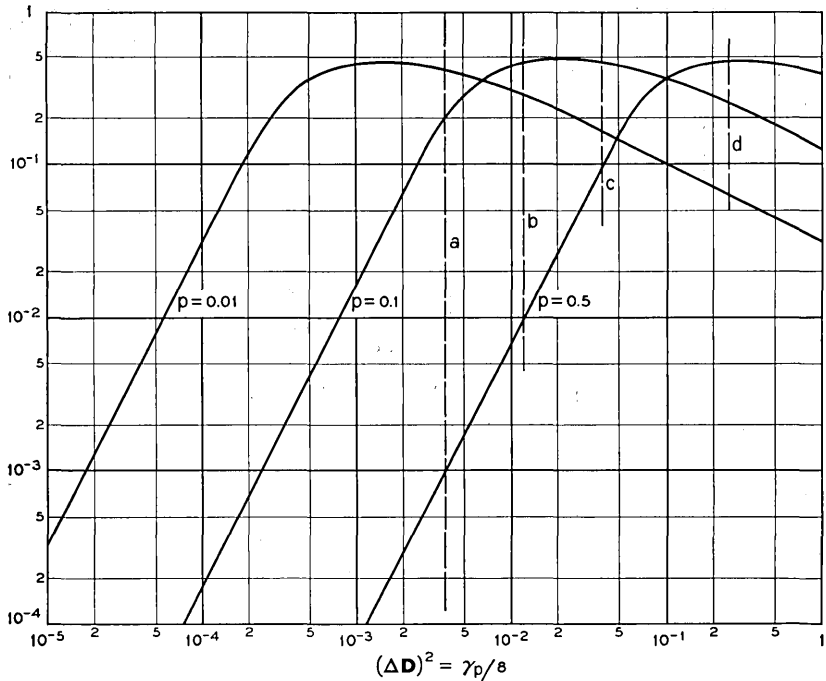
Fig. 10 — Function $H(\gamma_p)$ for various probabilities p.

tions.[2] Because of the approximations involved in the present first-order distortion theory, the above probability distribution curves should be considered illustrative and may not be accurate enough for certain engineering applications.

XI. PREDICTION OF INTERMODULATION DISTORTION

The present first-order intermodulation theory indicates that intermodulation distortion depends on the delay difference Δ, and this would apply also for an exact theory. For various troposcatter links with different angles α_0 and θ, intermodulation distortion would be the same for equal values of Δ. This is exemplified by comparison of intermodulation noise in systems B and C as shown in Figs. 7 and 8. Though these systems have different angles α_0 and θ, intermodulation noise is virtually the same since Δ is the same. Thus, if Δ could be determined, the above first-order theory, in conjunction with the above experimental data, would permit determination of intermodulation distortion for a variety
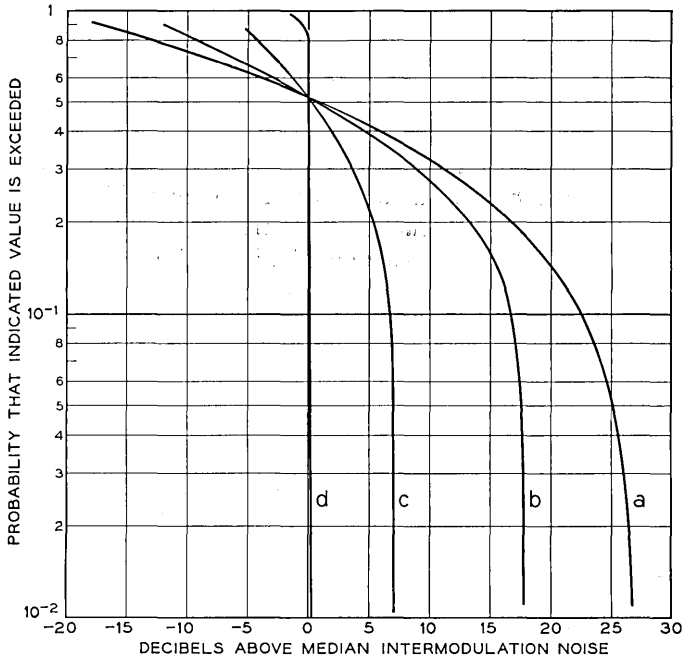
Fig. 11 — Probability distributions of intermodulation noise for various values of $(\Delta D)^2$ corresponding to dashed lines a, b, c and d in Fig. 10.

of conditions other than those in the tests. The above experimental data were confined to intermodulation noise in the top channel, i.e., for $a = 1$ in Fig. 3, and for a particular pre-emphasis, $c = 16$. The expression for $G(c,a)$, or the curves in Fig. 3, permit approximate determination of intermodulation noise at other frequencies, and also for other kinds of pre-emphasis. For example, for $a = 0.3$, intermodulation noise would be greater than for $a = 1$ by an approximate factor $0.32/0.19 \approx 1.7$. If pure FM($c = 0$) had been used in the tests, intermodulation noise at $a = 1$ would have been increased by an approximate factor $1/0.19 \approx 5.2$.

At present there is a principal obstacle to prediction of intermodulation distortion for other values of α_0 than in the above experimental systems. This is the lack of comprehensive experimental data on the beam angle α as affected by scatter for troposcatter links of various lengths. When and if such data become available, it will be possible to determine Δ and in turn intermodulation distortion in the manner indicated above for any kind of system.

XII. APPLICATION TO DIGITAL MULTIBAND TRANSMISSION

The distributions in Fig. 10 apply for average intermodulation noise over brief time intervals, as determined by changes in phase distortion with time. During each such interval the instantaneous amplitudes of intermodulation noise will fluctuate about the average value. For a signal with the properties of random noise, as considered here, the probability distribution of this fluctuation is approximated by the normal law. The distribution of instantaneous amplitudes or intermodulation noise is important in transmission by FM of a number of digital channels in frequency division multiplex, as discussed below.

In digital transmission over troposcatter paths, the error probability for a given signal-to-noise ratio of the receiver depends on the transmission rate, as discussed elsewhere.[1] As the transmission rate is increased, the error probability is ultimately determined by intersymbol interference owing to selective fading, and may be excessively high. The error probability can in this case be reduced, for a given total transmitter power, by transmitting at a slower rate over each of a number of narrower channels in frequency division multiplex. This could be accomplished by individual transmission over each channel, which would entail a number of independent transmitters. An alternative method would be to use a common amplifier and to transmit the combined digital signal by frequency modulation of a common carrier, as now used for transmission of voice frequency channels in frequency division multiplex. In the latter case, it is necessary to consider the possibility of additional transmission impairments owing to intermodulation noise.

With a sufficiently large number of digital channels in frequency division multiplex, the combined wave will have virtually a Gaussian amplitude distribution, like random noise. Hence the probability distribution of average intermodulation noise amplitudes would be as indicated in Fig. 11 for various conditions. The instantaneous amplitude will fluctuate with respect to the above average values, as noted in Section X.

In binary transmission it is often assumed that the error probability will not be excessive if the average noise power from all sources is about 12 db below the average signal power, or 18 db below the peak signal power in on-off binary pulse transmission. From the previous curves and expressions it appears that intermodulation noise power averaged over short intervals will be at least 10 db below the average signal power, with a small probability that it exceeds -15 db. It thus appears that intermodulation noise will not be a limiting or predominant factor even when a large number of binary channels are combined in frequency

division multiplex for transmission by frequency modulation of a common carrier.

XIII. SUMMARY

In broadband transmission over troposcatter paths, selective fading will be encountered with resultant transmission impairments, depending on the modulation method. A previous analysis has been made of such selective fading, based on an idealized model of a troposcatter path. It indicated that selective fading will be accompanied by phase distortion that in a first approximation can be regarded as quadratic, and a probability distribution curve for such quadratic phase distortion was derived. On the premise of such quadratic phase distortion, the error probability owing to selective fading was determined for digital transmission by various methods of carrier modulation.

In the present study the same basic premise of quadratic phase distortion has been used in determining intermodulation distortion for a signal with the properties of random noise, based on the same idealization of a troposcatter path. An approximate relation for intermodulation noise owing to quadratic phase distortion has been derived, applying for any frequency pre-emphasis in FM. In turn, median intermodulation noise as well as the probability distribution of intermodulation noise has been determined, as related to certain basic system parameters.

Median intermodulation noise predicted on basis of free-space antenna beam angles conforms well with observations on links 185 and 194 miles in length. For links 340 and 440 miles long it is necessary to use antenna beam angles that are greater than the free-space angles by factors of about 1.35 and 2.15, respectively. On long links employing narrow-beam antennas, beam broadening is expected because of scatter. Thus if the beam angles had been determined by independent observations or by more elaborate theory, it is probable that predicted intermodulation noise would conform reasonably well with observations.

The results of intermodulation noise measurements thus appear to confirm the conclusion in a previous theoretical analysis of troposcatter transmittance, which indicated that phase distortion owing to selective fading could in a first approximation be represented by a component of quadratic phase distortion, with a probability distribution that can be determined from certain basic system parameters. This affords a simplified first-order theoretical model of selective fading in troposcatter paths that is applicable to evaluation of resultant transmission impairments in both analog and digital transmission.

It can be shown analytically, and it is confirmed by observations, that the above first-order distortion theory yields intermodulation noise that in the case of large signal bandwidths and frequency deviations will be greater than observed or obtained with a more exact distortion theory. An empirical curve presented here permits determination of the expected correction for large bandwidths and frequency deviations.

It has also been demonstrated that the first-order multipath distortion theory presented here affords a significant improvement over single-echo distortion theory applied to random multipath transmission, in that it is simpler and accounts for the probability distribution of intermodulation noise without certain contradictions that are inherent in single-echo theory. Taken in conjunction with presently available data on observed intermodulation noise on certain troposcatter links, as discussed herein, it affords a means of predicting intermodulation noise on any system when more comprehensive experimental data become available on antenna beam broadening by scatter.

APPENDIX

Intermodulation Noise from Quadratic Phase Distortion in Pre-Emphasized FM

General

To facilitate analysis of intermodulation noise in FM owing to attenuation and phase distortion, it is customary to introduce two basic approximations. One is the use of "quasistationary theory" in conjunction with the concept of instantaneous frequency, which is permissible when the signal bandwidth B is negligible in comparison with the carrier frequency, so that the frequency changes imperceptibly over a signal interval $T = 1/2B$. The other customary approximation is that distortion $\alpha(\omega) + i\beta(\omega)$ is sufficiently small to permit the approximation exp $[-\alpha(\omega) - i\beta(\omega)] \approx 1 - \alpha(\omega) - i\beta(\omega)$ over the bandwidth of the modulated carrier wave. The latter is a legitimate approximation for most transmission systems, and greatly simplifies the analysis, but may lead to appreciable errors in applications to tropospheric paths where pronounced attenuation and phase distortion can be encountered. For this reason an alternative approximate analysis is adopted herein to determine intermodulation noise from quadratic phase distortion, in which no limitation is placed on the phase distortion.

Two limiting cases are considered, from which it is possible to make an approximate determination of intermodulation noise as related to phase

distortion, rms frequency deviation, and bandwidth of the baseband signal. In the first case, phase distortion is assumed adequately small, such that the maximum phase distortion in the carrier signal band is less than π radians. Under this condition it is possible by use of "quasistationary" theory to determine the power spectrum of intermodulation noise without much difficulty. In the second case, no limitation is placed on phase distortion, in which case determination of the power spectrum becomes excessively difficult or laborious. It is possible, however, to determine total intermodulation noise power at the detector output, prior to post-detection low-pass filtering. From the manner in which total intermodulation noise power behaves with increasing phase distortion, it is possible to obtain an approximate evaluation of intermodulating noise in a narrow band, such as a voice channel.

A.1 *Power Spectrum of Phase Modulation*

In FM the transmitted wave is of the general form

$$V = \cos [\omega_0 t + \psi(t)] \qquad (25)$$

where the phase $\psi(t)$ is related to the modulating wave $m(t)$ by

$$\psi(t) = k \int_0^t m(t) \, dt \qquad (26)$$

where k is a constant.

The instantaneous frequency deviation is accordingly

$$\Omega(t) = \psi'(t) = km(t). \qquad (27)$$

If the original signal wave has a power spectrum $s(\omega)$ and power pre-emphasis $p(\omega)$ is used, the power spectrum of the modulating wave is

$$W_m(\omega) = s(\omega)p(\omega). \qquad (28)$$

The squared rms frequency deviation $\psi'(t)$ is

$$\underline{\Omega}^2 = k^2 \int_0^\infty s(\omega)p(\omega) \, d\omega. \qquad (29)$$

In accordance with (26), $\psi(t)$ is the integral of $m(t)$. Hence the power spectrum of $\psi(t)$ is given by

$$W_\psi(\omega) = k^2 s(\omega)p(\omega)/\omega^2. \qquad (30)$$

From (29) and (30)

$$W_\psi(\omega) = \underline{\Omega}^2 \frac{s(\omega)p(\omega)/\omega^2}{\int_0^\infty s(\omega)p(\omega) \, d\omega}. \qquad (31)$$

The power spectrum of $\psi'(t)$ is $\omega^2 W_\psi(\omega)$.

A.2 *Autocorrelation Function of Phase Modulation*

The autocorrelation function of $\psi(t)$ is

$$R_\psi(\tau) = \int_0^\infty W_\psi(\omega) \cos \omega\tau \, d\omega \tag{32}$$

$$= k^2 \int_0^\infty \frac{s(\omega)p(\omega)}{\omega^2} \cos \omega\tau \, d\omega. \tag{33}$$

When the constant k is determined from (29), the following relation is obtained

$$R_\psi(\tau) = \underline{\Omega}^2 \left[\int_0^\infty \frac{s(\omega)p(\omega)}{\omega^2} \cos \omega\tau \, d\omega \right] \bigg/ \left[\int_0^\infty s(\omega)p(\omega) \, d\omega \right]. \tag{34}$$

When the baseband power spectrum $s(\omega)$ has a bandwidth $\underline{\Omega}$, (34) can be written

$$R_\psi(\tau) = \mu^2 \left[\Omega^2 \int_0^\Omega \frac{s(\omega)p(\omega)}{\omega^2} \cos \omega\tau \, d\omega \right] \bigg/ \left[\int_0^\Omega s(\omega)p(\omega) \, d\omega \right] \tag{35}$$

where μ is the rms deviation ratio

$$\mu = \underline{\Omega}/\Omega = D/B. \tag{36}$$

In the special case of a flat power spectrum, $s(\omega) = s$ and (35) yields

$$R_\psi(\tau) = \mu^2 \left[\Omega^2 \int_0^\Omega \frac{p(\omega)}{\omega^2} \cos \omega\tau \, d\omega \right] \bigg/ \left[\int_0^\Omega p(\omega) \, d\omega \right]. \tag{37}$$

With pure FM, $p(\omega) = p = $ constant and (37) reduces to

$$R_\psi(\tau) = \mu^2 \int_0^1 \frac{\cos \Omega\tau x}{x^2} \, dx \tag{38}$$

where $x = \omega/\Omega$. From (38) it follows that

$$R_\psi(0) - R_\psi(\tau) = \mu^2 \int_0^1 \frac{1 - \cos \Omega\tau x}{x^2} \, dx$$

$$= \mu^2 [\Omega\tau \, \text{Si}(\Omega\tau) + \cos \Omega\tau - 1] \tag{39}$$

$$= \mu^2 \frac{(\Omega\tau)^2}{2} \left[1 - \frac{(\Omega\tau)^2}{36} + \cdots \right]$$

where Si is the sine integral function.

With pure PM, $p(\omega) = \omega^2$ and (37) yields

$$R_\psi(\tau) = 3\mu^2 \int_0^1 \cos \Omega\tau x \, dx$$

$$= 3\mu^2 \sin \Omega\tau/\Omega\tau \tag{40}$$

$$R_\psi(0) - R_\psi(\tau) = 3\mu^2[1 - \sin \Omega\tau/\Omega\tau]$$

$$= \mu^2 \frac{(\Omega\tau)^2}{2}\left[1 - \frac{(\Omega\tau)^2}{20} + \cdots\right]. \tag{41}$$

A.3 *Intermodulation from Phase Distortion*

It will be assumed that the phase characteristic is of the form

$$\varphi(u) = b_0 + b_1 u + b_2 u^2 + b_3 u^3 + \cdots . \tag{42}$$

Phase distortion is then represented by the term

$$\beta(u) = b_2 u^2 + b_3 u^3 + \cdots \tag{43}$$

where $u = \omega - \omega_0$ is the frequency relative to the carrier frequency ω_0.

When the transmitted wave is of the form (25), the instantaneous frequency deviation is

$$u(t) = \psi'(t) \tag{44}$$

and the corresponding variation in phase distortion with time is

$$\beta[u(t)] = b_2[\psi'(t)]^2 + b_3[\psi'(t)]^3 + \cdots . \tag{45}$$

In the above relation $\psi'(t)$ is given by (27) and the power spectrum of $\psi'(t)$ by (28) multiplied by k^2 or

$$W_{\psi'}(u) = k^2 s(\omega)p(\omega). \tag{46}$$

In determining intermodulation distortion it must be recognized that distortion increases in the range $0 < \beta[u(t)] \leqq \pi$, diminishes in the range $\pi < \beta[u(t)] < 2\pi$, increases in the range $2\pi < \beta[u(t)] < 3\pi$, etc., as illustrated in Fig. 12.

To determine intermodulation distortion it is thus necessary to evaluate the distortion obtained when a wave with the power spectrum (46) is applied to a device with the output vs input characteristic illustrated in Fig. 12. Two limiting cases will be considered below.

A.4 *Intermodulation Spectrum for Small Quadratic Phase Distortion*

With quadratic phase distortion only, (45) becomes

$$\beta[u(t)] = b_2[\psi'(t)]^2. \tag{47}$$

It will be assumed that the probability that $\beta[u(t)]$ exceeds π is so small that it is permissible to assume $\beta[u(t)] < \pi$, and furthermore that $u(t)$ changes at a sufficiently slow rate such that $\beta'[u(t)] = 2b_2\psi''(t) \ll \pi$. For signals with the properties of random noise, these assumptions are
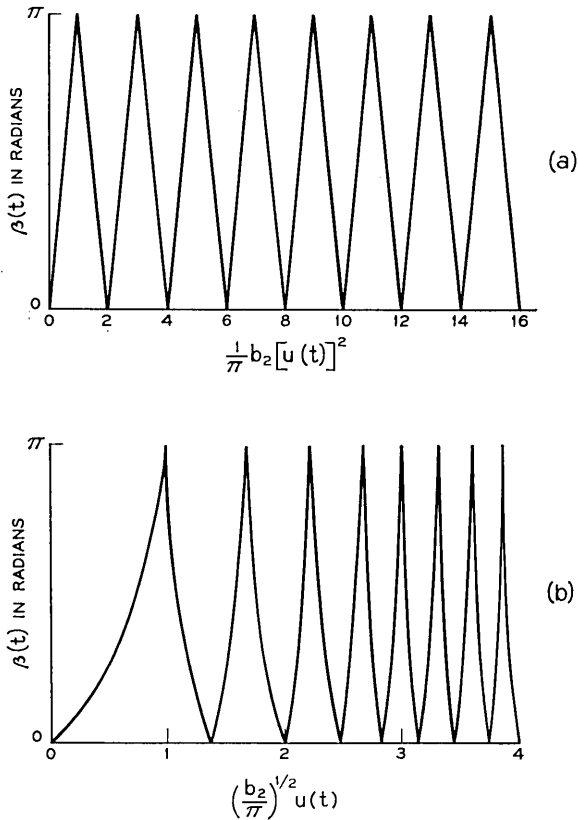
Fig. 12 — Instantaneous phase distortion $\beta(t)$ vs instantaneous frequency deviation $u(t)$ of signal.

permissible provided the rms phase error γ defined by (12) and appearing in Fig. 4 is much less than 1. With these assumptions, the autocorrelation function of the output phase distortion is the same as for a square law device and is given by (Ref. 10, Equation 4.10-1)

$$b_2{}^2[R_{\psi'}{}^2(0) + 2R_{\psi'}{}^2(\tau)].\tag{48}$$

The first term can be identified with a dc component that does not give rise to noise. The power spectrum of the nonlinear output phase distortion is obtained from the second component in (48) and is given by

$$W_\psi{}^{(2)}(\omega) = 2b_2{}^2 \int_0^\infty R_{\psi'}{}^2(\tau) \cos \omega\tau \ d\tau.\tag{49}$$

The ratio of average intermodulation noise power at the frequency ω to the average signal power becomes

$$\rho(\omega) = \frac{W_\psi^{(2)}(\omega)}{W_\psi(\omega)} = \frac{2b_2{}^2 \int_0^1 R_{\psi'}{}^2(\tau) \cos \omega\tau \, d\tau}{k^2 p(\omega)s(\omega)/\omega^2} . \tag{50}$$

In view of (46) the following relation applies

$$R_{\psi'}(\tau) = k^2 \int_0^\infty s(\omega)p(\omega) \cos \omega\tau \, d\omega. \tag{51}$$

Expression (50) can be written

$$\rho(\omega) = \frac{2b_2{}^2 k^4 \int_0^\infty k^{-4} R_{\psi'}{}^2(\tau) \cos \omega\tau \, d\tau}{k^2 p(\omega)s(\omega)/\omega^2} \tag{52}$$

$$= \frac{2\gamma^2 a^2}{\mu^2} \frac{\int_0^\infty k^{-4} R_{\psi'}{}^2(\tau) \cos \omega\tau \, d\tau}{p(\omega)s(\omega) \int_0^\infty s(\omega)p(\omega) \, d\omega} \tag{53}$$

where

$$a = \omega/\Omega = f/B \tag{54}$$

$$\gamma = b_2\mu^2\Omega^2 = b_2\Omega^2 = (2\pi)^2 b_2 \, D^2. \tag{55}$$

The following relation applies[*]

$$\int_0^\infty R_{\psi'}{}^2(\tau) \cos \omega\tau \, d\tau = \frac{1}{2} \int_0^\infty W_{\psi'}(u)W_{\psi'}(\omega - u) \, du \tag{56}$$

where $W_{\psi'}(u)$ is the power spectrum given by (46).

In view of (56) and (46), expression (53) can be written

$$\rho(\omega) = \frac{a^2\gamma^2/\mu^2}{p(\omega)s(\omega) \int_0^\infty p(\omega)s(\omega)d\omega} \tag{57}$$

$$\cdot \int_{-\infty}^\infty s(\omega)p(\omega)s(\omega - u)p(\omega - u)du.$$

In the special case of a flat power spectrum $s(\omega) = s$ of bandwidth

---

[*] Ref. 10, Eq. (4C-6). In this reference the autocorrelation function is defined differently from the definition used here and has a factor 4 in integral (51), so that an additional factor $\frac{1}{4}$ appears in (56).

$\Omega = 2\pi B$, (57) becomes

$$\rho(\omega) = \frac{a^2\gamma^2/\mu^2}{p(\omega)\dfrac{1}{\Omega}\displaystyle\int_0^\Omega p(\omega)d\omega}\frac{1}{\Omega}\int_{\omega-\Omega}^\Omega p(u)p(\omega - u)du \tag{58}$$

$$= \frac{\gamma^2}{\mu^2}\frac{a^2}{p(a)\displaystyle\int_0^1 p(x)dx}\int_{a-1}^1 p(x)p(a - x)dx. \tag{59}$$

When $p(x)$ is of the form

$$p(x) = 1 + c(u/\Omega)^2 = 1 + cx^2 \tag{60}$$

relation (59) becomes

$$\rho(\omega) = \frac{3a^2\gamma^2}{\mu^2(1 + ca^2)(3 + c)}\int_{a-1}^1 (1 + cx^2)[1 + c(a - x)^2]dx \tag{61}$$

$$= \frac{\gamma^2 3a^2}{\mu^2(1 + ca^2)(3 + c)}F(c,a)$$

where

$$F(c,a) = 2 - a + \frac{2c + c^2a^2}{3}[1 + (1 - a)^3]$$
$$- \frac{c^2a}{2}[1 - (1 - a)^4] + \frac{c^2}{5}[1 + (1 - a)^5]. \tag{62}$$

In the particular case of pure FM, $c = 0$ and $F(c,a) = 2 - a$, so that (61) yields

$$\rho(\omega) = \frac{\gamma^2 a^2}{\mu^2}(2 - a)$$
$$= \left(\frac{B}{D}\right)^2\gamma^2 a^2(2 - a) \tag{63}$$

where $a = \omega/\Omega = f/B$, $D = \Omega/2\pi$ and $\gamma = b_2\Omega^2 = b_2(2\pi D)^2$.

The above result (63) conforms with an expression derived by Rice for this limiting case (Ref. 11, Equation 5.6).

A.5 Total Intermodulation from Quadratic Phase Distortion

The previous analysis of the power spectrum of intermodulation noise was based on the assumption that the maximum phase distortion in the transmission band is substantially less than 180°. Without this limitation, numerical determination of the power spectrum becomes very diffi-

cult, though a formal solution may be feasible. However, it is possible to determine total intermodulation distortion without too much difficulty, without limitation on the phase distortion, as shown below.

Let x designate the instantaneous amplitude of $\psi'(t) = km(t)$, and let x have a probability density

$$p(x) = \left(\frac{2}{\pi\sigma_x^2}\right)^{\frac{1}{2}} \exp\left(-x^2/2\sigma_x^2\right). \tag{64}$$

For large instantaneous frequency deviations $\psi'(t)$ the derivative $\psi''(t)$ is on the average sufficiently small to be neglected. The total intermodulation distortion in the received signal prior to post-detection low-pass filtering is then for a nonlinear characteristic as illustrated in Fig. 12.

$$I = \int_0^{L_1} (b_2 x^2)^2 p(x) dx + \int_{L_1}^{L_3} (2\pi - b_2 x^2)^2 p(x) dx$$
$$+ \int_{L_3}^{L_5} (4\pi - b_2 x^2)^2 p(x) dx + \cdots + \int_{L_{2n-1}}^{L_{2m+1}} (2\pi m - b_2 x^2)^2 p(x) dx \tag{65}$$

where

$$L_j = (j\pi/b_2)^{\frac{1}{2}}.$$

With $b_2 x^2 = u^2$, $\gamma = b_2 \sigma_x^2$ and

$$p(u) = \left(\frac{2}{\pi\gamma}\right)^{\frac{1}{2}} \exp\left(-u^2/2\gamma\right) \tag{66}$$

expression (65) can be written

$$I = \int_0^{l_1} u^4 p(u) du + \int_{l_1}^{l_3} (2\pi - u^2)^2 p(u) du$$
$$+ \int_{l_3}^{l_5} (4\pi - u^2)^2 p(u) du + \cdots \tag{67}$$

where

$$l_j = (j\pi)^{\frac{1}{2}}. \tag{68}$$

Writing $2m\pi - u^2 = -\tau$, $2u\,du = d\tau$, expression (67) can be transformed into

$$I = \int_0^{\pi} \tau^{\frac{1}{2}} p(\tau) d\tau + e^{-\pi/\gamma} \int_{-\pi}^{\pi} \frac{\tau^2}{(2\pi + \tau)^{\frac{1}{2}}} p(\tau) d\tau$$
$$+ e^{-2\pi/\gamma} \int_{-\pi}^{\pi} \frac{\tau^2}{(4\pi + \tau)^{\frac{1}{2}}} p(\tau) d\tau + \cdots \tag{69}$$

where

$$p(\tau) = \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} e^{-\tau/2\gamma}. \qquad (70)$$

Total distortion I includes a mean or dc power component I_0 that must be subtracted from I to obtain the nonlinear component. The mean amplitude component $I_0^{\frac{1}{2}}$ ls given by

$$I_0^{\frac{1}{2}} = \int_0^{L_1} b_2 x^2 p(x) dx + \int_{L_1}^{L_3} (2\pi - b_2 x^2) p(x) dx$$
$$+ \int_{L_3}^{L_5} (4\pi - b_2 x^2) p(x) dx + \cdots \qquad (71)$$

where L_m and $p(x)$ are defined as before.

With the same notation as before, (71) can be transformed into

$$I_0^{\frac{1}{2}} = \int_0^\pi \tau^{\frac{1}{2}} p(\tau) d\tau + e^{-\pi/\gamma} \int_{-\pi}^\pi \frac{|\tau|}{(2\pi + \tau)^{\frac{1}{2}}} p(\tau) dt$$
$$+ e^{-2\pi/\gamma} \int_{-\pi}^\pi \frac{|\tau|}{(4\pi + \tau)^{\frac{1}{2}}} p(\tau) d\tau + \cdots . \qquad (72)$$

In the above relations γ is the phase distortion corresponding to the rms frequency deviation as given by

$$\gamma = b_2 \sigma_x^2 = b_2 \underline{\Omega}^2 = b_2 \mu^2 \Omega^2. \qquad (73)$$

The last relations follow from (29) since σ_x^2 is the variance of $\psi'(t)$.

The total average signal power is

$$S = R_\psi(0) = \mu^2 \left[\Omega^2 \int_0^\Omega \frac{p(\omega)}{\omega^2} d\omega \right] \Big/ \left[\int_0^\Omega p(\omega) d\omega \right] = \mu^2/C \quad (74)$$

where C is a constant depending on $p(\omega)$.

The ratio of total nonlinear intermodulation noise to total average signal power becomes

$$\rho = \frac{I - I_0}{S} = C \frac{I - I_0}{\mu^2}$$
$$= C(I - I_0) \left(\frac{B}{D}\right)^2. \qquad (75)$$

A.6 *Total Intermodulation for Small Phase Distortion*

For sufficiently small values of $\gamma = b_2 \underline{\Omega}^2$, such that $\pi/\gamma \gg 1$, only the

first integral in (69) needs to be considered. Hence

$$I \approx \int_0^\pi \tau^{\frac{3}{2}} p(\tau) d\tau$$

$$= 3\gamma^2 \text{ erf } (z) - 3 \cdot 2^{\frac{1}{2}} \gamma^{\frac{3}{2}} \exp(-z^2) - 2^{\frac{1}{2}} \pi \gamma^{\frac{1}{2}} \exp(z^2) \qquad (76)$$

where

$$z^2 = \pi/2\gamma. \qquad (77)$$

With a similar approximation (72) yields

$$I_0^{\frac{1}{2}} \approx \int_0^\pi \tau^{\frac{1}{2}} p(\tau) d\tau \qquad (78)$$

$$= \gamma \text{ erf } (z) - 2^{\frac{1}{2}} \exp (-z^2).$$

For $z \geq 2$, or $\gamma \leq \pi/8$:

$$I \approx 3\gamma^2 \qquad \text{and} \qquad I_0^{\frac{1}{2}} = \gamma.$$

Hence $I - I_0 = 2\gamma^2$ and (75) becomes

$$\rho = C \cdot 2 \frac{\gamma^2}{\mu^2} \qquad (79)$$

where the constant C is defined through (74).

It will be noted that (79) is of the same basic form as (61) for the ratio $\rho(\omega)$ at the frequency ω. In (61) the multiplier of γ^2/μ^2 is a constant, as is the case in (79).

A.7 *Total Intermodulation for Large Phase Distortion*

When $\gamma \gg 1$, it is permissible to approximate $p(\tau)$ as given by (70) with

$$p(\tau) \approx \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}}. \qquad (80)$$

This approximation is valid in evaluation of the various integrals in (69) and (72) provided that for the minimum value of $\tau = \pi$, $\exp (-\tau/2\gamma) \ll 1$. This is the case if

$$\pi/2\gamma \ll 1 \qquad \text{or} \qquad \gamma \gg \pi/2.$$

With (80) in (69)

$$I = \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \left[\int_0^\pi \tau^{\frac{3}{2}} d\tau + \sum_{m=1}^\infty e^{-m\pi/\gamma} \int_{-\tau}^\pi \frac{\tau^2 d\tau}{(2\pi m + \tau)^{\frac{1}{2}}} \right]. \qquad (81)$$

In (81),

$$\int_{-\pi}^{\pi} \frac{\tau^2 d\tau}{(2m\pi + \tau)^{\frac{1}{2}}} = \frac{2 \cdot \pi^{\frac{5}{2}}}{15} [(2m + 1)^{\frac{1}{2}}(32m^2 - 8m + 3)$$

$$- (2m + 1)^{\frac{1}{2}}(32m^2 + 8m + 3)] \tag{82}$$

$$\approx \frac{1}{m^{\frac{1}{2}}} \frac{2^{\frac{1}{2}}\pi^{\frac{5}{2}}}{3} \quad \text{for} \quad m \geqq 1. \tag{83}$$

For $m = 1$, (82) gives about 0.5 and (83) about 0.47. Hence (83) represents a good approximation of (82).

With (83) in (81)

$$I = \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \left[\int_0^{\pi} \tau^{\frac{1}{2}} d\tau + \frac{2^{\frac{1}{2}}\pi^{\frac{5}{2}}}{3} \sum_{m=1}^{\infty} \frac{e^{-m\pi/\gamma}}{m^{\frac{1}{2}}}\right]. \tag{84}$$

As a first approximation the summation can be replaced by an integral, in which case

$$I \approx \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \left[\frac{2}{5} \pi^{\frac{5}{2}} + \frac{2^{\frac{1}{2}}\pi^{\frac{5}{2}}}{3} \int_{m=1}^{\infty} \frac{e^{-m\pi/\gamma} dm}{m^{\frac{1}{2}}}\right]. \tag{85}$$

With $m = u^2$

$$I = \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \pi^{\frac{5}{2}} \left[\frac{2}{5} + \frac{2^{\frac{1}{2}} \cdot 2}{3} \int_{u=1}^{\infty} e^{-u^2 \pi/\gamma} du\right]$$

$$= \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \pi^{\frac{5}{2}} \left[\frac{2}{5} + \frac{2^{\frac{1}{2}}\gamma^{\frac{1}{2}}}{3} \operatorname{erfc}\left(\sqrt{\frac{\pi}{\gamma}}\right)\right] \tag{86}$$

$$= \frac{\pi^2}{3} \left[\operatorname{erfc}\left(\sqrt{\frac{\pi}{\gamma}}\right) + \frac{6}{5 \cdot 2^{\frac{1}{2}}\gamma^{\frac{1}{2}}}\right] \tag{87}$$

$$\approx \frac{\pi^2}{3} \quad \text{for} \quad \gamma \gg 4\pi. \tag{88}$$

By a similar approximation $I_0^{\frac{1}{2}}$ as given by (72) becomes

$$I_0^{\frac{1}{2}} = \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \left[\int_0^{\pi} \tau^{\frac{1}{2}} d\tau + \sum_{m=1}^{\infty} e^{-m\pi/\gamma} \int_{-\pi}^{\pi} \frac{|\tau| d\tau}{(2m\pi + \tau)^{\frac{1}{2}}}\right] \tag{89}$$

$$= \left(\frac{1}{2\pi\gamma}\right)^{\frac{1}{2}} \left[\frac{2}{3} \pi^{\frac{3}{2}} + \frac{\pi\sigma}{2^{\frac{1}{2}}} \operatorname{erfc}\left(\sqrt{\frac{\pi}{\gamma}}\right)\right] \tag{90}$$

$$= \frac{\sqrt{\pi}}{2} \left[\operatorname{erfc}\left(\sqrt{\frac{\pi}{\gamma}}\right) + \frac{2^{\frac{1}{2}}}{3\gamma^{\frac{1}{2}}} \pi^{\frac{1}{2}}\right] \tag{91}$$

$$\approx \frac{\pi^{\frac{1}{2}}}{2} \quad \text{for} \quad \gamma \gg 4\pi. \tag{92}$$

The ratio ρ is obtained from (75) with I as given by (87) and $I_0^{\frac{1}{2}}$ by (91). In the limit of $\gamma \to \infty$ the ratio becomes

$$
\rho = C \frac{I - I_0}{\mu^2} = \frac{C}{\mu^2} \pi \left(\frac{\pi}{3} - \frac{1}{4} \right)
$$

$$
\approx 2.5 \frac{C}{\mu^2}.
$$

(93)

A.8 Approximation for Total Intermodulation

The general expression for the ratio ρ of total intermodulation noise power to average signal power can be written in the form

$$
\rho = \frac{2C}{\mu^2} h(\gamma).
$$

(94)

For the limiting case of $\gamma \to 0$, the function h is in accordance with (79)

$$
h = \gamma^2.
$$

(95)

For the other limiting case in which $\gamma \to \infty$, the function h is in accordance with (93)

$$
h = 2.5/2 = 1.25.
$$

(96)

In Fig. 13 are shown the above two limiting cases, together with the function h obtained from (75) as $\eta = I - I_0$, when I and I_0 are determined from (86) and (91). The approximate function $h(\gamma)$ is obtained by drawing a transition curve between the above two limiting cases, as in Fig. 13.

A.9 Approximation for Intermodulation Spectrum

The function $h(\gamma)$ in Fig. 13 is proportional to the total intermodulation noise power and can be related to the power spectrum $W_i(\omega)$ of intermodulation noise by

$$
h(\gamma) = c_0 \int_0^\infty W_i(\omega) \, d\omega
$$

(97)

where c_0 is a constant. Relation (94) can thus be written

$$
\rho = \frac{2c_0 C}{\mu^2} \int_0^\infty W_i(\omega) \, d\omega.
$$

(98)

For $\gamma \to 0$, (98) must conform with (95), which is possible provided the power spectrum is of the general form

$$W_i^0(\omega) = c_1\gamma^2 \frac{1}{\Omega} F_0(\omega/\Omega) \tag{99}$$

where F_0 is any functional relation dependent only on the ratio $a = \omega/\Omega$. With (99) in (98)

$$\begin{aligned}
\rho &= \frac{\gamma^2}{\mu^2} 2c_0c_1C \frac{1}{\Omega} \int_0^\infty F_0(\omega/\Omega)\ d\omega \\
&= \frac{\gamma^2}{\mu^2} 2c_0c_1C \int_0^\infty F_0(u)\ du.
\end{aligned} \tag{100}$$

This yields relation (95) provided

$$c_0c_1C \int_0^\infty F_0(u)\ du = 1. \tag{101}$$
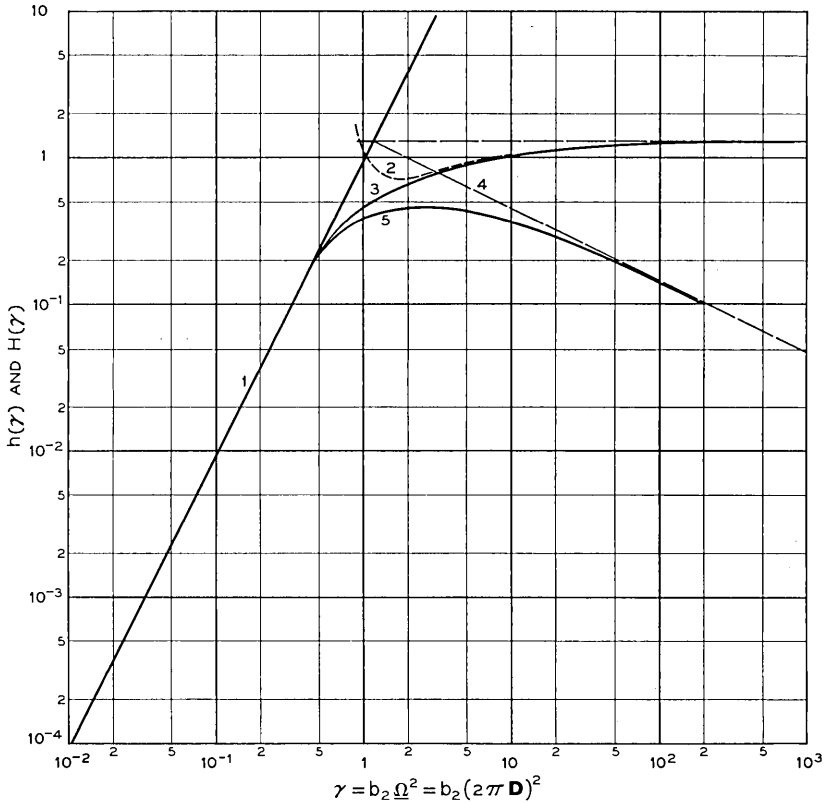
Fig. 13 — Functions $h(\gamma)$ and $H(\gamma)$: 1, functions $h(\gamma)$ and $H(\gamma)$ for $\gamma \ll 1$; 2, function $h(\gamma)$ for $\gamma \gg 1$; 3, approximate interpolated function $h(\gamma)$; 4, function $H(\gamma)$ for $\gamma \gg 1$; 5, approximate interpolated function $H(\gamma)$.

From (100) it is apparent that the ratio of intermodulation noise power to average signal power in a narrow band $d\omega$ at ω is

$$\rho(\omega) \;=\; \frac{\gamma^2}{\mu^2}\, 2c_0c_1C\, \frac{1}{\Omega}\, F_0(\omega/\Omega)\,. \tag{102}$$

Comparison of (102) with (61) shows that in this case

$$2c_0c_1C\, \frac{1}{\Omega}\, F_0(\omega/\Omega) \;=\; \frac{3a^2}{(1 + ca^2)(3 + c)}\, F(c,a) \tag{103}$$

where $F(c,a)$ is given by (62).

