An Introduction to Matched Filters*

GEORGE L. TURIN†

Summary—**In a tutorial exposition, the following topics are discussed: definition of a matched filter; where matched filters arise; properties of matched filters; matched-filter synthesis and signal specification; some forms of matched filters.**

I. Foreword

In this introductory treatment of matched filters, an attempt has been made to provide an engineering insight into such topics as: where these filters arise, what their properties are, how they may be synthesized, etc. Rigor and detail are purposely avoided, on the theory that they tend, on first contact with a subject, to obscure fundamental concepts rather than clarify them. Thus, for example, although it is not assumed that the reader is conversant with statistical estimation and hypothesis-testing theories, the pertinent results of these are invoked without mathematical proof; instead, they are justified by an appeal to intuition, starting with simple cases and working up to greater and greater complexity. Such a presentation is admittedly not sufficient for a completely thorough understanding: it is merely a prelude. It is hoped that the interested reader will fill in the gaps himself by consulting the cited references at his leisure.

Of course, one must always start somewhere, and here it is with the assumption that the reader is already familiar with the elements of probability theory and linear filter theory—that is, with such things as probability density functions, spectra, impulse response functions, transfer functions, and so forth. If he is not, reference to Chapters 1 and 2 of [57] and Chapter 9 of [7] will probably suffice.

The bibliography, although lengthy, is not meant to be complete, nor could it be. Aside from the inevitable inability of the author to be familiar with the entire unclassified literature on the subject, there is an extensive body of classified literature, much of it precedent to the unclassified literature, which of course could not be cited. Of the latter, it must be said regretfully, large portions should not have been classified in the first place, or should long since have been declassified.

No bibliography can satisfactorily reflect the influence of personal conversations with colleagues on an author's thoughts about a subject. The present author would especially like to acknowledge the many he has had over the years with Drs. W. B. Davenport, Jr., R. M. Fano, P. E. Green, Jr., and R. Price; this, without attributing to them in any way the inadequacies of what follows.

* Manuscript received by the PGIT, January 23, 1960.
† Hughes Research Laboratories, Malibu, Calif.

II. Definition of a Matched Filter

If $s(t)$ is any physical waveform, then a filter which is matched to $s(t)$ is, by definition, one with impulse response

$$h(\tau) = ks(\Delta - \tau), \tag{1}$$

where k and Δ are arbitrary constants. In order to envisage the form of $h(t)$, consider Fig. 1, in part (a) of which is shown a wave train, $s(t)$, lasting from t_1 to t_2. By reversing the direction of time in part (a), *i.e.*, letting $\tau = -t$, one obtains the reversed train, $s(-\tau)$, of part (b). If this latter waveform is now delayed by Δ seconds, and its amplitude multiplied by k, the resulting waveform—part (c) of Fig. 1—is the matched-filter impulse response of (1).[1]

The transfer function of a matched filter, which is the Fourier transform of the impulse response, has the form

$$\begin{aligned} H(j2\pi f) &= \int_{-\infty}^{\infty} h(\tau)e^{-i2\pi f\tau}\, d\tau \\ &= k\int_{-\infty}^{\infty} s(\Delta - \tau)e^{-i2\pi f\tau}\, d\tau \\ &= ke^{-i2\pi f\Delta}\int_{-\infty}^{\infty} s(\tau')e^{i2\pi f\tau'}\, d\tau', \end{aligned} \tag{2}$$

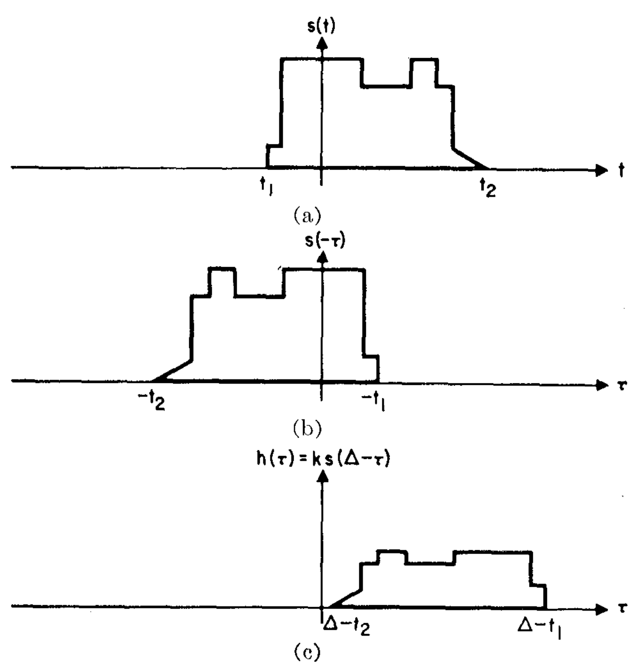

Fig. 1—(a) A wave train; (b) the reversed train; (c) a matched-filter impulse response.

[1] For some types of synthesis of $h(\tau)$—for example, as the impulse response of a passive, linear, electrical network—Δ is constrained by realizability considerations to the region $\Delta \geq t_2$. If $t_2 = \infty$, approximations are sometimes necessary. The problems of realization will be considered more fully later.

where the substitution $\tau' = \Delta - \tau$ has been made in going from the third to the fourth member of (2). Now, the spectrum of $s(t)$, *i.e.*, its Fourier transform, is:[2]

$$S(j2\pi f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}\,dt. \tag{3}$$

Comparison of (2) and (3) reveals, then, that

$$H(j2\pi f) = kS(-j2\pi f)e^{-j2\pi f\Delta} = kS^*(j2\pi f)e^{-j2\pi f\Delta}. \tag{4}$$

That is, except for a possible amplitude and delay factor of the form $ke^{-j2\pi f\Delta}$, the transfer function of a matched filter is the complex conjugate of the spectrum of the signal to which it is matched. For this reason, a matched filter is often called a "conjugate" filter.

Let us postpone further study of the characteristics of matched filters until we have gained enough familiarity with the contexts in which they appear to know what properties are important enough to investigate.

III. Where Matched Filters Arise

A. Mean-Square Criteria

Perhaps the first context in which the matched filter made its appearance [31], [55] is that depicted in Fig. 2.

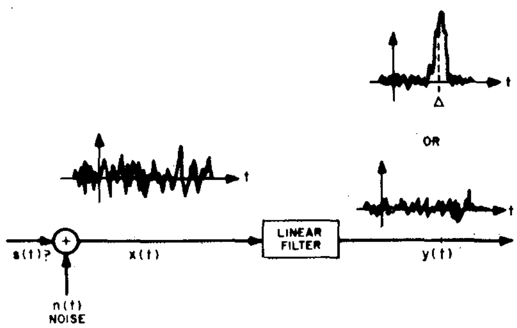

Fig. 2—Pertaining to the maximization of signal-to-noise ratio.

Suppose that one has received a waveform, $x(t)$, which consists either solely of a white noise, $n(t)$, of power density $N_0/2$ watts/cps;[3] *or* of $n(t)$ *plus* a signal, $s(t)$—say a radar return—of known form. One wishes to determine which of these contingencies is true by operating on $x(t)$ with a linear filter in such a way that if $s(t)$ *is* present, the filter output at some time $t = \Delta$ will be considerably greater than if $s(t)$ is absent. Now, since the filter has been assumed to be linear, its output, $y(t)$, will be composed of a noise component $y_n(t)$, due to $n(t)$ only, and, if $s(t)$ is present, a signal component $y_s(t)$, due to $s(t)$ only. A simple way of quantifying the requirement that $y(\Delta)$ be "considerably greater" when $s(t)$ is present than when $s(t)$ is absent is to ask that the filter make the instantaneous power in $y_s(\Delta)$ as large as possible compared to the average power in $n(t)$ at time Δ.

