THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XLIII — JANUARY 1964 — NUMBER 1, PART 2

Demodulation of Wideband, Low-Power FM Signals*

By SIDNEY DARLINGTON

(Manuscript received October 3, 1963)

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-

* Parts of the material of this article were discussed by the author in lectures at the University of California at Berkeley during May, 1963.

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 determined. The Appendix outlines very briefly some analytical and circuit details, which may be needed for an appreciation of some of the conclusions.

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 discrete frequencies. The distinction has important repercussions throughout 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 Nyquist 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 modulator has been called PAM-FM.[3] It is a known means of adapting time

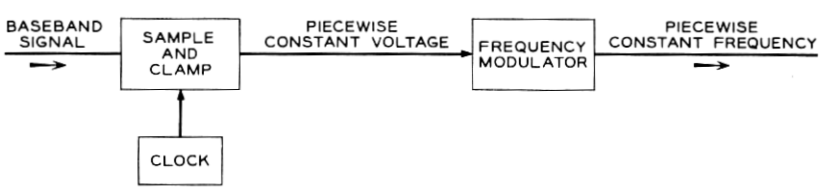

Fig. 1 — Block diagram of a PAM-FM modulator.

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.

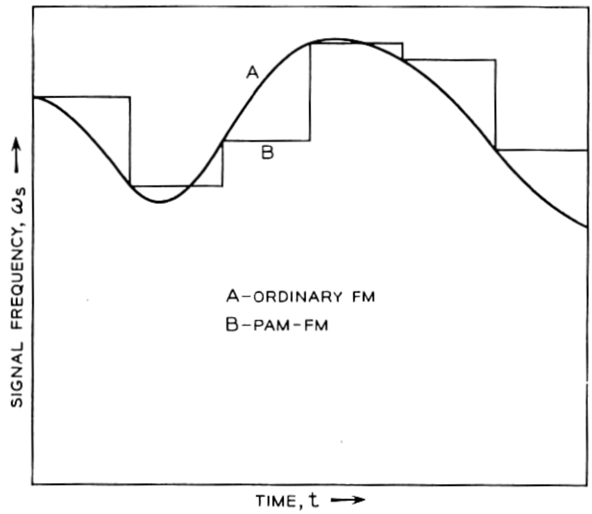

Fig. 2 — Instantaneous signal frequencies.

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.

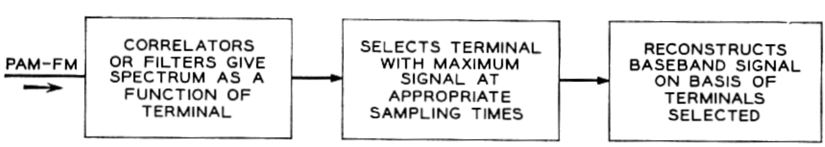

Fig. 3 — PAM-FM demodulation by correlators or filters.

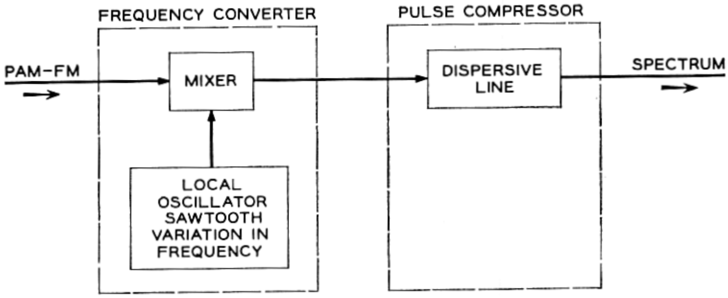

Fig. 4 — Spectrum generation.

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

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

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

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

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.