In summary, for $\gamma \to 0$ the power spectrum has a fixed shape independent of γ and an amplitude proportional to γ^2.

Consider next the limiting case in which $\gamma \to \infty$. In accordance with (96) h then approaches a constant, which is possible for various power spectra of the general form

$$W_i^{(\infty)}(\omega) \;=\; \frac{c_1}{\gamma^n}\, F_\infty(\omega/\gamma^n) \tag{104}$$

where $F_\infty(\omega/\gamma^n)$ is any functional relation dependent only on the ratio (ω/γ^n). In this case (104) in (98) yields

$$\begin{aligned}
\rho &= \frac{2c_0c_1C}{\mu^2}\, \frac{1}{\gamma^n} \int_0^\infty F_\infty(\omega/\gamma^n)\, d\omega \\
&= \frac{2c_0c_1C}{\mu^2} \int_0^\infty F(u)\, du
\end{aligned} \tag{105}$$

where $u = \omega/\gamma^n$.

The exponent n can be determined from consideration of the input vs output characteristic shown in Fig. 12. If b_2 is increased by a factor k, the intervals between zero points are multiplied by a factor $k^{-\frac{1}{2}}$, as indicated in Fig. 14 for $k = 4$. For a given frequency deviation, the bandwidth of the power spectrum is then multiplied by a factor $k^{\frac{1}{2}}$ and the amplitude of the spectrum at each frequency multiplied by a factor $k^{-\frac{1}{2}}$. Hence in the case of quadratic phase distortion as considered here, $n = \frac{1}{2}$ in (104).

Based on the above considerations, the power spectrum at any frequency ω for the above two limiting cases would vary with γ as indicated in Fig. 13. The shape of the curves between these two limiting cases would in a first approximation be represented by the function $H(\gamma)$ shown in Fig. 13.
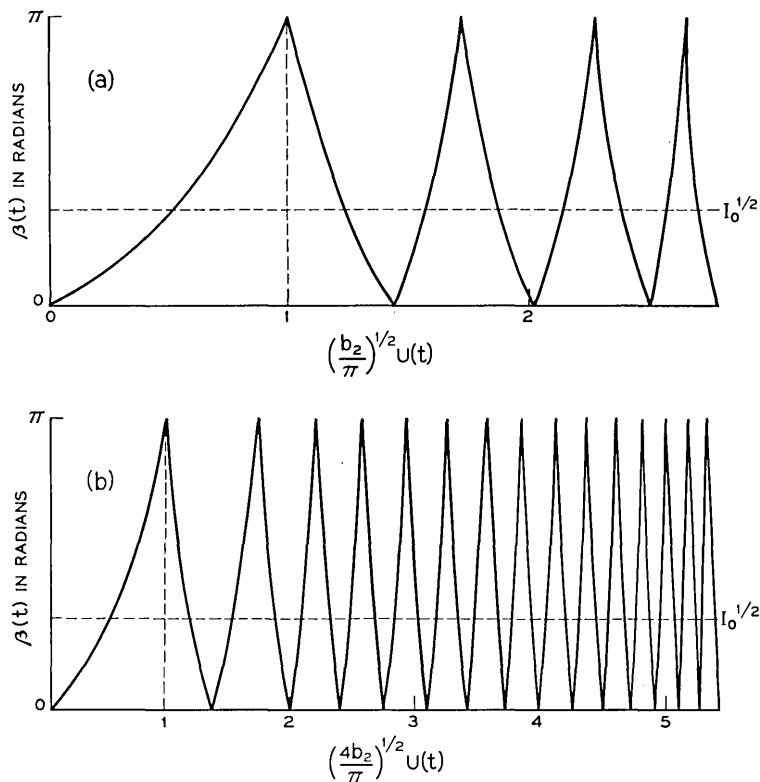
Fig. 14 — (a) Relation of instantaneous phase distortion $\beta(t)$ to instantaneous frequency deviation $u(t)$ for a given b_2; (b) relation of instantaneous phase distortion to instantaneous frequency deviation with fourfold increase in b_2.

A.10 Approximation for $\rho(\omega)$

The ratio $\rho(\omega)$ of intermodulation noise power in a narrow band at ω to average signal power in the same narrow band can be written

$$\rho(\omega) = \frac{2C(\omega)}{\mu^2} H(\gamma). \tag{106}$$

This relation differs from (94) in that $h(\gamma)$ as shown in Fig. 13 is replaced by $H(\gamma)$ shown in the same figure, and C is replaced by $C(\omega)$. The constant C defined through (74) depends on the frequency pre-emphasis $p(\omega)$. The function $C(\omega)$ depends both on the frequency pre-emphasis $p(\omega)$ and the frequency under consideration.

For the particular type of frequency pre-emphasis represented by

(60), expression (106) must conform with (61). This results in the following approximate relation

$$\rho(\omega) = \left(\frac{B}{D}\right)^2 G(c,a)H(\gamma) \tag{107}$$

where $H(\gamma)$ is the function shown in Fig. 13 and

$$G(c,a) = \frac{3a^2}{(1 + ca^2)(3 + c)} F(c,a) \tag{108}$$

where $F(c,a)$ is given by (62).

In the particular case in which $c = 16$ and $a = f/B = 1$

$$G(c,a) \approx 0.192 \tag{109}$$

and (107) yields

$$\rho(B) = \left(\frac{B}{D}\right)^2 \times 0.192H(\gamma) \tag{110}$$

$$= \left(\frac{B}{D}\right)^2 \times 0.192 \cdot \gamma^2 \quad \text{for} \quad \gamma \ll 1. \tag{111}$$

REFERENCES

1. Sunde, E. D., Digital Troposcatter Transmission and Modulation Theory, this issue, Part 1, p. 143.
2. Clutts, C. E., Kennedy, R. N., and Trecker, J. M., Results of Bandwidth Tests on the 185-Mile Florida-Cuba Scatter Radio Systems, IRE Trans. on Communication Systems, **9**, December, 1961, p. 434.
3. Beach, C. D., and Trecker, J. M., A Method for Predicting Interchannel Modulation Due to Multipath Propagation in FM and PM Tropospheric Radio Systems, B.S.T.J., **42**, January, 1963, p. 1.
4. Bennett, W. R., Curtis, H. E., and Rice, S. O., Interchannel Interference in FM and PM Systems under Noise Loading Conditions, B.S.T.J., **34**, May, 1955, p. 601.
5. Medhurst, R. G., and Small, G. F., Distortion in Frequency-Modulation Systems Due to Small Sinusoidal Variations of Transmission Characteristics, Proc. IRE, **44**, November, 1956, p. 1608.
6. Medhurst, R. G., and Small, G. F., An Extended Analysis of Echo Distortion in FM Transmission of Frequency-Division Multiplex, Proc. IEE, **103**, Pt. B, March, 1956, p. 190.
7. Carson, J. R., and Fry, T. C., Variable Frequency Electric Circuit Theory with Applications to the Theory of Frequency Modulation, B.S.T.J., **16**, October, 1937, p. 513.
8. van der Pohl, B., The Fundamental Principles of Frequency Modulation, Jour. IEE, Part III, May, 1946, p. 153.
9. Albersheim, W. J., and Schafer, J. P., Echo Distortion in the FM Transmission of Frequency Division Multiplex, Proc. IRE, **40**, March, 1952, p. 316.
10. Rice. S. O., Mathematical Analysis of Random Noise-I and -II, B.S.T.J., **23**, July, 1944, p. 282, and **24**, January, 1945, p. 46.
11. Rice, S. O., Distortion Produced by a Noise Modulated Signal by Nonlinear Attenuation and Phase Shift, B.S.T.J., **36**, July, 1957, p. 879.

Cutoff Frequencies of the Dielectrically Loaded Comb Structure as Used in Traveling-Wave Masers*

By S. E. HARRIS, R. W. DEGRASSE and E. O. SCHULZ-DuBOIS

(Manuscript received June 27, 1963)

The subject of traveling-wave maser design is reviewed and a first step towards an analytical design procedure is presented. A method is derived for calculating the upper and lower cutoff frequencies of a comb-type slow-wave structure of simple geometry. It is based on the electromagnetic field pattern and the equivalent impedances which are calculated for these frequencies, both for the dielectrically loaded and the empty comb structure. The design procedure resulting from these calculations permits the prediction of a dielectric loading geometry that shifts the upper and lower cutoff frequency of the empty comb to new, lower values which can be arbitrarily specified within certain limitations. Frequencies calculated by this procedure are compared with the results of measurements, and it is found that cutoff frequencies can be predicted to better than 10 per cent.

I. INTRODUCTION

In the early development of the traveling-wave maser (TWM),[1] the design procedures used were largely empirical. Short TWM model sections were built, tested and modified in order to meet the desired performance specifications. By this cut-and-try method, a satisfactory design was finally derived which was applied in the construction of full-length TWM's.

However, a more satisfying approach is possible if the relevant theoretical aspects regarding the maser active material, the ferrimagnetic isolator and the electromagnetic behavior of the slow-wave structure are known, either rigorously or approximately. Then a TWM can be designed on the basis of analysis before actual construction. Most attractive in the analytical approach is the inherent flexibility and versa-

tility. Thus, a large number of design ideas may be explored and a near optimum configuration can be found before any hardware is built.

The present paper is a step in the direction of a more analytical approach. Using reasonably accurate approximations to the field pattern at both cutoff frequencies, the equivalent TEM line impedances, the "effective" dielectric constants and, finally, the cutoff frequencies are calculated. This results in a numerical design procedure for the TWM structure. The analysis is made for a comb having fingers of rectangular cross section and for dielectric loading with maser material in the form of one or two rectangular parallelepipeds as shown in Fig. 1. Comparison of cutoff frequencies calculated by this method with experiment shows agreement to usually better than 5 per cent.

TOP CROSS SECTION   SIDE CROSS SECTION   END CROSS SECTION

PERSPECTIVE VIEW
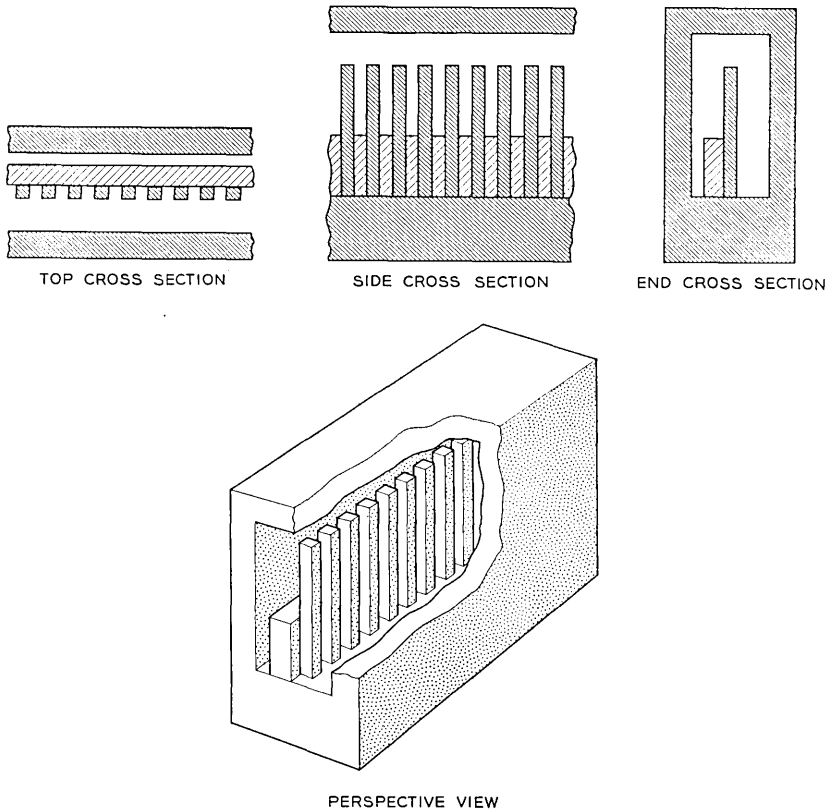
Fig. 1 — Typical comb structure.

1.1 *The Significance of Cutoff Frequencies in TWM Design*

Consider the TWM electronic gain formula[1]

$$G(\text{db}) = 27.3(-\chi'')fFl/v_g. \tag{1}$$

Here $(-\chi'')$ is the inverted susceptibility of the maser active material, f the signal frequency, F the filling factor, l the length of the maser structure and v_g the group velocity. The TWM net gain is obtained by subtracting from (1) the slow-wave structure loss (copper loss) and the ferrimagnetic isolator loss (ferrite loss).

In the development of a practical TWM, the design frequency f and the structure length l are generally determined by the application. The susceptibility $(-\chi'')$ is a property of the active material which cannot be theoretically predicted and must be experimentally determined. $(-\chi'')$ is redefined as

$$-\chi'' = I\chi_0'' \tag{2}$$

and the quantities I and χ_0'' are determined by two independent measurements. Here, I is the inversion ratio, i.e., the ratio of electronic gain from the activated maser material to electronic loss in the same material at thermal equilibrium. J. E. Geusic and W. J. Tabor have carried out inversion measurements for ruby maser material in a helix test structure, and the method and results will be described in a forthcoming paper.[2] The susceptibility at thermal equilibrium, χ_0'', is measured by standard resonance techniques[3] or may be calculated from the material composition and linewidth. In this way, $-\chi''$ can be determined to about 10 per cent, which is adequate for the present design procedure. Complications can arise in practice, however, if nominally identical crystals show variations in the active ion concentration or in the crystalline perfection.

The filling factor F may be factorized into two expressions

$$F = F_p F_v \tag{3}$$

where

$$F_p = \left[\int_M | \mu \cdot H^* |^2 \, dA \right] \bigg/ \left[| \mu |^2 \int_M | H |^2 \, dA \right] \tag{4}$$

and

$$F_v = \left[\int_M | H |^2 \, dA \right] \bigg/ \left[\int_A | H^2 | \, dA \right]. \tag{5}$$

Here, μ is the magnetic dipole moment associated with the maser signal

transition and H is the RF magnetic field in the TWM structure. The asterisk * denotes the conjugate complex time dependence. The integration is performed in the cross-sectional plane where M denotes the cross section of the maser material and A the total structure cross section. F_p may be called the polarization efficiency factor and F_v the volume filling factor. F_p expresses the excitation efficiency of the signal transition by the RF magnetic field present in the maser material. For example, if both μ and H are of circular polarization in the same direction, then F_p is unity. Similarly, for maser material symmetrically loaded on both sides of the comb and with a circular transition perpendicular to the finger direction, a symmetry argument shows that $F_p = \frac{1}{2}$. F_v indicates what fraction of the total magnetic field energy is contained within the maser material. F_p and F_v are functions of frequency across the passband of the comb structure. Usually, however, it is sufficient to consider F at some midband frequency where it is only a slowly varying function of frequency.

Experience suggests that it is possible to estimate F to fair accuracy from the TWM geometry and a qualitative estimate of the RF magnetic field pattern. For example, it is estimated that in TWM's designed in this laboratory for 5.6, 4.2, 2.4 and 1.4 gc the filling factor F varies over the relatively limited range from 25 to 45 per cent. Thus, from the viewpoint of the analytical design of the TWM, a detailed computation of the RF magnetic field configuration is of no great value unless the other factors entering the TWM gain formula are known with comparable accuracy.

Up to the present time, this was not the case, the factor least amenable to analytical prediction being the group velocity v_g. It is well known that a wave traveling through a slow-wave structure has field components varying like $\exp[i(\omega t - \beta z)]$, where $\omega = 2\pi f$, t is the time, β the phase propagation constant and z the length coordinate along the structure. In the comb structure, each finger is an energy storage element capable of resonant storage in the same way as a quarter-wavelength coaxial resonator. As a general rule, the phase shift between adjacent elements may assume values between 0 and $\pm\pi$ as the frequency is varied across the passband. The phase shift values 0 and $\pm\pi$ are associated with the cutoff frequencies. The comb structure is normally a forward-wave structure, where $+\pi$ is the phase shift at the upper cutoff frequency and 0 that at the lower. It is possible (although not of practical importance in TWM design) to make the comb a backward-wave structure, in which case $-\pi$ is the phase shift at the lower cutoff frequency and 0 that at the upper. In the normal forward-wave comb

structure, the phase propagation constant then varies from $\beta = 0$ to $\beta = (N - 1)\pi/l$ across the passband, where N is the number of fingers and l is the structure length measured between centers of the first and last finger. The group velocity is given by

$$v_g = d\omega/d\beta. \tag{6}$$

Typical diagrams of β as a function of ω are shown in Fig. 2(a). As the curves approach the cutoff points, they assume infinite slope, corresponding to zero group velocity. There is a range at midband, however, where the group velocity is fairly constant. These graphs are typical of most of the structures studied but exceptions occasionally were found, as indicated in Fig. 2(b). These exceptions include backward-wave structures where phase and group propagation take place in opposite directions. They also include "mongrel" structures where, over part of the band, β is a double-valued function of ω; these, therefore, are forward and backward at the same time. This latter case is a very undesirable one; as discussed in Ref. 4, the existence of two propagation modes at the same frequency, one a forward wave, the other a backward wave, allows for propagation with gain in both directions despite the presence of an isolator. As a result, the maser will oscillate instead of offering stable gain. Empirically, however, this situation can be easily diagnosed and there are remedies to rectify it. Therefore, double-valued ω-β relations may be excluded from the present considerations.

With this proviso, it can be seen from Fig. 2(a) that the midband group velocity can be estimated reasonably well from a knowledge of the two cutoff frequencies alone, viz.

$$v_g = 2a\Delta fl/(N - 1) = 2a\Delta f\Delta l. \tag{7a}$$

Here Δf is the frequency width of the passband, Δl is the center-to-center spacing between comb fingers and a is a numerical factor which takes into account the detailed shape of the ω-β curve. Equation (7a) may be rearranged in terms of the group velocity slowing

$$S = \frac{c}{v_g} = \frac{1}{a}\frac{\lambda/2}{\Delta l}\frac{f}{\Delta f} \tag{7b}$$

indicating that slowing is partly a geometric effect, i.e., the compression of a half wavelength into one period of the structure, and partly the effect of compression in the frequency domain, sometimes expressed by a loaded Q. a assumes values of one for a straight line ω-β relation, 1.57 for an inverse cosine, and may in practice be as high as four for a "sagging" ω-β curve. In other words, the uncertainty in estimating the
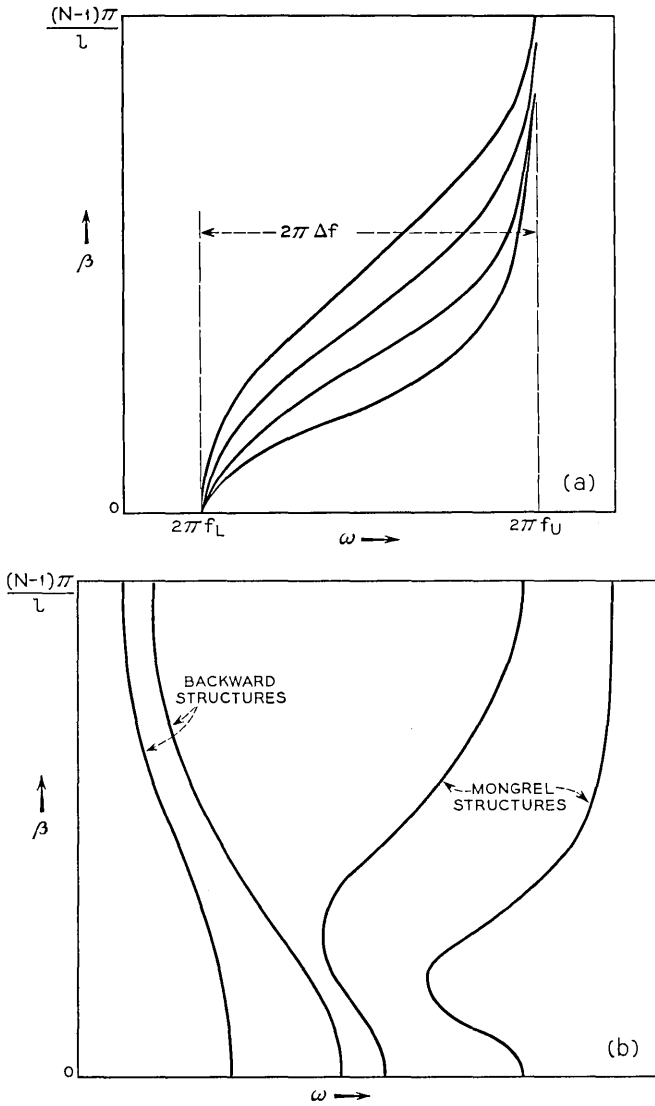
Fig. 2 —(a)Typical forward ω-β diagrams of loaded comb structures with normalized cutoff frequencies. (b) Exceptional ω-β diagrams found in comb structures with extreme dielectric loading.

group velocity or slowing from the cutoff frequencies is not very large, usually less than a factor of two.

Thus, it is clear that a method for calculating the two cutoff frequencies would be an important first step towards an analytical design procedure. Such a mathematical method should be carried out as rigorously as possible. The reason for this may be demonstrated in the following way. If fringe capacity at the finger tips and dielectric loading effects are neglected, the comb structure is electrically equivalent to the Easitron[*] structure. In this approximation, the comb would be a zero passband structure with identical cutoff frequencies like the Easitron. In reality, they differ only because fringe capacity and dielectric loading affect both frequencies to different degrees. Thus, the width of the passband Δf is obtained as a small difference between large numbers, the upper and lower cutoff frequencies, f_U and f_L. To obtain Δf with *fair* accuracy, f_U and f_L must be known with *good* accuracy. Similarly, a small change in the dielectric loading may change f_U and f_L each by a small percentage, but Δf by an appreciable factor. Experience has shown that comb structures with different dielectric loadings may have a passband width Δf anywhere between 1 and 50 per cent of the midband frequency. In other words, as long as the cutoff frequencies are not known, the uncertainty in an estimate of v_g may be almost two orders of magnitude. The computation of the cutoff frequencies would be very useful if it could reduce this uncertainty to about a factor of two. Besides determining the group velocity and hence, indirectly the electronic gain, the cutoff frequencies also define the center frequency and the tunable bandwidth of the TWM. Since it is impossible to match a structure right up to the cutoff frequency, the useful tunable band is well inside the structure passband Δf. An analytical design procedure that allows a reasonably accurate prediction of the cutoff frequencies would clearly be desirable, as center frequency and tunable bandwidth are among the primary TWM specifications.

1.2 *The Function of Slow-Wave Structures in Electron Beam Tubes and TWM's*

A considerable amount of work, both theoretical and experimental, has gone into the study of slow-wave structures for tubes. It would be gratifying if this knowledge could be used in TWM work. Unfortunately,

---

[*] The Easitron was analyzed by L. R. Walker, unpublished manuscript, quoted in Ref. 1. This structure consists of a rectangular waveguide with an array of uniform, identical conductors in the H plane connecting both short walls. It has zero passband, nonpropagating resonances of frequencies where the conductor length is one or more half wavelengths.

this work has only limited applicability to the TWM. This is more readily understood if slow-wave structures for electron beam tubes and for TWM's are compared.

In tubes, slowing factors between 10 and 100 are typical, while in the TWM, slowing of 50 to 1000 is used. This difference influences primarily the mechanical tolerances, which are tighter for higher slowing.

A more fundamental distinction concerns the applicable slowing concept. In an electron beam tube there must be synchronism between the electromagnetic mode propagated on the slow-wave structure and the interacting mode characterized by a charge distribution on the beam. Therefore, the analysis of tubes is concerned with the phase velocity of the slow-wave structure mode. Similarly, in a traveling-wave parametric amplifier there must be synchronism between pump, idler and signal propagation, requiring a phase velocity relation for these three frequencies. In filter circuits the condition of synchronism is usually satisfied only over a small fraction of the total structure bandwidth. By contrast, the amplification by the maser material does not depend on the existence of phase relations along the TWM structure. The maser material may be considered as an incoherent, long-time energy reservoir from which energy is withdrawn upon stimulation by an incident signal and added to the incident signal in a coherent phase preserving fashion. The function of the slowing is merely to "give the signal more time" to interact with the energy stored in the maser material, i.e., to enhance the stimulating gain interaction. Thus, the analysis of TWM's is concerned with the signal group velocity in the structure rather than phase velocity. It is not necessary that v_g be constant over the tunable band. If the gain over the tunable band is required to be constant, then the product $-\chi'' F f / v_g$ (neglecting copper and ferrite losses) should be constant over the band. Experience has shown that this condition can be met over almost the entire passband.

Another point is the interaction mechanism between the active element and the slowing structure. An electron beam interacts with a structure mode via the RF electric field, and the interaction is conventionally represented by an interaction impedance. The interaction of the inverted spins in the maser material with the structure mode takes place via the RF magnetic field, and its strength is measured by the filling factor.

All the differences mentioned have no bearing on the question whether the knowledge of slow-wave structures accumulated in studies directed towards electron beam interaction can be applied to TWM structures. For example, the degree of slowing is not essential for a theoretical anal-

ysis, the group slowing is easily derived by differentiation from the phase propagation, and electric and magnetic interaction terms can be obtained equally well from the field analysis.

The chief difference in slow-wave structures for these two applications lies in their relation to dielectric loading. In a tube, dielectric loading is undesirable and is usually avoided as far as possible. By virtue of its dielectric constant, the glass envelope of a TWT, for example, drags away from the beam some of the electric field energy carried by the helix and thus reduces the gain interaction. In fact, most studies of slow-wave structures for beam tubes pertain to metal structures surrounded by vacuum.

Dielectric loading, being an undesirable side effect for tubes, is an essential and rather beneficial feature in maser structures. Since the gain interaction is magnetic in nature, the interaction of the electric field with dielectrics may be used to advantage without deteriorating the gain interaction. Indeed, it is being used for reducing the over-all maser size, tuning the band center frequency, adjusting the tunable bandwidth or increasing the gain by increased slowing (of course, the items mentioned are not independent). Thus, a high degree of design flexibility can be obtained, even with the identical copper comb, merely by changing the dielectric loading.

For this reason, dielectric loading must be included in any treatment of TWM structures. The present paper is a first contribution to the theoretical treatment of maser structures taking dielectric loading into account. To keep the mathematics reasonably simple, the maser comb geometry, including the dielectric loading, was chosen to be fairly simple. In the laboratory, dielectric loading techniques were developed in which the loading consists of more than one dielectric and has more complex shapes. Work to be published by F. S. Chen has generalized the analysis to take these modifications into account. It also expands the present analysis of the cutoff frequencies into a more general one which allows the prediction of the entire ω-β diagram. This will be particularly valuable in finding criteria to avoid structures having a double-valued "foldover" or "mongrel" ω-β diagram.

II. GENERAL PROBLEM AND APPROACH TO SOLUTION

The problem is to find by analysis the upper and lower cutoff frequencies of the comb-type slow-wave structure as used in a traveling-wave maser (TWM). In particular, this implies taking into account the dielectric effect resulting from loading the comb with maser material or

possibly some other dielectric material and the effect of the fringe capacity at the tips of the comb finger. It was pointed out before that, in a zero-order approximation neglecting both effects, the comb is a zero passband structure.

In the course of this treatment it will be necessary to introduce a number of restrictions and approximations. These are mostly required in order to keep the mathematics manageable. Some other restrictions are introduced in order to have the geometry underlying the calculations correspond to the type of TWM geometry which is presently investigated in the laboratory. These various restrictions and approximations are labeled with lower-case roman numerals for reference in this discussion.

(*i*) *The first restriction pertains to the cross section of the comb fingers. The treatment used here is applicable only to combs with fingers of rectangular cross section.*

This means that it is not possible to apply this type of analysis to a comb having round fingers as used in the original TWM's. It may be mentioned here, however, that it is possible to treat the round-finger comb as long as certain simple frequency or impedance data are available from measurements on scale models, resistance cards or measurements in the electrolytic tank.

Besides being better suited for mathematical analysis, there is another justification for treating combs with rectangular fingers. This has to do with fabrication of combs. There is indication that it is possible to fabricate combs with rectangular fingers not only with greater ease but also with greater perfection. The subject of these fabrication techniques may be discussed at some later date.

A typical comb structure as treated here is shown in Fig. 1. The fingers shown are of square cross section and are spaced by a finger width. It should be emphasized that the general method used here is applicable to any rectangular cross section and spacing, although a great many of the computations are concerned with square fingers spaced by a finger width.

(*ii*) *The next restriction is that maser material (or some other dielectric) is inserted into the comb in the shape of a single rectangular parallelepiped.*

The restriction to parallelepipeds is rather definite. There is a possibility, however, of considering more than one slab of maser material loading the comb. No change in the general analysis is required if two identical slabs are considered which are loaded symmetrically on both sides of the comb. This is shown in Fig. 3(a). The analysis could be carried out also for the case shown in Fig. 3(b) where the maser material is in-
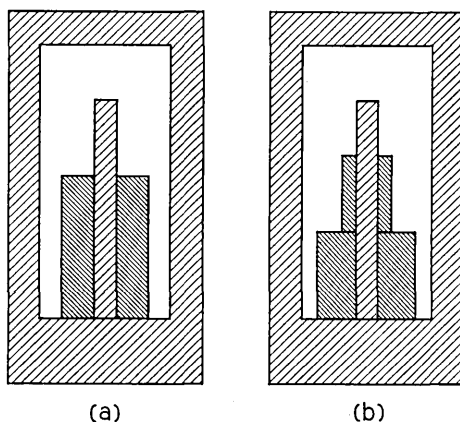
Fig. 3 — Loading geometries.

serted in the form of two pairs of identical slabs. It should be mentioned, however, that the calculation will be appreciably more cumbersome in this case. Although it will not be described in detail, it will be fairly obvious to the reader how the calculations have to be modified to take into account geometries like the one of Fig. 3(b).

(iii) *A further simplifying assumption is that the dielectric loading is assumed to have an isotropic dielectric constant, at least for field components perpendicular to the finger direction.*

This assumption is not too restrictive. An effective dielectric constant may be estimated in the case of a tensor dielectric constant. This estimate will usually be different for either cutoff frequency, since it depends on the electric field configuration. As the tensor components are always of the same order of magnitude, the estimated effective dielectric constant should turn out to be sufficiently accurate for most cases of practical interest.

No provisions have to be made for magnetic permeability. Outside the maser signal line, $\mu' = 1$ for the maser material. Even within the frequency range of the signal line, the deviation of μ' from unity is so small that it can be neglected for all practical purposes as a factor influencing the cutoff frequencies. A similar reasoning applies to the ferrimagnetic isolator. Even though the values of $\mu' - 1$ are larger there, they are less effective due to the very small ferrimagnetic filling factor.

The starting point for the calculation is the phase shift. At one cutoff frequency the phase shift between fingers is zero. This has the consequence that an instantaneous electric field pattern within the comb may

look like Fig. 4(a). Usually, although not necessarily so, this is the case at the lower cutoff frequency, f_L. Throughout the paper this case will be referred to as the "lower cutoff," although the term "zero phase shift case" would be more appropriate. The field pattern is repetitive and shows no field lines from finger to finger since they are on the same potential. It is symmetric with respect to a cross-sectional plane in the structure which contains either the center line of a finger or the center line in the space between two fingers. Therefore the same field pattern is obtained with a single finger if the section of the comb containing this finger is enclosed by a "magnetic wall." A magnetic wall is a fictitious plane on which the electromagnetic field components obey boundary conditions such that the electric field is tangential and the magnetic field normal to the plane. These boundary conditions are opposite from those on a perfect conductor. The perfect conductor is closely approximated
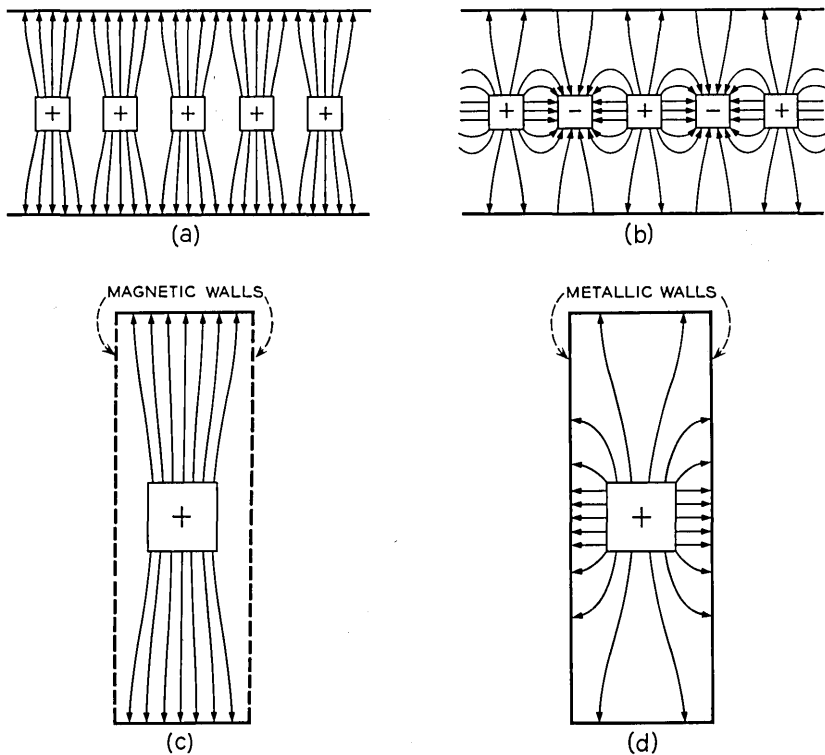

Fig. 4 — Field patterns showing phase shift conditions.

in experiments by high-conductivity metals, whereas the magnetic wall is a mathematical model only. Since the field patterns of Figs. 4(a) and 4(c) are identical, the frequencies will be the same, too. Thus the lower cutoff frequency of the comb can be found as the resonant frequency of the one-finger structure in Fig. 4(c).

A similar reasoning applies to the upper cutoff frequency f_U . Here the phase shift is π between adjacent fingers. An instantaneous field pattern will therefore look like Fig. 4(b). Since adjacent fingers are subject to opposite potential, there are strong electric field components going from finger to finger. The field pattern is symmetric with respect to a cross-sectional plane in the structure which contains the center line of a finger, but antisymmetric with respect to a cross-sectional plane which contains the center line in the space between two fingers. Thus the same field pattern can be realized on a single finger if the section of the comb containing the finger is enclosed by a perfectly conducting (or metallic) wall. This wall will take the place of the plane of antisymmetry in the comb. This is illustrated in Fig. 4(d). Again, identical field patterns require the same frequency. Thus the upper cutoff frequency of the comb can be found as the resonant frequency of the one-finger structure in Fig. 4(d).

The method of determining the resonance frequency of either one-finger model, that of Fig. 4(c) or 4(d), is suggested by Fig. 5. The finger acts essentially as a quarter-wave TEM resonator. At the comb base, this TEM line is terminated in a short. At the finger tip the TEM line is terminated by a nearly perfect "open." This is only slightly modified by fringing electric fields between the finger tip and the surrounding walls. The effect of these fields can be lumped into a fringe capacity C. In principle, C will be different for both cutoff frequencies.

Unfortunately, both capacities C_U and C_L cannot be calculated easily. Therefore, measurements have been made in an analog electrolytic tank setup. A scale model having the cross section of the one-finger lines in Figs. 4(c) and 4(d) was built. This cross section is shown in Fig. 6(a) for the upper cutoff frequency and in Fig. 6(b) for the lower. In the electrolytic tank the electric field lines of the object under study are simulated by the current lines in the tank fluid. No approximation is involved in this analogy. In particular, it is possible to simulate a magnetic wall like that of Fig. 3(c) by an insulating wall. This is done in the cross section used in the lower cutoff analog measurement shown in Fig. 6(b). In the analog measurements, the metal configuration was first lowered to the insulating bottom of the tank as indicated in Fig. 6(c). The resistance measured between electrodes in this fashion is proportional to the impedance of the corresponding TEM mode of the one-finger line; it is
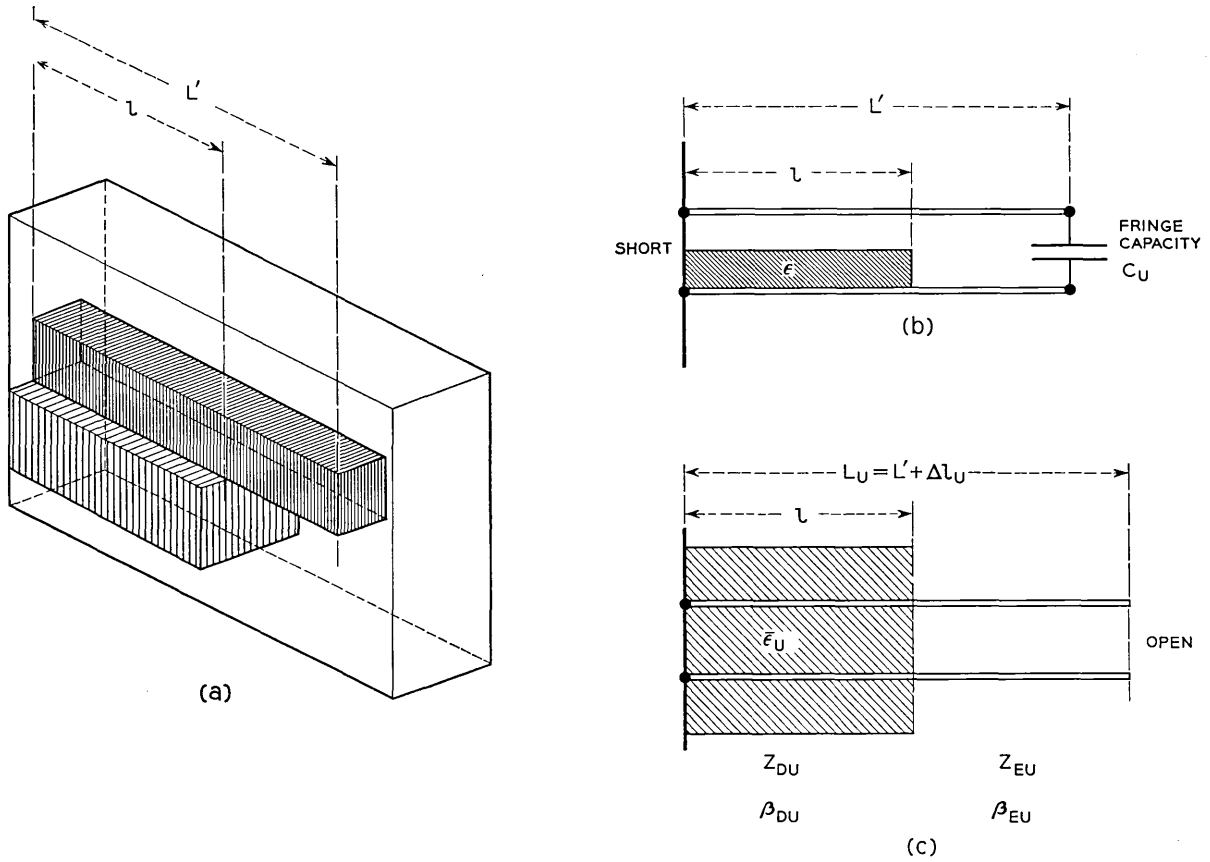
Fig. 5 — Equivalent one-finger line: (a) geometry considered, (b) two-wire line model, (c) simplified equivalent TEM mode line model; use either subscript U for upper, or L for lower, cutoff.
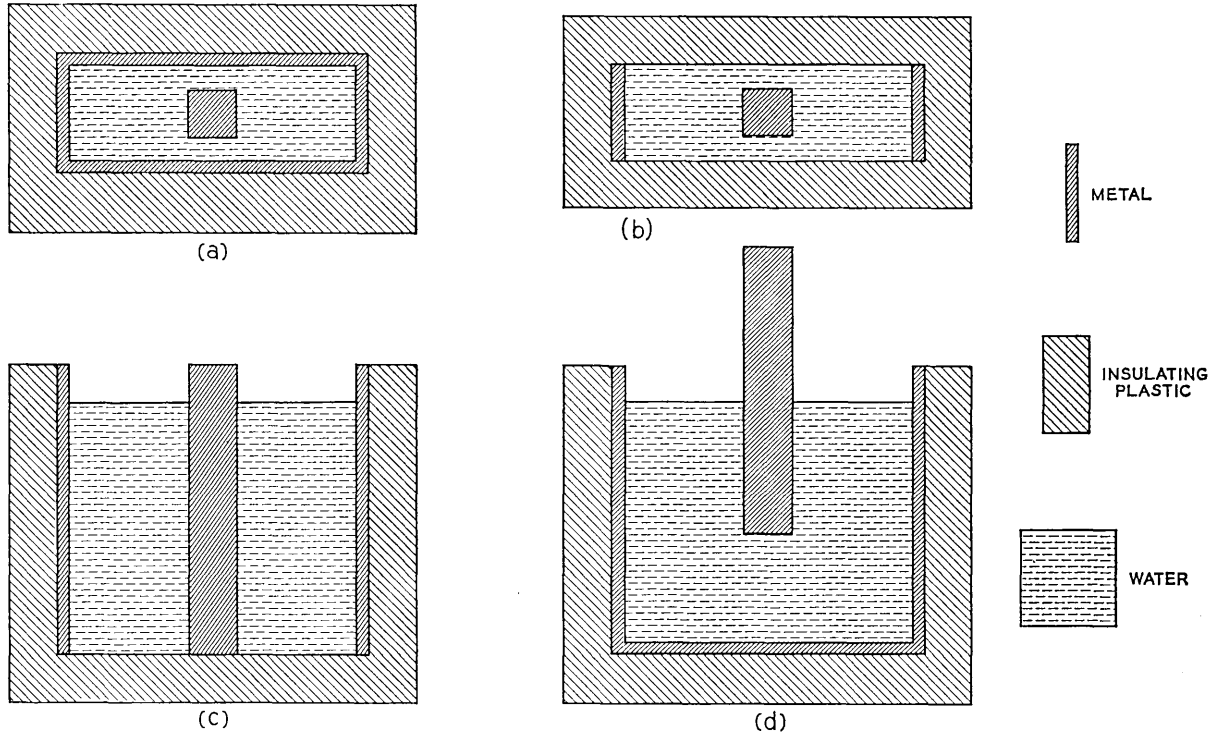
Fig. 6 — Electrolytic tank measurements.

inversely proportional to the capacitance of the line. For a second measurement, the finger was raised to the proper scaled height and a metal plate was placed on the bottom of the tank. The inverse of the resistance so measured is proportional to the capacity of the appropriate length of one-finger TEM line plus the fringe capacity arising from the diverging field pattern beyond the end of the finger. [See Fig. 6(d).] The conductivity of the tap water used was measured also. From these measurements it is possible then to evaluate the fringe capacity as well as impedance and capacitance of the TEM mode on the one-finger line.