Assuming that $n(t)$ is stationary, the average power in $n(t)$ at any instant is the integrated power under the noise power density spectrum at the filter output. If $G(j2\pi f)$ is the transfer function of the filter, the output noise power density is $(N_0/2)\ |G(j2\pi f)|^2$; the output noise power is therefore

$$\frac{N_0}{2}\int_{-\infty}^{\infty} |G(j2\pi f)|^2\,df. \tag{5}$$

Further, if $S(j2\pi f)$ is the input signal spectrum, then $S(j2\pi f)G(j2\pi f)$ is the output signal spectrum, and $y_s(\Delta)$ is the inverse Fourier transform of this, evaluated at $t = \Delta$; that is,

$$y_s(\Delta) = \int_{-\infty}^{\infty} S(j2\pi f)G(j2\pi f)e^{j2\pi f\Delta}\,df. \tag{6}$$

The ratio of the square of (6) to (5) is the power ratio we wish to maximize:

$$\rho = \frac{2\left[\int_{-\infty}^{\infty} S(j2\pi f)G(j2\pi f)e^{j2\pi f\Delta}\,df\right]^2}{N_0\int_{-\infty}^{\infty} |G(j2\pi f)|^2\,df}. \tag{7}$$

Recognizing that the integral in the numerator is real [it is $y_s(\Delta)$], and identifying $G(j2\pi f)$ with $f(x)$ and $S(j2\pi f)e^{j2\pi f\Delta}$ with $g(x)$ in the Schwarz inequality,

$$\left|\int f(x)g(x)\,dx\right|^2 \le \int |f(x)|^2\,dx \int |g(x)|^2\,dx, \tag{8}$$

one obtains from (7)

$$\rho \le \frac{2}{N_0}\int_{-\infty}^{\infty} |S(j2\pi f)|^2\,df. \tag{9}$$

Since $|S(j2\pi f)|^2$ is the energy density spectrum of $s(t)$, the integral in (9) is the total energy, E, in $s(t)$. Then

$$\rho \le \frac{2E}{N_0}. \tag{10}$$

It is clear on inspection that the equality in (8), and hence in (9) and (10), holds when $f(x) = kg^*(x)$, *i.e.*, when

$$G(j2\pi f) = kS^*(j2\pi f)e^{-j2\pi f\Delta}. \tag{11}$$

Thus, when the filter is matched to $s(t)$, a maximum value of ρ is obtained. It is further easily shown that the equality in (8) holds *only* when $f(x) = kg^*(x)$, so the matched filter of (11) represents the only type of linear filter which maximizes ρ.

Notice that we have assumed nothing about the statistics of the noise except that it is stationary and white, with power density $N_0/2$. If it is not white, but has some arbitrary power density spectrum $|N(j2\pi f)|^2$, a derivation similar to that given above [9], [15] leads to the solution

$$G(j2\pi f) = \frac{kS^*(j2\pi f)e^{-j2\pi f\Delta}}{|N(j2\pi f)|^2}. \tag{12}$$

[2] This is a *density* spectrum: that is, if $s(t)$ is, *e.g.*, a voltage waveform, $S(j2\pi f)$ is a voltage density, and its integral from f_1 to f_2 (plus that from $-f_2$ to $-f_1$) is the part of the voltage in $s(t)$ originating in the band of frequencies from f_1 to f_2.

[3] This is the "double-ended" density, covering positive and negative frequencies. The "single-ended" physical power density (positive frequencies only) is thus N_0.

One can convince himself of this intuitively in the following manner. If the input, $x(t)$, of Fig. 2 is passed through a filter with transfer function $1/N(j2\pi f)$, the noise component at its output will be white; however, the signal component will be distorted, now having the spectrum $S(j2\pi f)/N(j2\pi f)$. On the basis of our previous discussion of signals in white noise, it seems reasonable, then, to follow the noise-whitening filter with a filter matched to the distorted signal spectrum, *i.e.*, with the filter $kS^*(j2\pi f)e^{-j2\pi f\Delta}/N^*(j2\pi f)$. The cascade of the noise-whitening filter and this matched filter is indeed the solution (12).[4]

So far we have considered only a detection problem: is the signal present or not? Suppose, however, we know that the signal is present, but has an unknown delay, t_0, which we wish to measure (*e.g.*, radar ranging). Then the first of the output waveforms in Fig. 2 obtains, but the peak in it is delayed by the unknown delay. In order to measure this delay accurately, we should not only like the output waveform, as before, to be large at $t = \Delta + t_0$ but also to be very small elsewhere.

More generally, we may frame the problem in the manner depicted in Fig. 3. A sounding signal of known form, $s(t)$, is transmitted into an unknown filter, the impulse response of which is $u(\tau)$. [In Fig. 2, this filter is merely a pair of wires with impulse response $\delta(t)$, the Dirac delta function.] At the output of the unknown filter, stationary noise is added to the signal, the sum being denoted by $x(t)$. We desire to operate on $x(t)$ with a linear filter, whose output, $y(t)$, is to be as faithful as possible an estimate of the unknown impulse response, perhaps with some delay Δ.

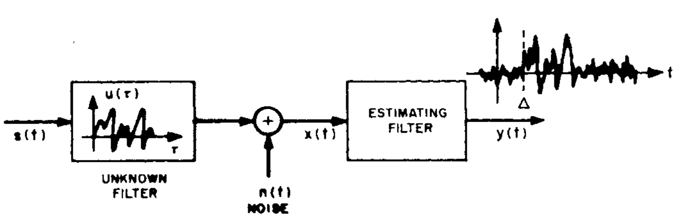

Fig. 3—Pertaining to mean-square estimation of an unknown impulse response.

The unknown filter may be some linear or quasi-linear transmission medium, such as the ionosphere, the characteristics of which we wish to measure. Again, it may represent a complex of radar targets, in which case $u(\tau)$ may consist of a sequence of delta functions with unknown delays (ranges) and strengths.

A reasonable mean-square criterion for faithfulness of reproduction is that the average of the squared difference between $y(t)$ and $u(t - \Delta)$, integrated over the pertinent range of t, be as small as possible: the average must be taken over both the ensemble of possible impulse responses of the unknown filter, and the ensemble of possible noises. Using such a criterion [51], one arrives at an optimum estimating filter which is in general relatively complicated. However, when the signal-to-noise ratio is small, and it is assumed that nothing whatever is known about $u(\tau)$ except possibly its maximum duration, the optimum filter turns out to be matched to $s(t)$ if the noise is white, and has the form of (12) for nonwhite noise. Further, in the important case when the sounding signal, $s(t)$, is also optimized to minimize the error in $y(t)$, the optimum estimating filter is matched to $s(t)$ for all signal-to-noise ratios and for all degrees of *a priori* knowledge about $u(\tau)$, provided only that the noise is white.[5]

B. Probabilistic Criteria

In the preceding discussion, we have confined ourselves to mean-square criteria—maximization of a signal-to-noise power ratio or minimization of a mean-square difference. The use of such simple criteria has the advantage of not requiring us to know more than a second-order statistic of the noise—the power-density spectrum. But, although mean-square criteria often have strong intuitive justifications, we should prefer to use criteria directly related to performance ratings of the systems in which we are interested, such as radar and communication systems. Such performance ratings are usually probabilistic in nature: one speaks of the probabilities of detection, of false alarm, of error, etc., and it is these which we wish to optimize. This brings us into the realm of classical statistical hypothesis-testing and estimation theories.