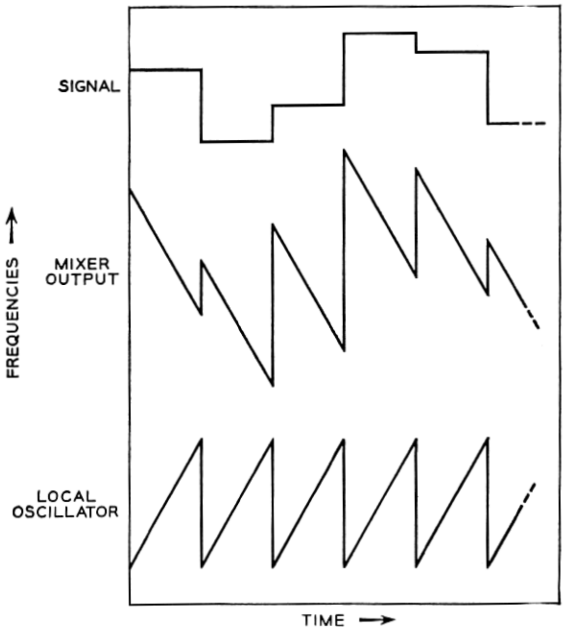

Fig. 5 — Frequency conversion.

The corresponding output of the pulse compression line is approximately

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

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

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

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

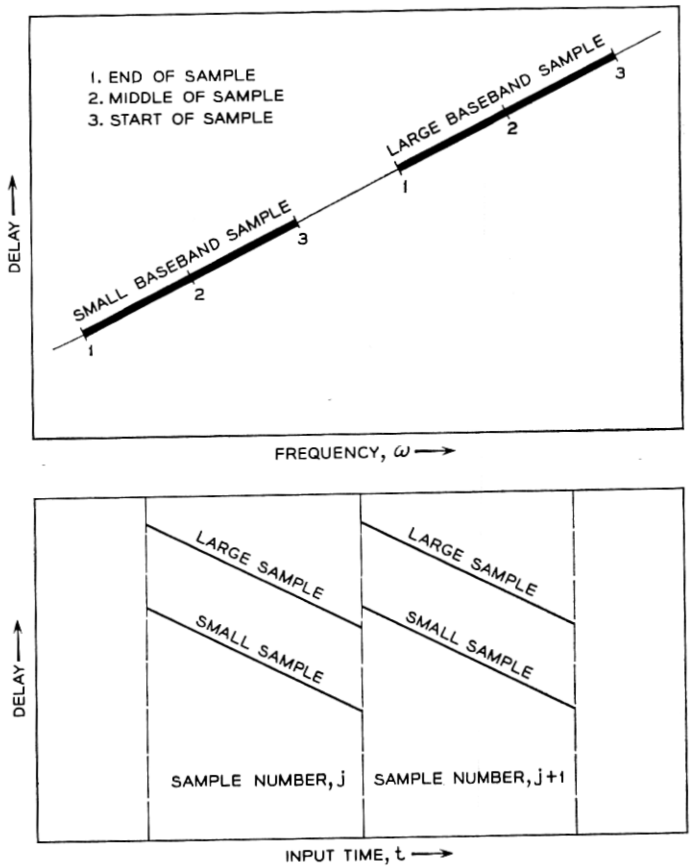

Fig. 6 — Delay vs frequency and time.

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:

$$S(t) = \sqrt{2P_s}\, F(\omega_s - \omega_k) \cos \beta_s \cos \left(\omega_c t + \tfrac{1}{2} q t^2 - \beta_c\right) - \sqrt{2P_s}\, F(\omega_s - \omega_k) \sin \beta_s \sin \left(\omega_c t + \tfrac{1}{2} q t^2 - \beta_c\right). \tag{4}$$

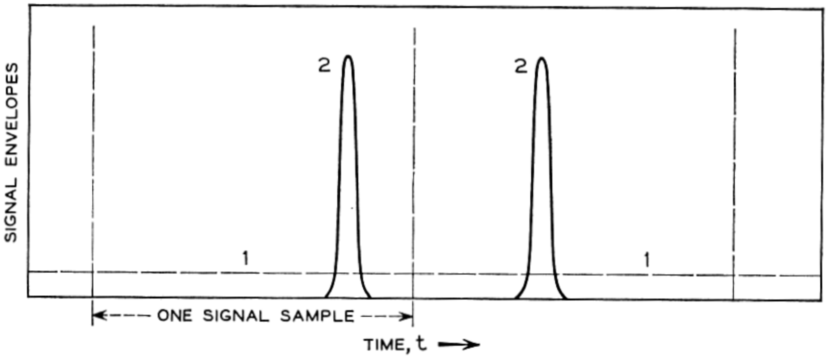

Fig. 7 — Signals at terminals of dispersive line: 1, input signal, before pulse compression; 2, output signal, after pulse compression.