In Fig. 7, the fringe capacitance values C_L for the lower cutoff frequency and C_U for the upper cutoff frequency obtained from the tank
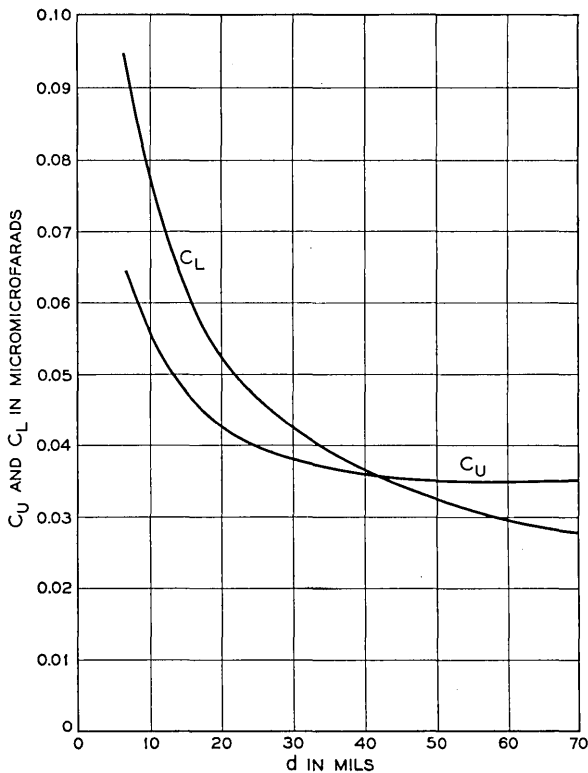
Fig. 7 — Fringing capacitance C_L for lower cutoff frequency and C_U for upper cutoff frequency. The geometry of the comb used includes fingers of cross section 0.040 x 0.040 inch, spaced 0.080 inch on center in a housing 0.240 inch wide (ratio $W_U/D_U = 1.25$). Capacitance is plotted vs spacing d between finger tips and the opposing housing wall.

measurements are shown as a function of the distance d between the finger tips and the opposite waveguide wall. The data are valid for fingers of square cross section, $D_U/2 \times D_U/2 = 0.040 \times 0.040$ inch, spaced center-to-center by $D_U = 0.080$ inch and contained in a housing of width $2W_U + D_U/2 = 0.240$ inch (aspect ratio $W_U/D_U = 1.25$). The W_U and D_U are the dimensions of the empty comb, as shown in Fig. 9 in Section III. It should be mentioned here that these data can be applied to dimensions other than those indicated if one observes two facts. First, if all linear dimensions are scaled simultaneously by some factor, the capacity is scaled by the same factor. Second, experience has shown that the fringe capacity is a very slow function of the ratio W_U/D_U ; no noticeable errors were found when these capacity values were used for W_U/D_U values ranging from 0.75 to 1.5.

Another experimental method to determine the fringe capacity is based on the availability of either a comb structure of exact size or a scale model. The upper and lower cutoff frequencies, f_{EU} and f_{EL} , of the empty, unloaded structure are measured by direct measurement. The result of the measurement can best be expressed in terms of a new effective finger length, L_U or L_L . This is based on the fact that a transmission line of physical length L', shorted at one end and terminated with a small capacity at the other, is electrically equivalent to a somewhat longer transmission line which is shorted at one end and open at the other. The effective finger lengths are different for both cutoff frequencies

$$L_U = \frac{c}{4f_{EU}} = L' + \Delta l_U \qquad (8a)$$

$$L_L = \frac{c}{4f_{EL}} = L' + \Delta l_L. \qquad (8b)$$

Here c is the velocity of light.

The relation between these length dimensions and the fringe capacity involves the characteristic impedance of the line. The fringe capacity follows from

$$\frac{1}{2\pi f_{EU}C_U} = Z_{EU} \tan \frac{2\pi f_{EU}L'}{c} = Z_{EU} \cot \frac{2\pi f_{EU}\Delta l_U}{c} \qquad (9a)$$

and

$$\frac{1}{2\pi f_{EL}C_L} = Z_{EL} \tan \frac{2\pi f_{EL}L'}{c} = Z_{EL} \cot \frac{2\pi f_{EL}\Delta l_L}{c}. \qquad (9b)$$

For the particular structure geometry investigated here in detail, the fringe capacity was determined using these equations and the characteristic impedances derived later in this section. The values of C_L and C_U obtained agree well with these from the tank measurement.

For the subsequent calculations it is assumed that the new effective lengths, L_U and L_L, are known. If, instead, the fringe capacities, C_U and C_L, are known, the new effective lengths can be calculated using the impedances Z_{EU} and Z_{EL}. Since the capacities are small, (9a) and (9b) can be approximated by

$$\Delta l_U = Z_{EU}C_Uc \tag{10a}$$

$$\Delta l_L = Z_{EL}C_Lc. \tag{10b}$$

In this fashion, the problem of Fig. 5(b) is reduced to that of Fig. 5(c). The cutoff frequencies, f_L and f_U, are found as resonance frequencies of a transmission line L_L or L_U long, where one end is shorted, the other open, and a length l is partially loaded with dielectric.

The field pattern in the unloaded part of the transmission line is rigorously a TEM mode. Therefore, the impedance of this line can be found by a resistance card or an electrolytic tank technique. The electrodes are shaped for the model in the same way as the conductors in the unloaded TEM line. Then the impedance of the line is simply equal to the resistance measured in the model provided the resistance per square is adjusted to or scaled to 377 ohms. In addition to this measuring technique, these impedances, Z_{EU} and Z_{EL}, will be determined analytically below. This involves a calculation with good accuracy of the electric field pattern.

The dielectrically loaded section of the transmission line would, if treated with the same rigor, require a much more involved procedure. Therefore, at this point an approximation is introduced.

(iv) *The field configuration in the loaded part of the transmission line can be treated as a TEM mode.*

In reality, this is not true. An exact solution of Maxwell's equations for a TEM-type transmission line having a cross section partly filled with dielectric is not a TEM mode. Instead, the process of matching boundary conditions requires the presence of longitudinal field components. It can be seen, however, that these longitudinal components will become smaller with decreasing frequency and vanish in the zero frequency limit. Thus this approximation implies the representation of a dynamic field configuration by its static analog. The accuracy of such an approximation, therefore, tends to be better the shorter the linear dimensions involved are with respect to the wavelength. In the range of

dimensions used here it is expected that no appreciable loss of accuracy is incurred in this connection.

The consequences of treating the field configuration in the loaded part of the one-finger model as a TEM mode are far-reaching and very helpful for the subsequent analysis. Considering the same metal boundaries as in the unloaded part, the field configuration in the loaded part has to be the same. This follows from the fact that the TEM fields are given as a unique solution to Laplace's equation for the appropriate geometry. Thus one way to treat the loaded part of the one-finger model, consistent with a TEM mode in the same geometry, is by an effective dielectric constant. This allows for a reformulation of approximation (iv):

The part of the transmission line loaded partially by a high dielectric constant material can be treated as if it were loaded uniformly throughout the cross section with a material of a lower "effective" dielectric constant.

This effective dielectric constant will, of course, be different for the upper and lower cutoff frequencies. Using these effective dielectric constants, $\bar{\epsilon}_U$ and $\bar{\epsilon}_L$, the impedances and propagation constants of the loaded section are related to those of the empty section by

$$Z_{DL} = Z_{EL}/\sqrt{\bar{\epsilon}_L} \qquad Z_{DU} = Z_{EU}/\sqrt{\bar{\epsilon}_U} \qquad (11)$$

$$\beta_{DL} = \sqrt{\bar{\epsilon}_L}\,\beta_{EL} \qquad \beta_{DU} = \sqrt{\bar{\epsilon}_U}\,\beta_{EU}. \qquad (12)$$

Here the first indices E and D refer to the empty and dielectrically loaded line, the second indices L and U to the lower and upper cutoff frequencies. The propagation constants in the empty TEM line are, of course, identical to that in vacuum

$$\beta_{EL} = (2\pi f_L/c) \qquad \beta_{EU} = (2\pi f_U/c). \qquad (13)$$

Assuming for the moment that the effective finger lengths, L_U and L_L, the characteristic impedances of the empty line, Z_{EU} and Z_{EL}, and the effective dielectric constants, $\bar{\epsilon}_U$ and $\bar{\epsilon}_L$, are known, the cutoff frequencies, f_U and f_L, can be calculated. The procedure is to match voltage and current at the boundary between the loaded and unloaded section of the line. This results in impedance equations

$$Z_{EU}\cot\beta_{EU}(L_U - l) = Z_{DU}\tan\beta_{DU}l \qquad (14a)$$

$$Z_{EL}\cot\beta_{EL}(L_L - l) = Z_{DL}\tan\beta_{DL}l. \qquad (14b)$$

These are rewritten in a more convenient form

$$\sqrt{\bar{\epsilon}_U} = \tan\left[\frac{\pi}{2}\sqrt{\bar{\epsilon}_U}\,\frac{l}{L_U}\frac{f_U}{f_{EU}}\right]\tan\left[\frac{\pi}{2}\left(1 - \frac{l}{L_U}\right)\frac{f_U}{f_{EU}}\right] \qquad (15a)$$

$$\sqrt{\bar{\epsilon}_L} = \tan\left[\frac{\pi}{2}\sqrt{\bar{\epsilon}_L}\,\frac{l}{L_L}\frac{f_L}{f_{EL}}\right]\tan\left[\frac{\pi}{2}\left(1 - \frac{l}{L_L}\right)\frac{f_L}{f_{EL}}\right]. \qquad (15b)$$

These equations are identical for lower and upper cutoff frequencies. They do not contain the characteristic impedances explicitly. They are solved in the following way.

$\sqrt{\bar{\epsilon}_U}$ or $\sqrt{\bar{\epsilon}_L}$ is considered a given parameter. Then the frequency ratio f_U/f_{EU} or f_L/f_{EL} is a function of l/L_L or l/L_U. This function requires the solution of transcendental equation (15a) or (15b). Numerical values were obtained by machine computations using the IBM 7090. The results are plotted in Fig. 8.

This graph can then be used to determine the upper and lower cutoff frequencies of the loaded comb structure. It is assumed here that the upper and lower cutoff frequencies of the empty comb, f_{EU} and f_{EL}, and connected with them, the effective finger lengths, L_U and L_L, are known. They are best determined by measurement, but they could also be calculated from the fringe capacity and the characteristic impedance. The quantity yet to be evaluated is the effective dielectric constant, $\bar{\epsilon}_U$ and $\bar{\epsilon}_L$, before the cutoff frequencies can be read from the graph in Fig. 8. It will be necessary, however, to work out the electric field pattern within the unloaded comb, then in the loaded section, including the respective characteristic impedances, before the effective dielectric constant can be obtained.

III. FIELD PATTERN AND CHARACTERISTIC IMPEDANCE OF UNLOADED COMB

3.1 *Upper Cutoff Frequency*

The electric field pattern of the unloaded one-finger model will look about like Fig. 9(a). This geometry is, unfortunately, too complicated for a closed analytical treatment. On the basis of the geometry and the mathematical tools at hand, the following approach may be suggested. The area available to the electric field is divided into four regions, two equivalent regions of type A and two equivalent regions of type B, as shown in Fig. 9(b). Two further approximations are then necessary.

(v) *The electric field in the regions A can be represented as a homogeneous, parallel plate condenser field.*

(vi) *The electric field in the regions B can be represented by the field produced by an infinitely thin metal fin inserted in a rectangular enclosure of corresponding dimensions.*

These approximations are illustrated in Figs. 9(c) and 9(d). Along the joints of regions A and B the field thus assumed is discontinuous. In reality, it is inhomogeneous near the boundary of region A, and it is less inhomogeneous than assumed near the boundary of region A because there is only a 90° bend, not a 180° bend as in the model used. These
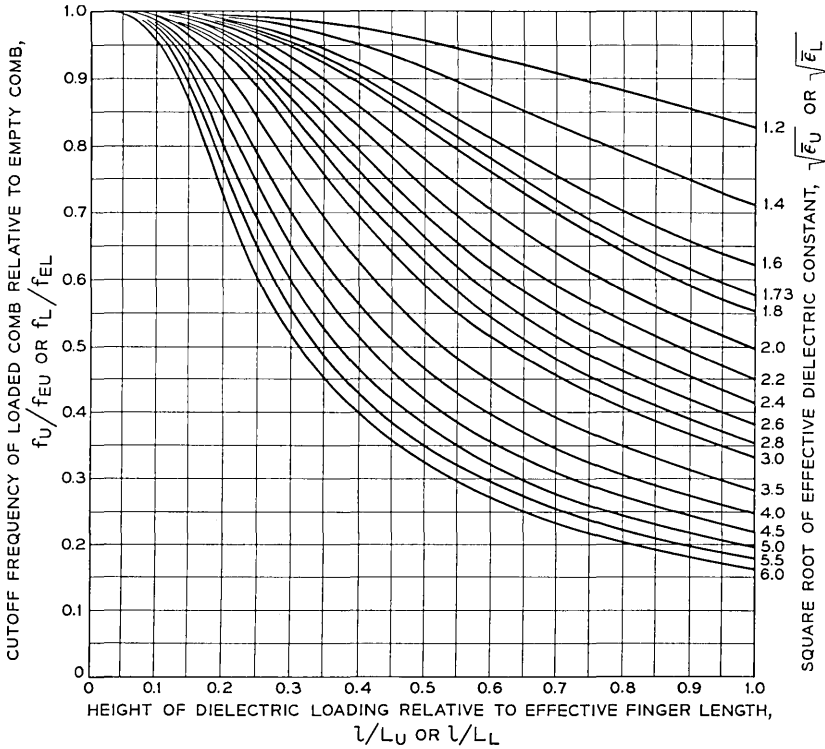
Fig. 8 — Plot of numerical values obtained by machine computations.

discrepancies of the field model from what would be expected should in reality be very small, particularly if the gap between finger and wall, the dimension $W_A = \frac{1}{2}(1 - r)D_U$ defined in Figs. 9(a) and 9(e), is small compared to other dimensions. This is so in cases of practical interest.

As far as the impedance is concerned, the two regions A and the two regions B are in parallel. The impedance of a region A is simply the ratio of its dimensions multiplied by free-space impedance. The impedance of region B is not as easily found. It is possible, however, to use a conformal transformation which maps the region B into a parallel plate geometry. This is schematically indicated in Fig. 9(e). The transformation actually utilized consists of the consecutive application of two transformations, each using elliptical functions. The procedure, including the mathematical details of the conformal transformation by elliptical functions, is outlined in the Appendix.

It is known from the theory of conformal mapping by functions of
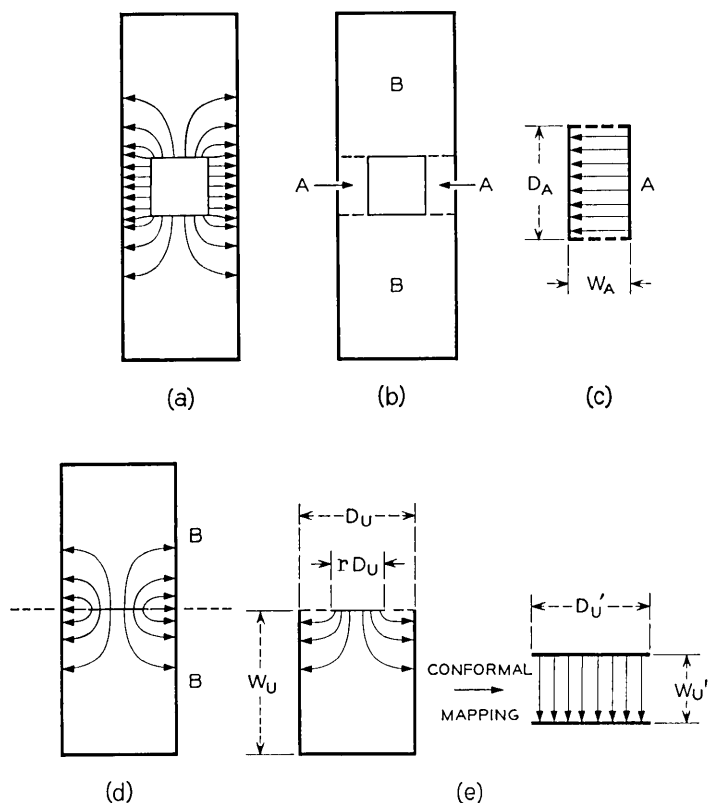
Fig. 9 — Analysis for upper cutoff frequency: (a) real field patterns, (b) regions used for analysis, (c) homogeneous field assumed in region A, (d) fin field assumed in region B, (e) fin field equivalent to homogeneous field.

complex variables that the geometry is preserved in infinitesimal regions. In particular, it is clear that infinitesimal squares with boundaries formed by field lines and equipotential lines continue to be squares. Since the impedance can be thought of as composed of the impedance of these infinitesimal squares, partly in parallel and partly in series as indicated by the over-all geometry, it follows finally that the impedance of the two transmission lines of Fig. 9(e) is the same.

The geometry before transformation is characterized by the two ratios: W_U/D_U and r. Thus W_D'/D_U' will be a function of both of these ratios. So far only combs with $r = \frac{1}{2}$ have been investigated in practice. For convenience, therefore, the subsequent calculations are carried out for this value of r. This implies a further restriction.

(vii) In the numerical calculations to follow, only comb geometries with the finger width as large as the gap between fingers are considered.

From a mathematical point of view, this restriction is somewhat arbitrary. Any other choice of r, the ratio of finger width to length of period, however, would necessitate another application of the elliptic integral conformal transformation.

With $r = \frac{1}{2}$, W_U'/D_U' is a single-valued function of W_U/D_U. This function is plotted in Fig. 10, An interesting feature of this graph is that W_U'/D_U' goes asymptotically to $\frac{1}{2}$; it reaches this value to within 2 per cent at $W_U/D_U = 0.65$. The physical interpretation of this observation is as follows. For $W_U/D_U > 0.65$, essentially all the field lines originating at the center fin in Fig. 9(d) terminate on the side wall; none reach the opposite end wall. Therefore, this wall can be moved out toward infinity with no noticeable effect on the impedance at the upper cutoff frequency.

The characteristic impedance of the empty structure at the upper cutoff frequency can now be given. It is

$$Z_{0U} = 377 \text{ ohms} \bigg/ \left(2 \frac{D_A}{W_A} + 2 \frac{D_U'}{W_U'} \right). \qquad (16a)$$

An important special case is one where, first, W_U/D_U is greater than 0.65 so that the asymptotic value $W_U'/D_U' = \frac{1}{2}$ applies and where, second, the fingers have a square cross section so that $W_A/D_A = \frac{1}{2}$. Then the characteristic impedance is simply

$$Z_{0U} = \frac{1}{8} \times 377 \text{ ohms} = 47.1 \text{ ohms}. \qquad (16b)$$
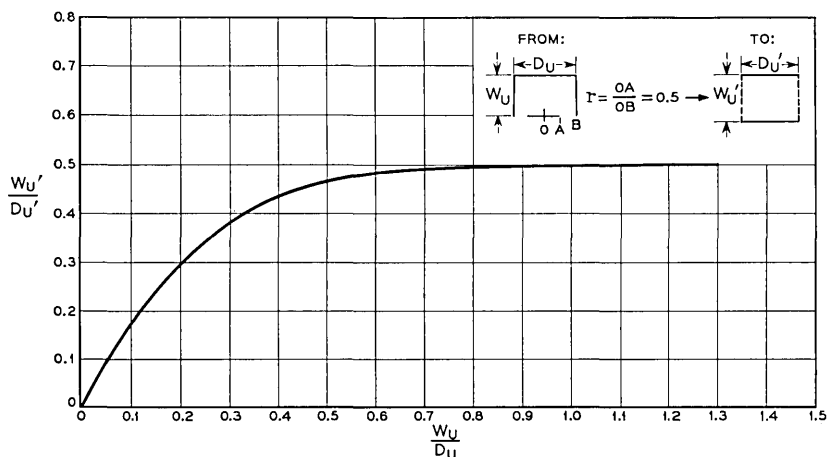
Fig. 10 — Conformal transformation for upper cutoff.

Since the partial impedances are equal, it also follows in this case that the total stored energy is equally distributed between the four regions A, A, B, B. This remark may be helpful in estimating the filling factor.

3.2 Lower Cutoff Frequency

The procedure here is quite similar to that in the case of the upper cutoff frequency. The field pattern is illustrated in Fig. 11(a). The cross-section area available to the electric field is divided into four regions, two electrically equivalent regions of type A and two regions of type B, as shown in Fig. 11(b). Again two approximations are required.

(viii) *The electric field in region A is so small that it can be neglected.*
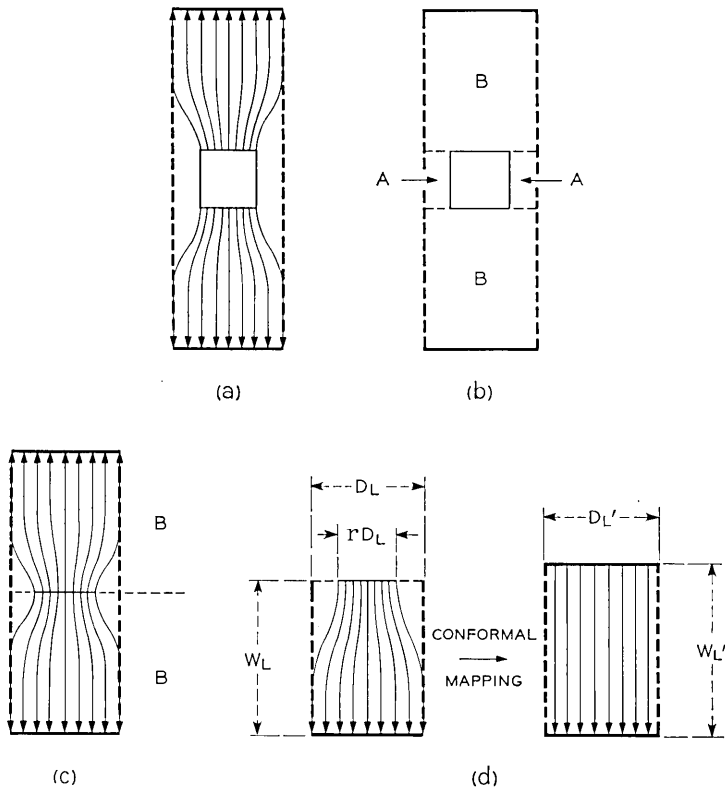
Fig. 11 — Analysis for lower cutoff frequency: (a) typical electric field pattern, (b) regions used for analysis, region A assumed field-free, (c) fin field assumed in region B, (d) fin field equivalent to homogeneous field.

(ix) *The electric field in region B can be represented by the field produced by an infinitely thin metal fin inserted into a rectangular enclosure with appropriate dimensions and boundary conditions.*

It is apparent that the approximation (viii) is justified. Only very small fringing fields will exist in region A. The implication of approximation (ix) is indicated in Fig. 11(c). It should also be very well justified, since there is no essential difference between the idealized field pattern and the real one. Region B can be transformed into a simple parallel plate geometry. This is indicated in Fig. 11(d). The transformation again consists of two consecutive conformal mappings by means of elliptic functions. The procedure is outlined in the Appendix. The impedance of region B is simply given by the aspect ratio W_L'/D_L' of the parallel plate geometry resulting from the transformation, multiplied by the free-space impedance. This resulting ratio W_L'/D_L' is a function of two ratios, r and W/D. For mathematical convenience and because of practical importance, only comb geometries with $r = \frac{1}{2}$ are considered in the subsequent calculations. For other ratios r, a new evaluation of the elliptical transformation is necessary. Thus restriction (vii) is invoked here, too.

(vii) *In the numerical calculations which follow, only comb geometries with the finger width equal to the gap width between fingers are considered.*

The single-valued function W_L'/D_L' of W_L/D_L with the parameter $r = \frac{1}{2}$ is shown in Fig. 12. The characteristic impedance of the empty structure at the lower cutoff frequency is then given by

$$Z_{0L} = \frac{W_L'}{2D_L'} \times 377 \text{ ohms.} \qquad (17a)$$

As long as $W_L/D_L > 0.2$, it is seen from the graph that this can be approximated by

$$Z_{0L} = \frac{1}{2}\left(\frac{W_L}{D_L} + 0.11\right) 377 \text{ ohms.} \qquad (17b)$$

The asymptotically linear curve in Fig. 12 and this last equation suggest an almost obvious interpretation. The electrical behavior of region B is essentially the same as that of a parallel plate geometry having the same width $D_L' = D_L$, but a slightly greater distance between plates, $W_L' > W_L$. Also, the asymptotic slope for the curve is unity. Considering a geometry with $W_L/D_L > 0.2$, this would mean the following. If W_L is increased further, the electric field pattern near the fin stays the same, while the added volume away from the finger is taken up by a homogeneous electric field.
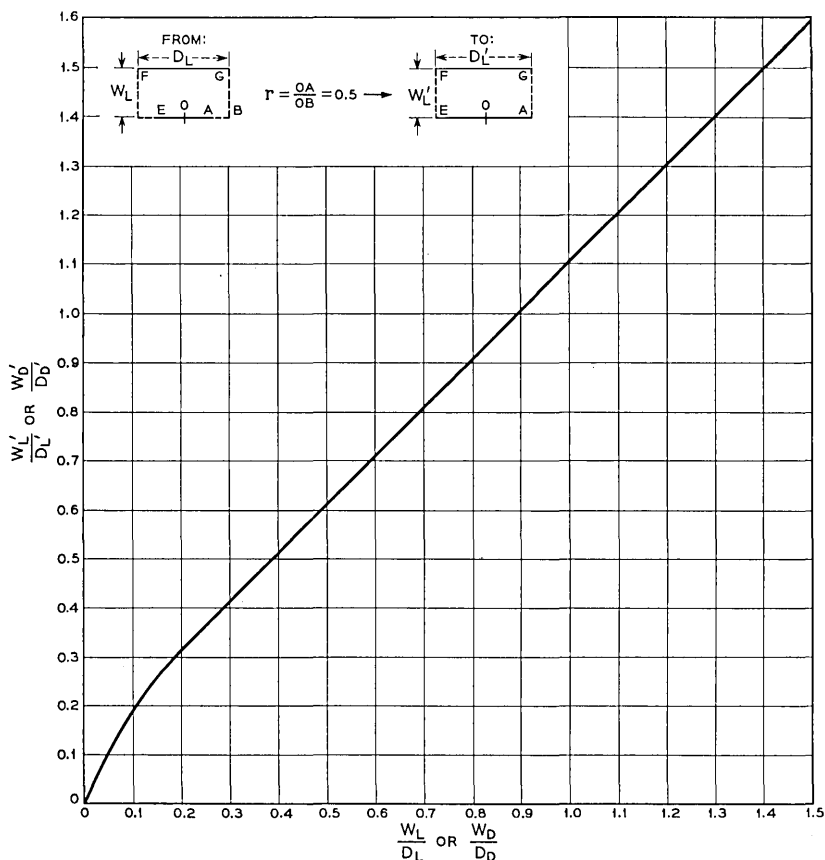
Fig. 12 — Conformal transformation for lower cutoff.

IV. CAPACITANCE AND EFFECTIVE DIELECTRIC CONSTANT OF COMB PARTIALLY FILLED WITH DIELECTRIC

It was mentioned that the electromagnetic field configuration in the comb line partially loaded with dielectric should be treated as a TEM mode. It was pointed out that this is equivalent to finding a static solution of the electric field problem. Thus the problem here is to find the static value of the capacitance per unit length of the loaded finger line. The difference in electrical behavior of the loaded line compared to the unloaded line is then fully expressed by an effective dielectric constant. This effective dielectric constant is simply the ratio of the static capacitance of the loaded line to the capacitance of the unloaded line.

4.1 *Upper Cutoff Frequency*

The field pattern in the presence of one dielectric slab is illustrated in Fig. 13. It is seen that the dielectric is present in one of the regions called B before. The usual boundary conditions for the continuity of the tangential E vector and of the normal D vector have to be observed in fitting together the electric field pattern inside and outside the dielectric.

At first sight it seems that no difficulty is incurred in this respect at the boundary of the dielectric. In the model chosen for the field configuration, the field lines run parallel to the boundary both in regions A and B. The boundary condition for tangential electric field seems to apply, with the consequence that the field pattern remains the same in the dielectric as before in the unloaded region B. Calculations are based on this assumption, and they are presumably of sufficient accuracy for present purposes.

There is a small error in this assumption. It was pointed out before that the two models chosen to represent the field in regions A and B do not match at the boundary. In the models, the field in A is homogeneous, that in B strongly inhomogeneous. The real field at the boundary of A and B should be somewhere between these two extremes. It is expected,
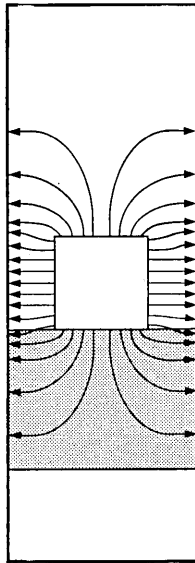
Fig. 13 — Electric field pattern at upper cutoff with dielectric loading present.

therefore, that the error in the impedance calculation of the empty comb at the upper cutoff frequency is negligible. The same is not necessarily true in the presence of dielectric. The real, inhomogeneous field in the region near the boundary of A and B will be disturbed by the insertion of dielectric. The deviations of the real field from that used in the calculations — homogeneous in A, elliptic function field in B — are now accentuated by a high dielectric constant rather than evened out as in the empty comb. This will lead to an error in the calculation of the capacitance and the effective dielectric constant. Hence it is not trivial that the approximations (v) and (vi) are still reasonably good in the presence of dielectric. Fortunately, it can be argued that the error incurred by this approximation is still negligible within the accuracy sought for here and with respect to typical structure geometries and dielectric constants considered. A formulation of the approximation follows.

(x) *In the presence of dielectric loading, the static electric field can still be represented by a homogeneous, parallel plate field in region A and the field of a metal fin inside a rectangular enclosure in region B filled by the dielectric.*

The next concern is the other boundary of the dielectric away from the finger. Here the field lines cross the boundary at all directions between tangential and perpendicular. It would be very difficult to apply boundary conditions to this field pattern. Therefore another restriction is introduced.

(xi) *The calculation is restricted to dielectric loadings thick enough so that essentially the total electric field energy of region B is contained within the dielectric.*

The numerical implication of this restriction follows directly from Fig. 10. It is assumed that the fingers are as wide as the gap between them. From the graph the following fact can be deduced. If a geometry is considered where W_U is considerably larger than D_U, then 98 per cent of the electric field energy is concentrated in a rectangle near the finger, D_U wide and 0.65 D_U deep. Restriction (xi) thus implies that only dielectric slabs which have a thickness of at least 0.65 times the length of a period of the comb are considered.

Fortunately, this restriction does not exclude any cases of practical interest. Since the field configuration on the finger is treated here as a TEM mode, the filling factor in the plane perpendicular to the finger is the same for the dielectric and the magnetic field energy. Thus, slabs thinner than indicated by restriction (xi) would also have a reduced gain interaction near the upper cutoff frequency, since not all of the magnetic field energy of region B would be contained in the maser material. Gain

is still at a premium in present TWM development, and thus it does not seem to be necessary to treat cases other than those restricted by (xi).

It is now possible to write down the capacitance and the effective dielectric constant. By comparison with (16a), it is seen that the capacitance per unit length of the empty one-finger line is

$$c_{EU} = \epsilon_0 \left[2\frac{D_A}{W_A} + 2\frac{D_U'}{W_U'} \right]. \tag{18}$$

(Lower case c is used to distinguish this quantity from the fringe capacity C_U.) With dielectric loading on one side of the finger

$$c_{DU} = \epsilon_0 \left[2\frac{D_A}{W_A} + (\epsilon + 1)\frac{D_U'}{W_U'} \right] \cdots . \tag{19a}$$

Similarly, if the dielectric is loaded on both sides of the finger

$$c_{DU} = \epsilon_0 \left[2\frac{D_A}{W_A} + 2\epsilon\frac{D_U'}{W_U'} \right] \cdots . \tag{19b}$$

The effective dielectric constant is then simply, for loading on one side

$$\bar{\epsilon}_U = [(\epsilon + 1) + 2b]/[2 + 2b] \tag{20a}$$

and for loading on both sides

$$\bar{\epsilon}_U = (\epsilon + b)/(1 + b) \tag{20b}$$

with

$$b = D_A W_U'/W_A D_U'. \tag{21}$$

Most important perhaps for present applications is the case where, first, the fingers are square so that (16b) applies and where, second, the dielectric is ruby with an isotropic average dielectric constant of $\epsilon = 9$. In that case, for loading on one side

$$\bar{\epsilon}_U = 3 \tag{22a}$$

and for loading on both sides

$$\bar{\epsilon}_U = 5. \tag{22b}$$

4.2 *Lower Cutoff Frequency*

The field pattern in the presence of one dielectric slab is illustrated in Fig. 14. The dielectric fills part of the region called B before. For the evaluation of the capacitance it is significant that restriction (xi) is applied here, too. Then the following approximation can be made.
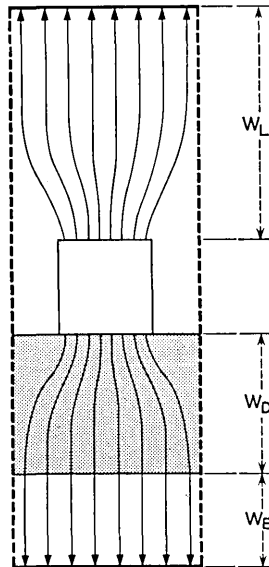
Fig. 14 — Electric field pattern at lower cutoff with dielectric loading present.

(xii) *In the presence of dielectric loading, the static electric field can be represented in the following way: There is zero field in region A; in the dielectric there is a field like that produced by a metal fin in a rectangular enclosure, having the dimensions of the dielectric and subject to appropriate boundary conditions. The field past the dielectric is a homogeneous parallel plate field.*

It can be argued that these approximations are well justified. There is no potential difference between fingers; hence region A should be field-free except perhaps for some very small fringe fields. In connection with 17(b) it was shown that the field has its inhomogeneities near the finger, whereas the field region near the wall is reasonably homogeneous.

The capacitance per unit length of the loaded one-finger model can now be given. For the empty line it is

$$c_{EL} = 2\epsilon_0 (D_L'/W_L').$$ (23)

For the loaded line, the capacitance is obtained from two contributions in parallel, one from each side of the finger. The capacitance of the loaded side comes from two contributions in series: one from the dielectric, involving an elliptical transformation using the dimensions of the dielectric, and one a parallel plate contribution from the space behind the dielectric. Thus, for dielectric loading on one side

$$c_{DL} = \epsilon_0 \left[\frac{D_L'}{W_L'} + \left(\frac{W_D'}{\epsilon D_D'} + \frac{W_E}{D_E} \right)^{-1} \right] \tag{24a}$$

and for loading on both sides

$$c_{DL} = 2\epsilon_0 \left/ \left(\frac{W_D'}{\epsilon D_D'} + \frac{W_E}{D_E} \right) \right. . \tag{24b}$$

Here W_D and D_D are the physical dimensions of the dielectric cross section per one-finger line, W_E and D_E the dimensions of the empty space behind the dielectric. W_D'/D_D' is obtained from W_D/D_D by means of the elliptical transformation illustrated in Fig. 11.

The effective dielectric constant can now be evaluated as the ratio c_{DL}/c_{EL}. The formulas, however, turn out to be fairly long. They are given here, therefore, only for the case that the approximation in (17b) is valid both for the empty structure and the dielectric. It is further observed that

$$D_L = D_D = D_E = D$$

and

$$W_L = W_D + W_E .$$

Then the effective dielectric constant for dielectric loading on one side is

$$\bar{\epsilon}_L = \tfrac{1}{2} \frac{2\epsilon W_L - (\epsilon - 1)W_D + (\epsilon + 1)0.11D}{\epsilon W_L - (\epsilon - 1)W_D + 0.11D} \tag{25a}$$

and similarly, for loading on both sides

$$\bar{\epsilon}_L = \epsilon \frac{W_L + 0.11D}{\epsilon W_L - (\epsilon - 1)W_D + 0.11D} . \tag{25b}$$

It is seen that the effective dielectric constant is a function of ϵ, W_D/W_L and D/W_L. Once a particular structure geometry has been picked, then D/W_L is known. If a particular maser material is selected, ϵ is known. Then $\bar{\epsilon}_L$ is a unique function of the relative loading thickness, W_D/W_L. One example of such a function is given in Fig. 15. For convenience in using the graph of Fig. 8, the square root $\sqrt{\bar{\epsilon}_L}$ is given instead of $\bar{\epsilon}_L$. Curves for effective constants based on other parameters can easily be calculated using either (25a) or (25b).

V. EXAMPLE FOR DESIGN PROCEDURE

In Sections III and IV the empty and the dielectrically loaded comb structure were evaluated. Field pattern, impedance and propagation constants were obtained for both the upper and lower cutoff frequencies.
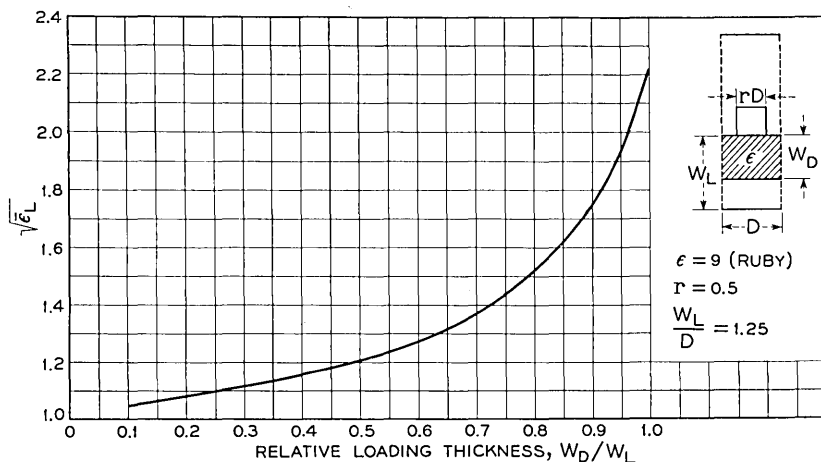
Fig. 15 — $\sqrt{\epsilon_L}$ vs relative loading thickness, one side only, loaded at lower cutoff frequency for parameters indicated.

With this information at hand, it is now possible to arrive at a numerical design procedure. The aim is to predict the cross-sectional dimensions of a dielectric parallelepiped which will simultaneously tune the upper and lower cutoff frequencies of the comb structure to some predetermined values. Of course, it is not possible to ask for completely arbitrary design cutoff frequencies. Obviously there are limits to the amount of tuning which can be achieved by a given dielectric material within a given comb geometry. These limits can also be determined easily by the analysis.

The design procedure follows the outlines given briefly at the end of Section II. It can now be described in general terms. Perhaps it is advantageous, however, to illustrate the procedure by means of an example. The example to be described is a case of a "design on paper." That is to say, the design calculations can be made entirely on the basis of calculable values. It is not necessary to fabricate a size or scale model of the comb structure under consideration in order to determine certain values by measurement. The only empirical value required is the fringe capacity between finger tip and the structure enclosure; this may be obtained from Fig. 7.