Let us first examine perhaps the simplest hypothesis-testing problem, the one posed at the start of the section on mean-square criteria: the observed signal, $x(t)$, is either due solely to noise, or to both an exactly known signal and noise.[6] Such a situation could occur, for example, in an on-off communication system, or in a radar detection system. Adopting the standard parlance of hypothesis-testing theory, we denote the former hypothesis, noise only, by H_0, and the alternative hypothesis by H_1. We wish to devise a test for deciding in favor of H_0 or H_1.

There are two types of errors with which we are concerned: a Type I error, of deciding in favor of H_1 when H_0 is true, and a Type II error, of deciding in favor of H_0 when H_1 is true. The probabilities of making such errors are denoted by α and β, respectively. For a choice of criterion, we may perhaps decide to minimize the average of α and β (*i.e.*, the over-all probability of error); this would require a knowledge of the *a priori* probabilities of H_0 and H_1, which is generally available in a communication system. On the other hand, perhaps one type of error is more costly than the other, and we may then wish to minimize an average cost [29]. When, as is often the case in radar detection, neither the *a priori*

[4] The weak link in this heuristic argument is, of course, that it is not obvious that an optimization performed on the output of the noise-whitening filter is equivalent to one performed on its input, the observed waveform; it can be shown, however, that this is so.

[5] Another approach to this problem of impulse-response estimation appears in [25].

[6] Here, however, we shall not initially restrict ourselves as before solely to additive combinations of signal and noise.

probabilities nor the costs are known, or even definable, one often alternatively accepts the criterion of minimizing β (in radar: maximization of the probability of detection) for a given, predetermined value of α (false-alarm probability)—the Neyman-Pearson criterion [7].

What is important for our present considerations is that all these criteria lead to the same generic form of test. If one lets $p_0(x)$ be the probability (density) that *if* H_0 is true, the observed waveform, $x(t)$, could have arisen; and $p_1(x)$ be the probability (density) that *if* H_1 is true, $x(t)$ could have arisen; then the test has the form [7], [29], [33]:

$$\left.\begin{array}{lll}\text{accept} & H_1 & \text{if } \dfrac{p_1(x)}{p_0(x)} > \lambda \\ \text{accept} & H_0 & \text{if } \dfrac{p_1(x)}{p_0(x)} \leq \lambda \end{array}\right\}. \tag{13}$$

Here λ is a constant dependent on *a priori* probabilities and costs, if these are known, or on the predetermined value of α in the Neyman-Pearson test; most importantly, it is *not* dependent on the observation $x(t)$. The test (13) asks us to examine the possible causes of what we have observed, and to determine whether or not the observation is λ times more likely to have occurred if H_1 is true than if H_0 is true; if it is, we accept H_1 as true, and if not, we accept H_0. If $\lambda = 1$, for example, we choose the cause which is the more likely to have given rise to $x(t)$. A value of λ not equal to unity reflects a bias on the part of the observer in favor of choosing one hypothesis or the other.

Let us assume now that the noise, $n(t)$, is additive, gaussian and white with spectral density $N_0/2$,[3] and further that the signal, if present, has the known form $s(t - t_0)$, $t_0 \leq t \leq t_0 + T$, where the delay, t_0, and the signal duration, T, are assumed known. Then, on observing $x(t)$ in some observation interval, I, which includes the interval $t_0 \leq t \leq t_0 + T$, the two hypotheses concerning its origin are:

$$\left.\begin{array}{ll} H_0: & x(t) = n(t), \quad t \text{ in } I \\ H_1: & x(t) = s(t - t_0) + n(t), \quad t \text{ in } I \end{array}\right\}. \tag{14}$$

Now, it can be shown [57] that the probability density of a sample, $n(t)$, of white, gaussian noise lasting from a to b may be expressed as[7]

$$p(n) = k \exp\left[-\frac{1}{N_0}\int_a^b n^2(t)\, dt\right], \tag{15}$$

where $N_0/2$ is the double-ended spectral density of the noise, and k is a constant not dependent on $n(t)$. Hence the likelihood that, if H_0 is true, the observation $x(t)$ could arise is simply the probability (density) that the noise waveform can assume the form of $x(t)$, *i.e.*,

$$p_0(x) = k \exp\left[-\frac{1}{N_0}\int_I x^2(t)\, dt\right], \tag{16}$$

[7] The space on which this probability density exists must be carefully defined, but the details of this do not concern us here.

the region of integration being, as indicated, the observation interval, I. Similarly, the likelihood that, if H_1 is true, $x(t)$ could arise is the probability density that the noise can assume the form $n(t) = x(t) - s(t - t_0)$, *i.e.*,

$$\begin{aligned} p_1(x) &= k \exp\left[-\frac{1}{N_0}\int_I [x(t) - s(t - t_0)]^2\, dt\right] \\ &= k \exp\left[-\frac{1}{N_0}\int_I x^2(t)\, dt \right.\\ &\qquad \left. + \frac{2}{N_0}\int_I s(t - t_0)x(t)\, dt - \frac{E}{N_0}\right], \end{aligned} \tag{17}$$

where we have denoted $\int_I s^2(t - t_0)\, dt$, the energy of the signal, by E.

On substituting (16) and (17) in (13) and taking the logarithm of both sides of the inequality, the hypothesis-testing criterion becomes

$$\left.\begin{array}{lll}\text{accept} & H_1 & \text{if } y(t_0) > \lambda' \\ \text{accept} & H_0 & \text{if } y(t_0) \leq \lambda' \end{array}\right\}, \tag{18}$$

where we have set

$$y(t_0) = \int_I s(t - t_0)x(t)\, dt \tag{19}$$

and

$$\lambda' = \frac{N_0}{2}\log \lambda + 2E. \tag{20}$$

Changing variable in (19) by setting $\tau = t_0 - t$, one obtains

$$y(t_0) = \int_{-T}^{0} s(-\tau)x(t_0 - \tau)\, d\tau. \tag{21}$$

But one immediately recognizes this last as the output, at time t_0, of a filter with impulse response $g(\tau) = s(-\tau)$, in response to an input waveform $x(t)$. The filter thus specified is clearly matched to $s(t)$, and the optimum system called for by (18) therefore takes on the form of Fig. 4.[8]

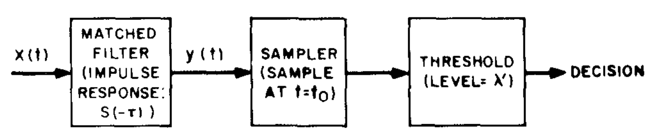

Fig. 4—A simple radar detection system.

[8] Notice that $s(-\tau)$, hence $g(\tau)$, is nonzero only in the interval $-T \leq \tau \leq 0$; hence the limits of integration. In order to realize $g(\tau)$, a delay of $\Delta \geq T$ must be inserted in $g(\tau)$—see (1) and footnote 1. This introduces an equal delay at the filter output, which must then be sampled at time $t_0 + \Delta$, rather than at t_0.