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-

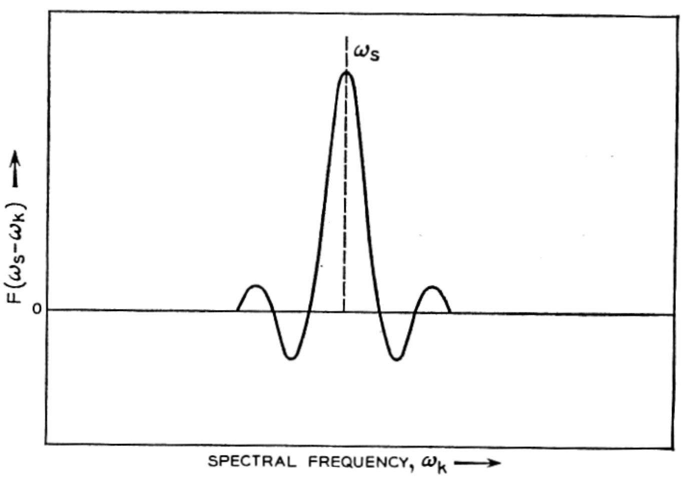

Fig. 8 — The function $F(\omega_s - \omega_k)$.

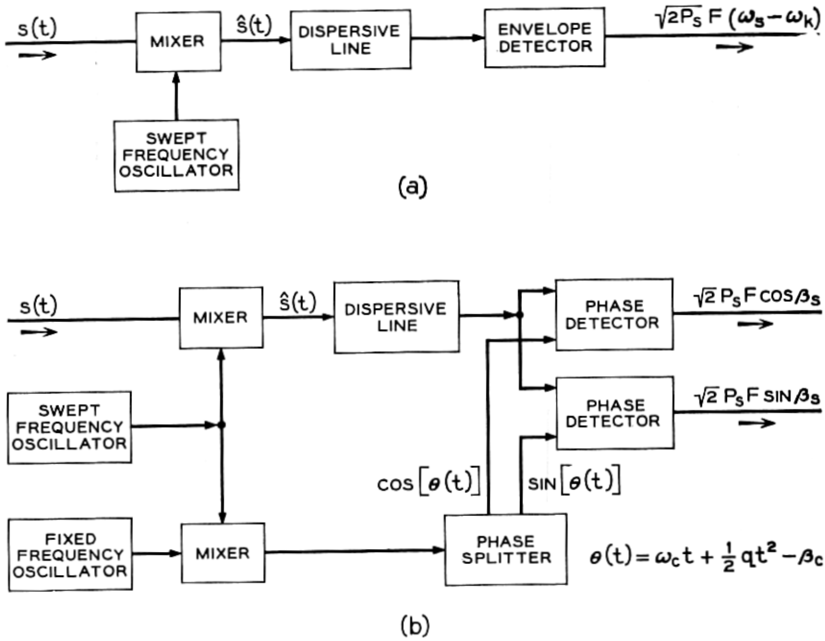

Fig. 9 — Detection of envelope and components: (a) envelope, (b) components.

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)

$$r = \omega_r/\omega_b = \text{bandwidth expansion ratio} \tag{6}$$

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

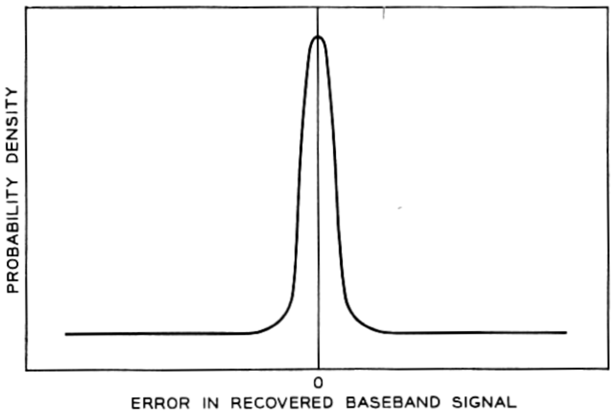

Fig. 10 — Qualitative form of error distribution.