One interesting and valuable feature of the design procedure is that of independently setting the upper and lower cutoff frequencies. This is possible because the upper cutoff frequency can be controlled by adjusting the height l of the dielectric loading alone, and because it is not dependent in any way on the dielectric thickness W_D as long as W_D ex-

ceeds a certain small minimum value. Then the dimension W_D can be used to control the lower cutoff frequency independently.

As an example for the design procedure, a comb structure is considered with the following dimensions:

(a) finger length 0.400 inch

(b) spacing between fingers 0.040 inch

(c) finger cross section 0.040 square inch

(d) wall-to-wall spacing of enclosure 0.240 inch.

As further information, the fringe capacity was measured in an electrolytic tank model and was found to be (see Fig. 7 for gap spacing greater than 70 mils):

(e) fringe capacity $C_L = 0.025$ $\mu\mu F$, $C_U = 0.035$ $\mu\mu F$.

The problem considered is that of finding the dimensions for a single ruby parallelepiped which brings the upper cutoff frequency to 4200 mc and the lower cutoff frequency to 3210 mc.

First step: Find effective finger length at the upper cutoff frequency of the empty comb.

Equation (10a) applies for the increase in length and (16b) applies for the impedance; thus

$$\Delta l_U = Z_{EU} C_U c$$

$$= 47.1 \times 0.025 \times 10^{-12} \times 3 \times 10^{10}$$

$$= 0.035 \text{ cm.}$$

The effective length for upper cutoff is then [see (8a)]

$$L_U = L' + \Delta l_U$$

$$= 2.54 \times 0.400 + 0.035 = 1.051 \text{ cm.}$$

This corresponds to an upper cutoff frequency for the empty comb [see (8a)]

$$f_{EU} = c/4L_U$$

$$= 7150 \text{ mc.}$$

Thus the design specification

$$f_U = 4200 \text{ mc}$$

is equivalent to specifying a ratio of

$$f_U/f_{EU} = 0.587.$$

Second step: Find in an analogous way the effective finger length at the lower cutoff frequency of the empty comb.

Equation (17b) applies for the impedance. From the dimensions given, $W_L = 0.100''$, $D_L = 0.080''$, hence

$$Z_{EL} = \frac{1}{2}\left(\frac{W_L}{D_L} + 0.11\right) 377 \text{ ohms}$$

$$= 256 \text{ ohms.}$$

The addition to length is given by (10b)

$$\Delta l_L = Z_{EL}C_L c$$

$$= 256 \times 0.035 \times 10^{-12} \times 3 \times 10^{10}$$

$$= 0.268 \text{ cm.}$$

The effective length for lower cutoff becomes

$$L_L = L' + \Delta l_L$$

$$= 0.400 \times 2.54 + 0.268$$

$$= 1.284 \text{ cm}$$

corresponding to a cutoff frequency for the empty comb

$$f_{EL} = c/4L_L$$

$$= 5840 \text{ mc.}$$

The design specification of

$$f_L = 3210 \text{ mc}$$

is thus equivalent to specifying a ratio

$$f_L/f_{EL} = 0.55.$$

Third step: Satisfy the upper cutoff frequency specification by choosing an appropriate dielectric height l without regard for W_D, the thickness of the loading. This is possible because, as mentioned before, the effective dielectric constant at the upper cutoff frequency is independent of loading thickness. The effective dielectric constant, $\bar{\epsilon}_U$, for one-sided loading with ruby is 3 from (22a); thus

$$\sqrt{\bar{\epsilon}_U} = 1.73.$$

Consulting Fig. 8 for the dielectric height which makes $f_U/f_{EU} = 0.59$ with the parameter $\sqrt{\bar{\epsilon}_U}$, it is seen that

$$l/L_U = 0.96.$$

Hence, the dielectric loading height should be

$$l = 0.96 \times 1.051 = 1.010 \text{ cm}$$

$$= 0.398''$$

In other words, the dielectric loading height turns out to be very nearly the same as the finger length.

Fourth step: Satisfy the lower cutoff frequency specification by choosing an appropriate thickness W_D of the dielectric loading. This is done by the following successive measures.

From the loading height l just determined find

$$l/L_L = 1.010/1.284$$

$$= 0.79.$$

Enter the graph of Fig. 8 with $l/L_L = 0.79$ and $f_L/f_{EL} = 0.55$. The value interpolated at the point having these two coordinates is

$$\sqrt{\bar{\epsilon}_L} = 2.22.$$

The graph of Fig. 15 is valid for present calculations; entering this last value into the graph it is found that

$$W_D = W_L$$

hence

$$W_D = 0.100 \text{ inch.}$$

The final answer, then, is that the comb described initially will have the specified cutoff frequencies if a slab of ruby of height 0.398 inch and of width 0.100 inch is inserted.

An experiment was carried out to check the results of this calculation. The two cutoff frequencies of a comb as specified above were measured after inserting a single slab of polycrystalline high density alumina (dielectric constant ≈ 9.3) with cross-sectional dimensions of 0.400 inch and 0.100 inch. The cutoff frequencies measured were 4200 mc and 3210 mc respectively. These frequencies were then specified as design frequencies for the above example. The close agreement between the actual dimensions of the alumina slab and those calculated by the present recipe is gratifying. It may be argued, however, that the obtained agreement is somewhat fortuitous. In particular, one should expect that the fringe capacity is altered if the dielectric loading extends all the way along the fingers up to the finger tips. To investigate the accuracy of the present analysis, a series of systematic measurements was made.

For this study a number of short sections of comb structures were built and tested. They all had finger dimensions of 0.040 × 0.040 × 0.445 inch, and the fingers were spaced 0.080 inch on center. The structures were loaded symmetrically with two slabs of high-density poly-crystalline alumina (dielectric constant quoted to be 9.3) of full finger height. The geometry and the result of the measurements are shown in Fig. 16. In two series of measurements, the fraction of the housing width filled by the alumina loading, W_D/W_L, was held at 0.90 and 0.95, respectively, while the gap width between the finger and the housing wall, $W_L = W_U$, was varied in the range $0.75D = 0.060$ inch, $D = 0.080$ inch, $1.25D = 0.100$ inch and $1.5D = 0.120$ inch. From the analysis, it is known that f_U should be independent of these dimensional changes. This is borne out by the experiment. Both the experimental points and the solid line for the theoretical value of f_U show the frequency independence. It is observed, however, that the experimental frequencies are 3.5 per cent higher. A somewhat greater disagreement is found for the lower cut-off frequency, which seems to indicate a systematic trend between theory and experiment. It can be said, however, that the largest deviations are
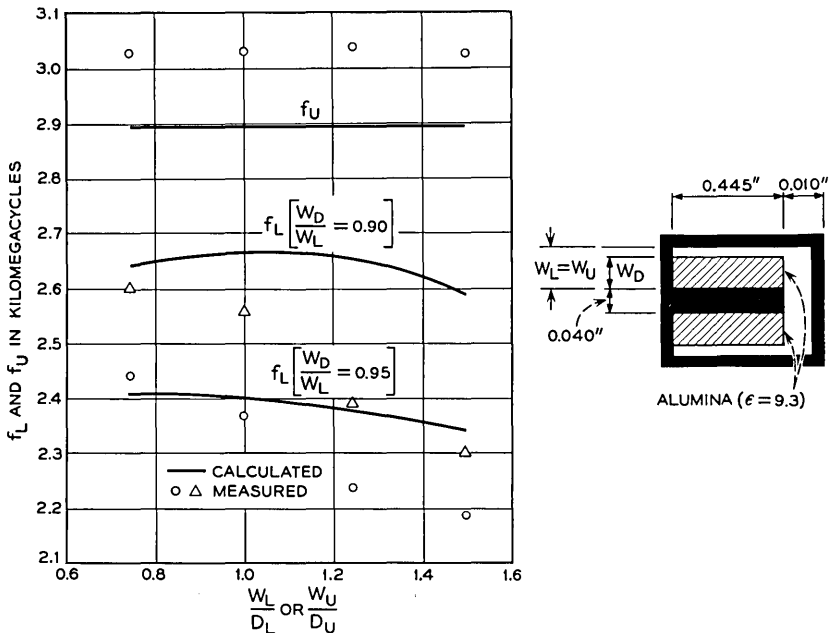
Fig. 16 — Examples of measured and calculated cutoff frequencies; the insert shows the comb geometry used.

10 per cent and that the typical discrepancy between theory and experiment is less than 5 per cent. The chance for greater systematic errors increases, of course, if comb and loading geometries are considered which comply less rigorously with the restrictions and approximations made in the text.

The numerical examples shown demonstrate that dielectric loading indeed decreases the fundamental passband frequency of the empty comb by a very appreciable factor. A one-sided loading with ruby may reduce the frequencies by a factor of 1.7, while double-sided loading may lead to a reduction by a factor 2.5. Still greater reductions may be obtained by using dielectric materials with higher dielectric constants and by modified comb geometries, in particular by changing the finger cross section from square to rectangular. It is also clear from this treatment that the shaping of the dielectric loading can be used to vary the degree of slowing within wide limits. These remarks may suffice here to illustrate the prominent role of dielectric loading techniques in the field of TWM development which was pointed out in the Introduction.

Since the original derivation of this analysis in 1960,[4] several TWM's have been developed in this laboratory. They include the TWM for the ground station receiver in the Telstar satellite communication experiment[6] and radio astronomy TWM preamplifiers for hydrogen line work at 1420 mc.[7] In these cases, the analysis has proved to be a valuable aid for arriving at a first-order design and similarly for providing guidelines in the subsequent improvements of these designs.

APPENDIX

The conformal mapping transformations are derived and evaluated, leading to the impedance transformation curves in Figs. 10 and 12. The mathematical treatment given here is not too extensive, because the type of transformation used is known from other areas of electrical engineering. Yet the description of the mathematical procedure is made reasonably complete so that it may be useful as a guide for treating other related problems: for example, traveling-wave masers where the finger width is not identical to the spacing between fingers.

A.1 *The Schwartz-Christoffel Transformation*

The particular conformal transformation used here is a special case of the more general Schwartz-Christoffel transformation. The theorem proved independently by these two mathematicians states that it is possible to find an analytical function which maps the inside of a polygon on the

complex plane into the upper half of this plane. The boundary of the polygon thus is mapped into the real axis. If two transformations are considered, one of the type mentioned, the other performing the inverse function, it follows that the inside of a polygon can be mapped into the inside of any other polygon.

The general Schwartz-Christoffel transformation is illustrated in Fig. 17. For purposes of discussion, it is perhaps easier to consider first the inverse transformation of the upper half of the complex plane into a polygon. The transformation will be accomplished by a function whose derivative is given by a product of the type

$$\frac{dz}{dw} = (w - a)^{(\alpha/\pi - 1)} (w - b)^{(\beta/\pi - 1)} (w - c)^{(\gamma/\pi - 1)} \cdots. \qquad (26)$$
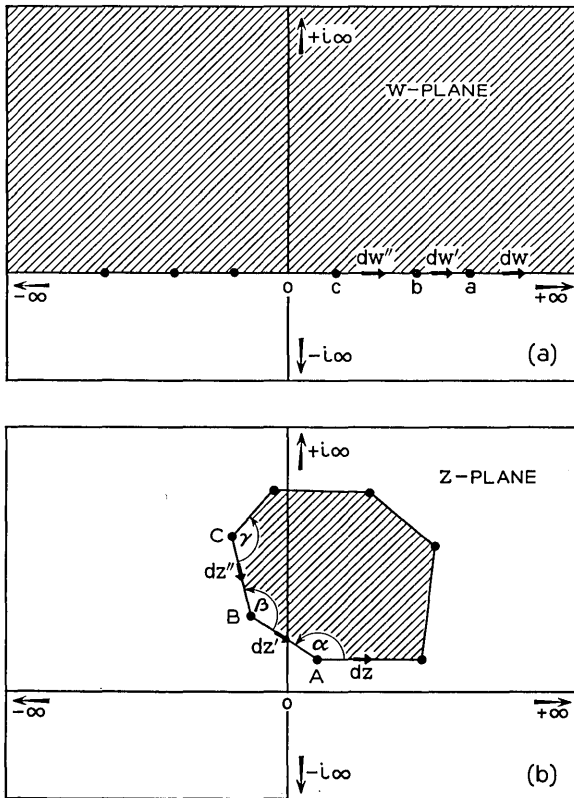
Fig. 17 — Illustration of the general mapping properties of the Schwartz-Christoffel transformation.

To demonstrate the transformation property, consider values of w and dw on the real axis. Also represent each factor in the form $r_k e^{i\phi_k}$ with a real number r_k for the magnitude and ϕ_k for the angle. It is seen then that for values w such that $w > a,b,c \cdots$ all the ϕ_k on the right-hand side of (26) vanish. Hence the angles of dz and dw are identical; that is, these line elements are parallel. Mathematically

$$\Delta dz = 0 \qquad \text{if } w \text{ and } dw \text{ are real and } w > a,b,c \cdots . \qquad (27a)$$

For values $a < w < b$ the first bracket changes sign; that is, its angle is π. The angle of the first factor becomes $\alpha - \pi$

$$\Delta dz' = \alpha - \pi \qquad \text{if } w \text{ and } dw' \text{ are real and } a > w > b,c \cdots . \qquad (27b)$$

That is to say, the real axis of the w plane near a is transformed in the z plane into a polygon corner at some as yet undetermined point A including an angle α. Similarly

$$\Delta dz'' = \alpha + \beta - 2\pi \qquad \text{if } w \text{ and } dw'' \text{ are real and } a,b > w > c \cdots \qquad (27c)$$

indicating another polygon corner at B including an angle β and corresponding to the point b on the real axis of the w plane.

In this fashion, it is shown that the transformation (26) indeed maps the upper half of the w plane into the inside of a polygon having specified angles $\alpha, \beta, \gamma \cdots$ at points in the z plane corresponding to $a,b,c \cdots$ in the w plane. While it is thus easy to satisfy conditions on the angles of the polygon, the difficulty is to find the points $A,B,C \cdots$ in the z plane which correspond to $a,b,c \cdots$ in the w plane. This requires an evaluation of the integral of (26).

Even more typical for engineering applications, and important in the present example, is the inverse situation. The corner points $A,B,C \cdots$ of the polygon are given. Then the problem is to find the real numbers $a,b,c \cdots$ which when inserted into (26) will transform this polygon into the upper half of the w plane. In most cases, this problem can only be solved numerically. The procedure would be to tabulate integrals of (26) for some range of values $a,b,c \cdots$. Numbering such tables with the given integral values $A,B,C \cdots$, the appropriate transformation parameters a,b,c could be picked.

To keep the need for tabulation down to a manageable chore, the number of significant parameters has to be restricted as much as possible. The example of importance in this connection is the mapping of a rectangle into the upper half of the complex plane. The number of significant parameters here can be reduced to one, the length ratio of two adjacent sides. Other parameters can be eliminated by trivial transformations

such as scaling and rotation of the coordinate system, which is accomplished simultaneously by a complex constant factor in (26) or a shift of the coordinate origin which corresponds to the integration constant of (26).

A.2 *Mapping of a Rectangle into the Upper Half of the Complex Plane*

It is now possible to write down the transformation equation for a rectangle. The conventional notation is illustrated in Fig. 18. The corners of the rectangle in the z plane are the complex numbers K, $K + iK'$, $-K + iK'$ and $-K$. In the w plane they correspond to the points 1, $1/k$, $-1/k$ and -1 on the real axis.
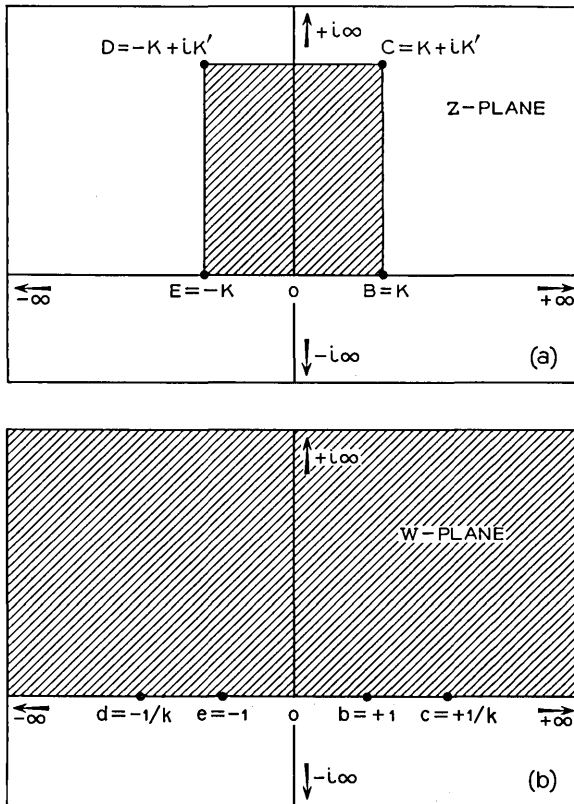
Fig. 18 — Illustration of the transformation of a rectangle in the z plane into the upper half of the w plane, introducing the conventional mathematical notation.

From (26) the transformation derivative is

$$\frac{dz}{dw} = A \left(w - \frac{1}{k} \right)^{-\frac{1}{2}} (w - 1)^{-\frac{1}{2}} (w + 1)^{-\frac{1}{2}} \left(w + \frac{1}{k} \right)^{-\frac{1}{2}}. \tag{28}$$

When the constant A is chosen appropriately ($A = -1/k$) this becomes

$$dz = \frac{dw}{(1 - w^2)^{\frac{1}{2}} (1 - k^2 w^2)^{\frac{1}{2}}} \tag{29}$$

and

$$z = \int_{\omega=0}^{w} \frac{d\omega}{(1 - \omega^2)^{\frac{1}{2}} (1 - k^1 \omega^2)^{\frac{1}{2}}}. \tag{30}$$

This integral is an elliptical integral of the first kind. It gives z as a function of w and k, where k is referred to as the modulus of the integral.

From the definition adapted in the figures it follows that

$$K = \int_{\omega=0}^{1} \frac{d\omega}{(1 - \omega^2)^{\frac{1}{2}} (1 - k^2 \omega^2)^{\frac{1}{2}}} \tag{31}$$

and

$$iK' = \int_{\omega=1}^{1/4} \frac{d\omega}{(1 - \omega^2)^{\frac{1}{2}} (1 - k^2 \omega^2)^{\frac{1}{2}}}. \tag{32}$$

K is called the complete elliptical integral. K' is the complete integral to the complementary modulus obeying the functional relationship

$$K'(k) = K(k') \tag{33}$$

where $k^2 + k'^2 = 1$ is used to define the modulus k' as complementary to k.

The definition of the elliptical integral of the first kind as given in in (30) is due to Jacobi. Many tables use also the notation of Legendre. This is obtained by setting

$$\begin{aligned} w &= \sin \phi, & dw &= \cos \phi \, d\phi \\ k &= \sin \theta, & k' &= \cos \theta. \end{aligned} \tag{34}$$

Then

$$z = \int_{\Psi=0}^{\phi} \frac{d\Psi}{(1 - \sin^2 \theta \sin^2 \Psi)^{\frac{1}{2}}} \tag{35}$$

$$K = \int_{\Psi=0}^{\pi/2} \frac{d\Psi}{(1 - \sin^2 \theta \sin^2 \Psi)^{\frac{1}{2}}} \tag{36}$$

$$K' = \int_{\Psi=0}^{\pi/2} \frac{d\Psi}{(1 - \cos^2 \theta \sin^2 \Psi)^{\frac{1}{2}}}. \tag{37}$$

From this discussion it is clear that the transformation of a rectangle into the upper half plane requires finding the modulus k or equivalently the modular angle θ of the elliptical integral *from the given geometry of the rectangle*. It is further clear that K and K' are not independent, but related through either (31) and (32) or (36) and (37). Therefore, it is not possible to specify both length dimensions of the rectangle of Fig. 18 but rather only their ratio. The problem thus is reduced to finding the dependence of the modulus k or θ from the aspect ratio $K'/2K$ of the rectangle.

This functional dependence was evaluated using the Smithsonian Elliptic Function Tables, in particular tables of complete elliptical integrals. The result is presented in Fig. 19.

It should be added that frequently, instead of the elliptical integral (30), its inverse is used. This inverse function is written

$$w = \text{sn } z \text{ modulo } k \tag{38}$$

which is defined to mean (30). This notation is reminiscent of the sine function, with which the sn function is indeed identical in the special case $k = 0$.

A.3 *Mapping of the Upper Cutoff Frequency Configuration*

It is now possible to carry out the mapping transformations used in the comb structure analysis. The initial geometry for the lower cutoff frequency is indicated in Fig. 20(a), where solid lines represent conducting electrodes. The final result is a parallel plate geometry like that of Fig. 20(d). This figure represents the cross section of an idealized transmission line for which the impedance is simply given by the ratio of the length dimensions times free-space impedance. The transformation makes use of two intermediate steps. The interior of the rectangle (Fig. 20a) is first mapped into the upper half of the complex plane (Fig. 20b). Then a readjustment of the scale leads to Fig. 20(c). Then the upper half plane is finally mapped into the inside of the desired rectangle (Fig. 20d) with electrodes only on opposite sides.

To keep track of these steps, the relevant points in the original geometry and their transforms are denoted by capital letters O,A,B $\cdots$. The first and second elliptical transformations are distinguished by indices 1 and 2 attached to the modulus and the complete integral values. The mapping then proceeds as follows.

(a) *From the z plane to the y plane.*

$$y = \text{sn } z \text{ modulo } k_1 = \sin \theta_1 \tag{39}$$
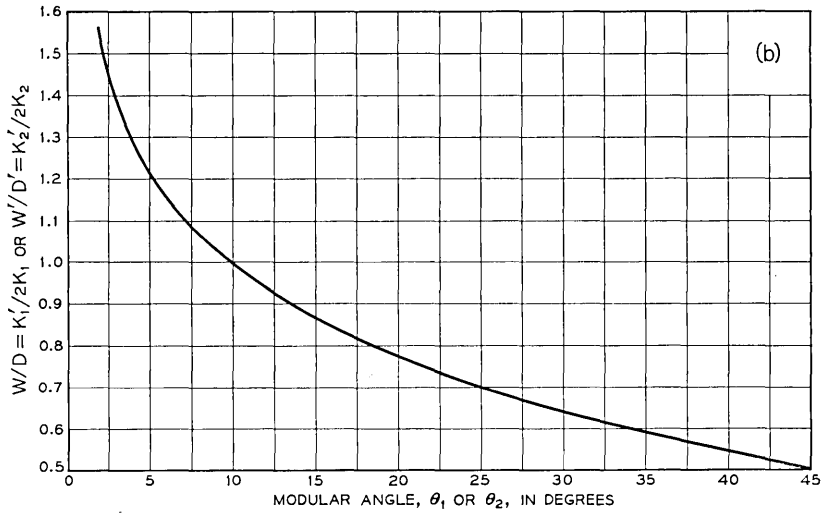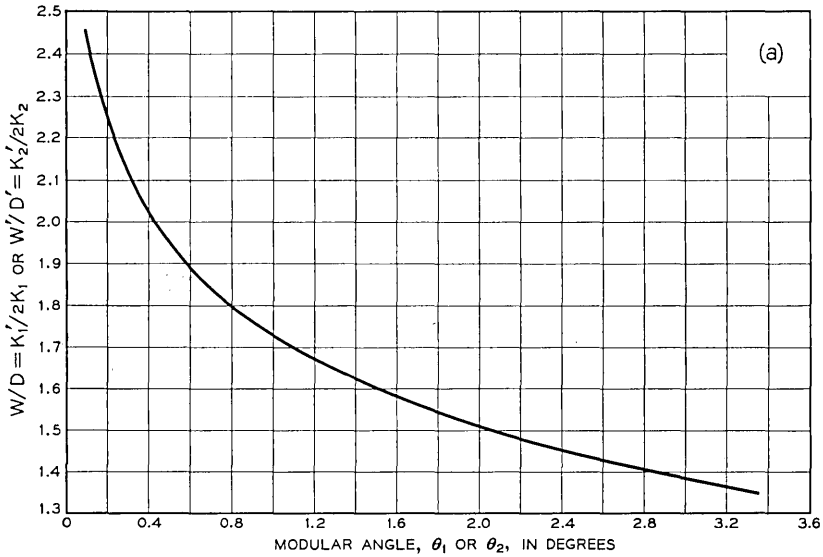
Fig. 19 — Relation between the ratio of the length dimensions of the rectangle to be transformed and the modular angle of the transforming elliptical function.

The modular angle θ_1 is found by entering the curves of Fig. 19 with the aspect ratio $W/D = K_r'/2K_1$ of the original rectangle. The corresponding coordinates in the z and y planes are given in Table I. The transformation of points O through D requires only the graph of Fig. 19. For points A and F, use has to be made of elliptic function tables. In the Smithsonian
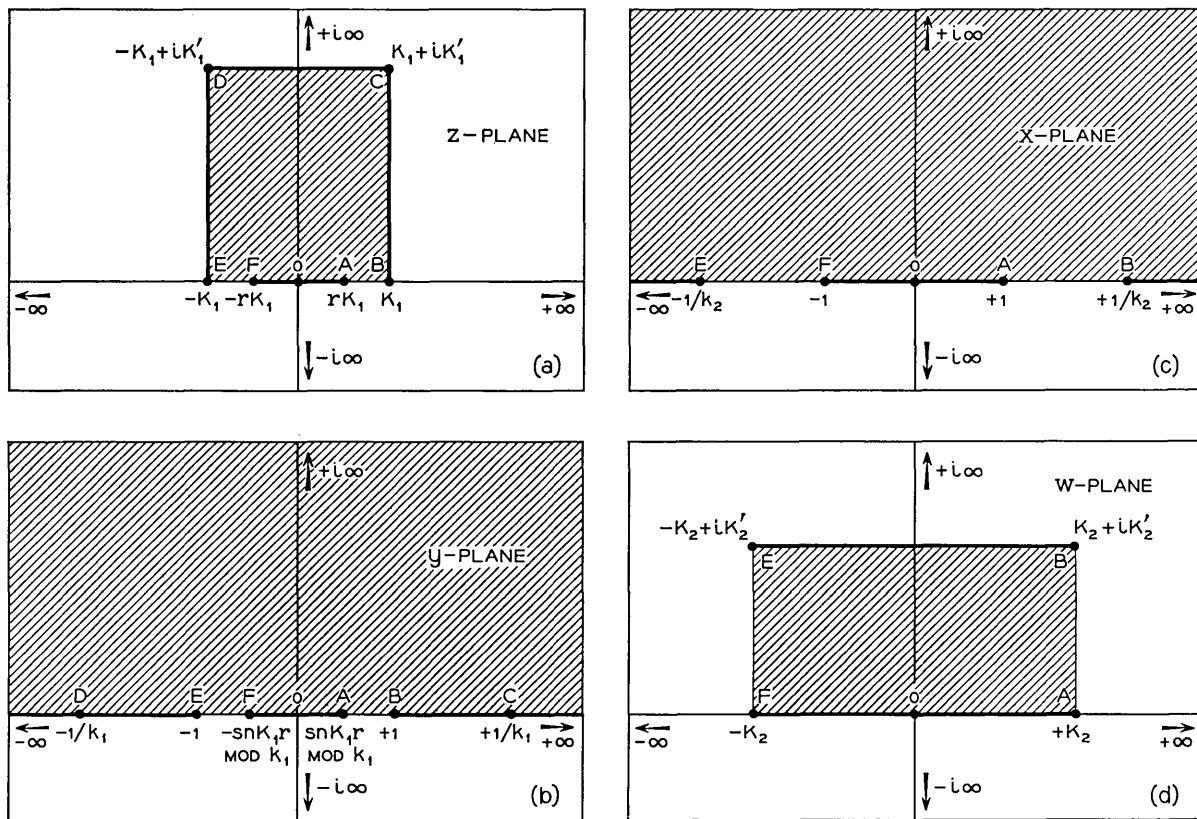
Fig. 20 — (a) Original electrode configuration of the upper cutoff frequency situation in the z plane. (b) Geometry after transformation into the upper half of the y plane. (c) Same geometry scaled in the x plane for subsequent transformation. (d) Electrode configuration transformed into a simple parallel plate geometry in the w plane.

Table I — Summary of Transformations
for Upper Cutoff Frequency Case

	z	y	x	w
O	0	0	0	0
B	K_1	1	$1/(\text{sn } K_1 r \text{ mod } k_1)$	$K_2 + iK_2'$
E	$-K_1$	-1	$-1/(\text{sn } K_1 r \text{ mod } k_1)$	$-K_2 + iK_2'$
C	$K_1 + iK_2'$	$1/k_1$	} not of interest	
D	$-K_1 + iK_1'$	$-1/k_1$		
A	$K_1 r$	$\text{sn } K_1 r \text{ mod } k_1$	1	K_2
F	$-K_1 r$	$-\text{sn } K_1 r \text{ mod } k_1$	-1	$-K_2$

$$k_1 = \sin \theta_1$$
$$k_2 = \sin \theta_2$$
$$\theta_2 = \phi = \sin^{-1} [\text{sn } K_1 r \text{ mod } k_1 \text{ or } \theta_1]$$

Tables the Legendre notation (34), (35), (36), and (37) is used. Entering these tables with $z = K_1 r$ and the angle θ_r, a value of ϕ in radians is found. This value ϕ is converted to degrees and renamed θ_2.

(b) *From the y plane to the x plane.*

This is a change of scale and is accomplished by dividing all values by

$$\sin \phi = k_2 = \text{sn } K_1 r \text{ mod } \theta_r. \qquad (40)$$

After this step the arrangement of the points OBEAF on the real axis is the standard one for transformation of the upper half plane into a rectangle.

(c) *From the x plane to the w plane.*

This transformation finally shapes the original electrode geometry into the desired parallel plane geometry. The transformation is indicated in Table I. However, since the interest centers only on the impedance — that is, the length-dimension ratio of this final rectangle — it is not necessary to carry out this transformation in detail. This ratio $W'/D' = K_2'/2K_2$ is obtained from Fig. 19 by entering it with the modular angle $\theta_2 = \phi = \sin^{-1} k_2$.

Following these steps in the case $r = \frac{1}{2}$, the curve of Fig. 10 was obtained.

A short-cut is possible if $\theta_1 < 30°$; that is, if $W/D > 0.65$. In that case the sn function can be approximated by a sine function and $K \approx \pi/2$. Then $\phi = \theta_2 = r\pi/2$; in particular, for $r = \frac{1}{2}$, $\phi = \theta_2 = 45°$ and $W'/D' = \frac{1}{2}$.

A.4 *Mapping of the Lower Cutoff Frequency Configuration*

The procedure is quite similar to that used for the upper cutoff frequency geometry. It is summarized in Table II and Fig. 21.

TABLE II — SUMMARY OF TRANSFORMATIONS
FOR LOWER CUTOFF FREQUENCY CASE

	z	y	x	w
O	0	0	0	0
B	K_1	$\left.\begin{array}{c}1\\-1\end{array}\right\}$	not of interest	
E	$-K_1$			
C	$K_1 + iK_1$	$1/k_1$	$1/(k_1 \text{ sn } K_1 r \text{ mod } k_1)$	$K_2 + iK_2'$
D	$-K_1 + iK_1'$	$-1/k_1$	$-1/(k_1 \text{ sn } K_1 r \text{ mod } k_1)$	$-K_2 + iK_2'$
A	rK_1	$\text{sn } K_1 r \text{ mod } k_1$	1	K_2
F	$-rK_1$	$-\text{sn } K_1 r \text{ mod } k_1$	-1	$-K_2$

$$k_1 = \sin \theta_1$$
$$k_2 = \sin \theta_2$$
$$\theta_2 = \sin^{-1} [(\text{sn } K_1 r \times \sin \theta_1) \text{ mod } k_1 \text{ or } \theta_1]$$

(a) *From the z plane to the y plane.*

This step is identical to the first transformation of the upper cutoff frequency configuration.

(b) *From the y plane to the x plane.*

This scaling is also the same as that used before. The difference is, however, that now the points C and D are of interest, whereas before the points considered were B and E.

(c) *From the x plane to the w plane.*

Here the transformation differs; now a different modulus

$$k_2 = k_1 \text{ sn } K_1 r \text{ mod } k_1$$

is used. The resulting complete integral values K_2 and K_2' are not to be confused with those obtained for the upper cutoff frequency case. Since the interest centers only on the impedance value $K_2'/2K_2 = W'/D'$ of the resulting rectangle, it is not necessary to evaluate this transformation in detail. The numerical evaluation is quite similar to the one of the upper cutoff situation. Using Fig. 19, one finds the first modular angle θ_1 from $K_1'/2K_1 = W/D$ of the original geometry. Entering the tables with $z = K_1 r$ and θ_1, an integral value ϕ is found. This value is obtained in radians. Then form

$$\begin{aligned}k_2 &= \sin \theta_2 = \sin \theta_1 \times (\text{sn } K_1 r \text{ mod } k_1)\\ &= \sin \theta_1 \times (\sin \phi \text{ mod } \theta_1).\end{aligned} \quad (41)$$

Using this formula, the angle θ_2 is evaluated in degrees. Then the graphs (Fig. 19) can be used again to obtain from θ_2 the length dimension ratio W'/D' of the transformed rectangle.

Following this procedure for the case $r = \frac{1}{2}$, the graph of Fig. 12 was obtained.
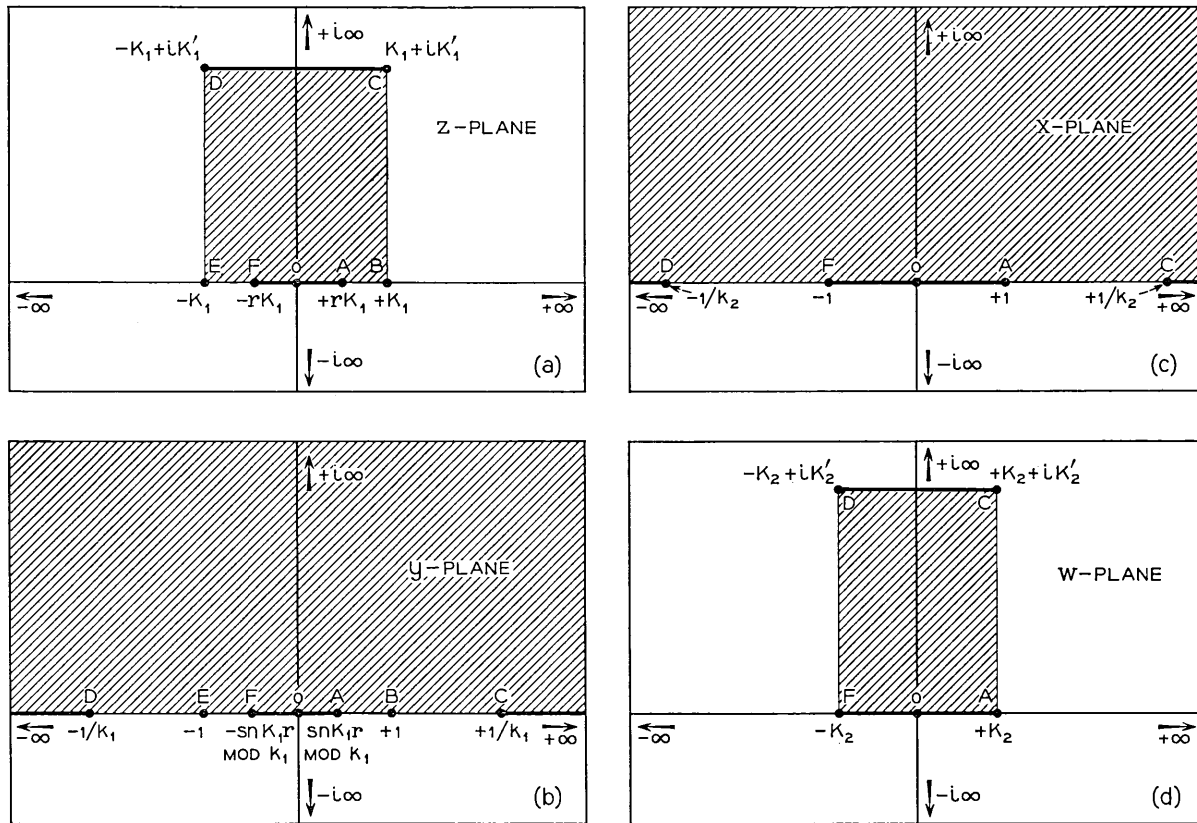
Fig. 21 — (a) Original electrode configuration corresponding to the lower cutoff frequency situation, shown in the z plane. (b) Geometry after transformation into the upper half of the y plane. (c) Same geometry scaled up in the x plane for subsequent transformation. (d) Electrode configuration transformed into a simple parallel plate geometry in the w plane.

REFERENCES

1. DeGrasse, R. W., Schulz-DuBois, E. O., and Scovil, H. E. D., Three-Level Solid-State Traveling Wave Maser, B.S.T.J., **38,** March, 1959, pp. 305–334.
2. Geusic, J. E., to be published.
3. Feher, G., Sensitivity Considerations in Microwave Paramagnetic Resonance Absorption Techniques, B.S.T.J., **36,** March, 1957, pp. 449–484.
4. DeGrasse, R. W., Kostelnick, J. J., and Scovil, H. E. D., Dual Channel 2390-mc Traveling-Wave Maser, B.S.T.J., **40,** July, 1961, pp. 1117–1128.
5. Harris, S. E., DeGrasse, R. W., and Schulz-DuBois, E. O., Solid State Maser Research, U. S. Signal Corps Report under Contract No. DA 36-039 SC-85357, First Quarterly Report, 20 September 1960. Available through ASTIA. Unpublished.
6. Tabor, W. J., and Sibilia, J. T., Masers for the *Telstar* Satellite Communications Experiment, B.S.T.J., **42,** July, 1963, pp. 1863–1886.
7. Hensel, M. L., and Treacy, E. B., to be published.

Permutation Decoding of
Systematic Codes

By JESSIE MACWILLIAMS

A symmetry of a systematic code is a permutation of bit positions in each code word (the same permutation is applied to all code words) which preserves the code as a whole. Permutation decoding makes use of these symmetries to build up a decoding algorithm for the code.

It is difficult to find an appropriate set of symmetries for a code picked at random. For cyclic codes the problem is somewhat easier, and for some special cyclic codes it is solved completely in this paper. For these codes, at least, it is evident that permutation decoding is easy to implement and inexpensive compared with other decoding schemes.

Permutation decoding as a means of error control is evaluated for the binary symmetric channel and for the switched telephone network as represented by experimental data. It is found to be extremely effective on the binary symmetric channel and of very doubtful value on the present telephone network.

INTRODUCTION

A systematic code of block length n is a subspace of the vector space of all possible rows of n symbols chosen from a finite field. In this paper such a code will be called an alphabet,[1] and the sequences belonging to the alphabet will be called letters.

The parameters used to describe an alphabet are block length, n, number of information places, k, and error correcting capability, e: n is the number of symbols in each letter, k is the dimension of the alphabet as a vector space, and e is defined by the property that the minimum Hamming distance between two letters is either $2e + 1$ or $2e + 2$.

It is well known[1] that an alphabet with parameters n,k,e is theoretically capable of correcting all occurrences of $\leq e$ errors in a block of length n. However for $e > 1$, the process of error correction by decoding is complicated, and likely to require expensive equipment. In this paper we

describe a new decoding scheme, permutation decoding, which is conceptually simple and quite easy to implement.

The decoding procedure consists of a sequence of permutations of the received block of symbols, each of which is followed by a parity check calculation. We can thus make a rough comparison between the complexity of the equipment required for encoding and decoding. The encoder uses one parity check register, and the decoder uses r (or uses one r times), where r is the number of permutations in the decoding sequence. Real time operation with a constant time delay is possible and perhaps not too expensive.

Permutation decoding owes much to the previous work of Peter Neumann[2] and Eugene Prange.[3] It depends essentially on the symmetries of the alphabet. A symmetry of an alphabet means a permutation of digit positions which preserves the alphabet as a whole. The same permutation is applied to the digits of every letter, and each letter is changed, if at all, into another letter of the same alphabet. Very little is known about symmetries of alphabets in general, but it will be shown that even this little is enough to enable us to apply the decoding scheme to a large class of alphabets.

Permutation decoding differs from previous schemes in two important ways. First, it becomes easier as the redundancy of the alphabet increases; it is most useful for alphabets with high error correcting capabilities. Secondly, it cannot correct more than e errors in n places. A received sequence containing more than e errors either will be "corrected" wrongly or will emerge unchanged from the decoder.

It will become apparent in Section III that permutation decoding produces many more undetected errors than does error control by detection and retransmission. The simpler scheme of detection and retransmission should be used when it is at all feasible.

The plan of this paper is as follows: Section I contains a description of permutation decoding in general, without reference to the particular alphabet we wish to decode. Section II is an example; it contains a detailed account of a particular permutation group which will suffice to decode many binary cyclic alphabets. Section III describes how the probability of improper correction and of detection without correction may be estimated.