Notice also that $y(t_0)$ could be obtained by literally following the edicts of (19). That is, one could multiply the incoming waveform, $x(t)$, by a stored replica of the signal waveform, $s(t)$, delayed by t_0; the product, integrated over the observation interval, is $y(t_0)$, and is to be compared with the threshold, λ'. Such a detector is called a correlation detector. We shall in this paper, however, refer solely to the matched-filter versions of our solutions, with the understanding that these may be obtained by correlation techniques if desired.

The similarity between the solution represented in Fig. 4 and the one we obtained in connection with Fig. 2 is apparent: the matched filter in Fig. 4 in fact maximizes the signal-to-noise ratio of $y(t)$ at $t = t_0$, the sampling instant. That a mean-square criterion and a probabilistic criterion should lead to the same result in the case of additive, gaussian noise is no coincidence; there is an intimate connection between the two types of criteria in this case.

So far, we have considered that the signal component of $x(t)$, if it is present, is known exactly; in particular, we have assumed that the delay t_0 is known. Suppose now that the envelope delay of the signal is known, but not the carrier phase, θ [50]. Assuming that a probability density distribution is given for θ, (13) reduces to

$$\left.\begin{array}{ll}\text{accept } H_1 & \text{if } \dfrac{\int_0^{2\pi} p_1(x/\theta)p(\theta)\, d\theta}{p_0(x)} > \lambda'' \\ \text{accept } H_0 & \text{otherwise}\end{array}\right\}. \qquad (22)$$

Here $p_1(x/\theta)$ is the conditional probability, given θ, that if H_1 is true $x(t)$ will arise; $p(\theta)$ is the probability density of θ. If θ is completely random, so $p(\theta)$ is flat, then carrying through the computations for the case of white, gaussian noise leads to the intuitively expected result that an envelope detector should be inserted in Fig. 4 between the matched filter and the sampler [38], [57]. Note, however, that the matched filtering of $x(t)$ is still the core of the test.

If the envelope delay is also unknown, and no probability distribution is known for it *a priori* other than that it must lie in a given interval, Ω, a good test is [7]:

$$\left.\begin{array}{ll}\text{accept } H_1 & \text{if } \dfrac{\max\limits_{\Omega} \int_0^{2\pi} p_1(x/\theta)p(\theta)\, d\theta}{p_0(x)} > \lambda''' \\ \text{accept } H_0 & \text{otherwise}\end{array}\right\}, \qquad (23)$$

where it is implicit that the integral in the numerator depends on a hypothesized value of the envelope delay.[9] For a flat distribution of θ, (23) yields a circuit in which the envelope detector just inserted in Fig. 4 remains, but is now followed not by a sampler, but by a wide gate, open during the interval Ω. For all values of t_0 within this interval, this gate passes the envelope of $y(t_0)$ to the threshold; if the threshold is exceeded at any instant, then the maximum value of the gate output will also surely exceed it, and, by (23), H_1 must then be accepted. As before, a matched filtering operation is the basic part of the test.[10]

We have been discussing, mostly in the radar context, the simple binary detection problem, "signal plus noise or noise only?" Clearly, this is a special case of the general binary detection problem, more germane to communications, "signal 1 plus noise or signal 2 plus noise?", where we have taken one of the signals to be identically zero. Let us now address ourselves to the even more general digital communications problem, "which of M possible signals was transmitted?"

Let us first consider the case in which the forms of the signals, *as they appear at the receiver*, are known exactly. In this situation, a good hypothesis test, of which (13) is a special case, is of the form:

$$\left.\begin{array}{l}\text{accept the hypothesis, } H_m, \text{ for which} \\ \mu_m p_m(x)\ (m = 1, \cdots, M) \text{ is greatest}\end{array}\right\}. \qquad (24)$$

Here $p_m(x)$ is the probability that if the mth signal was sent, $x(t)$ will be received. The μ_m's are constants, independent of $x(t)$, which are determined solely by the criterion of the test. If the criterion is, for example, minimization of the over-all probability of error, μ_m is the *a priori* probability of transmittal of the mth signal [56], [57]. The μ_m's may possibly also be related to the costs of making various types of errors. If neither *a priori* probabilities nor costs are given, the μ_m's are generally all equated to unity, and the test is called a maximum-likelihood test [7].

If the noise is additive, gaussian, and white with spectral density $N_0/2$, then following the reasoning which led to (17), we have

$$p_m(x) = k \exp\left[-\frac{1}{N_0}\int_I x^2(t)\, dt + \frac{2}{N_0} y_m(0) - \frac{E_m}{N_0}\right], \qquad (25)$$

where, letting $s_m(t)$ be the mth signal,

$$y_m(t) = \int_{-T}^{0} s_m(-\tau)x(t-\tau)\, d\tau. \qquad (26)$$

We have arbitrarily assumed that the signals are received with no delay (this amounts to choosing a time origin), and that they all have the same duration, T. E_m is the energy in the mth signal.

Taking logarithms for convenience, test (24) reduces to:

$$\left.\begin{array}{l}\text{accept the hypothesis, } H_m, \text{ for which} \\ y_m(0) + \left(\dfrac{N_0}{2}\log\mu_m - 2E_m\right) \text{ is greatest}\end{array}\right\}. \qquad (27)$$

[9] Here, for lack of knowing the true envelope delay, we have essentially used test (22), in which we have assumed that the envelope delay has that value which maximizes the integral in the numerator. This procedure is part and parcel of the estimation problem, which we have already considered briefly in the discussion of mean-square criteria, and which we shall later bring into the present discussion on probabilistic criteria.

[10] Note that if a correlation detector (see footnote 8) is to be used here, it must compute the envelope of $y(t_0)$ separately for each value of t_0 in Ω. That is, (19) *and its quadrature component* must be obtained, squared and summed *for each* t_0 by a separate multiplication and integration process; t_0 is here a parameter (parametric time). In general, the advantages of using a matched filter to obtain $y(t_0)$, for all t_0 in Ω, as a function of real time are obviously great. In a few circumstances, however—such as range-gated pulsed radar—the correlation technique may be more easily implemented, to a good approximation.

$y_m(0)$ is, as we have seen, the output at $t = 0$ of a filter matched to $s_m(t)$, in response to the input $x(t)$. Therefore, the optimum receiver has the configuration of Fig. 5, where the biases B_m are the quantities $(N_0/2) \log \mu_m - 2E_m$. The output of the decision circuit is the index m which corresponds to the greatest of the M inputs.

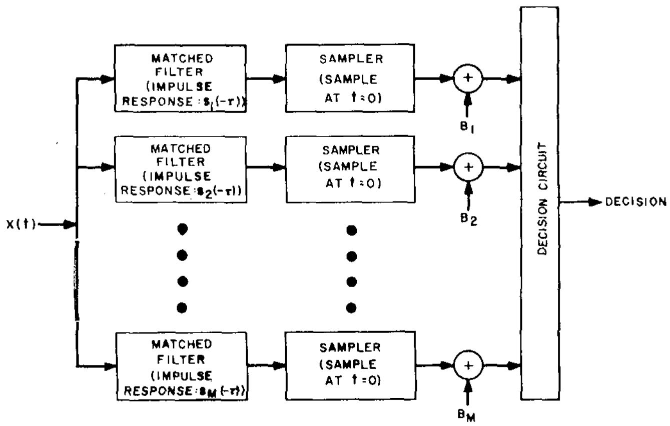

Fig. 5—A simple M-ary communications receiver.