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]

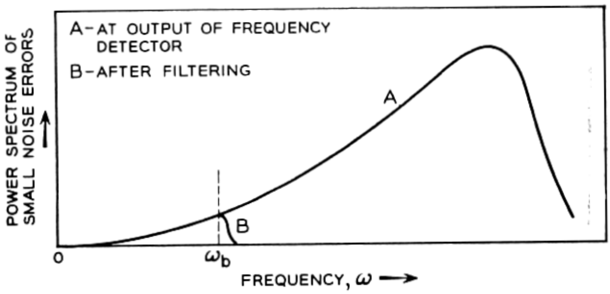

Fig. 11 — Small noise errors in frequency detection.

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.

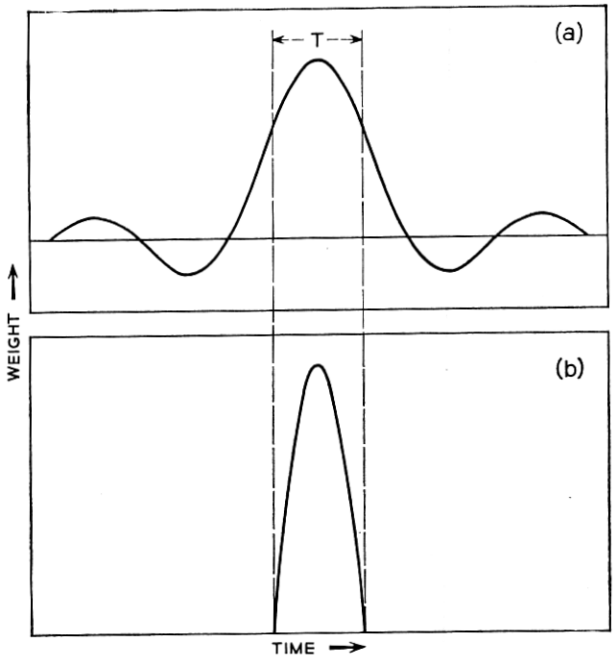

Fig. 12 — Filter weight functions: (a) ideal band-limiting filter; (b) optimum when constrained to one sample interval.

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

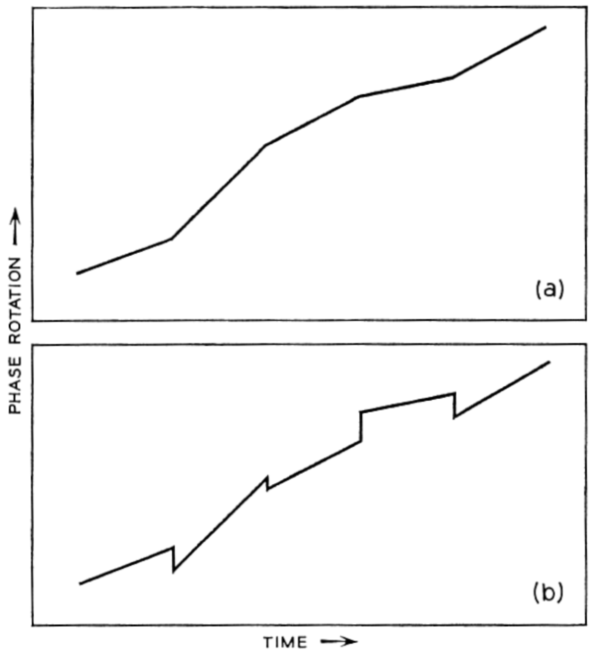

Fig. 13 — Phase rotations: (a) differential phase coherence; (b) differential phase incoherence.