I. ERROR CONTROL AND PERMUTATION DECODING

Let V^n be the set of all possible binary sequences of length n.* The distance between two sequences is the number of places in which they

---

* The method will work equally well for multilevel codes. All that is needed is to find a euphonious substitute for the term "binary sequence."

differ; the distance between v_1 and v_2 is the minimum number of bits we must change in v_1 in order to convert it into v_2.

For purposes of error control, some sequences of V^n are designated as the sequences which will be transmitted. This subset of V^n is called a code. Error detection consists in finding out whether a received sequence belongs to the code. Every method of error correction consists in mapping a received sequence onto the nearest member of the code, where nearness is defined in terms of the distance function defined above. If there are several nearest members, the correction procedure chooses one in some arbitrary fashion or indicates that an uncorrectable error has been found.

The strategy for choosing a code is usually to place its members as far apart as possible in V^n. It will then take a relatively large number of errors to cause a transmitted code sequence to be received as a different code sequence. If the distance between any two code sequences is $\geqq 2e + 1$, it is theoretically possible to correct all single, double, $\cdots$ e-fold errors. This may be restated as follows: If v is a received sequence in which $\leqq e$ errors have occurred, there is a unique code sequence α at distance $\leqq e$ from v. Every other member of the code is at distance $>e$ from v.

The business of decoding is to find α, given v. To do this expeditiously we need some additional structure in the code, and from now on we restrict our choice of codes to the kind described in the next paragraph.

An alphabet* (systematic code, group code) is one in which a fixed number of fixed bit positions are designated as information places, and the other bit positions contain parity checks, which are linear combinations of the contents of the information places. For convenience, the first k bit positions are taken to be the information places. From any k-place binary sequence h we obtain a unique letter of the alphabet by adding $n - k$ parity checks. This letter will be denoted by $m(h)$.

Let $\bar{\alpha}$ stand for the first k coordinates of the n-place sequence α. α is a letter of the alphabet if and only if $\alpha = m(\bar{\alpha})$. Let π be a permutation of bit positions in V^n which preserves the alphabet; if α is a letter, so is $\alpha\pi$. The first k positions of $\alpha\pi$ are information places, and $\alpha\pi = m(\overline{\alpha\pi})$.

Let v be a received sequence containing $\leqq e$ errors. If no errors have occurred in the first k places of v, $\alpha_0 = m(\bar{v})$ is the unique letter of the alphabet at distance $\leqq e$ from v, and is the corrected version of v. On the other hand, if one or more errors have occurred in the first k places of v, the letter $m(\bar{v})$ is not the corrected version of v, since it is the same as v in the first k places. In this case $m(\bar{v})$ is at distance $>e$ from v.

* It has been shown[1] that every systematic code is a group code and that every group code is a systematic code.

The first step in the decoding procedure is to form $\alpha_0 = m(\bar{v})$, find the distance between v and α_0, and take α_0 as the corrected version of v if this distance is $\leqq e$.

Let α denote the unique letter of the alphabet at distance $\leqq e$ from v. Let π be, as before, a permutation of bit positions which preserves the alphabet. Clearly the distance between $\alpha\pi$ and $v\pi$ is the same as that between α and v.

Suppose that we can find a permutation π_i which preserves the alphabet and which moves the errors in v out of the first k positions. Then $\alpha_i = m(\overline{v\pi_i})$ is at distance $\leqq e$ from $v\pi_i$, and is the unique letter of the alphabet with this property. Consequently $\alpha = \alpha_i\pi_i^{-1}$ is the corrected version of v.

This suggests the following decoding procedure: Let I (the identity), $\pi_1, \pi_2, \cdots$ be a sequence of permutations which preserve the alphabet. Form the letters

$$\alpha_0 = m(\overline{vI}), \qquad \alpha_1 = m(\overline{v\pi_1}), \qquad \alpha_2 = m(\overline{v\pi_2}), \cdots$$

and at each step find the distance between α_i and $v\pi_i$. Continue until a letter α_i is found which is at distance $\leqq e$ from $v\pi_i$. Then $\alpha_i\pi_i^{-1}$ is the corrected version of v.

A received vector which is at distance $> e$ from all letters of the alphabet will be detected as an error but not corrected by this procedure. Some provision must be made for this eventuality. This is discussed in Section III.

It is also possible for the decoder to make an incorrect "correction." This will happen if an error pattern (of more than e errors) causes the transmitted letter α to be received as a sequence v which is at distance $\leqq e$ from a different letter α'. The probability of this occurrence is calculated in Section III.

In order for permutation decoding to work, we must be sure that one of the sequences $v\pi_i$ is correct in the first k places. If $v = \alpha + f, f$ being the error sequence, the permutation π_i must move all nonzero coordinates of f out of the first k places. In order for the procedure to be practical, it must be possible to move all sets of $\leqq e$ errors out of the first k places with a fairly short sequence of permutations.

To correct all sets of $\leqq e$ errors in a block of length n, we need the following: (i) an alphabet of block length n, dimension k, and minimum distance $\geqq 2e + 1$; and (ii) a set of permutations, $\pi_1, \pi_2, \cdots$ which preserve the alphabet and at the same time move any set of $\leqq e$ errors out of the first k places.

We emphasize that the reason for insisting that the permutations π_i shall preserve the alphabet is to keep the parity check calculation always the same. The encoder (the parity check calculator) is a complicated and expensive piece of equipment; it is desirable to use only one encoder in the decoding scheme. If for any reason (such as real time operation) it is necessary to have more than one encoder, we can, to a certain extent, relax the restriction on the permutations π_i.

It may seem to be quite a trick to find at the same time both a suitable alphabet and a suitable set of permutations; really the chief difficulty is that neither alphabets nor permutation groups have been studied from this point of view. It is shown in Section II that a very simple permutation group will do for many cyclic alphabets.

We conclude this section with an example of permutation decoding applied to the Hamming alphabet with $n = 7$, $k = 4$, $e = 1$.[*] The alphabet is written out in Table I; it is seen to be invariant under cyclic permutation.

TABLE I — A CYCLIC ALPHABET WITH $n = 7$, $k = 4$, $e = 1$

0	0	0	0	0	0	0
0	1	1	0	1	0	0
0	0	1	1	0	1	0
0	0	0	1	1	0	1
1	0	0	0	1	1	0
0	1	0	0	0	1	1
1	0	1	0	0	0	1
1	1	0	1	0	0	0
1	1	1	1	1	1	1
1	0	0	1	0	1	1
1	1	0	0	1	0	1
1	1	1	0	0	1	0
0	1	1	1	0	0	1
1	0	1	1	1	0	0
0	1	0	1	1	1	0
0	0	1	0	1	1	1

Let T denote the cyclic permutation. Clearly at most four applications of T will move any single error out of the first four places. The decoding sequence consists of the permutations I, T, T^2, T^3, T^4.[†]

Let the received vector be 1110100 (the first nonzero vector of Table I with an error in the first place). The successive stages of the decoding process are shown in Table II.

---

[*] The example is chosen for simplicity. Permutation decoding is *not* the most efficient way of correcting single errors.

[†] E. R. Berlekamp has pointed out that the shorter decoding sequence I, T^3, T^6 is sufficient.

TABLE II — DECODING PROCEDURE FOR THE ALPHABET OF TABLE I

V = 1110100	
$m(\bar{V})$ = 1110010	distance = 2
VT = 0111010	
$m(\overline{VT})$ = 0111001	distance = 2
VT^2 = 0011101	
$m(\overline{VT^2})$ = 0011010	distance = 3
VT^3 = 1001110	
$m(\overline{VT^3})$ = 1001011	distance = 2
VT^4 = 0100111	
$m(\overline{VT^4})$ = 0100011	distance = 1

Thus α = 0100011 is the unique letter at distance ≤ 1 from VT^4, and the corrected version of V is $\alpha T^{-4} = \alpha T^3$ = 0110100.

II. PERMUTATION DECODING OF CYCLIC ALPHABETS*

The coordinate places in V^n are labeled by the numbers 0, 1, 2, $\cdots$, $n - 1$. This notation is convenient for describing permutations. If ω stands for one of these numbers, the cyclic permutation is

$$T: \omega \rightarrow \omega + 1 \text{ (addition mod } n).$$

The powers of the cyclic permutation are

$$T^2: \omega \rightarrow \omega + 2; \qquad T^3: \omega \rightarrow \omega + 3, \cdots, \qquad T^n: \omega \rightarrow \omega + n = \omega.$$

A cyclic alphabet in V^n is an alphabet which is invariant under T, hence also invariant under T^2, T^3, etc. We assume that we wish to decode a cyclic alphabet with parameters n,k,e.

Successive cyclic shifts will eventually bring any k consecutive bits to the first k positions, and hence will move out of the first k positions any error pattern in which there is a gap of length $\geq k$. In particular, the sequence $I, T, \cdots, T^{n-1}$ will always correct all single errors.†

This sequence will not correct an error pattern in which there is no gap of length $\geq k$. Suppose, for example, that n = 23, k = 12. The error pattern shown below cannot be corrected by cyclic shifts alone.

X 1 2 3 4 5 6 7 8 X 10 11 12 13 14 15 16 17 18 X 20 21 22

To deal with such cases we introduce another permutation $U: \omega \rightarrow 2\omega$ (multiplication mod n), and its powers, $U^2: \omega \rightarrow 4\omega$, $U^3: \omega \rightarrow 8\omega$, etc. If n is odd, there exists a least integer t such that $2^t \equiv 1 \bmod n$; and $U^t = I$. The choice of U is motivated by the following theorem.

---

* It is to be emphasized that this section is only an example. The permutations described here will not suffice to decode all cyclic alphabets of odd block length. A method of finding other permutations is given in Appendix A.

† It will also correct all double errors if $k < n/2$, and so on.

Theorem 1: Every binary cyclic alphabet of odd block length is invariant under U and the powers of U.

The proof of this theorem is given in Appendix A.

The error pattern 0, 9, 19 above is changed by U into 0, 18, 15. This pattern is moved out of the first twelve places by 21 cyclic shifts.

The permutation group on $0, 1, \cdots, n - 1$ generated by T and U will be called G_n. It is easy to check that $TU = UT^2$; hence we may represent every permutation in G_n in the form U^iT^j, with $0 \leq i \leq t - 1$, $0 \leq j \leq n - 1$. Now every power of U leaves 0 fixed, and no power of T (except the identity) leaves 0 fixed; thus $U^iT^j = U^hT^k$ if and only if $i = h \bmod t$ and $j = k \bmod n$. It follows that the group G_n is of order nt and consists of the permutations:

$$\begin{array}{ccccc}
I, & T, & T^2, & \cdots, & T^{n-1} \\
U, & UT, & UT^2, & \cdots, & UT^{n-1} \\
U^2, & U^2T, & U^2T^2, & \cdots, & U^2T^{n-1} \\
\cdot & \cdot & \cdot & \cdots, & \cdot \\
U^{t-1}, & U^{t-1}T, & U^{t-1}T^2, & \cdots, & U^{t-1}T^{n-1}.
\end{array}$$

Let $0 \leq u_1 < u_2 < \cdots < u_s \leq n - 1$ (n odd), be a set of integers; we suppose that errors occur in places $u_1, \cdots, u_s$. Let $g(u_1, \cdots, u_s, n)$ be the length of the maximum gap which can be inserted in this sequence by repeated multiplication by 2 mod n.

If $u_{\nu k}$ denotes that integer less than n which is congruent to $2^k u_\nu$ mod n, then

$$g(u_1, \cdots, u_s, n) + 1 = \operatorname*{Max}_{i,j,k} \mid u_{ik} - u_{jk} \mid,$$

under the condition that the interval $[u_{ik}, u_{jk}]$ contain no other u_{vj}. Let $g(s,n)$ be the minimum value of this maximum gap for all possible choices of the s values $u_1, \cdots, u_s$.

$$g(s,n) = \operatorname*{Min}_{u_1,\cdots,u_s} g(u_1, \cdots, u_s, n).$$

The group G_n then contains a permutation which moves any set of s errors out of the first $g(s,n)$ places. Clearly $s' < s$ implies $g(s',n) \geq g(s,n)$. Hence a binary cyclic n,k,e alphabet with n odd may be decoded by G_n if and only if $k \leq g(e,n)$. The quantity $g(e,n)$ has a few obvious properties.

The numbers $0, 1, \cdots, n - 1$ can be partitioned into subsets which are invariant under U.[*] For example, for $n = 15$, these subsets are

---

[*] The number of cyclic alphabets of block length n is determined by the number of these subsets; the dimensions of the cyclic alphabets are determined by the sizes of the invariant subsets; see, for example, Ref. 4.

$$(0), (1,2,4,8), (3,6,12,9), (5,10), (7,14,13,11).$$

The union of any number of invariant subsets is also invariant under U; from these subsets we may obtain upper bounds on $g(e,n)$ by inspection. In the example above we obtain:

$g(2,15) \leqq 9$, since the invariant set $(5,10)$ gives us a maximum gap 11,12,13,14,0,1,2,3,4 of length 9.

$g(3,15) \leqq 4$, since the invariant set $(0,5,10)$ gives us a maximum gap 1,2,3,4 of length 4.

$g(5,15) \leqq 2$, since the invariant set $(0,3,6,9,12)$ gives us a maximum gap of length 2.

These upper bounds limit the usefulness of the group G_n. However it is still sufficiently useful to be of interest.

The value of $g(e,n)$ for various choices of e,n have been computed on the IBM 7090. These are tabulated in Table III together with the parameters n,k,e for several cyclic alphabets.

TABLE III — EFFECT OF PERMUTATION U FOR
DIFFERENT BLOCK LENGTHS

n	Code k	e	Gap Length g(2)	g(3)	g(4)	g(5)
47	24	5				26
31	21	2	25			
31	16	3		19		
31	11	4			12	
23	12	3		17		
21	12	2	13			
21	9	3		6		
21	5	4			6	
17	9	2	13			
15	7	2	9			
15	5	3		4		

The gap length $g(s)$ is the maximum number of consecutive error-free positions which can be inserted into an arbitrary pattern of s errors in n places by successive applications of the permutation $w \rightarrow 2w \bmod n$. If $k \leqq g(e)$, the group generated by T and U contains a sequence of permutations which will suffice to decode an n, k, e alphabet.

It is desirable, if possible, to use only part of G_n in the decoding sequence. As an example we consider the alphabet with $n = 23$, $k = 12$, $e = 3$.[*] In this case one of the permutations, U, U^2, U^{11} will always create a gap of length at least 12 in any set of 3 errors. The decoding sequence is:[†]

---

[*] This alphabet is described in detail in Table V.

[†] It has been shown by E. R. Berlekamp that a subsequence of length 40 is all that is really necessary.

$$I, \quad T, \quad T^2, \quad \cdots, \quad T^{22}$$
$$U, \quad UT, \quad UT^2, \quad \cdots, \quad UT^{22}$$
$$U^2, \quad U^2T, \quad U^2T^2, \quad \cdots, \quad U^2T^{22}$$
$$U^{11}, \quad U^{11}T, \quad U^{11}T^2, \quad \cdots, \quad U^{11}T^{22}.$$

The decoding procedure for this particular alphabet has been simulated on the IBM 7090. A diagram of the logical program is given in Fig. 1, and Table IV traces a particular sequence through the decoder. In practice it is convenient to add one more permutation (in this case TU) to the decoding permutations, so that a sequence passing through the entire

Fig. 1 — Logic of decoder for (23,12,3) alphabet. The contents of the boxes are Fortran-type instructions; for example, $i = i + 1$ is to be interpreted as "replace i by $i + 1$."

Table IV — Decoding Procedure for the (23,12,3) Alphabet

													Parity Check											Distance
v	3	5	7	11	13	14	15	18	19	20	21	22	3	5	7	11	13	16	18	21				6
vT	0	4	6	8	12	14	15	16	19	20	21	22	0	4	6	8	16	18	19					7
vT^{22}	2	4	6	10	12	13	14	17	18	19	20	21	2	4	6	10	12	15	17	20				6
vU	3	5	6	7	10	13	14	15	17	19	21	22	3	5	6	7	10	12	15	16	17	19	20	7
vUT	0	4	6	7	8	11	14	15	16	18	20	22	0	4	6	7	8	11	12	13	14	20	21	22 · 6
vUT^{22}	2	4	5	6	9	12	13	14	16	18	20	21	2	4	5	6	9	13	14	17	18	19	21	22 · 6
vU^2	3	5	6	7	10	11	12	14	15	19	20	21	3	5	6	7	10	11	12	13	19	20	21	22 · 4
vU^2T	4	6	7	8	11	12	13	15	16	20	21	22	4	6	7	9	11	13	16	17	18	20	21	5
vU^2T^{12}	0	1	3	4	8	9	10	15	17	18	19	22	0	1	3	4	8	9	10	12	15	18	19	3
αU^2T^{12}	0	1	3	4	8	9	10	12	15	18	19←													
αU^2T^{22}	2	5	6	9	12	14	16	18	20	21	22													
αU^{10}	0	9	10	11	13	14	17	18	19	21	22													
$\alpha U^{10}T$	0	1	10	11	12	14	15	18	19	20	21													
$\alpha U^{10}T^{22}$	8	9	10	12	13	16	17	18	20	21	22													
α	0	3	5	11	13	15	18	19	20	21	22													

The numbers indicate the positions of the ones in the 23-bit sequence. For example, the first sequence v is 00010101000101110011111.

set emerges in its original form. The final output of the permuting register is then the corrected form of the received sequence.

The operation of the 7090 program is of course sequential—it employs one subroutine to simulate the parity check calculation. It is clear from the logic diagram that it is quite convenient to split the encoder into four parallel sections, each of which contains a register capable of making a cyclic permutation and an encoder to calculate parity checks. This idea can be applied to speed up the decoding of any cyclic alphabet.

III. EVALUATION OF PERMUTATION DECODING AS A MEANS OF ERROR CONTROL

Permutation decoding of an n,k,e alphabet $\mathcal{C}$ will map a received sequence v onto the nearest letter of $\mathcal{C}$ provided that this letter is unique. This is the case if v lies at distance $\leq e$ from some letter of $\mathcal{C}$. If v is at distance $> e$ from every letter of $\mathcal{C}$ the decoder will detect an error but will be unable to correct it.[*]

The decoder will also make mistakes. If f is an error sequence of more than e errors, and α the transmitted letter, the received sequence $\alpha + f$ may lie at distance $\leq e$ from some other letter α' of $\mathcal{C}$. The decoder will then interpret $\alpha + f$ as α'. The error sequences which cause such incorrect decoding are characterized by the following theorem.

Theorem 2: The error sequences which cause the decoder to "correct" incorrectly are exactly those sequences of weight $> e$ which lie at distance $\leq e$ from some letter of $\mathcal{C}$.

Proof: Let f be a sequence of weight $> e$ such that $f = \beta + f'$, $\beta \ \epsilon \ \mathcal{C}$, f' of weight $\leq e$.

For any transmitted letter α

$$\alpha + f = \alpha + \beta + f',$$

and the decoder will interpret $\alpha + f$ as $\alpha + \beta$.

Conversely, suppose that f is an error sequence such that $\alpha + f$ is decoded as $\alpha' \neq \alpha$. Then

$$\alpha + f = \alpha' + f' \qquad \alpha' \ \epsilon \ \mathcal{C}, f' \text{ of weight } \leq e,$$

and

$$f = \alpha' - \alpha + f'.$$

Hence f is at distance $\leq e$ from the letter $\alpha' - \alpha$ of $\mathcal{C}$.

---

[*] One is tempted to suggest that the decoder ask for retransmission of such detected errors. This idea is of dubious value; error correction by decoding should not be used at all if error correction by detection and retransmission is a possible alternative. We must assume that retransmission is extremely awkward, if not completely infeasible.

Let $A(s), s = 0, 1, \cdots, n$ be the number of letters of $\mathfrak{A}$ of weight s. Let $C(s)$, $s = 0, 1, \cdots, n$ be the number of sequences of V^n of weight s which lie at distance $\leq e$ from letters of $\mathfrak{A}$. The $C(s)$ are uniquely determined by the $A(s)$, and may be obtained from them by a simple calculation; the exact formula is given in Appendix B. The values of $A(s)$ for a number of binary cyclic alphabets are tabulated in Table V, and the values of $C(s)$ for these alphabets are given in Table VI.

For $s \leq e$, $C(s) = \binom{n}{s}$ and is the number of sequences of weight s which are properly corrected by the decoder. For $s > e$, $C(s)$ is the number of sequences of weight s which are improperly corrected by the decoder.

Let $D(s), s = e + 1, \cdots, n$ be the number of error sequences of weight s which cause the decoder to detect an error that it cannot correct. Clearly $D(s) = 0$ for $s \leq e$, and $D(s) = \binom{n}{s} - C(s)$ for $s > e$.

P_E denotes the probability that a received sequence will be "corrected" incorrectly by the decoder; P_D denotes the probability that a received sequence will be detected as an error but not corrected. We consider first a binary symmetric memoryless channel with bit error probability p. The probability of s specific errors in a block of length n is then $p^s(1 - p)^{n-s}$, and this probability is independent of the location of the errors. Hence

$$P_E(\text{B.S.}) = \sum_{s=e+1}^{n} C(s)p^s(1 - p)^{n-s},$$

$$P_D(\text{B.S.}) = \sum_{s=e+1}^{n} D(s)p^s(1 - p)^{n-s}.$$

It is to be noted that if error correction by detection and retransmission is used, the probability of an undetected error is

$$\sum_{s=d}^{n} A(s)p^s(1 - p)^{n-s}; \quad d = 2e + 1 \quad \text{or} \quad 2e + 2.$$

This sum starts with the first nonzero value of $A(s)$ (for $s > 0$), i.e. with $s = 2e + 1$ or $s = 2e + 2$. It is obvious that for the values of p,n currently in use in the Bell System, error correction by detection and retransmission is the preferable scheme.

The values of $P_E(\text{B.S.})$ and $P_D(\text{B.S.})$ for a number of alphabets are tabulated in Table VII; p is taken to be 3.22×10^{-5}, the over-all bit

error rate on the telephone network obtained from the Alexander, Gryb and Nast task force data.[5]

Except for the first and last example, the alphabets in Tables V, VI and VII occur in pairs. The second alphabet in each pair consists of the letters of even weight in the first. Since the minimum distance of the second alphabet is $2e + 2$, its value of $C(e + 1)$ is zero. In other words, every error of weight $e + 1$ will be detected. If this is important, it is advantageous to use the second alphabet.

The error rates of Table VI are fantastically low; unfortunately the fantasy resides in the binary symmetric channel. The situation is very different on the real telephone channel.

Let $P(s,n)$ be the probability of s errors in n consecutive bits. [For the binary symmetric channel $P(s,n) = \binom{n}{s} p^s (1 - p)^{n-s}$.] Tables of $P(s,n)$ for the telephone network have been calculated from the Alexander, Gryb and Nast task force data.

The decoder will either detect or correct wrongly every error of weight $> e$. Hence

$$P_E + P_D = \sum_{s=e+1}^{n} P(s,n).$$

It is impossible to obtain exact formulae for P_E and P_D separately. Using the methods of Ref. 5 we obtain the approximate formulae

$$P_E(\text{T.N.}) = \sum_{s=e+1}^{n} C(s) \frac{P(s,n)}{\binom{n}{s}}$$

$$P_D(\text{T.N.}) = \sum_{s=e+1}^{n} D(s) \frac{P(s,n)}{\binom{n}{s}}.$$

These numbers have been computed and are tabulated in Table VIII. This shows rather clearly that on the channel described by the Alexander, Gryb and Nast data the word error rate with error correction is greater than the bit error rate with no encoding.

Of course the average error rates of Table VIII conceal something. Examination of the $P(s,n)$ tables for the individual calls shows that about half of the calls in which errors occurred would be handled successfully by permutation decoding.

TABLE V — VALUES OF $A(s)$

n,k,e	47,24,5		31,21,2	31,20,2		31,16,3	31,15,3
s	$A(s)/47$	s	$A(s)/31$	$A(s)/31$	s	$A(s)/31$	$A(s)/31$
0	1/47	0	1/31	1/31	0	1/31	1/31
11	92	5	7	0	7	5	0
12	276	6	27	27	8	15	15
15	3795	7	75	0	11	168	0
16	7590	8	245	245	12	280	280
19	35420	9	655	0	15	589	0
20	49588	10	1387	1387	16	589	589
23	81720	11	2640	0	19	280	0
24	81720	12	4480	4480	20	168	168
27	49588	13	6510	0	23	15	0
28	35420	14	8310	8310	24	5	5
31	7590	15	9489	0	31	1/31	0
32	3795	16	9489	9489			
35	276	17	8310	0			
36	92	18	6510	6510			
47	1/47	19	4480	0			
		20	2640	2640			
		21	1387	0			
		22	655	655			
		23	245	0			
		24	75	75			
		25	27	0			
		26	7	7			
		31	1/31	0			

$n,k,e =$	23,12,3	23,11,3		(21,12,2)	(21,11,2)	(17,9,2)	(17,8,2)		(15,7,2)
s	$A(s)/23$	$A(s)/23$	s	$A(s)$	$A(s)$	s	$A(s)$	$A(s)$	s	$A(s)$
0	1/23	1/23	0	1	1	0	1	1	0	1
7	11	0	5	21	0	5	34	0	5	18
8	22	22	6	168	168	6	68	68	6	30
11	56	0	7	360	0	7	68	0	7	150
12	56	56	8	210	210	8	85	85	8	150
15	22	0	9	280	0	9	85	0	9	30
16	11	11	10	1008	1008	10	68	68	10	18
23	1/23	0	11	1008	0	11	68	0	15	1
			12	280	280	12	34	34		
			13	210	0	17	1	0		
			14	360	360					
			15	168	0					
			16	21	21					
			21	1	0					

TABLE VI — VALUES OF $C(s)$

$n,k,e =$	47,24,5*	31,21,2*	31,20,2	31,16,3	31,15,3
$s = 1$	$C(s) =$ 47	31	31	31	31
2	1081	465	465	465	465
3	16215	2170	0	4495	4495
4	178365	13640	12555	5425	0
5	1533939	82274	5022	29295	26040
6	1997688	360964	339047	92225	13020
7	11700743	1276115	81685	329375	303180
8	58503719	3829585	3591040	1248525	86025
9	253516120	9788250	604655	3190675	2861455
10	1094459500	21506932	20159981	6790333	690525
11	3681363800	41087771	2569497	12963363	11812395
12	10764415000	68535730	64275400	21284445	1987720
13	28981118000	100106900	6245260	31108034	28201320
14	70307802000	128661310	120616350	40561485	3675360
15	154677160000	145890120	9085914	45969682	41569263
16	310193350000	145890120	136804210	45969682	4400419
17	565646700000	128661310	8044965	40561485	36886124
18	942103590000	100106900	93361645	31108034	2906715
19	1437947500000	68535730	4260330	21284445	19296724
20	2012287600000	41087771	38518275	12963363	1150968
21	2587226900000	21506932	1346950	6790333	6099808
22	3059047600000	9788250	9183595	3190675	329220
23	3325051800000	3829585	238545	1248525	1162500
24		1276115	1194430	329375	26195
25		360964	21917	92225	79205
26		82274	77252	29295	33255
27	+ symmetric	13640	1085	5425	5425
28	terms	2170	2170	4495	0
29		465	0	465	0
30		31	0	31	0
31		1	0	1	0

$n,k,e =$	23,12,3	23,11,3	21,12,2	21,11,2	17,9,2	17,8,2	15,7,2
$s = 1$	$C(s) =$ 23	23	21	21	17	17	15
2	253	253	210	210	136	136	105
3	1771	1771	210	0	340	0	180
4	8855	0	2625	2520	1190	1020	540
5	33649	28336	10269	1008	3910	408	1413
6	100947	14168	24024	21168	7820	6936	2355
7	245157	216568	52440	4200	11560	1428	3135
8	490314	61226	92610	85050	14450	13005	3135
9	817190	715990	131530	12810	14450	1445	2355
10	1144066	138138	161196	146748	11560	10132	1413
11	1352078	1180774	161196	14448	7820	884	540
12	1352078	171304	131530	118720	3910	3502	180
13	1144066	1005928	92610	7560	1190	170	105
14	817190	101200	52440	48240	340	340	15
15	490314	429088	24024	2856	136	0	1
16	245157	28589	10269	9261	17	0	
17	100947	86779	2625	105	1	0	
18	33649	5313	210	210			
19	8855	8855	210	0			
20	1771	0	21	0			
21	253	0	1				
22	23	0					
23	1	0					

* This table is correct to eight significant figures.

500

TABLE VII — ERROR RATES FOR THE BINARY SYMMETRIC CHANNEL

n,k,e	P_E	P_D	$P_E + P_D$
47,24,5	2.23×10^{-21}	4.56×10^{-20}	4.78×10^{-20}
31,21,2	7.23×10^{-11}	0.77×10^{-10}	1.50×10^{-10}
31,20,2	1.342×10^{-12}	1.50×10^{-10}	1.50×10^{-10}
31,16,3	5.83×10^{-15}	2.80×10^{-14}	3.38×10^{-14}
31,15,3	9.01×10^{-19}	3.38×10^{-14}	3.38×10^{-14}
23,12,3	9.59×10^{-15}	0	9.59×10^{-15}*
23,11,3	9.81×10^{-19}	9.59×10^{-15}	9.59×10^{-15}
21,12,2	7.00×10^{-12}	3.72×10^{-11}	4.42×10^{-11}
21,11,2	2.70×10^{-15}	4.42×10^{-11}	4.42×10^{-11}
17,9,2	4.54×10^{-12}	1.80×10^{-11}	2.27×10^{-11}
17,8,2	1.11×10^{-13}	2.27×10^{-11}	2.27×10^{-11}
15,7,2	6.092×10^{-12}	1.078×10^{-10}	1.078×10^{-10}

* This is a close-packed alphabet; every sequence of 23 binary bits is at distance $\leqq 3$ from some letter of the alphabet.

CONCLUSION

Permutation decoding is a simple and feasible scheme for error correction without retransmission. It is particularly suitable for use with a highly redundant alphabet. Like any such scheme it produces many more undetected errors than error correction by detection and retransmission,

TABLE VIII — ERROR RATES FOR THE TELEPHONE NETWORK*

n,k,e	P_E	P_D	$P_E + P_D$
47,24,5	4.65×10^{-6}	3.51×10^{-5}	4.08×10^{-5}
31,21,2	3.17×10^{-5}	3.73×10^{-5}	6.9×10^{-5}
31,20,2	1.23×10^{-5}	5.67×10^{-5}	6.9×10^{-5}
31,16,3	7.34×10^{-6}	3.99×10^{-5}	4.72×10^{-5}
31,15,3	3.06×10^{-6}	4.41×10^{-5}	4.72×10^{-5}
23,12,3	3.69×10^{-5}	0	3.69×10^{-5}
23,11,3	1.52×10^{-5}	2.17×10^{-5}	3.69×10^{-5}
21,12,2	1.82×10^{-5}	3.13×10^{-5}	5.05×10^{-5}
21,11,2	8.72×10^{-6}	4.18×10^{-5}	5.05×10^{-5}
17,9,2	2.36×10^{-5}	1.89×10^{-5}	4.25×10^{-5}
17,8,2	0.98×10^{-5}	3.27×10^{-5}	4.25×10^{-5}
15,7,2	1.83×10^{-5}	2.0×10^{-5}	3.83×10^{-5}

* P_E and P_D are approximate values. $P_E + P_D$ is exact.

but it is quite adequate for a channel in which $P(s,n)$ decreases rapidly as s increases. It is of very doubtful value on the telephone network as described by the Alexander, Gryb and Nast data.

APPENDIX A

Idempotents and Automorphisms of Cyclic Codes

We give first a short summary of the properties of cyclic alphabets.

Let V be a finite field of characteristic q. Let V^n denote the direct sum of n copies of V.

Denote by $V[y]$ the ring of polynomials in y over the field V. Let $V[x] = V[y]/(y^n - 1)$ be the residue class ring of $V[y]$ mod $y^n - 1$. $V[x]$ consists of all polynomials of degree $\leq n - 1$ with coefficients in V. Addition of polynomials is done as usual; to multiply two polynomials we multiply in the usual way and then reduce exponents mod n.

A subset α of polynomials of $V[x]$ is called an ideal if

$$(i) \quad g_1, g_2 \in \alpha \Rightarrow \alpha_1 g_1 + \alpha_2 g_2 \in \alpha, \qquad \alpha_1, \alpha_2 \in V$$
$$(ii) \qquad g \in \alpha \Rightarrow xg \in \alpha.$$

A polynomial is completely determined by its coefficients; it is possible, in fact, to identify $V[x]$ and V^n. However, it is convenient to regard them as separate entities, related by the (1-1) mapping

$$\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1} x^{n-1} \rightleftarrows \alpha_0, \alpha_1, \cdots, \alpha_{n-1}.$$

An ideal in $V[x]$ is, by property (i), a linear subspace of V^n. By property (ii) it is invariant under a cyclic permutation of coordinates; hence it is a cyclic alphabet in V^n. Conversely, a cyclic alphabet in V^n is an ideal in $V[x]$. We represent the ideal and the alphabet by the same letter, α.

The ring $V[x]$ may be regarded as the group algebra of the cyclic group $1, x, \cdots, x^{n-1}$ over V. The group algebra is semi-simple provided that q does not divide n (Ref. 6, Section 10.8). In this case it is known that every ideal contains a polynomial e with the following properties

   (i) e is idempotent. $(e^2 = e)$

   (ii) e is a unit for $\mathcal{Q}$. $(a \, \epsilon \, \mathcal{Q} \Rightarrow ae = a)$

   (iii) e generates $\mathcal{Q}$.   ($\mathcal{Q}$ consists of all polynomials

                                        $fe, f \, \epsilon \, V[x]$).

e will be called the generating idempotent of $\mathcal{Q}$.

An automorphism σ of $V[x]$ is a (1-1) mapping of $V[x]$ onto itself which respects both addition and multiplication. If v_1, $v_2 \, \epsilon \, V[x]$, then

$$(v_1 + v_2)\sigma = v_1\sigma + v_2\sigma$$

$$(v_1 v_2)\sigma = (v_1\sigma)(v_2\sigma).$$

Lemma 1.1: An automorphism σ of $V[x]$ preserves an ideal $\mathcal{Q}$ if and only if σ preserves the generating idempotent e of $\mathcal{Q}$.

Proof: Suppose that σ preserves e. Then $\mathcal{Q}\sigma = V[x] \cdot e\sigma = V[x] \cdot e = \mathcal{Q}$.

Suppose that σ preserves $\mathcal{Q}$. Then $e\sigma \, \epsilon \, \mathcal{Q}$, and $e\sigma e = e\sigma$ by property (ii).

Let b be the element of $\mathcal{Q}$ such that $b\sigma = e$. Then

$$e\sigma e = e\sigma b\sigma = (eb)\sigma = b\sigma = e.$$

Hence $e = e\sigma e = e\sigma$, and σ preserves e.

Lemma 1.2: If q (the characteristic of V) is relatively prime to n (the block length of $\mathcal{Q}$), then the mapping $\sigma: x^i \rightarrow x^{iq}$ is an automorphism of $V[x]$.

Proof: Clearly this mapping respects addition and multiplication in $V[x]$. We have only to show that it is 1-1.

If $x^{iq} = x^{jq}$, then $(i - j)q \equiv 0 \bmod n$. Since q is prime to n, this implies that $i - j \equiv 0 \bmod n$ or $x^i = x^j$.

This proves the lemma.

Theorem 1.3: If q is prime to n, every ideal in $V[x]$ is preserved by the mapping $\sigma: x^i \rightarrow x^{iq}$.

Proof: Let $\mathcal{Q}$ be an ideal in $V[x]$ and $e = \sum_{i=0}^{n} \alpha_i x^i$, the generating idempotent of $\mathcal{Q}$.

$$\sigma e = \sum_{i=0}^{n} \alpha_i x^{iq} = \left(\sum_{i=0}^{n} \alpha_i x^i \right)^q = e^q,$$

since q is the characteristic of V.

$e = e^2 = e^3 = \cdots = e^q$; hence σ preserves e, and by Lemmas 1.1 and 1.2, σ preserves $\mathcal{Q}$.

Let the coordinate places of V^n be labeled $0, 1, \cdots, n - 1$; the map-

ping $\sigma: x^i \rightarrow x^{iq}$ in $V[x]$ corresponds to the permutation $U_q: \omega \rightarrow q\omega$ of coordinate places in V^n.

Corollary: Every binary cyclic alphabet of odd block length is preserved by the permutation $U: \omega \rightarrow 2\omega$.

This is Theorem 1 of Section II. The proof given here contains more machinery than is necessary to prove Theorem 1. This is done on purpose, in order to get Lemma 1.1, which suggests a method of finding other automorphisms of $V[x]$ which preserve a particular cyclic alphabet.

APPENDIX B

Distribution of Weights in the Cosets of a Group Code

Let $\mathcal{A}$ be an (n,k,e) binary alphabet, and let $\mathcal{B}$ be the orthogonal complement (dual alphabet) of $\mathcal{A}$ in V^n. Let $A(i)$, $B(i)$ denote the number of letters of weight i in $\mathcal{A}$, $\mathcal{B}$ respectively. The quantities $A(i)$, $B(i)$ are connected by the generating function.[7]

$$\sum_{i=0}^{n} A(i)(1 + z)^{n-i}(1 - z)^i = 2^k \sum_{i=0}^{n} B(i)z^i.$$

Since the $A(i)$ are known, the $B(i)$ may be calculated from this relationship.
Set

$$(1 + z)^{n-i}(1 - z)^i = \sum_{j=0}^{n} \psi(i,j)z^j.$$

Let $C(s,j)$ denote the number of sequences of weight s in V^n which are at distance j from some letter of $\mathcal{A}$. Then if $j \leqq e$, $C(s,j)$ and $B(i)$ are related by the generating function.[7]

$$\sum_{i=0}^{n} B(i)\psi(i,j)(1 + x)^{n-i}(1 - x)^i = 2^{n-k} \sum_{s=0}^{n} C(s,j)x^s.$$

Since $B(i)$ and $\psi(i,j)$ are known, $C(s,j)$ may be calculated from this relation.
Clearly

$$C(s) = \sum_{j=0}^{e} C(s,j).$$

REFERENCES

1. Slepian, D., A Class of Binary Signaling Alphabets, B.S.T.J., **35**, January, 1956, p. 634.

2. Neumann, P. G., A Note on Cyclic Permutation Error-Correcting Codes, Information and Control, **5**, March, 1962, p. 72.
3. Prange, E., The Use of Information Sets in Decoding Cyclic Codes, IRE Trans., IT-**8**, September, 1962, p. S-5-9.
4. Prange, E., An Algorithm for Factoring $x^n - 1$ over a Finite Field, Air Force Cambridge Research Center, AFCRC-TN-59-775.
5. Elliott, E. O., Estimates of Error Rates for Codes on Burst-Noise Channels, B.S.T.J., **42**, September, 1963, p. 1977.
6. Curtis, C. W., and Reiner, I., *Representation Theory of Finite Groups and Associated Algebras*, Interscience Publishers, New York, 1962.
7. MacWilliams, J., A Theorem on the Distribution of Weights in a Systematic Code, B.S.T.J., **42**, January, 1963, p. 79.

Optical Maser Oscillators and Noise

By EUGENE I. GORDON

The transmission line matrix formalism so useful for describing the transfer properties of microwave networks is extended to the electromagnetic fields associated with optical masers. The spontaneous emission noise of the optical maser is examined and shown to be amenable to a thermal description. Taking the point of view, well accepted at microwave frequencies, that a weakly nonlinear oscillator is a saturated amplifier of noise, the power and linewidth of the noise radiation emitted by the optical maser is calculated using the transmission line formalism. The significant parameters for any optical maser are shown to be the frequency, the single-pass gain of the maser medium, the effective mirror reflectivity and the population ratio. The pre-oscillation characteristics of the maser are examined and the reason for the extremely sharp oscillation threshold of the gas masers is discussed. Some observations concerning semiconductor optical masers are also made.

I. INTRODUCTION

This paper represents an attempt to describe the optical maser or laser from a microwave circuit point of view and is largely tutorial, since many of the results obtained from a circuit viewpoint are already known. The generality of the method of approach enlarges their area of validity, however.

Many of the people working on optical masers who do not have a background in microwave theory and techniques may find a fresh point of view. In particular, they may find a very modest introduction to an extremely well developed store of computational techniques which are applicable to optical masers. This may save them the trouble of inventing their own.

On the other hand, those who have previously been working in the field of microwaves may find that the analogies between optical masers and more conventional microwave devices are more cogent than they had appreciated. Finally, it is hoped that some of the distinctions be-

tween oscillating and pre-oscillating or subthreshold masers will be clarified.