If the forms of the signals $s_m(t)$, as they appear at the receiver, are not known exactly, but are dependent on, say, L random parameters, $\theta_1, \cdots, \theta_L$, whose joint density distribution $q(\theta_1, \cdots, \theta_L)$ is known, then (24) takes the form:

$$\left.\begin{array}{l}\text{accept the hypothesis, } H_m \text{, for which} \\ \mu_m \int \cdots \int q(\theta_1, \cdots, \theta_L) \\ \quad \cdot p_m(x/\theta_1, \cdots, \theta_L)\, d\theta_1 \cdots d\theta_L \quad \text{is greatest}\end{array}\right\}, \tag{28}$$

where $p_m(x/\theta_1, \cdots, \theta_L)$ is the conditional probability, given $\theta_1, \cdots, \theta_L$, that if H_m is true, the observed received signal, $x(t)$, will arise.

Suppose there is only one unknown parameter, the carrier phase, θ, which is assumed to have a uniform distribution. Then, as in the on-off binary case, it turns out that the only modification required in the configuration of the optimum receiver is that envelope detectors of a particular type[11] must be inserted between the matched filters and samplers in Fig. 5.

Test (28) may also be applied to the case of multipath communications. For example, one may consider the case of a discrete, slowly varying, multipath channel in which the paths are independent, and each is characterized by a random strength, carrier phase-shift, and modulation delay; if P is the number of paths, there are then $3P$ random parameters on which the signals depend. It can be shown [50] that, for a broad class of probability distributions of path characteristics, an optimum-receiver arrangement similar to Fig. 5 still obtains. Instead of a simple sampler following the matched filter, however, there is in general a more complicated nonlinear device, the form of which depends on the form of the distribution, $q(\theta_1, \cdots, \theta_L)$, inserted in (28). If, for instance, it is assumed that the modulation delays of the paths are known, but that the path strengths and phase-shifts share joint distributions of a fairly general form [50], the device combines nonlinear functions of several samples, taken at instants corresponding to the expected path delays, of both the output of the matched filter and of its envelope. Again, if all the paths are assumed *a priori* to have identical strength distributions, their carrier phase-shifts assumed uniformly distributed over $(0, 2\pi)$, and their modulation delays assumed uniformly distributed over the interval (t_a, t_b),[12] the form of the nonlinear device which should replace the mth sampler in Fig. 5 is given, to a good approximation, by Fig. 6 [12], [50]. The exact form of the nonlinear device, F_m, is determined by the assumed path-strength distribution.

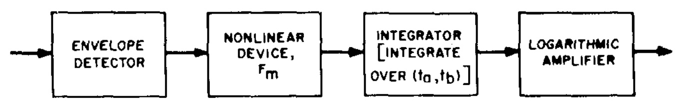

Fig. 6—A nonlinear device for use in Fig. 5 when the path delays are unknown.

The case in which the path modulation delays are known, and the carrier phase-shifts and strengths vary randomly *with time* (*i.e.*, are not restricted, as before, to "slow" variations with respect to the signal duration), has also been considered for Rayleigh-fading paths [36]. Again, at least to an approximation, the results can be interpreted in terms of matched filters. The operations performed on the matched filter outputs now vary with time, however.

In order to use test (28), it is necessary to have a parameter distribution, $q(\theta_1, \cdots, \theta_L)$. It may be, however, that a complete distribution of all the parameters is not available. Then it is often both theoretically and practically desirable to return to a test based on the tenet that the parameters for which no distribution is given are exactly known; in place of the true values of these parameters called for by the test, however, estimates are used. We have already come across such a technique in connection with (23); in that case, an *a priori* distribution of one parameter, phase, was used, but another parameter, modulation delay, was estimated to be that value which maximized the integral in the numerator.[9]

[11] Unlike the on-off binary case, in which any monotone-increasing transfer characteristic can be accommodated in the envelope detector by an adjustment of the threshold, in this case, except when all the biases are equal, the transfer characteristic must be of the log I_0 type [7], [38], [50]. In the region of small signal-to-noise ratios, however, such a characteristic is well approximated by a square-law detector.

[12] These distributions of phase shift and modulation delay may not, of course, be totally due to randomness in the channel, but may partially reflect a distribution of error in phase-base and time-base synchronization of transmitter and receiver; the original assumption of knowledge of the *exact* forms of the signals as they appear at the receiver implied perfect synchronization.

Let us, as a final topic in our discussion of probabilistic criteria, then, consider this problem of estimation of parameters. Clearly, estimation may be an end in itself, as in the case of radar ranging briefly considered in the section of mean-square criteria; on the other hand, it may be to the end discussed in the preceding paragraph.

The step from hypothesis testing to estimation may in a sense be considered identical to that from discrete to continuous variables. In the hypothesis test (24), for example, one inquires, "To which of M discrete causes (signals) may the observation $x(t)$ be ascribed?" If one is alternatively faced with a continuum of causes, such as a set of signals, identical except for a delay which may have any value within an interval, the equivalent test would clearly be of the form:

$$\left.\begin{array}{l}\text{accept the hypothesis,}\quad H_m\,,\quad \text{for which}\\ \mu(m)p_m(x)\ (m_1 \leq m \leq m_2)\quad \text{is greatest}\end{array}\right\}, \tag{29}$$

where the notation is the same as that used in (24), except that m is now a continuous parameter. Only one parameter has been explicitly shown in (29), but clearly the technique may be extended to many parameters.

To consider but one application of (29), let us again investigate the radar-ranging problem, in which the observation is known to be given by

$$x(t) = s(t - t_0) + n(t) \quad \begin{array}{c} a \leq t_0 \leq b \\ t \ \text{in}\ I \end{array}, \tag{30}$$

where $s(t)$ is of known form, nonzero in the interval $0 \leq t \leq T$. Again let us imagine $n(t)$ to be stationary, gaussian and white with spectral density $N_0/2$. Then a little thought will convince one that the probability that $x(t)$ will be observed, if the delay has some specific value t_0 in the interval (a, b), is of the form of the right-hand side of (17). Placing this in (29), identifying m with t_0, and letting $\mu(m)$ be independent of m (maximum-likelihood estimation) [7], it becomes clear that we seek the value of t_0 for which the integrals (19) and (21) are maximum. Remembering our previous interpretation of (21), it follows that a system which estimates delay in this case is one which passes the observed signal, $x(t)$, into a filter matched to the signal $s(t)$, and estimates, as the delay of $s(t)$, the delay (with respect to a reference time) of the maximum of the filter's output. Thus, in optimum radar systems, at least in the case of additive, gaussian noise, a matched filter plays a central part in both the detection and ranging operations. This is a dual rôle we have previously had occasion to note heuristically.[9]

The application of the technique of estimation of unknown signal parameters for use in a hypothesis test has been considered for ideal multipath communication systems [21], [36], [37], [41], [48], [50]. In fact, even if such estimation techniques are not postulated explicitly to start with, but some *a priori* distribution of the unknown parameters is assumed, it is on occasion found that operations which may be interpreted as implicit *a posteriori* estimation procedures appear in the hypothesis-testing receiver [21], [36]. Again, matched filters, or operations very close to matched filtering operations, arise prominently in the solutions.