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

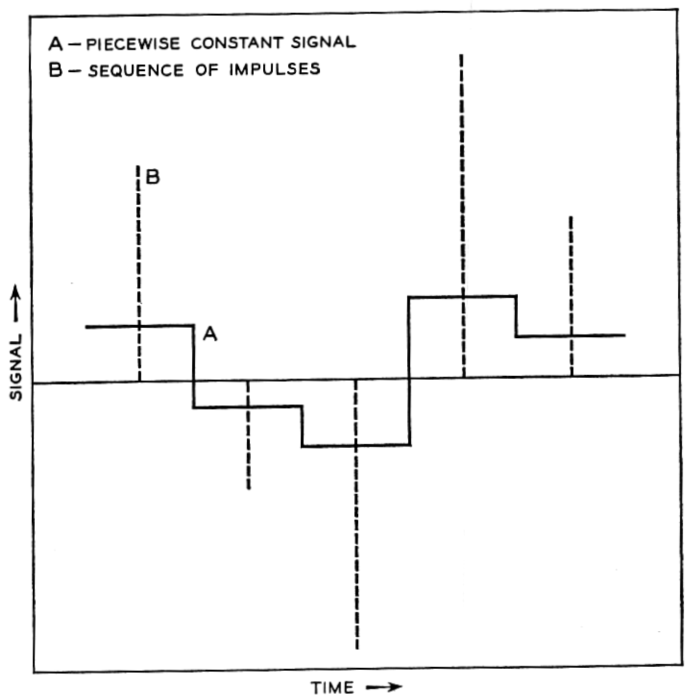

Fig. 14 — Filter inputs.

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

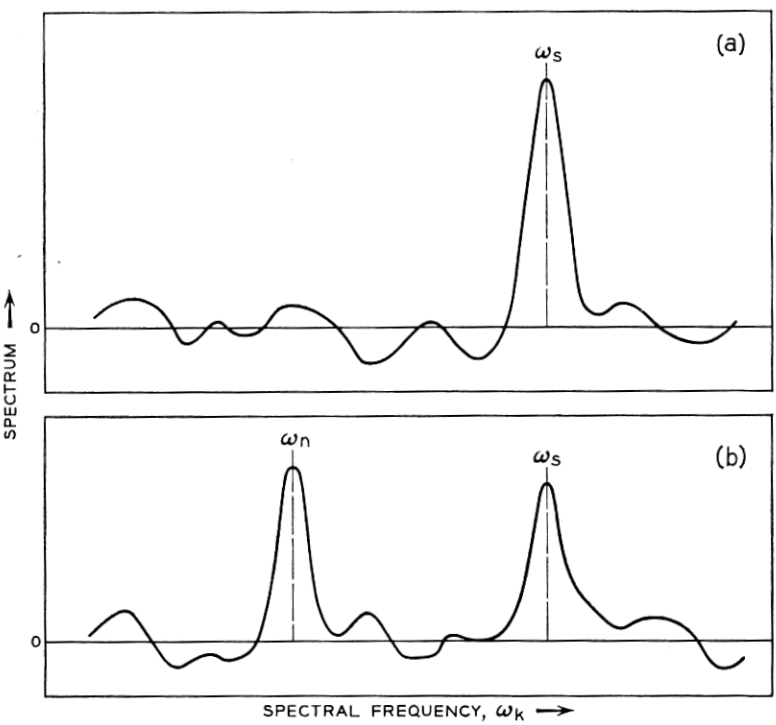

Fig. 16 — Spectrum of a single signal-plus-noise sample: (a) the usual sample; (b) the occasional sample which blocks.