II. THE CONVENTIONAL OSCILLATOR

Excluding strongly nonlinear oscillators with periodic but non-sinusoidal waveforms, it is often stated that an oscillator is a device having an internal gain which exceeds its total losses. Supposedly, noise triggers it off and it then continues to put out oscillatory power at a level determined only by saturation effects. The steady-state saturation level is defined as that for which the internal gain just equals the loss. An extensive discussion of microwave oscillators, based on this point of view, is given by Slater.[1]

Although this point of view often constitutes a good working definition of a feedback oscillator, it is incomplete in that it neglects the continuing presence of the noise. As a result, when the internal gain of the oscillator exactly equals the losses, so that the effective lifetime of a photon in the feedback loop is infinite, the noise power output of the oscillator must increase without limit. Similarly, the associated linewidth of the noise output must be zero. Since this situation is physically unrealizable, it is clear that the noise must be taken into account and that the steady-state gain could never exactly equal the loss and must always saturate at a slightly lower value. As a result, there could be no continuing, self-sustained oscillation which starts from noise.

From these considerations, it appears that a better, but still incomplete, description of an oscillator would be to say that a steady-state regenerative oscillator is a feedback amplifier driven to saturation by a noise input. Internally produced noise is usually the driving force; however, an additional source may be the noise entering through the output port. The external gain of the amplifier, that is, the gain experienced by any input signal, is usually extremely large unless the amplifier is saturated by the input signal. Since the amplification is obtained regeneratively, that is, by the use of feedback, the bandwidth of the gain is limited; the higher the gain the more limited it is. The amplified, narrow-band noise output is the output signal of the oscillator. When the large gain can be obtained without regeneration, the noise output need not be narrow-band.

This concept of the oscillator as a saturated amplifier of noise is not new and is well known in the microwave art. More recently, this concept has been employed by Gordon, Zeiger and Townes[2] in their treatment of the microwave maser oscillator, and by Wagner and Birnbaum,[3]

Schawlow and Townes,[4] Shimoda,[5] Blaquier[6] and Fleck[7] in their treatments of the optical maser oscillator.

While this definition of an oscillator is somewhat more satisfying, it implies that an oscillator is merely an extremely narrow-band filter with gain. As a result, the statistical properties of the noise input should be preserved, except for the spectral narrowing. For example, a filtered Gaussian noise input[8] would remain Gaussian. Since a narrow-band Gaussian noise process shows amplitude fluctuations with a time constant approximately the inverse of its spectral range,[8] the output of a filter with gain should exhibit this property also. The fact that a true oscillator does not indicates that a correlating mechanism is operative.

Gain saturation is one mechanism that operates to eliminate fluctuations in the output intensity. The action is similar to that of a limiter. In an optical maser, gain saturation arises almost entirely from depletion of the population inversion rather than from any nonlinearity in the stimulated emission process. To a very high degree, the output waveform is sinusoidal and the saturation depends only upon the time-average power, the time average extending over a time long compared to the period of the output waveform but short compared to any relaxation mechanism or pumping rate.

Suppose now that the filtered output has a spectral range $\Delta\nu$ so that the input noise has power fluctuations with a time constant approximately $\Delta\nu^{-1}$. If the gain of the maser medium has a relaxation time $\tau_g < \Delta\nu^{-1}$, then the input fluctuations will be virtually absent in the output. On the other hand, if $\tau_g > \Delta\nu^{-1}$ the input fluctuations will appear in the output.

Thus, for a true oscillator

$$\tau_g \Delta\nu < 1 \tag{1a}$$

while for an amplifying filter

$$\tau_g \Delta\nu > 1. \tag{1b}$$

As will be seen later, for a weak optical maser $\Delta\nu$ can be quite large, while for a strong maser $\Delta\nu$ becomes vanishingly small. The cavity bandwidth $\Delta\nu_c$ represents an upper limit for $\Delta\nu$. For example, in a gas maser $\Delta\nu_c \approx 10^6$ cps and $\tau_g \approx 10^{-6}$, so that for unsaturated gains less than the loss, $\Delta\nu \approx \Delta\nu_c$, $\Delta\nu\tau_g \approx 1$, and the device acts like an amplifying filter. When the unsaturated gain exceeds the loss, $\Delta\nu$ becomes much less than $\Delta\nu_c$ and the device becomes an oscillator.

In the semiconductor maser, the lowest reported value is $\Delta\nu \approx 3 \times 10^9$

cps, corresponding to 0.1Å. Unless $\tau_g < 3 \times 10^{-10}$, it is quite probable that the device acts like an amplifying filter rather than a true oscillator.

The interested reader will find a short discussion of the effect of noise on oscillators in a book by van der Ziel.[9] His discussion indicates a method of approach to obtaining a quantitative solution to a very complicated nonlinear problem. However, the concept of the amplifying filter probably yields a good first approximation to the spectral width of the oscillator output.

In the following sections, the foregoing concepts will be exploited to exhibit the circuit formalism and to study linewidth and threshold behavior. The basic results are applicable to any uniformly pumped, single-mode maser or any multimode maser for which nonlinear mixing or coupling of modes is not significant. Some remarks concerning the lack of extremely narrow linewidths in the semiconductor maser will also be made.

III. SINGLE MODE REPRESENTATION FOR AN OPTICAL MASER

The electromagnetic fields associated with the optical maser are very close to being plane waves. To a very good approximation, each mode of the electromagnetic field can be represented by field quantities, $E(z)$ and $I(z)$. These quantities can be normalized to have the dimensions of voltage and current, respectively. The relationship between $E(z)$ and $I(z)$ is obtained by specifying the value of the function $Z(z) = E(z)/I(z)$ at some point z. In the content of this paper, a mode is one member of a complete set of transverse eigenfunctions which are appropriate for the geometry in question. No orthogonality with respect to the z coordinate is implied.

If the various modes of the electromagnetic field are uncoupled and E and I are represented as complex quantities, then a linear relationship of the form

$$\begin{vmatrix} E(z_1) \\ I(z_1) \end{vmatrix} = \begin{vmatrix} A & Z_0 B \\ Z_0^{-1}C & D \end{vmatrix} \begin{vmatrix} E(z_2) \\ I(z_2) \end{vmatrix} \tag{2a}$$

or

$$\mathbf{\Psi}(z_1) = \mathbf{T}(z_1, z_2) \cdot \mathbf{\Psi}(z_2). \tag{2b}$$

can exist for each mode.[10,11] The input side is taken at z_1 and the output side at z_2, as in Fig. 1. The quantity Z_0 is the characteristic impedance associated with the transmission medium. The directions of E and I are defined so that when the phase difference between E and I falls in
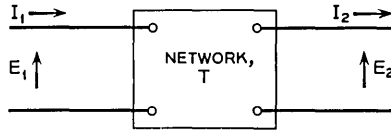
Fig. 1 — Generalized linear two-port network for single mode representation.

the first and fourth quadrant, power is flowing in the direction $z_1 \rightarrow z_2$, while for the second and third quadrant, the direction $z_2 \rightarrow z_1$. The choice of directions makes it possible to determine the result of cascading a number of sections by writing

$$\Psi(z_1) = \mathbf{T}(z_1, z_2) \cdot \mathbf{T}(z_2, z_3) \cdots \mathbf{T}(z_{n-1}, z_n) \cdot \Psi(z_n). \tag{3}$$

The complex quantities A, B, C and D are dimensionless and are independent of time in the steady state. In general, they are functions of frequency. For convenience, the characteristic impedance of the transmission medium will be taken as unity, i.e., $Z_0 = 1$. The transmission medium between planes z_n and z_{n+1} will be referred to as a "network." The properties of the network are described uniquely by the quantities A, B, C and D. General relations among these quantities can be determined by specifying the transfer properties of the network. These are reviewed in detail in Appendix A and are described below.[†] For example, a matched network which produces no reflection from side z_n when terminated by the characteristic impedance on side z_{n+1} can be written

$$\begin{vmatrix} A & B \\ B & A \end{vmatrix} \tag{4}$$

in which A and B are, in general, independent complex parameters.

A reciprocal network is characterized by the relation $AD - BC = 1$. A reactive network has the transformation

$$\begin{vmatrix} \alpha & j\beta \\ j\gamma & \delta \end{vmatrix} \tag{5}$$

in which the four independent parameters α, β, γ and δ are real. A reciprocal reactive network, in addition, has $\alpha\delta + \beta\gamma = 1$, so that only three of the parameters are independent. A matched reciprocal reactive network has $\alpha = \delta$ and $\beta = \gamma$, so that only one parameter is independent. The network of this type can be written

---

† The Appendix is included because there is no single convenient reference.

$$\begin{vmatrix} \cos \varphi & j \sin \varphi \\ j \sin \varphi & \cos \varphi \end{vmatrix} \tag{6}$$

in which the real parameter φ is the phase shift from z_n to z_{n+1}. A length of transmission medium is an example of a matched reciprocal reactive network. For this case, $\varphi = 2\pi\nu(z_{n+1} - z_n)/c'$, in which c' is the phase velocity of the radiation.

The most general unmatched reciprocal reactive network, which has three independent parameters, can always be characterized as

$$\begin{vmatrix} \alpha & j\beta \\ j\gamma & \delta \end{vmatrix} = \begin{vmatrix} \cos \varphi_1 & j \sin \varphi_1 \\ j \sin \varphi_1 & \cos \varphi_1 \end{vmatrix} \begin{vmatrix} N & 0 \\ 0 & N^{-1} \end{vmatrix} \begin{vmatrix} \cos \varphi_2 & j \sin \varphi_2 \\ j \sin \varphi_2 & \cos \varphi_2 \end{vmatrix}. \tag{7}$$

The network

$$\begin{vmatrix} N & 0 \\ 0 & N^{-1} \end{vmatrix} \tag{8}$$

is known as an ideal transformer of turns ratio N. Thus, the most general unmatched reactive reciprocal network, aside from phase shifts φ_1 and φ_2, is an ideal transformer.

A resistive network produces no phase shift and can be characterized as having A, B, C and D all real. A matched reciprocal resistive network has only one independent parameter and can be written

$$\begin{vmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{vmatrix}. \tag{9}$$

The matched reciprocal resistive network is known as an attenuator. It is shown in Appendix A that the power attenuation is given by $\exp -2\theta$. The parameter θ is referred to as the attenuator line length. The most general matched reciprocal network (resistance plus reactance) has two independent variables and can be written

$$\begin{vmatrix} \cos (\varphi - j\theta) & j \sin (\varphi - j\theta) \\ j \sin (\varphi - j\theta) & \cos (\varphi - j\theta) \end{vmatrix}$$

$$= \begin{vmatrix} \cos \varphi & j \sin \varphi \\ j \sin \varphi & \cos \varphi \end{vmatrix} \begin{vmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{vmatrix}. \tag{10}$$

Therefore, the most general matched reciprocal network consists of an attenuator with phase shift. Note that the attenuation and phase shift commute. As a result, a matched network with distributed attenuation and phase shift can be lumped into a network with attenuation in cascade with a network having only phase shift.

The transmission factor or transmissivity of the network, denoted as L, is shown in Appendix A to have the value

$$L = \frac{4}{|A + B + C + D|^2}.$$ (11)

For a matched reactive reciprocal network as in (6), $L = 1$. The factor L is also known as the gain of the network. The transmission factor equals the ratio of power transmitted to power incident when the network is preceded and followed by matched terminations.

The reflection factor or reflectivity of a network is given by

$$R = \tfrac{1}{4} L |A + B - C - D|^2.$$ (12)

The reflection factor is the ratio of power reflected to power incident when the input and output terminals are matched. For a matched network, $(A = D, B = C)$, $R = 0$.

The noise generated in a network can be represented by suitably chosen current and voltage generators i and e at the input to the network as in Fig. 2.[12] The network itself can then be considered as noiseless. For example, for a series resistor r, the noise generators are appropriately $i = 0$ and $|e| = [4p(\nu)d\nu r]^{\frac{1}{2}}$ in which

$$p(\nu) \, d\nu = \frac{h\nu d\nu}{\exp h\nu/kT - 1}$$ (13)

is the thermally generated noise power in a frequency range $d\nu$ centered at frequency ν when the resistor is at temperature T.[13] The phase of e is a random variable. For a shunt resistor, r, the noise generators are $e = 0$ and $|i| = [4p(\nu)d\nu/r]^{\frac{1}{2}}$. For reactive networks, $e = i = 0$.

For an arbitrary network containing lossy elements, the appropriate values of e and i can be expressed in terms of the components of the transformation matrix, A, B, C and D. Since an arbitrary passive network can always be matched by suitable use of transformers and line lengths, placed on either side of the network, the arbitrary network can
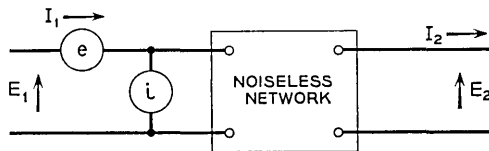
Fig. 2 — Equivalent external current and voltage generators for noiseless network representation.

always be made to appear matched and resistive. It follows that an arbitrary lossy network can always be represented as a resistive matched network imbedded in reactive networks.[14] The reactive networks on either side of the imbedded network are the inverse in reverse order of the networks required for matching the original arbitrary network.

It follows, too, that if the appropriate values of e and i can be determined for any resistive matched network, then the values can be determined for the imbedded network. These values can then be transformed through the reactive networks at the input side of the imbedded network to represent the appropriate values at the input side of the original network.

Thus, it is only necessary to have a general formula for e and i pertinent to a resistive matched network. In Appendix B it is shown that the appropriate values of e and i for any matched resistive network are given by[15]

$$\begin{aligned}
|e|^2 = |i|^2 &= AB4p(\nu)d\nu \\
ei^* = e^*i &= \tfrac{1}{2}(A^2 + B^2 - 1)4p(\nu)d\nu.
\end{aligned} \tag{14}$$

The values of e and i without the factor $4p(\nu)d\nu$ will be referred to as the normalized values. The quantities, $|e|^2$, $|i|^2$ and ei^* commute with a phase shift network (a length of transmission line). A value e following a transformer of turns ratio N becomes, at the input side of the transformer, Ne, while a current i becomes i/N. Thus, the procedure in finding the values of e and i preceding any given network, $\mathbf{T}$, is to find the appropriate reactive transformations of the form

$$\mathbf{T} = \mathbf{T}(\varphi_1) \cdot \mathbf{T}(N) \cdot T(\varphi_2) \cdot \mathbf{T}(r) \cdot \mathbf{T}(\varphi_3) \cdot \mathbf{T}(M) \cdot \mathbf{T}(\varphi_4) \tag{15}$$

in which $\mathbf{T}(r)$ is the imbedded matched resistive network, and find the values of e and i appropriate to $\mathbf{T}(r)$ using (14). The values of e' and i' appropriate to the input side of $\mathbf{T}$ are $e' = Ne$ and $i' = i/N$.

In Appendix B, it is shown that the noise power into a matched load following any given network at uniform temperature T (considering only the noise arising from the given network and ignoring the noise originating in the matched loads at the input and output side) is given by

$$dP = L|e + i|^2 p(\nu)d\nu \tag{16}$$

in which L is the transmission or gain factor for the network, given by (11), and e and i are the normalized noise generators at the input to the network. For example, for an attenuator with transmission factor

L, (16) yields

$$dP = (1 - L)p(\nu)d\nu \qquad (17)$$

as is shown in Appendix B.

Since the noise parameters commute with a matched phase shift network, attenuation and phase shift can be lumped with respect to noise properties as well as transfer properties.

IV. THE OPTICAL MASER

The maser medium has the property that light passing through the medium once is amplified by a factor $G_1(\nu)$. In addition, the light undergoes a phase shift, $\varphi(\nu)$. Thus, the matched maser medium can be characterized as an attenuator with a transmission factor G_1 in cascade with a matched reciprocal phase shift network.

Since the spontaneous emission from the maser medium can be considered as thermal noise,[16] the spontaneous emission power radiated into a given mode by a uniform maser medium should be given by

$$dP(\nu) = [1 - G_1(\nu)]p(\nu)d\nu \qquad (18)$$

as follows from (17). Since $p(\nu) = h\nu/(\exp h\nu/kT - 1)$, the question naturally arises as to what temperature to associate with the maser medium. In particular, one wonders whether a noise formula like (18), which is valid for passive networks in thermal equilibrium, can be used when the maser medium is active, i.e., when $G_1 > 1$. Normally, the radiation temperature of the uniform maser medium is defined by the Boltzmann factor[16]

$$n_2(\nu)/n_1(\nu) = \exp -h\nu/kT \qquad (19)$$

in which n_2 and n_1 are the densities of upper- and lower-state atoms, respectively. Using (19) as the definition of maser temperature, it follows that (18) is precisely correct for a uniform medium.

To illustrate, one can write for the emission power, dP, into a given mode in a frequency range $d\nu$, along the z-axis

$$\partial dP/\partial z = h\nu w_i dP(n_2 - n_1) + \tfrac{1}{2}dw_s h\nu n_2 \qquad (20)$$

in which w_i is the probability per unit intensity per unit time for stimulated emission into the mode and dw_s is the probability for spontaneous emission in either direction for the same mode into frequency range $d\nu$. The population densities for the upper and lower maser levels, n_2 and

n_1, are assumed constant with z.† Solving (20) subject to the initial condition $dP = 0$ at $z = 0$ yields

$$dP(z) = \frac{\frac{1}{2}(dw_s/w_i)(1 - G_1)}{(n_1/n_2) - 1} \tag{21}$$

in which

$$G_1 = \exp hvw_i(n_2 - n_1)z. \tag{22}$$

Since the probability for stimulated emission is related to the probability for spontaneous emission into frequency range dv by

$$dw_s = 2w_ihvdv \tag{23}$$

for a given mode, (21) and (18) are identical.‡ It follows that (18) correctly accounts for the spontaneous noise into a single mode, so long as one writes $p(v) = hv/(n_1/n_2 - 1)$. It also indicates the applicability of the formalism described in the preceding section to maser media.

The fact that the maser temperature T, as defined by (19), and $p(v)$, as defined by (13), are negative should not distract the reader from the more significant fact that (18) or (21) correctly predicts the noise power emitted by the maser medium. This quantity is never negative, and varies smoothly as T goes from positive to negative values.

The noise output from the maser can be calculated using (16), and this will be the aim of the following analysis. It is worth noting that the only significant parameters characterizing the medium, assuming that the maser medium is uniform and matched, are the total single-pass gain, G_1, the population ratio and the phase shift through the medium. The effective attenuator line length, θ, is given by the expression $G_1 = \exp -2\theta$.

V. THE REGENERATIVE OPTICAL MASER OSCILLATOR

The mirrors forming the optical cavity, being reactive (except for a small absorption loss which will be neglected) and reciprocal, can be characterized as ideal transformers. The phase shift associated with the mirrors on the cavity side can be added to the single-pass phase shift of the maser medium. The mirror phase shift external to the cavity is not significant to the problem at hand.

---

† This implies no saturation or uniform saturation.

‡ Equation (23) is equivalent to the statement that the total rate of spontaneous emission is related to the total rate of induced emission per unit intensity by $dw_s = w_i8\pi hv^3dv/c^3$, since the total number of modes per unit volume per unit frequency interval is given by $8\pi v^2/c^3$. (See Ref. 16.)

Since the reflection factor for an ideal transformer is, from (12) and (8)

$$R = (N^2 - 1)^2/(N^2 + 1)^2 \qquad (24)$$

it follows that the equivalent turns ratio for a mirror of reflectivity R is given by

$$N^2 = (1 + R^{\frac{1}{2}})/(1 - R^{\frac{1}{2}}). \qquad (25)$$

The transmission factor L_m for the maser cavity with unequal mirrors of reflectivity R and R' is obtained by combining the cascade of transformer with turns ratio N, attenuator with loss parameter $\theta = -\frac{1}{2} \ln G_1$, transmission line with phase shift $\varphi = 2\pi\nu L/c' + $ constant, in which $2L/c'$ is the equivalent round-trip time,† and transformer with turns ratio M^{-1}, $M^2 = (1 + R'^{\frac{1}{2}})/(1 - R'^{\frac{1}{2}})$, as in Fig. 3. The noise power
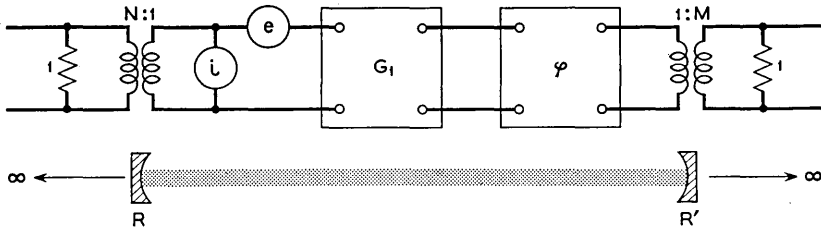
Fig. 3 — Equivalent circuit for maser cavity with mirrors of unequal reflectivity.

from one end of the maser is given by (16)

$$dP = L_m \,|\, Ne + N^{-1}i \,|^{\,2} p(\nu)d\nu \qquad (26)$$

where the normalized noise generators for the maser medium are given by

$$| e |^2 = | i |^2 = \cosh \theta \sinh \theta$$
$$ei^* = e^*i = \tfrac{1}{2}(\cosh^2 \theta + \sinh^2 \theta - 1) \qquad (27)$$

as follows from (14) and (9)

The cascade of mirror, maser medium and mirror takes the form

$$\begin{vmatrix} N & 0 \\ 0 & N^{-1} \end{vmatrix} \cdot \begin{vmatrix} \cos (\varphi - j\theta) & j \sin (\varphi - j\theta) \\ j \sin (\varphi - j\theta) & \cos (\varphi - j\theta) \end{vmatrix} \cdot \begin{vmatrix} M^{-1} & 0 \\ 0 & M \end{vmatrix}$$
$$= \begin{vmatrix} (N/M) \cos (\varphi - j\theta) & NMj \sin (\varphi - j\theta) \\ (NM)^{-1}j \sin (\varphi - j\theta) & (M/N) \cos (\varphi - j\theta) \end{vmatrix}. \qquad (28)$$

† The phase velocity c' is a function of ν and G_1 by virtue of the anomalous dispersion of the maser medium.

Using (11), the transmission factor is given by

$$
\begin{aligned}
L_m &= \frac{4}{|(N/M + M/N)\cos(\varphi - j\theta) + j(MN + [MN]^{-1})\sin(\varphi - j\theta)|^2} \\
&= \frac{1}{[(1 - R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1)^2/G_1(1 - R)(1 - R')] + [4R^{\frac{1}{2}}R'^{\frac{1}{2}}/(1 - R)(1 - R')]\sin^2\varphi}
\end{aligned}
\tag{29}
$$

Note that in the limit $R = R' = 0$ (no mirrors), $L_m = G_1$ as would be expected; while in the limit $R = R'$, $G_1 = 1$ (a transparent maser medium)

$$
L_m = \frac{1}{1 + [4R/(1 - R)^2]\sin^2\varphi}
\tag{30}
$$

which is the transmission factor for the Fabry-Perot or optical cavity surrounding the maser medium.[7,17] Equation (29), or versions of it with $R = R'$, has been derived before.[7] In these cases, however, it had been necessary to assume that the maser medium uniformly fills the region between the mirrors. No such restriction is necessary.

The noise power $|Ne + N^{-1}i|^2$ has the value, using (27)

$$
\begin{aligned}
|Ne + N^{-1}i|^2 &= (N^2 + N^{-2})|e|^2 + 2ei^* \\
&= (N^2 + N^{-2})\cosh\theta\sinh\theta \\
&\quad + (\cosh^2\theta + \sinh^2\theta - 1)
\end{aligned}
\tag{31}
$$

which after some manipulation yields

$$
|Ne + N^{-1}i|^2 = (1 - G_1)(1 + RG_1)/G_1(1 - R).
\tag{32}
$$

Combining (26), (29) and (32), the noise power in frequency range $d\nu$ leaving the maser cavity through the mirror R' can be written

$$
dP = \frac{(1 + RG_1)(1 - R')(1 - G_1)p(\nu)d\nu}{(1 - R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1)^2 + 4R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1\sin^2\varphi}.
\tag{33}
$$

In the vicinity of a cavity resonance at frequency ν_0, the phase shift φ differs from some multiple of π by an amount $\Delta\varphi = 2\pi(\nu - \nu_0)L/c'$. The free spectral range of the cavity mode at frequency ν_0 is the range $\nu_0 - c'/4L \leq \nu \leq \nu_0 + c'/4L$ as illustrated in Fig. 4; the mode spacing is $c'/2L$. Thus, the total noise power leaving the cavity through R', associated with one cavity mode, is obtained by integrating dP over the free spectral range of the mode, yielding

$$P = \int_{\nu_0 - c'/4L}^{\nu_0 + c'/4L} \frac{p(\nu)C_1 \, d\nu}{C_2 + C_3 \sin^2 2\pi(\nu - \nu_0)L/c'}. \qquad (34)$$

With the substitution $\varphi = 2\pi(\nu - \nu_0)L/c'$, (34) can be written

$$P = p(\nu_0) \frac{c'}{2\pi L} \int_{-\pi/2}^{+\pi/2} \frac{C_1 \, d\varphi}{C_2 + C_3 \sin^2 \varphi} \qquad (35)$$

in which the quantities C_1, C_2 and C_3 are given by

$$C_1 = [(1 + RG_1)(1 - R')(1 - G_1)],$$
$$C_2 = [1 - R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1]^2$$

and

$$C_3 = 4R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1.$$

Since the integrand is large only in a very small range of frequencies near ν_0, it can safely be assumed that $p(\nu)$, G_1 and R are constant with frequency and have their values at $\nu = \nu_0$.

With this approximation the integral has the value $\pi C_1/[C_2^2 + C_2 C_3]^{\frac{1}{2}}$ (see Ref. 18), yielding

$$P(R') = [(1 + RG_1)(1 - R')/(1 - RR'G_1^2)]$$
$$(1 - G_1)p(\nu_0)(c'/2L). \qquad (36)$$
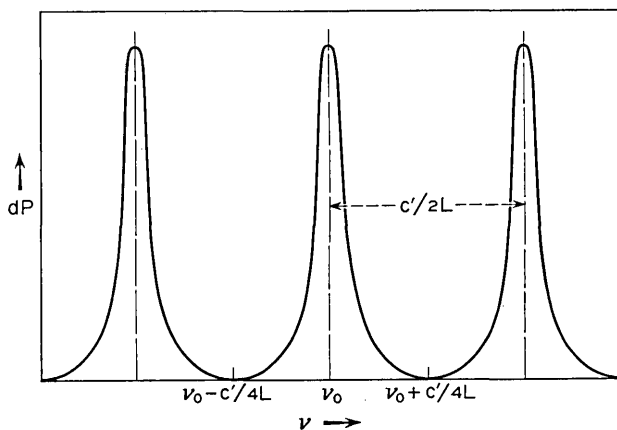
Fig. 4 — Free spectral range of cavity mode at frequency ν_0.

From (18) it may be noted that $[1 - G_1(\nu_0)]p(\nu_0)c'/2L$ is the spontaneous emission power that would be emitted by the maser medium, in the absence of the cavity, into the spectral range $c'/2L$ if the gain were constant over that range. When $G_1 > 1$, the spontaneous emission is enhanced by stimulated emission. The noise power

$$[1 - G_1(\nu_0)]p(\nu_0)c'/2L$$

will be denoted as $P_s(\nu_0)$. The noise power leaving the cavity through mirror R' is, therefore

$$P(R') = P_s(1 + RG_1)(1 - R')/(1 - RR'G_1^2). \tag{37}$$

The power leaving the cavity through mirror R is given by (37) with R and R' interchanged. The total power leaving the cavity is

$$P_t = 2P_s[1 - RR'G_1 + \tfrac{1}{2}(G_1 - 1)(R + R')]/[1 - RR'G_1^2] \tag{38}$$

so that the cavity enhances the spontaneous emission by the factor following $2P_s$.

It should be noted that the integration of (36), which leads to (37) and (38), is valid even if $G_1(\nu)$ varies over the range $c'/2L$, so long as the frequency range over which G_1 has a significant variation is large compared to the spectral range of the noise power. This will always be so, providing the natural linewidth of the transition is large compared to the spectral range of the noise, independent of whether the transition is homogeneously or inhomogeneously broadened. Equations (37) and (38) are valid even when the gain profile is saturated, providing the maser medium is uniformly saturated with respect to the axial direction. In general, the saturation will tend to be uniform, assuming uniform pumping, since the power in the cavity tends to be uniform with length. For very high-gain maser media the latter statement is not valid.

It should be noted that (38) contains an implication which is not immediately obvious. In the limit $n_1 = n_2$ with $G_1 = 1$ and P_s finite, (38) states that $P_t = 2P_s$ independent of R or R'. Thus the total spontaneous emission noise power leaving the optical cavity is independent of its bandpass. The spectral distribution of the noise is altered, however, to correspond to the bandpass.†

The preceding results can be used to determine the linewidth of the oscillator. The spectral range or bandwidth of the noise power is obtained from (33) by determining the frequencies at which dP falls to

---

† In the limit $c'/2L$ very much less than the inhomogeneously broadened linewidth, (38) implies that the total spontaneous emission power into all modes is independent of the mirror reflectivity since $G_1 = 1$ over the entire line.

half its value at $\nu = \nu_0$. This occurs at frequencies $\nu_{\frac{1}{2}}$ defined by

$$4R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1 \sin^2 2\pi(\nu_{\frac{1}{2}} - \nu_0)L/c' = (1 - R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1)^2 \qquad (39)$$

yielding a spectral width $\Delta\nu = 2(\nu_{\frac{1}{2}} - \nu_0)$ given by

$$\Delta\nu = (c'/L\pi) \sin^{-1} \tfrac{1}{2}(1 - R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1)/(RR'G_1^2)^{\frac{1}{4}}$$
$$\approx (c'/2L\pi)(1 - R^{\frac{1}{2}}R'^{\frac{1}{2}}G_1)/(RR'G_1^2)^{\frac{1}{4}}. \qquad (40)$$

Equation (40) can be rewritten by substituting for $1 - (RR')^{\frac{1}{2}}G_1$ the value given by (38)

$$1 - (RR'G_1^2)^{\frac{1}{2}} = 2(P_s/P_t)[1 - RR'G_1 + \tfrac{1}{2}(G_1 - 1)(R + R')]/ \qquad (41)$$
$$[1 + (RR'G_1^2)^{\frac{1}{2}}].$$

The resulting expression is a function of the single-pass gain G_1. For the strong oscillator the gain saturates at a value such that $RR'G_1^2$ differs from unity by a vanishingly small amount. Thus it is expedient, in the expression for $\Delta\nu$ found by substituting (41), to write for the saturated single-pass gain

$$G_1 \equiv [1 - (1 - (RR'G_1^2)^{\frac{1}{2}})]/(RR')^{\frac{1}{2}} \qquad (42)$$

and then to replace $1 - (RR'G_1^2)^{\frac{1}{2}}$ by its approximate value obtained from (41). In the reiteration, the approximate value can be written

$$1 - (RR'G_1^2)^{\frac{1}{2}} \approx \frac{(R^{\frac{1}{2}} + R'^{\frac{1}{2}})^2}{4(RR')^{\frac{1}{2}}} \frac{(1 - (RR')^{\frac{1}{2}})}{(RR')^{\frac{1}{4}}} \frac{2\pi h\nu\Delta\nu_c}{(1 - n_1/n_2)P_t} \qquad (43)$$

which follows by substituting in (41) the value $G_1 = (RR')^{-\frac{1}{2}}$ and replacing P_s by its value $(1 - G_1)(c'/2L)h\nu/(n_1/n_2 - 1)$. The cavity bandwidth $\Delta\nu_c$ is given by (40) with $G_1 = 1$.

Combining (40–43) and performing the necessary algebra yields

$$\Delta\nu = 2\pi(h\nu/P_t)(\Delta\nu_c)^2(1 - n_1/n_2)^{-1}z[1 + (\pi h\nu\Delta\nu_c/P_t) \qquad (44)$$
$$\cdot(1 - n_1/n_2)^{-1} \cdot (- z[x + 3] + 1 + x)/x^{\frac{1}{2}}]$$

in which

$$z = (R^{\frac{1}{2}} + R'^{\frac{1}{2}})^2/4(RR')^{\frac{1}{2}} \qquad (45)$$

is a term which is identically unity when $R = R'$ and remains close to unity even for R and R' differing by a factor of ten, and the quantity $x = (RR')^{\frac{1}{2}}$.

When $h\nu\Delta\nu_c/P_t \ll 1$ the linewidth is

$$\Delta\nu \approx 2\pi \frac{h\nu}{P_t} (\Delta\nu_c)^2 (1 - n_1/n_2)^{-1} \qquad (46)$$

and differs from the well-known Schawlow-Townes[4] formula by a factor of two. The correction term $(1 - n_1/n_2)^{-1}$ approaches unity for the ideal maser $(n_1/n_2 = 0)$. This term has also been found by Shimoda.[5] It should be noted that n_1/n_2 is the saturated value and in general may be very difficult to evaluate. An approximate value may be found by setting the single-pass gain given by (22) equal to the saturated gain given by (42). Extracting the appropriate value of n_1/n_2 requires a knowledge of the transition probability, the time constants for the upper and lower states and the mirror reflectivities. The degeneracy of the upper and lower levels should also be taken into account.

For the 6328 Å gas maser, typical values for a one-meter-long discharge are $P_t \approx 10^{-3}$ watts/mode, $\Delta\nu_c \approx 10^6$ cps and $n_1/n_2 = 0.98$, yielding a linewidth of 10^{-1} cps.

For the semiconductor optical maser, (46) predicts a linewidth of approximately 5×10^8 cps, taking $L = 0.06$ cm and $P_s/P_t \approx 10^{-2}$. The latter numbers are taken from the data of Quist et al.[19] The corresponding wavelength range is 10^{-2} Å. This estimate is probably a conservative lower limit because of the neglect of internal losses in the derivation and the fact that the internal losses are significant in the actual device. The lowest observed linewidths are about 10^{-1} Å. The relatively large linewidths arise from very short spontaneous emission lifetimes ($<10^{-9}$ sec). The relatively large amount of enhanced spontaneous emission available produces saturation at small values of P_t/P_s.

As has just been shown, it is possible to write formal expressions for the power output and spectral width of the noise emitted in a given mode. Since the power, and hence the spectral width, depend on the saturated single-pass gain, it is necessary to take the dynamic properties of the maser medium into account. This is illustrated graphically in Fig. 5, in which the curve of P vs G_1 as represented by (38) is shown. Also shown are a curve of $\Delta\nu$ and three dashed curves representing the dynamic properties of the maser medium. The latter curves may be found by solving the rate equations for the maser medium to obtain a relation between the power taken from the maser medium and the single-pass gain. When large power is taken from the maser medium, G_1 must approach unity, since the population inversion must approach zero. When the noise power taken from the maser medium approaches zero, the single-pass gain approaches its small signal value G_{10}. The curve representing the dynamics of the maser medium intersects the curve representing the static characteristics of the cavity at some value $G_1 < (1/RR')^{\frac{1}{2}}$. This is the operating point of the maser. The value of G_1 at the operating point determines the value of $\Delta\nu$.
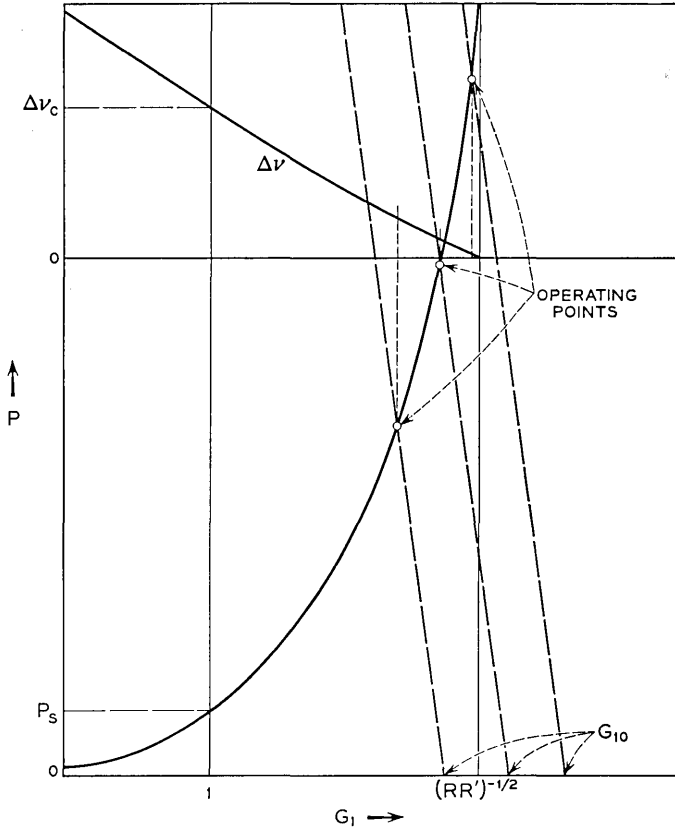
Fig. 5 — Noise power vs single-pass gain as represented by (38).

For a gas maser with small single-pass gain, the dynamic properties of the maser medium can be shown to be controlled by an equation of the form[20]

$$G_1 \approx 1 + k_0 l / (1 + \kappa P_c / P_s) \tag{47}$$

in which k_0 is the small signal gain parameter for the mode in question, l is the active length of the medium, κ/P_s is a saturation parameter dependent on the Einstein A and B coefficients for the maser levels and P_c is the power in the cavity. In most maser media $\kappa \ll 1$, so that G_1 varies quite slowly with P_c/P_s. It should be noted that P_s is essentially constant and can be evaluated for $G_1 = 1$.

The gain parameter k_0 varies with discharge current. For low cur-

rents, k_0 is proportional to current.[20] Thus, as the discharge current is increased from zero, the output power increases. From Fig. 5 it would be expected that as the discharge current approaches a value for which $1 + k_0 l$ approaches $(RR')^{-\frac{1}{2}}$, the output power would show an extremely sharp rise with current. This can be illustrated by writing for G_1 as a function of discharge current I, taking $R = R'$ for convenience,

$$G_1 \approx 1 + [(1 - R)/R](I/I_0)/(1 + \kappa P_c/P_s) \tag{48}$$

in which I_0 is the current for which $G_1 R = 1$ in the absence of saturation. From (38), taking $P_c = P_t/2(1 - R)$

$$P_c/P_s = 1/(1 - RG_1). \tag{49}$$

Solving (48) and (49) for P_c/P_s yields

$$P_c/P_s = [\Delta I/I_0 + \kappa + \sqrt{(\Delta I/I_0 + \kappa)^2 + 4\kappa}]/2\kappa \tag{50}$$

in which $\Delta I = I - I_0$. Note that when $(\Delta I/I_0) + \kappa = 0$, $P_c/P_s = \kappa^{-\frac{1}{2}}$. When $(\Delta I/I_0) + \kappa$ has the value $\kappa^{\frac{1}{2}}(f - f^{-1})$, P_c/P_s has the value $f\kappa^{-\frac{1}{2}}$. Therefore, the change in P_c/P_s by a factor f is larger than the change in $\Delta I/I_0$ of $\kappa^{\frac{1}{2}}(f - f^{-1})$ by a factor $\kappa^{-\frac{1}{2}}$. Thus, when $\kappa \ll 1$ the power output shows a very sharp threshold with current. The value of κ is of order 10^{-10} for the 6328 Å gas maser.

VI. THE NONREGENERATIVE OPTICAL MASER OSCILLATOR

Some maser media have such large single-pass gain that spontaneous emission originating at one end can be amplified sufficiently to saturate the maser medium at the opposite end. For example, the 3.39 μ transition in neon, in a helium-neon gas maser, has gains of order 50 db/meter in small-bore tubing.[21] The saturated output power is in the 1–10 mw range. Xenon at 3.5 μ has even larger gain.[22] Under these circumstances the maser can behave as a saturated oscillator without an optical cavity. The extreme line narrowing characteristic of the cavity oscillator will be lacking, but the power levels and directionality will be comparable.

This type of oscillator is characterized by a peak output always at the line center, independent of temperature and any physical dimension, line narrowing over the inhomogeneously broadened line and an extremely stable output. The power per steradian per cycle will be much greater than conventional light sources, and so these structurally simple oscillators may be very useful as frequency standards and for calibration purposes.

Some idea of the line narrowing possible with this type of oscillator can be obtained by reference to (18), which describes the noise power emanating from one end of the unsaturated maser medium in one mode with no mirrors at either end, while (33) with $R' = 0$ gives the noise power when one end has a mirror, but the other does not

$$dP = [1 + RG_1(\nu)][1 - G_1(\nu)]p(\nu)d\nu. \tag{51}$$

When $R = 0$, (51) properly yields (18); however, when $R = 1$

$$dP = [1 - G_1^2(\nu)]p(\nu)d\nu. \tag{52}$$

Since G_1^2 is the gain of a maser of twice the length of a maser of gain G_1, the perfectly reflecting mirror placed at one end serves to double the effective length of the maser medium.