It is perhaps not necessary to emphasize that the point of the foregoing discussion of where matched filters arise is not to paint a detailed picture of optimum detection and estimation devices; this has not been done, nor could it, in the limited space available. Rather, the point is to indicate that matched filters do appear as the core of a large number of such devices, which differ largely in the type of nonlinear operations applied to the matched-filter output. Having made this point, it behooves us to study the properties of matched filters and to discuss methods of their synthesis, to which topics the remainder of this paper will be devoted.

It seems only proper, however, to end this section with a caveat. The devices based on probabilistic criteria in which we have encountered matched filters were derived on the assumption that at least the additive part of the channel disturbance is stationary, white and gaussian. The stationarity and whiteness requirements are not too essential; the elimination of the former generally leads to time-varying filters, and the elimination of the latter, to the use of the noise-whitening pre-filters we have previously discussed [11], [30]. The requirement of gaussianness is not as easy to dispense with. We have seen that, within the bounds of linear filters, a matched filter maximizes the signal-to-noise ratio for any white, additive noise, and this gives us some confidence in their efficacy in nongaussian situations. But due care should nonetheless be exercised in implying from this their optimality from the point of view of a probabilistic criterion.

IV. Properties of Matched Filters

In order to gain an intuitive grasp of how a matched filter operates, let us consider the simple system of Fig. 7.

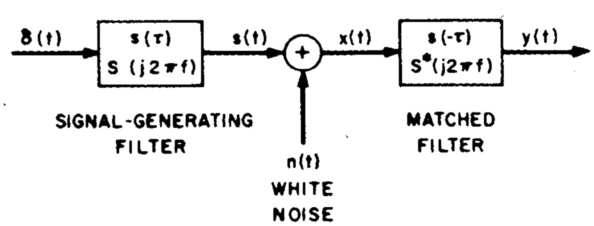

Fig. 7—Illustrating the properties of matched filters.

A signal, $s(t)$, say of duration T, may be imagined to be generated by exciting a filter, whose impulse response is $s(\tau)$, with a unit impulse at time $t = 0$. To this signal is added a white noise waveform, $n(t)$, the power density of which is $N_0/2$. The sum signal, $x(t)$, is then passed into a filter, matched to $s(t)$, whose output is denoted by $y(t)$.

This output signal may be resolved into two components,

$$y(t) = y_s(t) + y_n(t), \tag{31}$$

the first of which is due to $s(t)$ alone, the second to $n(t)$ alone. It is these two components which we wish to study. For simplicity of illustration, let us take all spectra to be centered around zero frequency; no generality is lost in our results by doing this.

The response to an input $s(t)$ of a linear filter with impulse response $h(\tau)$ is

$$\int_{-\infty}^{\infty} h(\tau)s(t - \tau)\, d\tau.$$

If $h(\tau) = s(-\tau)$, as in Fig. 7, then

$$y_s(t) = \int_{-\infty}^{\infty} s(-\tau)s(t - \tau)\, d\tau. \tag{32}$$

This is clearly symmetric in t, since

$$\begin{aligned} y_s(-t) &= \int_{-\infty}^{\infty} s(-\tau)s(-t - \tau)\, d\tau \\ &= \int_{-\infty}^{\infty} s(t - \tau')s(-\tau')\, d\tau' = y_s(t), \end{aligned} \tag{33}$$

where the second equality is obtained through the substitution $\tau' = t + \tau$. For the low-pass case we are considering, $y_s(t)$ may look something like the pulse in Fig. 8. The height of this pulse at the origin is, from (32),

$$y(0) = \int_{-\infty}^{\infty} s^2(\tau)\, d\tau = E, \tag{34}$$

where E is the signal energy. By applying the Schwarz inequality, (8), to (32), it becomes apparent that $| y_s(t) |$ cannot exceed $y(0) = E$ for any t.

The spectrum of $y_s(t)$, that is, its Fourier transform, is $| S(j2\pi f) |^2$. This is easily seen from the fact that $y_s(t)$ may be looked on, in Fig. 7, as the impulse response of the cascade of the signal-generating filter and the matched filter. But the over-all transfer function of the cascade, which is therefore the spectrum of $y_s(t)$, is just $| S(j2\pi f) |^2$. The spectrum corresponding to the $y_s(t)$ in Fig. 8 will look something like Fig. 9.

In Figs. 8 and 9 we have denoted the "widths" of $y_s(t)$ and $| S(j2\pi f) |^2$ by α and β, respectively. It has been shown [13] that for suitable definitions of these widths, the inequality,

$$\alpha\beta \geq \text{a constant of the order of unity,} \tag{35}$$

holds. The exact value of the constant depends on the definition of "width" and need not concern us here. The important thing is that the "width" of the signal component at the matched-filter output in Fig. 7 cannot be less than the order of the reciprocal of the signal bandwidth.

In view of the above results, one might question the optimal character of a matched filter. In the preceding section we saw, for example, that in the radar case with gaussian noise we were to compare $y(t)$ with a threshold to determine whether a signal is present or absent (see Fig. 4). Now, if a signal component *is* present in $x(t)$, we seemingly should require, for the purpose of assuring

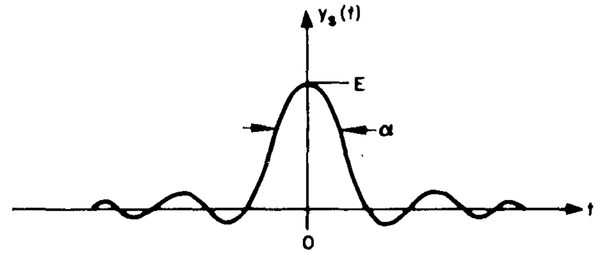

Fig. 8—The signal-component output of (32).

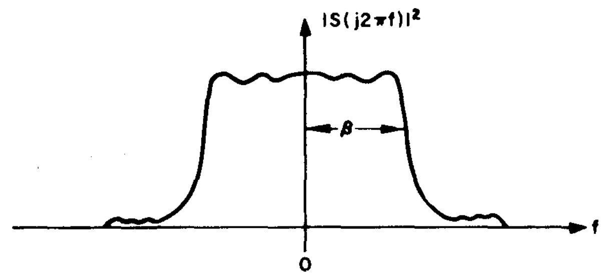

Fig. 9—The spectrum corresponding to $y_s(t)$ in Fig. 8.

that the threshold is exceeded, that the signal component of $y(t)$ be made as large as possible at some instant. Again, if $s(t)$ has been delayed by some unknown amount t_0, so that the signal component of $y(t)$ is also delayed, we saw that we could estimate t_0 by finding the location of the maximum of $y(t)$; in order to find this maximum accurately, one would imagine we would require that the signal component of $y(t)$ be a high, narrow pulse centered on t_0. In short, it appears that for efficient radar detection and ranging we really require an output signal component which looks very much like an *impulse*, rather than like the finite-height, nonzero-width, matched-filter output pulse of Fig. 8. Furthermore, we could actually obtain this impulse by using an inverse filter, with transfer function $1/S(j2\pi f)$, instead of the matched filter in Fig. 7. (To see this, note from Fig. 7 that the output signal component would then be the impulse response of the cascade of the signal-generating filter and the inverse filter, and since the product of the transfer functions of these two filters is identically unity, this impulse response would itself be an impulse.) We are thus led to ask, "Why not use an inverse filter rather than a matched filter?"