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:

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

For phase incoherent spectrum analysis:

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

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

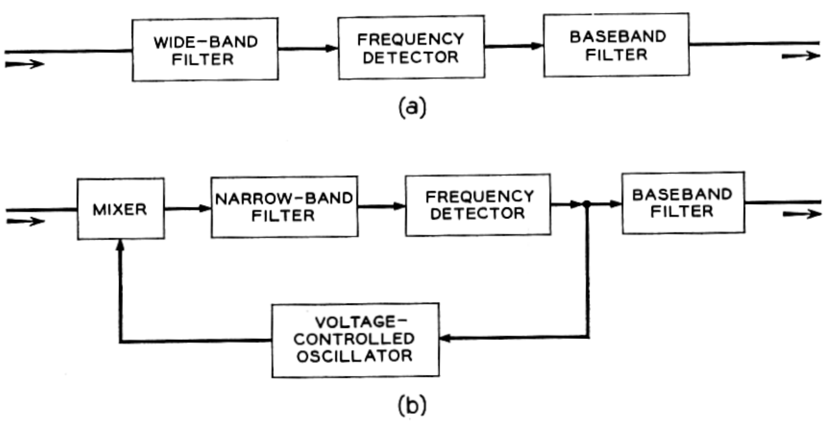

Fig. 17 — Demodulators using frequency detection: (a) conventional FM demodulator; (b) demodulator using FM feedback.

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 frequencies 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 modulation. 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	$\left[\frac{\text{Max. Demod. Signal}}{\text{rms Small Errors}}\right]^*$
		(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 pass-band 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 (db)	$\left[\frac{\text{Max. Demod. Signal}}{\text{rms Small Errors}}\right]^*$ (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.

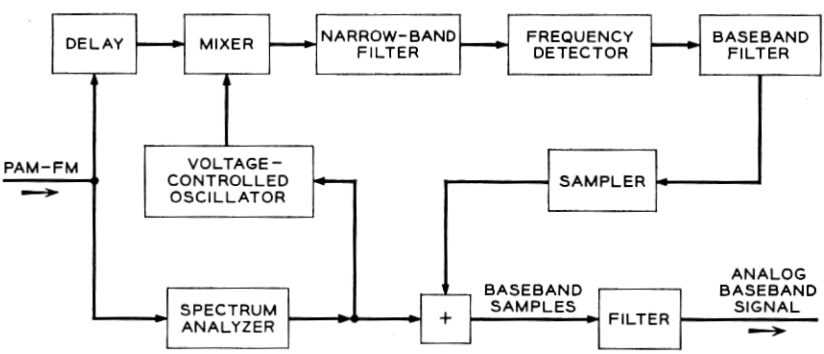

Fig. 18 — A combination of spectrum analysis and frequency detection.

V. ACKNOWLEDGMENTS

The author's work, reported here, has been strongly influenced by conversations with numerous colleagues in the data transmission, FM and FMFB, information theory, and radar pulse compression fields. These included V. E. Beneš, W. R. Bennett, L. H. Enloe, E. N. Gilbert, W. M. Goodall, C. W. Hoover, Jr., I. Jacobs, R. A. Kaenel, D. Mitchell, S. O. Rice, C. L. Ruthroff, and D. Slepian at Bell Telephone Laboratories, and T. Kailath, R. Price, and G. L. Turin while visiting the University of California at Berkeley. Several of the references were supplied by reviewers.

APPENDIX

A.1 *Spectrum Generation by Pulse Compression*

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

$$\begin{aligned} \hat{s}(t) &= \sqrt{2P_s}\, E(t) \cos\left(\omega_s t - \tfrac{1}{2} q t^2 + \beta_s\right) \\ E(t) &= 0 \text{ outside of interval } -T/2 \leqq t \leqq +T/2 \\ E(-t) &= E(t). \end{aligned}$$

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

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

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:

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

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.

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

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

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

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

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 $E(t) = 1, -T/2 \leqq t \leqq +T/2$.

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

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

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

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:

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

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_σ, y_σ, z_σ 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 $\sigma_n^2 = \text{Ave } n_\sigma^2$, $\sigma_m^2 = \text{Ave } m_\sigma^2$

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

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_λ = 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:

$$P \approx (\eta - 1)P\{x_\lambda > x_s\} = \frac{\eta - 1}{2}\left[1 - \mathrm{Erf}\left(\frac{R}{\sqrt{2}}\right)\right]$$
$$\approx \frac{\eta - 1}{\sqrt{\pi}R} \exp\left(-\frac{R^2}{4}\right).$$

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.