Assuming no saturation, so that the gain can be written

$$G_1(\nu) = \exp k(\nu)l \tag{53}$$

in which l is the effective length of the medium, and assuming further that the gain parameter $k(\nu)$ has a Doppler profile

$$k(\nu) = k_0 \exp - [2(\nu - \nu_0)/\Delta\nu_D]^2 \ln 2 \tag{54}$$

in which ν_0 is the frequency of the line center and $\Delta\nu_D$ is the full Doppler width, the spectral width at half power of the spontaneous emission $\Delta\nu$ can be determined by writing

$$1 - \exp [k_0 l \exp - [\Delta\nu/\Delta\nu_D]^2 \ln 2] = \tfrac{1}{2}[1 - \exp k_0 l]. \tag{55}$$

In the limit $k_0 l \gg 1$, (55) can be solved for $\Delta\nu/\Delta\nu_D$ to yield

$$\Delta\nu/\Delta\nu_D = (k_0 l)^{-\frac{1}{2}}. \tag{56}$$

The 3.39 μ maser is saturated by its own spontaneous emission by gains of order 80 db, so that the maximum possible value of $k_0 l$ before saturation is about 20, corresponding to $l \approx 2$ meters. This yields $\Delta\nu \approx \Delta\nu_0/4.5$. The Doppler width at 3.39 μ is about 300 mc, yielding $\Delta\nu \approx 70$ mc. It would be possible to decrease this value somewhat by using several lengths of maser media separated by attenuators to prevent saturation. The slow dependence on l exhibited by (56) indicates the impracticality of this scheme in achieving any more than a factor of four decrease in linewidth unless k_0 can be increased drastically. Fortunately, there is evidence that this can be done, and the nonregenerative oscillator may turn out to be an extremely interesting device. In addition, it can be tuned by application of magnetic fields.

VII. CONCLUSION

The output properties of an optical maser oscillator have been derived subject to the supposition that the oscillator is a saturated amplifier of spontaneous emission noise. The most significant new results concern an expression for the linewidth of the oscillating maser which differs from the commonly accepted value. Some techniques, well-known in the microwave art and used only in a limited way in optics (stratified media), have been generalized to apply to the transmission and noise properties of optical masers.

VIII. ACKNOWLEDGMENT

The author is indebted to H. Seidel, from whom he learned most of the circuit formalism herein, and to A. D. White, J. D. Rigden, and J. E. Geusic for many stimulating discussions.

APPENDIX A

Network Representations

The basic network of interest will be a linear two-port network which will be represented schematically by Fig. 1. The major concern here will be the relationships between the quantities I_1 and E_1 at port 1 and I_2 and E_2 at port 2. The linear relationship among these quantities will be represented by the transformation matrix

$$\begin{vmatrix} E_1 \\ I_1 \end{vmatrix} = \begin{vmatrix} A & B \\ C & D \end{vmatrix} \cdot \begin{vmatrix} E_2 \\ I_2 \end{vmatrix} \tag{57a}$$

which will be abbreviated by

$$\Psi_1 = \mathbf{T} \cdot \Psi_2 \tag{57b}$$

in which

$$\Psi_i = \begin{vmatrix} E_i \\ I_i \end{vmatrix} \qquad i = 1, 2, \cdots \tag{58}$$

and

$$\mathbf{T} = \begin{vmatrix} A & B \\ C & D \end{vmatrix}. \tag{59}$$

Note that the direction of positive current flow is defined to be into the network at port 1 and out of the network at port 2. This choice sim-

plifies the discussion of cascaded networks. It is clear that two networks in cascade can be represented by $\Psi_1 = T_1 \cdot \Psi_2 = T_1 \cdot T_2 \cdot \Psi_3$. The product $T_1 \cdot T_2$ follows the usual rules of matrix multiplication. In general, $T_1 \cdot T_2 \neq T_2 \cdot T_1$; that is, the two networks do not commute. For a cascade of n networks

$$\Psi_1 = T_1 \cdot T_2 \cdots T_n \cdot \Psi_{n+1}. \tag{60}$$

The determinant of the transformation will be a quantity of interest and will be defined as $\Delta = AD - CB$. The inverse of (57) is defined by

$$\Psi_2 = T^{-1} \cdot \Psi_1 \tag{61}$$

in which

$$T^{-1} = \frac{1}{\Delta} \begin{vmatrix} D & -B \\ -C & A \end{vmatrix} \tag{62}$$

is the inverse transformation matrix. Reference will sometimes be made to the exchange network $\underset{\sim}{T}$, which is merely the network T with its terminals exchanged so that port 2 becomes port 1 and vice versa. The components of $\underset{\sim}{T}$ are obtained by writing $\Psi_2 = T^{-1} \cdot \Psi_1$ and noting that for the exchange network the quantities E_2, I_2 become E_1', $-I_1'$ and E_1, I_1 become E_2', $-I_2'$. As a result, the correct description is given by

$$\begin{vmatrix} E_1' \\ I_1' \end{vmatrix} = \frac{1}{\Delta} \begin{vmatrix} D & B \\ C & A \end{vmatrix} \cdot \begin{vmatrix} E_2' \\ I_2' \end{vmatrix} \tag{63}$$

and

$$\underset{\sim}{T} = \frac{1}{\Delta} \begin{vmatrix} D & B \\ C & A \end{vmatrix}. \tag{64}$$

The net power *into* port 1 is given by

$$P_1 = \tfrac{1}{2}(E_1 I_1^* + E_1^* I_1) = \tfrac{1}{2}\overline{\Psi}_1 \cdot \Sigma \cdot \Psi_1 \tag{65}$$

in which $\overline{\Psi}$ is the complex transpose of Ψ and

$$\Sigma = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}. \tag{66}$$

Likewise, the net power *out* of port 2 is given by

$$P_2 = \tfrac{1}{2}\overline{\Psi}_2 \cdot \Sigma \cdot \Psi_2 \,. \tag{67}$$

Other transformations will be introduced as they are needed. In the following, certain types of networks will be characterized in terms of relationships among the components of the transformation matrix, i.e., A, B, C and D. These networks will form the basic "tools" of the analyses.

A.1 Matched Networks

If one of the ports of the network is terminated in a matched load and it is then observed that the input impedance of the other port is matched to the line, the network is said to be matched. First, consider the input impedance of the network in Fig. 6. The line is assumed to have unit characteristic impedance. The quantities B and C are thus normalized with respect to the characteristic impedance and admit-
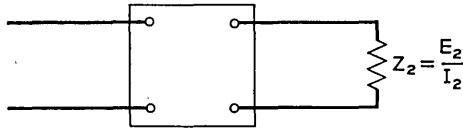
Fig. 6 — Terminating impedance for matched network.

tance of the line, respectively. Writing $E_1 = AE_2 + BI_2$, $I_1 = CE_2 + DI_2$ and

$$Z_{\text{input(1)}} = E_1/I_1 = (AZ_2 + B)/(CZ_2 + D) \tag{68}$$

in which $Z_2 = E_2/I_2$ is the terminating impedance. The network is then matched if $Z_{\text{input(1)}} = Z_2 = 1$, yielding the requirement $A + B = C + D$. Likewise, the exchange network must also be matched, requiring $D + B = C + A$. The two requirements yield the relations $A = D$, $B = C$, and for a matched network

$$\mathbf{T} = \begin{vmatrix} A & B \\ B & A \end{vmatrix}. \tag{69}$$

A.2 Reciprocal Networks

A reciprocal network has the same transmission properties in either direction. It follows that for a reciprocal network $\mathbf{T}$ and $\underset{\sim}{\mathbf{T}}$ must have the same transmission factor and phase shift, which from (84) and (72) implies $\Delta = 1$. The simplest nonreciprocal network can be written

$$\mathbf{T}_a = \begin{vmatrix} \Delta^{\frac{1}{2}} & 0 \\ 0 & \Delta^{\frac{1}{2}} \end{vmatrix} \tag{70}$$

with $\Delta \neq 1$; and all nonreciprocal networks can be written

$$\mathbf{T} = \begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} \Delta^{\frac{1}{2}} & 0 \\ 0 & \Delta^{\frac{1}{2}} \end{vmatrix} \cdot \begin{vmatrix} A/\Delta^{\frac{1}{2}} & B/\Delta^{\frac{1}{2}} \\ C/\Delta^{\frac{1}{2}} & D/\Delta^{\frac{1}{2}} \end{vmatrix} = \mathbf{T}_a \cdot \mathbf{T}_r . \tag{71}$$

The reciprocal network $\mathbf{T}_r$ is known as the reduced network; all the nonreciprocity resides in $\mathbf{T}_a$. The network $\mathbf{T}_a$, known as the abstracted network, commutes with all networks.

In general, Δ is complex and can be written $\Delta = |\Delta| \exp j \, 2\varphi$. Since

$$\mathbf{T}_a = \begin{vmatrix} |\Delta|^{\frac{1}{2}} \exp j\varphi & 0 \\ 0 & |\Delta|^{\frac{1}{2}} \exp j\varphi \end{vmatrix} \tag{72}$$

and the exchange network

$$\mathbf{\underset{\sim}{T}}_a = \begin{vmatrix} |\Delta|^{-\frac{1}{2}} \exp -j\varphi & 0 \\ 0 & |\Delta|^{-\frac{1}{2}} \exp -j\varphi \end{vmatrix} \tag{73}$$

it follows that argument φ is the nonreciprocal phase shift.

A.3 Reactive Networks

Since a reactive network is lossless, it would be expected that $P_2 = P_1$. It follows that $\overline{\Psi}_2 \Sigma \Psi_2 = \overline{\Psi}_2 \overline{T} \Sigma T \Psi_2$, so that for a reactive network $\overline{T} \Sigma T = \Sigma$. Performing the indicated matrix multiplication yields

$$\begin{vmatrix} A^*C + AC^* & A^*D + C^*B \\ AD^* + CB^* & BD^* + B^*D \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}. \tag{74}$$

It follows that A^*C and BD^* are imaginary and $A^*D + C^*B = 1$. Note that multiplication of $A^*D + C^*B = 1$ by CB^* yields $CB^* = (A^*C)(B^*D) + |C|^2 |B|^2$, which is real. Likewise,

$$AD^* = |A|^2 |D|^2 + (AC^*)(BD^*)$$

is real. If the reactive network is reciprocal so that $\Delta = AD - BC = 1 = (AC^*)(D/C^*) - D^*B(C/D^*) = (A/B^*)DB^* - (B/A^*)A^*C$, it follows that both D/C^* and A/B^* are imaginary. Since AC^* is imaginary and C^*B is real, A/B is imaginary. Thus B/B^* is real, which can only occur when B is real or imaginary. It follows that B and C are real (or imaginary) while A and D are imaginary (or real). The question of which set to choose real is decided by noting that the idemfactor (no network) is a reactive reciprocal network. Thus, the appropriate choice is A, D real and B, C imaginary, and the reactive reciprocal

network can be written

$$\begin{vmatrix} \alpha & j\beta \\ j\gamma & \delta \end{vmatrix} \tag{75}$$

for which $\alpha\delta + \beta\gamma = 1$. If the reactive reciprocal network is also matched so that $\alpha = \delta$ and $\beta = \gamma$, the network can be written

$$\begin{vmatrix} \cos\varphi & j\sin\varphi \\ j\sin\varphi & \cos\varphi \end{vmatrix} \tag{76}$$

in which the parameter φ is known as the angular length or phase shift. Note that the input impedance for this network is

$$Z_{\text{input}(1)} = \frac{Z_2 + j\tan\varphi}{1 + Z_2 j\tan\varphi} \tag{77}$$

which is recognizable as the impedance transformation for a transmission line of angular length φ and unit characteristic impedance.

Next, certain basic passive circuit elements of interest will be characterized.

A.4 *Series Impedance (Fig. 7)*

Since $I_1 = I_2$ and $E_1 = I_2 Z + E_2$, the transfer matrix is given by

$$\mathbf{T} = \begin{vmatrix} 1 & Z \\ 0 & 1 \end{vmatrix}. \tag{78}$$

A.5 *Shunt Admittance (Fig. 8)*

For the shunt admittance, $E_1 = E_2$ and $I_1 = E_2 Y + I_2$; thus

$$\mathbf{T} = \begin{vmatrix} 1 & 0 \\ Y & 1 \end{vmatrix}. \tag{79}$$

Both networks are reciprocal but not matched.
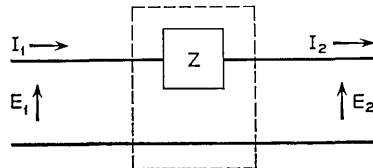
Fig. 7 — Series impedance.
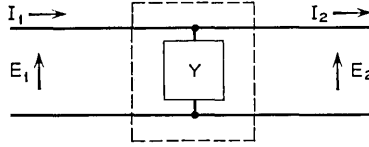
Fig. 8 — Shunt admittance.

A.6 *Ideal Transformer of Turns Ratio N :1 (Fig. 9)*

For the transformer $E_1 = NE_2$ and $I_1 = I_2/N_1$. Thus

$$T = \begin{vmatrix} N & 0 \\ 0 & N^{-1} \end{vmatrix}. \qquad (80)$$
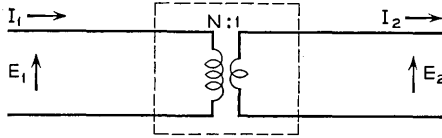
Fig. 9 — Ideal transformer.

A.7 *Network Transmissivity*

Consider a general two-port network which has a matched generator on side 1 and a matched load on side 2. The cascade of networks may be represented schematically as shown in Fig. 10, yielding

$$\begin{vmatrix} E_g \\ I_g \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} A & B \\ C & D \end{vmatrix} \cdot \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} \cdot \begin{vmatrix} E_2 \\ 0 \end{vmatrix}. \qquad (81)$$
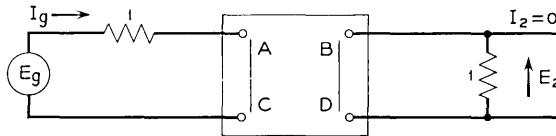
Fig. 10 — Circuit for two-port network with matched generator on side 1 and matched load on side 2.

Performing the indicated matrix multiplication yields

$$\begin{vmatrix} E_g \\ I_g \end{vmatrix} = \begin{vmatrix} A + B + C + D & B + D \\ C + D & D \end{vmatrix} \cdot \begin{vmatrix} E_2 \\ 0 \end{vmatrix}. \qquad (82)$$

The network transmission factor is defined as the ratio

$$L = |E_2|^2 / |E_{20}|^2 \tag{83}$$

in which E_{20} is the voltage that would be measured if the network were absent. In the latter case, the appropriate two-port transformation must be the identity matrix, so that $A = D = 1$, $B = C = 0$. Hence

$$L = \frac{|E_g|^2}{|A + B + C + D|^2} \Big/ \frac{|E_g|^2}{4} = \frac{4}{|A + B + C + D|^2}. \tag{84}$$

A transmission factor greater than one implies gain. The network transmission factor for a matched network is

$$L = 4|2A + 2B|^{-2} = |A + B|^{-2}. \tag{85}$$

A.8 *Network Reflectivity*

When a network is followed by a matched load, the input impedance to the network is given by $Z_{\text{input}} = (A + B)/(C + D)$. The power reflectivity, R, is given by

$$R = \left| \frac{(Z_{\text{input}} - 1)}{(Z_{\text{input}} + 1)} \right|^2 = \frac{|A + B - C - D|^2}{|A + B + C + D|^2} \tag{86}$$

$$= \tfrac{1}{4} L |A + B - C - D|^2$$

in which L is the transmission factor of the network. Note that for a matched network, $A = D$, $B = C$ and $R = 0$.

A.9 *Matched Attenuator*

An attenuator is a nonreactive reciprocal device. Thus, a matched attenuator must satisfy the conditions $A = D$, $B = C$ and $AD - BC = 1$, with A, B, C and D real. This is automatically satisfied by writing

$$\mathbf{T} = \begin{vmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{vmatrix}. \tag{87}$$

The parameter θ is referred to as the line length of the attenuator. The attenuator commutes with a matched reciprocal reactive network (a transmission line, for example) and the resultant network

$$\begin{vmatrix} \cos \varphi & j \sin \varphi \\ j \sin \varphi & \cos \varphi \end{vmatrix} \begin{vmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{vmatrix}$$

$$= \begin{vmatrix} \cos (\varphi - j\theta) & j \sin (\varphi - j\theta) \\ j \sin (\varphi - j\theta) & \cos (\varphi - j\theta) \end{vmatrix} \tag{88}$$

is the representation for an attenuator with phase shift or angular length φ. From the fact that a lossy transmission line would have a phase shift of the form $\exp -j(\varphi - j\theta) = \exp -\theta \exp -j\varphi$, it would be expected that the transmission factor is given by $\exp -2\theta$. For an attenuator the transmission factor is obtained by using (85) and (87)

$$L = |\cosh \theta + \sinh \theta|^{-2} = \exp -2\theta. \tag{89}$$

Thus, the attenuator matrix is specified completely by knowledge of its transmission factor.

A.10 Isolator

An isolator is by definition nonreciprocal although it is matched. Thus, an isolator with forward transmission unity and reverse transmission $\exp -2\theta$ can be synthesized by cascading an attenuator of line length $\theta/2$, having a transmission $\exp -\theta$, with a nonreciprocal network with transmission $\exp +\theta$ in the forward direction and transmission $\exp -\theta$ in the backward direction. In its most simple form, the nonreciprocal network (70) is given by

$$\begin{vmatrix} \exp -\theta/2 & 0 \\ 0 & \exp -\theta/2 \end{vmatrix}. \tag{90}$$

Thus, the isolator can be represented by

$$\begin{aligned} \mathbf{T} &= \begin{vmatrix} \exp -\theta/2 & 0 \\ 0 & \exp -\theta/2 \end{vmatrix} \begin{vmatrix} \cosh \theta/2 & \sinh \theta/2 \\ \sinh \theta/2 & \cosh \theta/2 \end{vmatrix} \\ &= \begin{vmatrix} \exp -\theta/2 \cosh \theta/2 & \exp -\theta/2 \sinh \theta/2 \\ \exp -\theta/2 \sinh \theta/2 & \exp -\theta/2 \cosh \theta/2 \end{vmatrix}. \end{aligned} \tag{91}$$

and the transmission factor is unity in the forward direction and $\exp -2\theta$ in the backward direction. An ideal isolator is one for which the line length θ approaches infinity, yielding

$$\mathbf{T} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix}. \tag{92}$$

APPENDIX B

Noisy Networks

In the following, the source of noise will be considered to be spontaneous fluctuations which arise because of the thermal properties of the material. At extremely high frequencies, the thermal noise is more

commonly called black-body radiation. The noise spectrum will be characterized by the following statement of Nyquist's theorem:[13] The noise power *available* per mode at frequency ν in a small frequency interval, $d\nu$, is given by

$$dP = p(\nu)d\nu$$
$$p(\nu) = h\nu(\exp h\nu/kT - 1)^{-1}$$

(93)

for a passive circuit at temperature T. The constant k is Boltzmann's constant (1.38×10^{-23} joules/degree) and h is Planck's constant (6.6×10^{-34} joules-sec).

It is sometimes convenient to write the noise power in terms of equivalent rms voltage and current generators e and i as shown in Fig. 11. The internal impedance of the voltage generator is zero and for the current generator it is infinite, and one can write

$$| e | = (4rp(\nu)d\nu)^{\frac{1}{2}}, \qquad | i | = (4gp(\nu)d\nu)^{\frac{1}{2}}.$$

(94)

In the following, both e and i will have the units of (power)$^{\frac{1}{2}}$ since r and g are normalized with respect to the line impedance and admittance.

A systematic method of handling the noise produced by a network will be developed next. First, the following theorem will be stated without proof: *Any passive noisy two-port network can be replaced by an equivalent noise-free network which has an added shunt current generator and series voltage generator at the input or output terminals which represent the noise contribution.*[12] Therefore, any passive noisy network can be replaced by the representation shown in Fig. 2, in which the network is now noise-free but the noise appears from equivalent voltage and current generators at one of the terminals. A proof for the theorem can be given, and although it is relatively simple, the proof is lengthy.

Next, a scheme for representing in a simple way the additional current and voltage appearing at the terminal will be described. The following technique is due to H. Seidel.[15] It is clear that one may always write

$$E_1 = AE_2 + BI_2 + e$$
$$I_1 = CE_2 + DI_2 + i.$$

(95)
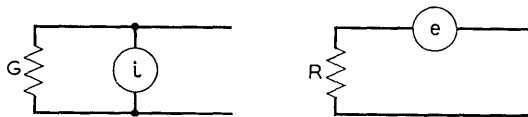
Fig. 11 — Noise-free representation of network with equivalent external voltage and current generators at one terminal.

The inclusion of the noise generators can be accomplished in an artificial but, as will be seen, highly useful way by writing

$$\begin{vmatrix} E_1 \\ I_1 \\ 1 \end{vmatrix} = \begin{vmatrix} A & B & e \\ C & D & i \\ 0 & 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} E_2 \\ I_2 \\ 1 \end{vmatrix}. \tag{96}$$

This representation results from adding the trivial equation $1 = 0 + 0 + 1$ to the set in (95). No new information has been added, but the bookkeeping advantages afforded by this change will become apparent shortly.

Equation (96) will be the general representation for a passive noisy two-port network. The problem now is to learn how to characterize the noise quantities e and i in terms of the properties of the network represented by A, B, C and D. It will be seen that this can be done by comparing the network to some simple network whose properties are known.

First, it will be demonstrated that for any passive network, one may take A, B, C and D either all real or all imaginary. This is equivalent to saying that the network may always be taken to appear purely resistive. This is clear since the input impedance is $Z_{\text{in}} = (AZ_2 + B)$ $(CZ_2 + D)$. If Z_2 is real, then it would be expected that a resistive network will have Z_{in} real also. The proof follows from the fact that the input impedance can always be made real with a suitable length of transmission line. In addition, with an ideal transformer one may match the input impedance to the line. It follows, therefore, that for noise calculations, one only need consider matched networks ($A = D$, $B = C$) with the ratio A/B real.

First, consider a matched network which is cascaded with a transmission line of arbitrary length θ. The latter network has no loss and has no source of noise. In addition, the transmission line can change only the phase but not the magnitude of the noise current and voltage generators. Consequently, one expects that for a matched network

$$\begin{vmatrix} \cos\theta & j\sin\theta & 0 \\ j\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} A & B & e \\ B & A & i \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} X & X & e\cos\theta + ij\cos\theta \\ X & X & ej\sin\theta + i\cos\theta \\ 0 & 0 & 1 \end{vmatrix}$$

has the same noise properties independent of θ. Thus, it is required that

$$|e|^2 = |e\cos\theta + i\sin\theta|^2 = |e|^2\cos^2\theta + |i|^2\sin^2\theta$$
$$+ e(ji)^* + e^*(ji) \sin\theta\cos\theta$$

or

$$(|e|^2 - |i|^2)\sin^2\theta + j(ei^* - e^*i)\sin\theta\cos\theta = 0.$$

Choosing $\theta = \pi/2$ yields

$$| e |^2 = | i |^2 \tag{97}$$

and it follows also that

$$e i^* = e^* i. \tag{98}$$

Next, the passive matched network at temperature T will be cascaded with a shunt conductance at the same temperature and the open-circuit noise voltage at the input terminals will be determined. The network representation for the shunt conductance in the new formalism follows from Nyquist's theorem as is shown in Fig. 11

$$| i' | = (4Gp(\nu)d\nu)^{\frac{1}{2}}$$

$$\begin{vmatrix} A & B & e \\ C & D & i \\ 0 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ G & 1 & i' \\ 0 & 0 & 1 \end{vmatrix}. \tag{99}$$

Thus, the network to be studied is as shown in Fig. 12 with

$$\begin{vmatrix} E_1 \\ 0 \\ 1 \end{vmatrix} = \begin{vmatrix} A & B & e \\ B & A & i \\ 0 & 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & 0 & 0 \\ G & 1 & i' \\ 0 & 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} E_2 \\ 0 \\ 1 \end{vmatrix}$$

$$= \begin{vmatrix} A + BG & B & Bi' + e \\ B + AG & A & Ai' + i \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} E_2 \\ 0 \\ 1 \end{vmatrix} \tag{100}$$

yielding

$$E_1 = (A + BG)E_2 + Bi' + e$$
$$0 = (B + AG)E_2 + Ai' + i \tag{101}$$

so that

$$E_2 = -\frac{(Ai' + i)}{(B + AG)}$$

$$E_1 = -\frac{(A + BG)}{B + AG} (Ai' + i) + Bi' + e. \tag{102}$$

At the open-circuited terminal at which the noise voltage E_1 is measured, one must have from Nyquist's theorem, (94)

$$| E_1 |^2 = 4Z_{\text{input}} p(\nu) d\nu \tag{103}$$
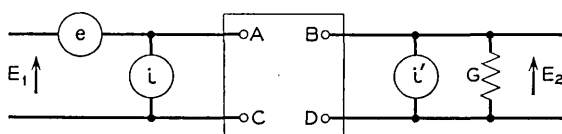
Fig. 12 — Passive matched network with shunt conductance for determination of open-circuit noise voltage.

since Z_{input} is resistive. Its value is

$$Z_{\text{input}} = \frac{A + BG}{B + AG}. \tag{104}$$

Hence, it is required that

$$\left(\frac{A + BG}{B + AG}\right) 4p(\nu)d\nu = \left| -\left(\frac{A + BG}{B + AG}\right)(Ai' + i) + Bi' + e \right|^2. \tag{105}$$

Remember also that $|e|^2 = |i|^2$ and ei^* is real. Since the noise arising from the shunt conductance G is independent of the noise produced by the network, cross terms like ei', i^*i' are zero, since their product represents a time average which must be zero. Solution of (105) yields uniquely the values[15]

$$|i|^2 = |e|^2 = AB$$
$$ei^* = \tfrac{1}{2}(A^2 + B^2 - 1) \tag{106}$$

which are the desired relations measured in units of $4p(\nu)d\nu$. The veracity of (106) can be established by direct substitution into (105), which yields an identity independent of the value of G.

The equivalent values of e and i for an attenuator are given by

$$|e|^2 = |i|^2 = AB = \cosh\theta\sinh\theta = \tfrac{1}{2}\sinh 2\theta$$
$$ei^* = \tfrac{1}{2}(A^2 + B^2 - 1) = \tfrac{1}{2}(\cosh^2\theta + \sinh^2\theta - 1) = \sinh^2\theta. \tag{107}$$

The phase angle between e and i^* is $\cos^{-1}\tanh\theta$. Since the phase angle is always positive, the implication is that the noise sources radiate more power toward the attenuator than away from it. This is reasonable, since the noise power leaving each end of the attenuator should be equal and the noise radiated toward the attenuator by the noise sources is attenuated before emerging from the output end. In the limit of large θ, the phase angle approaches zero.

The equivalent values for an isolator are given by

$$|e|^2 = |i|^2 = \exp -\theta \cosh\frac{\theta}{2} \sinh\frac{\theta}{2} = \tfrac{1}{2} \exp -\theta \sinh\theta$$

$$ei^* = \frac{1}{2}\left[\exp -\theta\left(\sinh^2\frac{\theta}{2} + \cosh^2\frac{\theta}{2}\right) - 1\right] = -|e|^2.$$

(108)

Therefore, the angle between e and i is always π, the implication being that the isolator radiates noise power away from the attenuator, i.e., only in the direction in which it attenuates.

The noise radiated into a matched resistor by any network is obtained by considering the network shown in Fig. 13
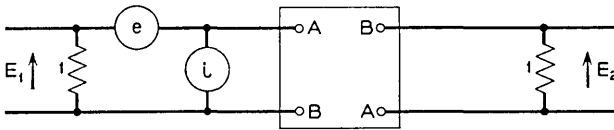
Fig. 13 — Noise radiated into matched resistor.

$$\begin{vmatrix} E_1 \\ 0 \\ 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} A & B & e \\ C & D & i \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} E_2 \\ 0 \\ 1 \end{vmatrix}$$

$$= \begin{vmatrix} A' & B' & e' \\ C' & D' & i' \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} E_2 \\ 0 \\ 1 \end{vmatrix}.$$ (109)

The noise contributed by the matched input and output resistors is neglected since this is additive. The primed quantities result from matrix multiplication. The noise output dP is evaluated by noting that

$$0 = C'E_2 + i' \qquad \text{and} \qquad dP = |E_2|^2 = |i'|^2/|C'|^2.$$

Performing the matrix multiplication yields $C' = A + B + C + D$ and $i' = e + i$. Thus, the output noise power is given by

$$dP = \frac{|e+i|^2\, 4p(\nu)\, d\nu}{|A+B+C+D|^2} = L\,|e+i|^2\, p(\nu)\, d\nu \qquad (110)$$

in which L is the network insertion loss [see (20)]. For a matched resistive network (attenuator)

$$|e+i|^2 = |e|^2 + |i|^2 + 2ei^* = (2AB + A^2 + B^2 - 1)$$

$$= [(A+B)^2 - 1] = L^{-1} - 1$$

which follows from (9) and (21). Thus,

$$dP = (1 - L)p(\nu)d\nu \tag{111}$$

is the noise power in frequency range $d\nu$ emanating from a matched resistive network. This result is well known in network theory. In optics, it is known as one form of Kirchhoff's law.

REFERENCES

1. Slater, J. C., *Microwave Electronics*, New York, D. Van Nostrand, 1950.
2. Gordon, J. P., Zeiger, H. J., and Townes, C. H., Maser — New Type of Microwave Amplifier, Frequency Standard, and Spectrometer, Phys. Rev., **99**, August, 1955, p. 1264.
3. Wagner, W. G., and Birnbaum, G., Theory of Quantum Oscillators in Multimode Cavity, J. Appl. Phys., **32**, July, 1961, p. 1185.
4. Schawlow, A. L., and Townes, C. H., Infrared and Optical Masers, Phys. Rev., **112**, December, 1958, p. 1940, and Townes, C. H., *Advances in Quantum Electronics*, ed. J. R. Singer, Columbia University Press, 1961.
5. Shimoda, K., Theory of Masers for Higher Frequencies, Inst. Phys. and Chem. Res. (Tokyo) — Sci. Papers, **55**, March, 1961, p. 1.
6. Blaquiere, A., Largeur de raie d'un oscillateur Laser, considéré comme le siège d'une reaction en chaine, Acad. des Sciences — CR, **255**, December, 1962, p. 3141.
7. Fleck, J. A., Jr., Linewidth and Conditions for Steady Oscillation in Single and Multiple Element Lasers, J. Appl. Phys., **34**, October, 1963, p. 2997.
8. Davenport, W. B., Jr., and Root, W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, 1958, p. 158.
9. Van der Zeil, A., *Noise*, Prentice-Hall, 1954, p. 283.
10. Guilleman, E. A., *Communication Networks*, Vol. II, John Wiley & Sons, Inc., 1935, Chap. 4.
11. Born, M., and Wolf, E., *Principles of Optics*, Pergamon Press, New York, 1959, p. 54.
12. Becking, A. G. T., Groendijk, H., and Knol, H. S., Noise Factor of Four-Terminal Networks, Philips Res. Repts., Vol. 10, No. 5, October, 1955, p. 349.
13. Van der Zeil, A., op. cit., p. 8.
14. Weissflock, A., Anwendung des Transformatorsatzes über verlustlose Vierpole auf die Hintereinanderscholtung von Vierpolen, Hochfrequenz and Elekakus, **61**, January, 1943, p. 19.
15. Seidel, H., Noise Properties of Four-Pole Networks with Application to Parametric Devices, IRE Trans., CT-8, December, 1961, p. 398.
16. Yariv, A., and Gordon, J. P., The Laser, Proc. IEEE, **51**, January, 1963, p. 4.
17. Born, M., and Wolf, E., op. cit., p. 324.
18. Peirce, B. O., and Foster, R. M., *A Short Table of Integrals*, 4th Ed., Ginn and Company, 1956.
19. Quist, T. M., Rediker, R. H., Keyes, R. J., Krag, W. E., Lax, B., McWhorter, A. L., and Zeigler, H. J., Semiconductor Maser of GaAs, Appl. Phys. Letters, **1**, December, 1962, p. 91.
20. White, A. D., Gordon, E. I., and Rigden, J. D., Output Power of the 6328-Å Gas Maser, Appl. Phys. Letters, **2**, March, 1963, p. 91.
21. Rigden, J. D., White, A. D., and Gordon, E. I., Visible HE-NE Maser and Some Developments, NEREM, TPM 14-4, November, 1962, p. 120.
22. Bridges, W. B., High Optical Gain at $3.5\,\mu$ in Pure Xenon, Appl. Phys. Letters, **3**, August, 1963, p. 45.

The 80 Diperiodic Groups in Three Dimensions

By ELIZABETH A. WOOD

(Manuscript received December 10, 1962)

The low-energy electron diffraction work of L. H. Germer, J. J. Lander, A. U. MacRae, J. Morrison and others is resulting in new information about surface structures. These three-dimensional structures have periodicity only in two dimensions. The 230 triperiodic space groups are not applicable to the solution of these structures. The 17 strictly two-dimensional groups do not admit the existence of a third dimension and may therefore not be appropriate for these structures which are not strictly planar. The useful space groups for these structures are the 80 diperiodic groups in three dimensions.

Nowhere in the literature have these been put into a form convenient for use, as have the other two sets of space groups. This has now been done and the tables are available on request from the Circulation Manager, Bell System Technical Journal, Bell Telephone Laboratories, Incorporated, 463 West Street, New York 14, N. Y. Sample tables are given in this paper.

I. BACKGROUND

Crystals grown under favorable conditions acquire an external shape whose symmetry has long attracted attention. Nineteenth century mineralogists systematically described the symmetry of these shapes in terms of *symmetry operations*. For example, the operation of *rotation* of a cube through 90° around an axis normal to a cube face brings the cube into a position indistinguishable from its original position. An operation that achieves this indistinguishability is called a symmetry operation. In this example the cube will present an identical appearance four times during a rotation of 360° around the axis, which is therefore called "an axis of 4-fold symmetry" or simply "a 4-fold axis." A cube has three 4-fold axes, four 3-fold axes (corner-to-corner) and six 2-fold axes (mid-edge-to-mid-edge) (Figs. 1a and b). Such axes are called *symmetry elements*. The terms "tetrad," "triad" and "diad" are also used for them.

Another type of symmetry element is a mirror plane, across which the operation of *reflection* produces an object indistinguishable from the original. Such a plane through the center of a cube parallel to two opposite faces reflects the left half into the right half and vice versa; that is, the two halves are mirror images of each other. Since there are also diagonal mirror planes in a cube there is a total of nine planes (Figs. 1c and d).

There is also a *center* of symmetry in the center of a cube which relates any feature located a given distance from it in one direction to an indistinguishable feature located the same distance away in the opposite direction. The operation is called *inversion*.
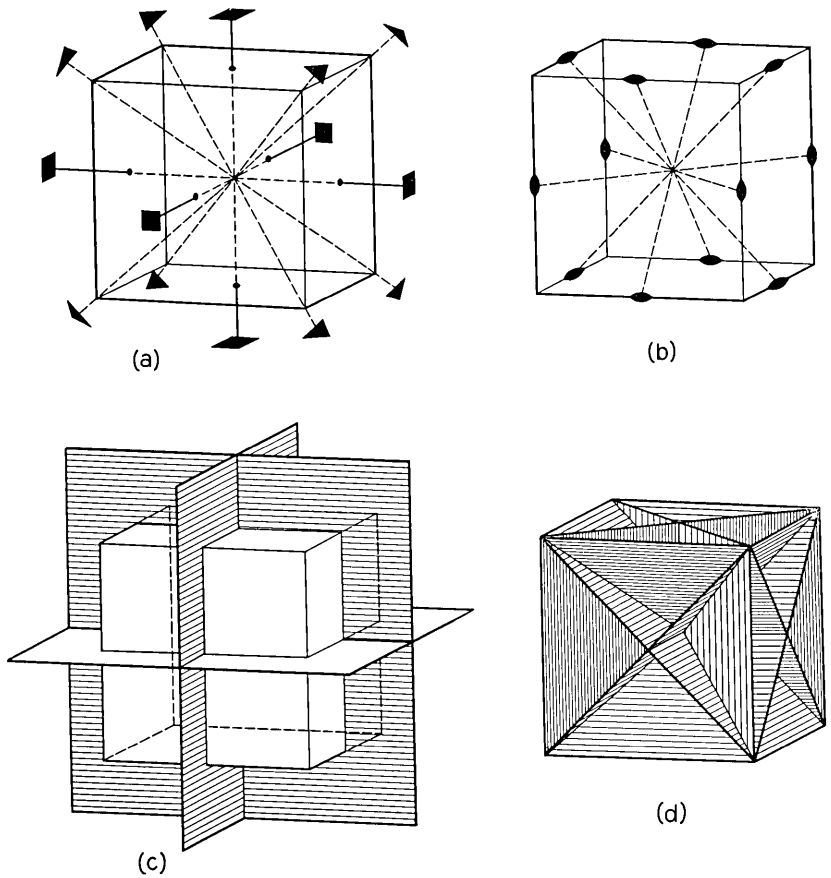
Fig. 1 — The symmetry elements of a cube: (a) the three 4-fold axes and the four 3-fold axes, (b) the six 2-fold axes, (c) three mirror planes parallel to the faces, and (d) six diagonal mirror planes.

An *inversion axis* combines the operation of rotation with that of inversion. The familiar regular tetrahedron which has neither a 4-fold axis nor a center of symmetry has a 4-fold inversion axis because after a rotation of 90° *plus* an inversion through the center point it occupies a position in space indistinguishable from its original position. A center of symmetry is equivalent to a one-fold inversion axis.

Note that during the entire group of "operations" on the cube, one point (the center of the cube) remains unmoved. Another way of saying this is to say that all of the symmetry elements pass through a single point. This group of operations or the symmetry elements which represent them therefore constitute the point group symmetry of the cube. When similar groups of operations are determined for all possible crystals, it is found that there are only 32 possible crystallographic point groups.

The symmetry of shape is the outward expression of the inner orderly atomic arrangement of the crystal. Any property of any piece of the crystal must obey the point group symmetry even though the piece be a ground sphere a few tenths of a millimeter in diameter.

When we consider in detail the *crystal structure* — that is, the positions of the atoms relative to each other — we find that the symmetry elements occur at well-defined positions in space and do not all go through the same point. This is readily illustrated by Fig. 2, the projection of the structure of calcite ($CaCO_3$) onto a plane normal to its 3-fold symmetry axis. Note that the 3-fold axis cannot be randomly placed, normal to the paper, but must pass through the black spots representing the carbon atoms, and further that there is a 3-fold axis through every carbon atom. There are also mirror planes in calcite. We could make a 3-dimensional model of the array of symmetry elements of calcite, and the operation of any symmetry element would shift every other symmetry element to a position indistinguishable from its original position.

Such a self-consistent array of symmetry elements in space is called a *space group*. Since location in space (not orientation alone) is of significance here, two other kinds of symmetry operations become meaningful: operations which combine *translation* with either rotation or reflection. The resulting symmetry elements are called, respectively, *screw axes* and *glide planes*.

As in the case of point groups, the space groups are limited in number. There are 230 possible space groups, i.e., 230 possible self-consistent arrangements in space of all the symmetry elements mentioned above.

Diagrams of these are given in the International Tables for X-ray Crystallography (edited by Henry and Lonsdale, 1952; see References).
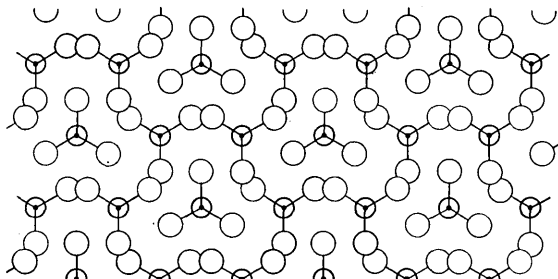
Fig. 2 — The structure of calcite projected onto a plane normal to its 3-fold symmetry axis.

The one appropriate to the structure of calcite is shown on the next page.* It is identified by the symbol $R\bar{3}c$ which states that the unit cell (the repeat unit of the structure) is rhombohedral in shape (R), that it has a 3-fold inversion axis $(\bar{3})$ with a glide plane parallel to it in which the translation is in the c direction.† Of course the 3-fold axis operating on this glide plane generates two more. Additional symmetry elements which are found to exist whenever the stated symmetry operations are performed are also shown in the calcite space group diagram.

The space group of a crystal can in many cases be uniquely determined directly from x-ray diffraction data. Since, in any given space group, the possible atom positions will be related in a well defined manner by the symmetry operations, a knowledge of the space group is a very powerful aid in determining the arrangement of atoms in the crystal, i.e., the crystal structure.

One could repetitiously extend the space-group symbols in the diagram, as we have the calcite structure in Fig. 2, by translation which would be in three dimensions if we were not limited to the printed page. (The translation vectors define the edges of the *unit cell*, the repeat unit of the three-dimensional structure.) The three-dimensional lattice of translation vectors which would represent this operation is called a *space lattice*. There are only 14 such lattices possible.

If we limit our attention strictly to two dimensions we find that, instead of 230 space groups, we have 17 *plane groups* and instead of 14 space lattices we have 5 *nets*. Here the periodicity no longer extends in three dimensions (triperiodicity) but only in two dimensions (diperiodicity).