The answer is fairly obvious once one considers the effect of the additive noise, which we have so far explicitly neglected in this argument. Any physical signal must have a spectrum, $S(j2\pi f)$, which approaches zero for large values of f, as in Fig. 9. The gain of the inverse filter will therefore become indefinitely large as $f \to \infty$. Since the input noise spectrum is assumed to extend over all frequencies, the power in the output noise component of the inverse filter will be infinite. Indeed, the output noise component will override the output signal-component impulse, as may easily be seen by a comparison of their spectra: the former spectrum increases without limit as $f \to \infty$, while the latter is a constant for all f.

Thus, one must settle for a signal-component output pulse which is somewhat less sharp than the impulse delivered by an inverse filter. One might look at this as a compromise, a modification of the inverse filter which, while keeping the output signal component as close to an impulse as possible, efficiently suppresses the noise "outside" the signal band. In order to understand the nature of this modification, let us write the signal spectrum in the form

$$S(j2\pi f) = |S(j2\pi f)| e^{-i\psi(f)}, \tag{36}$$

where $\psi(f)$ is the phase spectrum of $s(t)$. Then the inverse filter has the form

$$\frac{1}{S(j2\pi f)} = \frac{1}{|S(j2\pi f)|} e^{i\psi(f)}. \tag{37}$$

Now, since the input noise at any frequency has random phase anyway, we clearly can achieve nothing in the way of noise suppression by modifying the phase characteristic of the inverse filter—we would only distort the output signal component. On the other hand, it seems reasonable to adopt as the amplitude characteristic of the filter, not the inverse characteristic $1/|S(j2\pi f)|$ of (37), but a characteristic which is small at frequencies where the signal is small compared to the noise, and large at the frequencies where the signal is large compared to the noise. In particular, a reasonable choice seems to be $|S(j2\pi f)|$.[13] Then the compromise filter is of the form

$$H(j2\pi f) = |S(j2\pi f)| e^{i\psi(f)} = S^*(j2\pi f), \tag{38}$$

a matched filter. [The last equality in (38) follows from (36) and the fact that, for a physical signal, $|S(j2\pi f)|$ is even in f.]

The compromise solution of (38), of course, as we have already seen, maximizes the height of the signal-component output pulse *with respect to the rms noise output*. This maximum output signal-to-noise ratio turns out to be perhaps the most important parameter in the calculation of the performance of systems using matched filters [20], [28], [33]–[35], [39], [52], [53], [57]; it is given by the right-hand side of (10):

$$\rho_0 = \frac{2E}{N_0}. \tag{39}$$

Note that ρ_0 depends on the signal *only through its energy*, E; such features of the signal as peak power, time duration, waveshape, and bandwidth do not directly enter the expression. In fact, insofar as one is considering *only* the ability of a radar detection system or an on-off communication system to combat white gaussian noise, it follows from this observation that all signals which have the same energy are equally effective.[14]

[13] This is indeed the solution Brennan obtains in deriving weights for optimal linear diversity combination [2]; here we have essentially a case of coherent frequency diversity.

[14] As we shall see, for more complicated communication systems in which more than one signal waveform may be transmitted, parameters governing the "similarity" of the signals enter the performance calculations [20], [52].

One may relate the output signal-to-noise ratio, ρ_0, to that at the input of the filter. Let the noise bandwidth of the matched filter—*i.e.*, the bandwidth of a rectangular-band filter, with the same maximum gain, which would have the same output noise power as the matched filter—be denoted by B_N. Then, for simplicity, one may think of the amount of input noise power within the matched filter "band" as being given by $N_{in} = B_N N_0$. Further let the average signal power at the filter input be $P_{in} = E/T$, where T is the effective duration of the signal. Then, letting $\rho_i = P_{in}/N_{in}$, (39) becomes

$$\rho_0 = 2B_N T \rho_i \,. \tag{40}$$

In this formulation of ρ_0, it is apparent that the matched filter effects a gain in signal-to-noise *power* ratio of $2B_N T$.

This last result seems at first to contradict our previous observation that, in the face of white noise, the signal bandwidth and time duration do not directly influence ρ_0. There is no contradiction, of course. For a given total signal energy, the larger T is, the smaller the input average signal power P_{in} is, and hence the smaller ρ_i is. Any increase in the ratio of ρ_0 to ρ_i caused by "spreading out" a fixed-energy signal is thus exactly offset by a decrease in ρ_i. Similarly, any increase in the ratio of ρ_0 to ρ_i occasioned by increasing the signal bandwidth is also offset by a decrease in ρ_i; for, the larger the signal bandwidth, the larger B_N, and hence the larger N_{in}—that is, the greater the amount of input noise taken in through the increased matched-filter bandwidth.

However, in relation to the foregoing argument let us suppose that one is combating *band-limited* white noise of *fixed* total power, rather than true white noise, which has infinite total power. That is, suppose the interference is such that its total power N_{in} is always caused, malevolently, to be spread out evenly over the signal "bandwidth" B_N, *whatever* this bandwidth is. Then, in (40), ρ_i is no longer dependent on B_N, and it is clearly advantageous to make the signal and matched-filter bandwidths as large as possible; the larger the bandwidth, the more thinly the total interfering power must be spread, *i.e.*, the smaller the value of N_0 in (39) becomes. In this situation, of all signals with the same energy, the one with the largest bandwidth is the most useful.

An interesting way of looking at the functions of the pair of filters, $S(j2\pi f)$ and $S^*(j2\pi f)$, in Fig. 7 is from the point of view of "coding" and "decoding." The impulse at the input to the signal-generating filter has components at all frequencies, but their amplitudes and phases are such that they add constructively at $t = 0$, and cancel each other out elsewhere. The transfer function $S(j2\pi f)$ "codes" the amplitudes and phases of these frequency components so that their sum becomes some arbitrary waveform lasting, say, from $t = 0$ to $t = T$, such as is shown in Fig. 10. Now, what we should like to do at the receiver is to "decode" the signal, *i.e.*, restore all the amplitudes and phases to their original values. We have seen that we cannot do this, since it would entail an inverse filter; we compromise by restoring the phases, so that all frequency components at the filter output have zero phase at the same time ($t = 0$ in Fig. 8) and add constructively to give a large pulse. This pulse has a nonzero width, of not less than the order of the reciprocal of the signal bandwidth, because we are not able to restore the amplitudes of the components properly.

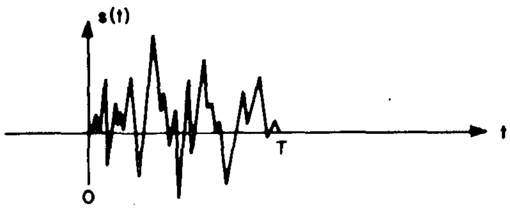

Fig. 10—A signal waveform.

In "coding," then, we have spread the signal energy out over a duration T; in "decoding," we are able to collapse this energy into a pulse of the order of βT times narrower, where β is some appropriate measure of the signal bandwidth. The "squashing" of the energy into a shorter pulse leads to an enhancement of signal-to-noise power ratio by a factor of the order of $B_N T$, already noted in connection with (40).