---

* Fractions on the diagram refer to positions of symmetry elements along the c direction (normal to the paper). The unit is the unit length of c, i.e., the c dimension of the unit cell.

† For a very brief discussion of space group symbols see the *Crystallographic Data* section of the American Institute of Physics Handbook.
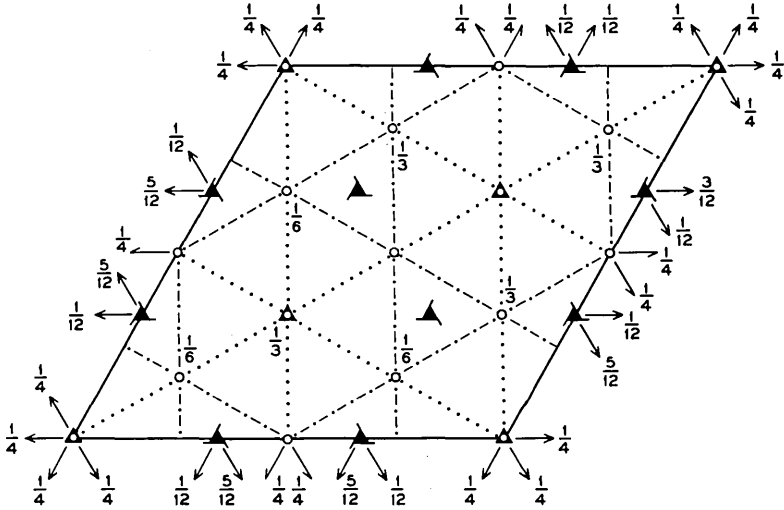
Symbols of Symmetry Planes

Symbol	Symmetry plane	Graphical symbol — Normal to plane of projection	Nature of glide translation
c	Axial glide plane	$\cdots\cdots\cdots$	$c/2$ along z-axis; or $(a+b+c)/2$ along [111] on rhombohedral axes.
n	Diagonal glide plane (net)	$-\cdot-\cdot-\cdot-\cdot-$	$(a+b)/2$ or $(b+c)/2$ or $(c+a)/2$; or $(a+b+c)/2$ (tetragonal and cubic).

Symbols of Symmetry Axes

Symbol	Symmetry axis	Graphical symbol	Nature of right-handed screw translation along the axis	Symbol	Symmetry axis	Graphical symbol	Nature of right-handed screw translation along the axis
1	Rotation monad	None	None	2_1	Screw diad	(normal to paper)	$c/2$
$\bar{1}$	Inversion monad	○	None			(parallel to paper)	$a/2$ or $b/2$ or $(a+b)/2$
2	Rotation diad	(normal to paper)	None			Normal to paper	
				3	Rotation triad	▲	None
		(parallel to paper)		3_1	Screw triads	▲	$c/3$
				3_2		▲	$2c/3$
				$\bar{3}$	Inversion triad	△	None

It is with still a third set of groups, the 80 diperiodic groups in three dimensions, that the present paper is concerned.

II. DISCUSSION OF THE 80 DIPERIODIC GROUPS

The International Tables for X-ray Crystallography (1952) ("ITXRC") give two different sets of space groups: the familiar 230 triperiodic space groups and the 17 two-dimensional space groups in which all operations are confined strictly to two dimensions. In the latter set, any operation which admits the existence of the third dimension, such as a two-fold axis lying in the plane, is forbidden.

The existence of a set of groups which admit such operations, but still refer to arrays that are infinitely periodic in only two dimensions, was recognized by several authors at about the same time (Speiser, 1927; C. Hermann, 1928; Alexander and K. Herrmann, 1928; L. Weber, 1929; Alexander and K. Herrmann, 1929). These and subsequent authors (see references at end of this paper) have used a wide variety of nomenclature, some giving some diagrams. C. Hermann gives point positions, but in many cases chooses a different origin and in some cases a larger cell than that given in ITXRC. This work and others contain errors and omissions and none of the authors has given the groups in the form currently used in the International Tables so that they could be conveniently used for structure determination. This has now been done.

Consideration of the restrictions imposed by the loss of periodicity in the third dimension leads to the exclusion of the following symmetry elements: (i) screw axes normal to the plane of diperiodicity, (ii) glide planes with glide directions out of this plane, and (iii) n-fold axes not normal to this plane, with $n > 2$. Since the upper side of our diperiodic array may be like or unlike the lower side, mirror planes, glide planes, two-fold rotation and screw axes may lie in the plane.

It is possible to choose the 80 diperiodic groups in three dimensions from the pages of the existing International Tables for X-ray Crystallography by using some of the "1st setting" monoclinic groups and some of the "2nd setting" monoclinic groups as well as various orientations of the orthorhombic groups, without deletion or addition of any symmetry operations. In the diperiodic-group case we always have a unique direction in the plane-normal. Placing this direction along each of two nonequivalent directions in a single (orthorhombic) triperiodic space group gives us two nonequivalent diperiodic groups. This, of course, requires the appropriate permutation of point coordinates and indices of forbidden reflections.

Special positions of atoms with a fixed coordinate expressed as a frac-

tion of the unit-cell length in the z-direction, other than zero, are not allowed since fractions of a period are meaningless in this nonperiodic direction.

The five nets (comparable to the 14 space lattices in three dimensions) for these diperiodic groups are the same as those for the 17 two-dimensional groups, namely, oblique ($a \neq b$, $\gamma \neq 90°$), primitive and centered rectangular ($a \neq b$, $\gamma = 90°$), square ($a = b$, $\gamma = 90°$), and hexagonal ($a = b, \gamma = 120°$), where γ is the angle between the a and b axes.

Alexander and Herrmann became interested in these groups because of their work with the smectic state in liquid crystals where only two-dimensional periodicity obtains. Cochran's interest in them grew out of his use of "generalized crystal-structure projections" (Cochran, 1952, b) and Holser's (1958, b) out of his investigation of the structure at the boundary between two parts of a twinned crystal (1958, a).

The interest of the writer in making these groups available in convenient form stems from cooperation with those members of Bell Laboratories who have been investigating surface structures by means of low-energy electron diffraction, in particular, L. H. Germer, J. J. Lander, A. U. MacRae and J. Morrison.* These structures are infinitely periodic in two dimensions but lack periodicity in the third dimension (normal to the surface).

Which set of diperiodic groups is appropriate for surface structures? Certainly the structures are not strictly planar: the atoms of the surface structure in many cases do not all lie in the same plane. But would an atom above some plane (parallel to the surface) be symmetrically related to an atom on the other side of the plane? Strictly speaking the atoms could not be symmetrically equivalent since one is closer to the substrate than the other and is therefore in a different force field. From this point of view one would say that only the seventeen strictly two-dimensional space groups would be useful. However, it is frequently so, in triperiodic crystallography, that the symmetry of a crystal structure closely approximates a symmetry that is higher than its true symmetry and that the use of the higher-symmetry space group is of great help in determining the structure. From this point of view one would say that the 80 diperiodic groups in three dimensions are likely to be useful in the solution of diperiodic surface structures. Their application to this field was suggested to the writer by A. L. Patterson.

There follow (i) a summary table, Table I; (ii) a diagram of net types, Fig. 3; (iii) an explanation of terms and symbols used in the

---

* For a survey of some of this work, see Low-energy Electron Diffraction, by A. U. MacRae, Science, **139**, 1963, pp. 379–388.

TABLE I—SUMMARY TABLE OF THE 80 DIPERIODIC GROUPS IN THREE DIMENSIONS

Net	Diperiodic Group (DG) Number	Full Hermann-Mauguin Symbols	Triperiodic-Group Schoenflies Symbol, ITXRC Number and Orientation, if other than that given in ITXRC		Symbol Proposed by A. Niggli	Weber Number*
Oblique	1	$P1$	C_1^1 - 1		$1P1$	1
	2	$P\bar1$	C_i^1 - 2		$1P\bar1$	2
	3	$P211$	C_2^1 - 3	1st setting	$1P2$	8
	4	$Pm11$	C_s^1 - 6	1st setting	$mP1$	3
	5	$Pb11$	C_s^2 - 7	1st setting	$aP1$	4
	6	$P2/m\,11$	C_{2h}^1 - 10	1st setting	$mP2$	12
	7	$P2/b\,11$	C_{2h}^4 - 13	1st setting	$aP2$	13
Rectangular	8	$P112$	C_2^1 - 3	2nd setting	$1P12$	9
	9	$P112_1$	C_2^2 - 4	2nd setting	$1P12_1$	10
	10	$C112$	C_2^3 - 5	2nd setting	$1C12$	11
	11	$P11m$	C_s^1 - 6	2nd setting	$1P1m$	5
	12	$P11a$	C_s^2 - 7	2nd setting $\bar cba$	$1P1g$	6
	13	$C11m$	C_s^3 - 8	2nd setting	$1C1m$	7
	14	$P11\,2/m$	C_{2h}^1 - 10	2nd setting	$1P12/m$	14
	15	$P11\,2_1/m$	C_{2h}^2 - 11	2nd setting	$1P12_1/m$	15
	16	$C11\,2/m$	C_{2h}^3 - 12	2nd setting	$1C12/m$	16
	17	$P11\,2/a$	C_{2h}^4 - 13	2nd setting $\bar cba$	$1P12/g$	18
	18	$P11\,2_1/a$	C_{2h}^5 - 14	2nd setting $\bar cba$	$1P12_1/g$	17
	19	$P222$	D_2^1 - 16		$1P222$	33
	20	$P222_1$	D_2^2 - 17	bca	$1P222_1$	34
	21	$P22_12_1$	D_2^3 - 18		$1P22_12_1$	35
	22	$C222$	D_2^6 - 21		$1C222$	36
	23	$P2mm$	C_{2v}^1 - 25		$1P2mm$	19
	24	$Pmm2$	C_{2v}^1 - 25	bca	$mP12m$	23
	25	$Pm2_1a$	C_{2v}^2 - 26	$\bar cba$	$mP12_1g$	24
	26	$Pbm2_1$	C_{2v}^2 - 26	$a\bar cb$	$aP12_1m$	25
	27	$Pbb2$	C_{2v}^3 - 27	$a\bar cb$	$aP12g$	26
	28	$P2ma$	C_{2v}^4 - 28		$1P2mg$	20
	29	$Pam2$	C_{2v}^4 - 28	$a\bar cb$	$bP12m$	27
	30	$Pab2_1$	C_{2v}^5 - 29	$a\bar cb$	$bP12_1g$	28
	31	$Pnb2$	C_{2v}^6 - 30	bca	$nP12g$	29
	32	$Pnm2_1$	C_{2v}^7 - 31	$a\bar cb$	$nP12_1m$	30
	33	$P2ba$	C_{2v}^8 - 32		$1P2gg$	21
	34	$C2mm$	C_{2v}^{11} - 35		$1C2mm$	22
	35	$Cmm2$	C_{2v}^{14} - 38	bca	$mC12m$	31
	36	$Cam2$	C_{2v}^{15} - 39	bca	$aC12m$	32
	37	$P2/m\,2/m\,2/m$	D_{2h}^1 - 47		$mP2mm$	37
	38	$P2/a\,2/m\,2/a$	D_{2h}^3 - 49	cab	$aP2mg$	38
	39	$P2/n\,2/b\,2/a$	D_{2h}^4 - 50		$nP2gg$	39
	40	$P2/m\,2_1/m\,2/a$	D_{2h}^5 - 51	$a\bar cb$	$mP2mg$	40
	41	$P2/a\,2_1/m\,2/m$	D_{2h}^5 - 51		$aP2mm$	41
	42	$P2/n\,2/m\,2_1/a$	D_{2h}^7 - 53	$a\bar cb$	$nP2mg$	42
	43	$P2/a\,2/b\,2_1/a$	D_{2h}^8 - 54	cab	$aP2gg$	43
	44	$P2/m\,2_1/b\,2_1/a$	D_{2h}^9 - 55		$mP2gg$	44
	45	$P2/a\,2_1/b\,2_1/m$	D_{2h}^{11} - 57	bca	$aP2gm$	45
	46	$P2/n\,2_1/m\,2_1/m$	D_{2h}^{13} - 59		$nP2mm$	46
	47	$C2/m\,2/m\,2/m$	D_{2h}^{19} - 65		$mC2mm$	47
	48	$C2/a\,2/m\,2/m$	D_{2h}^{21} - 67		$aC2mm$	48
Square	49	$P4$	C_4^1 - 75		$1P4$	58
	50	$P\bar4$	S_4^1 - 81		$1P\bar4$	·57
	51	$P4/m$	C_{4h}^1 - 83		$mP4$	61
	52	$P4/n$	C_{4h}^3 - 85		$nP4$	62
	53	$P422$	D_4^1 - 89		$1P422$	67
	54	$P42_12$	D_4^2 - 90		$1P42_12$	68
	55	$P4mm$	C_{4v}^1 - 99		$1P4mm$	59

TABLE I—CONTINUED

Net	Diperiodic Group (DG) Number	Full Hermann-Mauguin Symbols	Triperiodic-Group Schoenflies Symbol ITXRC Number and Orientation, if other than that given in ITXRC	Symbol Proposed by A. Niggli	Weber Number*
Square (cont.)	56	$P4bm$	C_{4v}^2 - 100	$1P4gm$	60
	57	$P\bar{4}2m$	D_{2d}^1 - 111	$1P\bar{4}2m$	63
	58	$P\bar{4}2_1m$	D_{2d}^3 - 113	$1P\bar{4}2_1m$	64
	59	$P\bar{4}m2$	D_{2d}^5 - 115	$1P\bar{4}m2$	65
	60	$P\bar{4}b2$	D_{2d}^7 - 117	$1P\bar{4}g2$	66
	61	$P4/m\,2/m\,2/m$	D_{4h}^1 - 123	$mP4mm$	69
	62	$P4/n\,2/b\,2/m$	D_{4h}^3 - 125	$nP4gm$	70
	63	$P4/m\,2_1/b\,2/m$	D_{4h}^5 - 127	$mP4gm$	71
	64	$P4/n\,2_1/m\,2/m$	D_{4h}^7 - 129	$nP4mm$	72
Hexagonal	65	$P3$	C_3^1 - 143	$1P3$	49
	66	$P\bar{3}$	C_{3i}^1 - 147	$1P\bar{3}$	50
	67	$P312$	D_3^1 - 149	$1P312$	54
	68	$P321$	D_3^2 - 150	$1P321$	53
	69	$P3m1$	C_{3v}^1 - 156	$1P3m1$	51
	70	$P31m$	C_{3v}^2 - 157	$1P31m$	52
	71	$P\bar{3}1\,2/m$	D_{3d}^1 - 162	$1P\bar{3}1m$	55
	72	$P\bar{3}\,2/m\,1$	D_{3d}^3 - 164	$1P\bar{3}m1$	56
	73	$P6$	C_6^1 - 168	$1P6$	76
	74	$P\bar{6}$	C_{3h}^1 - 174	$mP3$	73
	75	$P6/m$	C_{6h}^1 - 175	$mP6$	78
	76	$P622$	D_6^1 - 177	$1P622$	79
	77	$P6mm$	C_{6v}^1 - 183	$1P6mm$	77
	78	$P\bar{6}m2$	D_{3h}^1 - 187	$mP3m2$	74
	79	$P\bar{6}2m$	D_{3h}^3 - 189	$mP32m$	75
	80	$P6/m\,2/m\,2/m$	D_{6h}^1 - 191	$mP6mm$	80

* Useful for cross-comparison of this list with those of Weber (1929), C. Hermann (1928) and Alexander and Herrmann (1929) since the equivalence among these three is given in the last reference.

tables, Table II; (iv) samples of the systematic ITXRC "tables" for the 80 diperiodic groups in three dimensions, adapted from the three-dimensional space groups by making the appropriate modifications; and (v) an annotated list of references.

The full set of 80 diperiodic groups in three dimensions has been bound separately and is available from the Circulation Manager, Bell System Technical Journal, Bell Telephone Laboratories, Incorporated, 463 West Street, New York 14, N. Y. It is anticipated that these groups will be included in a later volume of the International Tables.

III. TABLES OF THE 80 DIPERIODIC GROUPS IN THREE DIMENSIONS

In the sample tables, the usage and notation of the International Tables for X-ray Crystallography for the three-dimensional space groups have been followed as closely as possible. Chosen directly from the
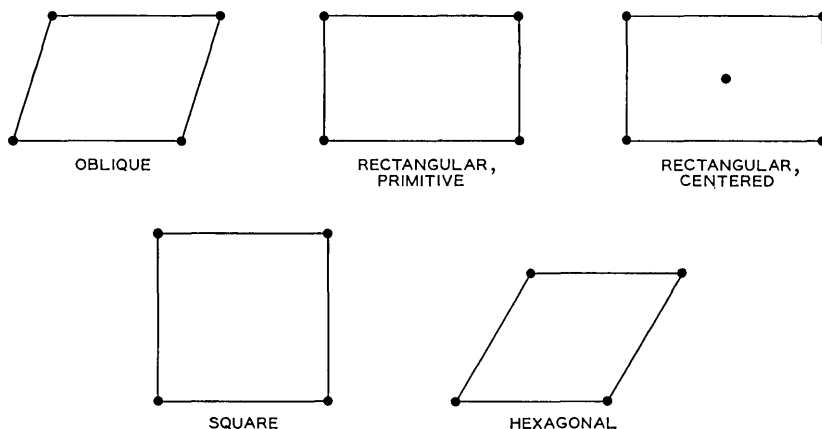
Fig. 3 — The five nets.

ITXRC Tables for the 230 space groups, these tables carry the same atom-position lettering. Letters of forbidden positions will therefore be missing.
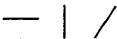
In the oblique and rectangular system, the ITXRC convention of listing the symmetry symbols in the order a, b, c has not been retained. Holser (1958,b) chose to permute these so that the first symbol referred to the c axis. The justification for this is that in the plane groups the c axis is unique and therefore should be put first as in, for example, the tetragonal system (e.g., $4mm$).

The possibility of confusion with the 230 three-dimensional groups will probably be avoided in all cases by the context. However, to aid in the distinction, the plane groups have been numbered, DG1, DG2, etc. The same letters could be used to distinguish $DGPmm2$, for example, from the three-dimensional C_{2v}^{1} - $Pmm2$, but since, in all cases, the two groups do in fact comprise the same symmetry operations, such a distinction may be undesirable.

The order of the DG list is that of the ITXRC which, in turn, is the Schoenflies order.

After this paper was in galley form a communication was received from A. Niggli to whom a manuscript copy had been sent. Niggli favors placing before the lattice symbol (P or C) that symbol referring to the glide plane or mirror plane which lies in the plane of diperiodicity and therefore occurs only once. This occurs in 37 of the 80 groups. This would be another way of distinguishing these groups from the triperiodic groups. The symbol proposed by Niggli is also listed in Table I.

TABLE II — SYMBOLS USED IN THE 80 DG TABLES

Symmetry Elements	Diperiodic Group Symbol	Symbol in the Symmetry Diagram	
		Normal to the Paper	Parallel to the Paper
mirror	m	— | /	⌐
glide plane*	a	⋮	⌐↓
	b	- - -	↼⌐
	n	not allowed	⤡⌐
2-fold rotation axis	2	⧫	← ↑
3-, 4- and 6-fold rotation axes	3, 4, 6	▲ ◆ ⬢	not allowed
2-fold screw axis	2_1	not allowed	— ↑
center of symmetry	$\bar{1}$	○	○
3-, 4- and 6-fold inversion axes†	$\bar{3}, \bar{4}, \bar{6}$	△ ◈ ⬡	not allowed
Center, on 2-fold axis		⧬	↼○— ⦵

* This operation combines reflection with translation of $\frac{1}{2}$ the length of the cell in the direction indicated by the letter. The diagonal glide, n, combines reflection with translation of $\frac{1}{2}$ of the length of the cell in both the a and b directions.
† Combined rotation through $360°/n$ (for $\bar{n}$) and inversion. Not equivalent to the two operations performed separately.

IV. EXPLANATION OF TERMS AND SYMBOLS USED ON THE 80 DG SHEETS
(These are the same as those used in the ITXRC)

1. *Top of sheet, left to right:* Net-type, full Hermann-Mauguin diperiodic group symbol, diperiodic-group (DG) number. The Hermann-Mauguin symbol begins with a letter which indicates whether the net is primitive or centered and is followed by symbols for symmetry elements that relate to the c, a and b axis, in turn. The c axis is normal to the paper in the diagrams, the a axis is directed toward the bottom of the page, and the b axis is directed toward the right. In DG 46 ($P\ 2/n\ 2_1/m\ 2_1/m$), for example, the lattice is primitive, there is a two-fold axis parallel to c with a diagonal-glide plane normal to c, a two-fold screw axis parallel to a with a mirror plane normal to a, and a two-fold screw axis parallel to b with a mirror plane normal to b. In DG 16 ($C\ 11\ 2/m$) we have a centered net with a two-fold axis parallel to b and a mirror plane normal to b.

2. *Diagrams:* On the right, the distribution of the symmetry elements in the unit mesh. On the left, the distribution in the unit mesh of the points in the "general position" (x, y, z and points symmetrically

equivalent to it). Here, the value of x is taken, arbitrarily, to be a very small distance and y, a slightly larger distance, except in the oblique groups where the reverse choice has been made. The sign of z is indicated beside the "point" (small circle). In both diagrams, the $+x$ direction (a) is down the page, $+y$ (b) toward the right. A comma within the circle indicates that that point is of opposite handedness to the points without commas, as when derived from these by mirror plane or inversion operation. Where two points are related by a mirror lying in the plane of the paper, half of the circular symbol is marked with a comma, half left blank.

Below the diagrams are the lists of all possible points in this diperiodic group and equivalent point positions.

First column: Number of positions that are symmetrically equivalent, given the first position in the series.

Second column: Arbitrary identifying letter, conventionally the same as that first used by Wyckoff for this position.

Third column: The symmetry of each point in the group (if each point lies on a two-fold axis, this will be "2"; if each point lies in a mirror plane this will be "m"; etc.). This will always be "1" for the "general position" which, by definition, is the position of a point not lying on any symmetry element.

Fourth column: Coordinates of equivalent positions. Note that not every group has "special positions." Special positions occur when a particular value of x, y, or z results in a reduction of the number of equivalent positions due to symmetry.
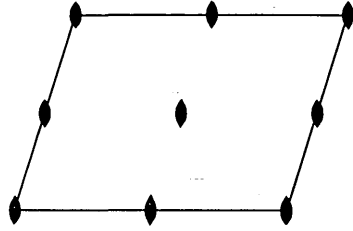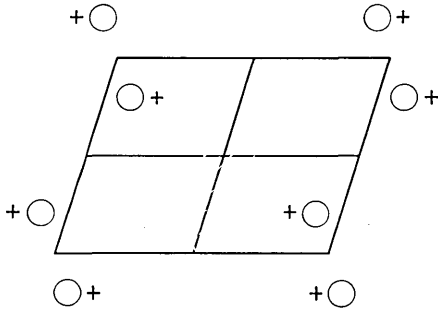
Fifth column: Conditions on hk which must be satisfied, for x-ray reflection to occur when the point positions in column 4 are occupied.

Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions	Conditions limiting possible reflections
2	e	1	$x,y,z;$ $\bar{x},\bar{y},z.$	General: $\left.\begin{array}{l} hk: \\ h0: \\ 0k: \end{array}\right\}$ No conditions
1	d	2	$\frac{1}{2},\frac{1}{2},z.$	Special: No conditions
1	c	2	$\frac{1}{2},0,z.$	
1	b	2	$0,\frac{1}{2},z.$	
1	a	2	$0,0,z.$	
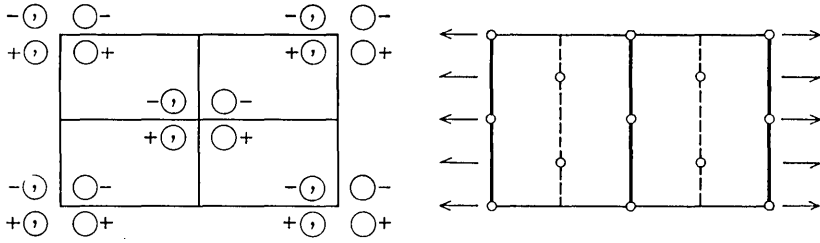
Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions				Conditions limiting possible reflections

$(0,0,0; \frac{1}{2},\frac{1}{2},0)+$

General:

8	*j*	1	$x,y,z;$	$x,\bar{y},z;$	$\bar{x},y,\bar{z};$	$\bar{x},\bar{y},\bar{z}.$	$hk: h + k = 2n$
							$h0: (h = 2n)$
							$0k: (k = 2n)$

Special: as above, plus

4	*i*	*m*	$x,0,z;$	$\bar{x},0,\bar{z}.$			No extra conditions
4	*g*	2	$0,y,0;$	$0,\bar{y},0.$			

4	*e*	$\bar{1}$	$\frac{1}{4},\frac{1}{4},0;$	$\frac{1}{4},\frac{3}{4},0.$			$hk: h = 2n; (k = 2n)$

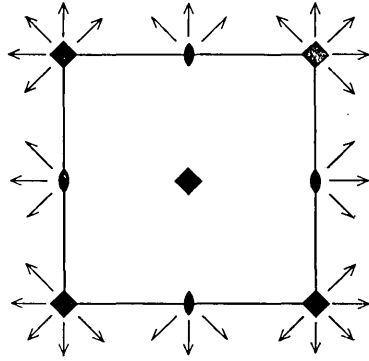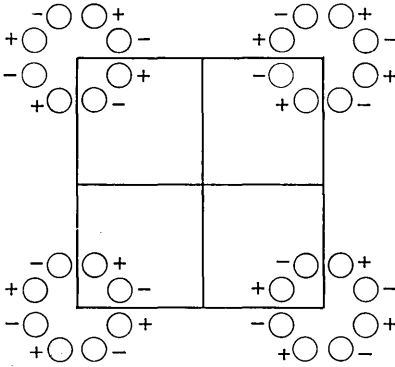| 2	*b*	2/*m*	$0,\frac{1}{2},0.$				No extra conditions
2	*a*	2/*m*	$0,0,0.$				

Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions				Conditions limiting possible reflections
8	g	1	$x,y,z;$ $\bar{x},\bar{y},z;$ $\tfrac{1}{2}-x,\tfrac{1}{2}-y,\bar{z};$ $\tfrac{1}{2}-x,\tfrac{1}{2}+y,\bar{z};$				General: $hk:\ h+k=2n$ $h0:\ (h=2n)$ $0k:\ (k=2n)$
			$\bar{x},y,z;$ $x,\bar{y},z;$ $\tfrac{1}{2}+x,\tfrac{1}{2}+y,\bar{z};$ $\tfrac{1}{2}+x,\tfrac{1}{2}-y,\bar{z}.$				Special: as above, plus
4	f	m	$x,0,z;$ $\bar{x},0,z;$ $\tfrac{1}{2}-x,\tfrac{1}{2},\bar{z};$ $\tfrac{1}{2}+x,\tfrac{1}{2},\bar{z}.$				} no extra conditions
4	e	m	$0,y,z;$ $0,\bar{y},z;$ $\tfrac{1}{2},\tfrac{1}{2}-y,\bar{z};$ $\tfrac{1}{2},\tfrac{1}{2}+y,\bar{z}.$				
4	c		$\tfrac{1}{4},\tfrac{1}{4},0;$ $\tfrac{3}{4},\tfrac{3}{4},0;$ $\tfrac{1}{4},\tfrac{3}{4},0;$ $\tfrac{3}{4},\tfrac{1}{4},0.$				$hkl:\ h=2n;\ k=2n$
2	b	mm	$0,\tfrac{1}{2},z;$ $\tfrac{1}{2},0,\bar{z}.$				} no extra conditions
2	a	mm	$0,0,z;$ $\tfrac{1}{2},\tfrac{1}{2},\bar{z}.$				

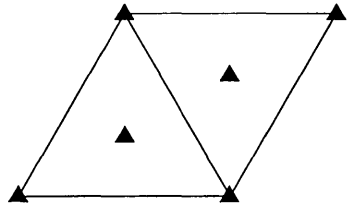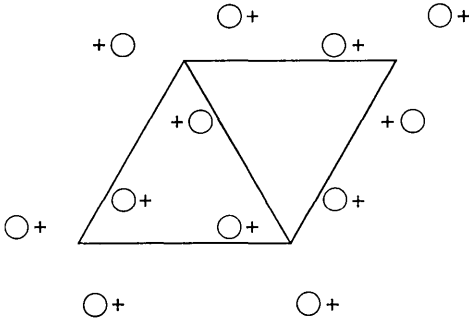| Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions				Conditions limiting possible reflections
8	*p*	1	$x,y,z;$ $\bar{y},\bar{x},\bar{z};$	$\bar{x},\bar{y},z;$ $y,x,\bar{z};$	$\bar{x},y,\bar{z};$ $y,\bar{x},z;$	$x,\bar{y},\bar{z};$ $\bar{y},x,z.$	**General:** No conditions
4	*o*	2	$x,\frac{1}{2},0;$	$\bar{x},\frac{1}{2},0;$	$\frac{1}{2},x,0;$	$\frac{1}{2},\bar{x},0.$	**Special:**
4	*l*	2	$x,0,0;$	$\bar{x},0,0;$	$0,x,0;$	$0,\bar{x},0.$	$\Big\}$No conditions
4	*j*	2	$x,x,0;$	$\bar{x},\bar{x},0;$	$\bar{x},x,0;$	$x,\bar{x},0.$	
4	*i*	2	$0,\frac{1}{2},z;$	$0,\frac{1}{2},\bar{z};$	$\frac{1}{2},0,z;$	$\frac{1}{2},0,\bar{z}.$	$hk:\ h+k=2n$
2	*h*	4	$\frac{1}{2},\frac{1}{2},z;$	$\frac{1}{2},\frac{1}{2},\bar{z}.$			$\Big\}$No conditions
2	*g*	4	$0,0,z;$	$0,0,\bar{z}.$			
2	*e*	222	$\frac{1}{2},0,0;$	$0,\frac{1}{2},0.$			$hk:\ h+k=2n$
1	*c*	42	$\frac{1}{2},\frac{1}{2},0.$				$\Big\}$No conditions
1	*a*	42	$0,0,0.$				

Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions	Conditions limiting possible reflections
				General: No conditions
3	d	1	$x,y,z;$ $\bar{y},x-y,z;$ $y-x,\bar{x},z.$	
1	c	3	$\frac{2}{3},\frac{1}{3},z.$	
1	b	3	$\frac{1}{3},\frac{2}{3},z.$	Special:
1	a	3	$0,0,z.$	No conditions
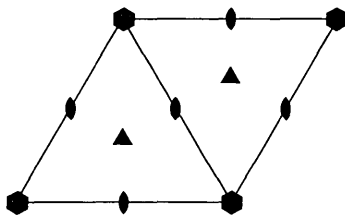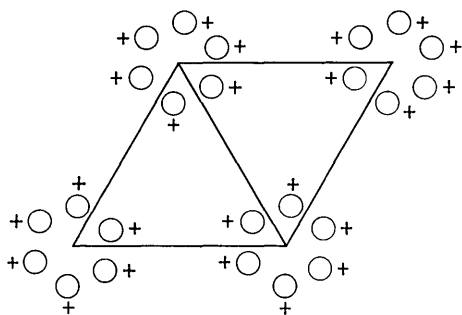
557

Number of positions, Wyckoff notation, and point symmetry			Co-ordinates of equivalent positions	Conditions limiting possible reflections
				General:
6	d	1	$x,y,z;\quad \bar{y},x-y,z;\quad y-x,\bar{x},z;$	No conditions
			$\bar{x},\bar{y},z;\quad y,y-x,z;\quad x-y,x,z.$	
3	c	2	$\frac{1}{2},0,z;\quad 0,\frac{1}{2},z;\quad \frac{1}{2},\frac{1}{2},z.$	
2	b	3	$\frac{1}{3},\frac{2}{3},z;\quad \frac{2}{3},\frac{1}{3},z.$	Special:
1	a	6	$0,0,z.$	No conditions

558

REFERENCES

The 80 diperiodic groups in three dimensions (in chronological order)

A. Speiser (1927), *Die Theorie der Gruppen von endlicher Ordnung*, 2nd edition, Berlin, Springer.

C. Hermann (1928), Zur systematischen Struktur theorie, III Ketten- und Netzgruppen, Z. f. Krist., **69,** p. 250. Lists point positions.

Ernst Alexander and Karl Herrmann (1928), Zur Theorie der flüssigen Kristalle, Z. f. Krist., **69,** p. 285.

C. Hermann (1928), Zur systematischen Struktur theorie, IV Untergruppen, Z. f. Krist., **69,** p. 533. Subgroups.

L. Weber (1929), Die Symmetrie homogener ebener Punktsysteme, Z. f. Krist., **70,** p. 309. Describes 80 plane groups.

Ernst Alexander and Karl Herrmann (1929), Die 80 zweidimensionalen Raumgruppen, Z. f. Krist., **70,** p. 328. Ibid, p. 460: Berichtung zu unserer Arbeit "Die 80 zweidimensionalen Raumgruppen." Comparative list of Weber, C. Hermann and A. and H. designations. Refers to H. Mark, Die Verwendung der Roentgenstrahlen for most point positions. Caution: Schoenflies symbols in this paper do not correspond with the Schoenflies symbols in ITXRC for many groups.

W. Cochran (1952, a), The Symmetry of Real Periodic Two-Dimensional Functions, Acta Cryst., **5,** p. 630. (Lists "46 reversal groups in two dimensions" using a *pgm* notation. Notes that Prof. Lonsdale has pointed out to him that these groups could be described by means of the symbols for the usual 230 space groups.)

W. Cochran and H. B. Dyer (1952, b), Some Practical Applications of Generalized Crystal-Structure Projections, Acta Cryst., **5,** p. 634.

N. F. M. Henry and Kathleen Lonsdale (1952), . . . Plane Groups in Three Dimensions, p. 56. International Tables for X-ray Crystallography, Vol. I, Kynoch Press. Calls attention to the existence of the 80 diperiodic groups in three dimensions.

K. Dornberger-Schiff (1956), On Order-Disorder Structures (OD-Structures), with an Appendix, "Proposal for International Symbols for the 80 Plane Groups in Three Dimensions," Acta Cryst., **9,** p. 593. (Compare Dornberger-Schiff, 1957, below.)

N. V. Belov and T. N. Tarkhova (1956), Groups with Color Symmetry, Kristallografiya, **1,** p. 4. Lists 46 groups, using color-group nomenclature, colored illustrations.

K. Dornberger-Schiff (1957), Zur OD-Struktur (Order-Disorder Structure) des Purpurogallin, Acta Cryst., **10,** p. 271. Gives table like the Tables 4.3 in the International Tables for the "Space Groupoid $P\{2_{\pm\frac{1}{2}}/b\}(\{2_1/n_{1,\pm\frac{1}{2}}\})\langle 2_1\rangle/a$."

W. T. Holser (1958, a), The Relation of Structure to Symmetry in Twinning, Z. Krist., **110,** p. 249.

W. T. Holser (1958, b), Point Groups and Plane Groups in a Two-Sided Plane and Their Subgroups, Z. f. Krist., **110,** p. 266. Lists ITXRC-type symbols, subgroups. Caution: Holser's first symbol always refers to the *c* axis.

N. V. Belov (1959), On the Nomenclature of the 80 Plane Groups in Three Dimensions, Translation in Soviet Physics: Crystallography **4,** p. 730. Original: Kristallografiya, **4,** p. 775. List of "Rational," Cochran, and Fedorov symbols. Groups arranged according to whether they are "two-color," "one-color" or "grey."

Contributors to This Issue

RICHARD R. ANDERSON, B.S.M.E., 1949, Northwestern University; M.S.E.E., 1960, Stevens Institute of Technology; Bell Telephone Laboratories, 1949—. Mr. Anderson first engaged in research on electronic switching systems for telephone central offices. In 1956 he joined the data transmission exploratory development department and made several prototype magnetic-tape transports for storing digital data. He has recently conducted theoretical studies of data transmission systems by computer simulation. Member, A.A.A.S., Sigma Xi, and Tau Beta Pi.

SIDNEY DARLINGTON, B.S., 1928, Harvard College; B.S. in E.E., 1929, Massachusetts Institute of Technology; Ph.D., 1940, Columbia University; Bell Telephone Laboratories, 1929—. He has been engaged in research in applied mathematics with emphasis on network theory and military and space electronics. He holds more than 20 patents in these fields. Fellow, IEEE; member, AIAA.

ROBERT W. DEGRASSE, B.S., 1951, California Institute of Technology; M.S., 1954, and Ph.D., 1958, Stanford University; Bell Telephone Laboratories, 1957–1960; Microwave Electronics Corp., 1960—. Mr. DeGrasse's work at Bell Laboratories was in research and development of solid state masers. He took part in the development of the ruby maser used in the Bell Laboratories receiving system for the Project Echo satellite communication experiments. Member, IEEE and Sigma Xi.

F.E. FROEHLICH, B.S., 1950, M.S., 1952, Ph.D., 1955, Syracuse University; Bell Telephone Laboratories, 1954—. Upon joining Bell Laboratories, Mr. Froehlich worked with magnetic core memory and magnetic core switching devices and circuits. In 1956 he became engaged in exploratory development of DATA-PHONE systems and subsequently had charge of groups conducting research and development in in the field of digital communications. He is now head of the high-speed data terminals department and is concerned with data transmission over wideband channels, digital coding and error control systems, and maintenance equipment for data services. Senior member, IEEE; member of Data and Telegraph Communication Committee and Communi-

cation Theory Committee; Chairman of the Monmouth County, N. J. subsection of the PTGCS. Member, American Physical Society, Sigma Xi, Phi Beta Kappa, Sigma Pi Sigma and Pi Mu Epsilon.

EUGENE I. GORDON, B.S., 1952, City College of New York; Ph.D., 1957, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1957—. A member of the electron device laboratory, he is engaged in research in optical masers and optical modulation techniques. Member, American Physical Society, Phi Beta Kappa, Sigma Xi and IEEE.

STEPHEN E. HARRIS, B.E.E., 1959, Rensselaer Polytechnic Institute; M.S.E.E., 1961, Stanford University; Bell Telephone Laboratories, 1959—. Mr. Harris was engaged in work on microwave noise generation and later in work on the development of traveling-wave masers. Since September, 1960, he has been on leave of absence from Bell Laboratories to pursue doctoral studies at Stanford University. He is also presently serving as an acting professor of electrical engineering at Stanford. Member, Tau Beta Pi, Sigma Xi, Eta Kappa Nu, American Physical Society, Optical Society of America and IEEE.

JESSIE MacWILLIAMS, B.A., 1939, M.A., 1941, Cambridge (England); Ph.D., 1962, Harvard; Bell Telephone Laboratories 1956—. Mrs. MacWilliams has been concerned with writing computer programs for the analysis and synthesis of transmission networks. She is now engaged in data systems studies, particularly the study of algorithms for decoding systematic error-correcting codes. Member, Mathematical Association of America and American Mathematical Society.

E. O. SCHULZ-DuBois, Dipl. phys., 1950, and Dr. Phil. nat., 1954, Johann Wolfgang Goethe University (Germany); Purdue University, 1954–1955; Raytheon Manufacturing Co., 1956–1957; Bell Telephone Laboratories, 1957—. At Purdue Mr. Schulz-DuBois was engaged in paramagnetic resonance studies of irradiated semiconductors. At Raytheon he was concerned with the development of ferrite materials and devices. After joining Bell Laboratories his work was with paramagnetic materials, slow-wave structures, and ferrimagnetic isolators for application to solid state maser devices. More recently he was responsible for a group engaged in advanced development of traveling-wave masers and in related exploratory studies. Since September 1963 he has been on sabbatical leave as visiting professor at Technische Hochschule, Karlsruhe (Germany).

ERLING D. SUNDE, Dipl. Ing., 1926, Technische Hochschule, Darmstadt, Germany; American Telephone and Telegraph Co., 1927–1934; Bell Telephone Laboratories, 1934—. He has made theoretical and experimental studies of inductive interference from railway and power systems, lightning protection of the telephone plant, and fundamental transmission studies in connection with the use of pulse modulation systems. He is the author of *Earth Conduction Effects in Transmission Systems*, a Bell Laboratories Series book. Fellow, IEEE; member, A.A.A.S. and American Mathematical Society.

ELIZABETH A. WOOD, B.A., 1933, Barnard College; M.A. 1934 and Ph.D., 1939, Bryn Mawr College; D.Sc., 1963, Wheaton College; Bell Telephone Laboratories, 1943—. Mrs. Wood's first work at Bell Laboratories had to do with the development of techniques for producing quartz oscillator plates. Since then she has been using X-ray diffraction and optical methods for studying a wide variety of crystals, most of them of interest because of being ferroelectric. She is the author of *Crystal Orientation Manual* (Columbia University Press, 1963) and *Crystals and Light* (D. Van Nostrand, 1963). Member and past President (1957), American Crystallographic Association; Fellow, American Physical Society and Mineralogical Society of America.