We thus see that the time-bandwidth product of teh signal or its matched filter, which we shall henceforth denote by TW, is a very important parameter in the description of the filter. It also turns out to be an index of the filter's complexity, as can be seen by the following argument. It is well known [43], that roughly $2TW$ independent numbers are sufficient (although not always necessary) to describe a signal which has an effective time duration T and an effective bandwidth W. It follows that the complete specification of a filter which is matched to such a signal requires, at least in theory, no more than $2TW$ numbers. Therefore, in synthesizing the filter there theoretically need be no more than $2TW$ elements or parameters specified, whence the use of the TW product as a measure of "complexity." (See footnote 20, however.)

Hopefully, this section has provided an intuitive insight into the nature of matched filters. It is now time to investigate the problems encountered in their synthesis.

V. Matched-Filter Synthesis and Signal Specification

In considering the synthesis of matched filters, one must take account of the edicts of two sets of constraints: those of physical realizability, and those of what might be called practical realizability. The first limit what one *could* do, at least in theory; the second are more realistic, for they recognize that what is theoretically possible is not always attainable in practice—they define the limits of what one *can* do.

The constraints of physical realizability are relatively easily given, and may be found in any good book on network synthesis [19]. Perhaps the most important for us, at least for electrical filters, is that expressed in footnote 1, that the impulse response must be zero for negative values of its argument.[15] If the signal to which the filter is to be matched "stops"—*i.e.*, falls to zero and remains there thereafter—at some finite time t_2 (see Fig. 1), then we have seen that by introducing a finite but perhaps large delay, $\Delta \geq t_2$, in the impulse response, we can render the impulse response physically realizable. We must of course then wait until $t = \Delta \geq t_2$ for the peak of the output signal pulse to occur; put another way, we cannot expect an output containing the full information about the signal at least until the signal has been fully received. Suppose, however, that t_2 is infinite, or it is finite but we cannot afford to wait until the signal is fully received before we extract information about it. Then it may easily be shown, for example, that in order to maximize the output signal-to-noise ratio at some instant $t < t_2$, we should use that part of the optimum impulse response which is realizable, and delete that part which is not [58]. Thus, if the signal of Fig. 1(a) is to be detected at $t = 0$, the *desired* impulse response is proportional to that of Fig. 1(b); the best we can do in rendering this physically realizable is to delete that part of it occurring prior to $\tau = 0$. The output signal-to-noise ratio at the instant of the output signal peak (in this case, $t = 0$) is still of the form of (39), but now E must be interpreted not as the total signal energy, but only as that part of the signal energy having arrived by the time of the output signal peak. Of course, we are no longer dealing with a true matched filter.

The constraints of practical realizability are not so easy to formulate: they are, rather, based on engineering experience and intuition. Let us henceforth assume that we are concerned with a true matched filter, *i.e.*, that the impulse response (1) is physically realizable. Even then, we are aware that the filter may not be practically realizable, because too many elements may be required to build it, or because excessively flawless elements would be needed, or because the filter would be too difficult to align or keep aligned, etc. In this light, the problem of realizing a matched filter changes from "Here is a desirable signal; match a filter to it" to "Here is a class of filters which I can satisfactorily build; which members of the class, if any, correspond to desirable signals?" In the former case we might here have neglected the question of how it was decided that the given signal is "desirable"; it would perhaps have sufficed merely to discuss how to realize the filter. But from the latter point of view the choice of a practical filter becomes inextricably interwoven with criteria governing the desirability of a signal, and one would therefore do well to design both the filter *and* the signal together to do the best over-all job. For this reason it is worthwhile to devote some time to answering the question, "What is a desirable signal?"

[15] Note that such a constraint is not necessary for optical filters [5], and filters—such as may be programmed on a computer—which use parametric rather than real time.

The answer, of course, depends on the application. Let us first consider the case of radar detection and ranging.

From the point of view of detection, we noted in the last section that in the face of white gaussian noise all signals with the same energy are equally effective, while if the interfering noise is band-limited and is constrained to have a given total power, then of all signals of the same energy those with the largest bandwidth are the most effective.

From the point of view of ranging there are at least three properties of the signal which we must consider: *accuracy*, *resolution*, and *ambiguity*. Let us first discuss these for the case of a low-pass signal. In this case, the matched-filter output has the form of the pulse in Fig. 8, but delayed by an unknown amount and immersed in noise. In order accurately to locate the position of the delayed central peak of the output signal pulse, we should like both to make ρ_0 of (39) large and also to make α, the width of this peak, small. This latter requirement implies making the bandwidth large [see (35)], whether the noise is truly white, or is of the band-limited, fixed-power variety. Further, if several targets are present, so that the signal component of the matched filter output consists of several pulses of the form of Fig. 8, delayed by various amounts (see Fig 11), then making α very small will allow closely adjacent targets to be resolved. That is, if two targets have nearly the same delay, as, for example, targets 3 and 4 in Fig. 11, making α small enough will lead to the appearance of two distinct peaks in the matched-filter output, rather than a broad hump.

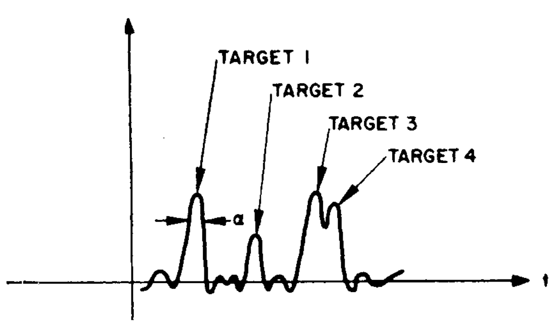

Fig. 11—The signal-component output of a matched filter for a resolvable multitarget or multipath situation.

The requirements of accuracy and resolution thus dictate a large-bandwidth signal. A more stringent requirement is that of lack of ambiguity. To explain this, let us again consider Fig. 8, the output signal component in response to a single target. As shown, there is but one peak, so even if the peak is shifted by an unknown delay there is hope that, despite noise, the amount of the delay can be determined unambiguously. However, suppose now that the waveform of Fig. 8 were to have many peaks of equal height, as would happen if the signal $s(t)$ were periodic [see (32)]. Then even without noise, it might not be totally clear which peak represented the unknown target delay; that is, there would be ambiguity in the target range, a familiar enough phenomenon in, say, periodically pulsed radars.

For the purposes of radar ranging then, what we require in the low-pass case is a large bandwidth signal, $s(t)$, for which, from (32),

$$y_s(t) = \int_{-\infty}^{\infty} s(\tau)s(t+\tau)\,d\tau \tag{41}$$

has a narrow central peak, and is as close to zero as possible everywhere else. The right-hand side of (41) may also be written as the real part of

$$2\int_0^{\infty} |\,S(j2\pi f)\,|^2\, e^{j2\pi f t}\,df, \tag{42}$$

where $S(j2\pi f)$ is, as usual, the spectrum of $s(t)$. We see, therefore, that we are here concerned with a shaping of the energy density spectrum of the signal.

In the case of a band-pass signal with random phase, all we have said still holds, but now not with respect to $y_s(t)$, but with respect to its envelope; $y_s(t)$ itself now has a fine oscillatory structure at the carrier frequency. This envelope is expressible as the magnitude of (42) [57]. In other words, we now want the *envelope* of the matched-filter output, in the absence of noise, to have a narrow central peak and be as close to zero as possible elsewhere. As before, the minimum width attainable by the central peak is of the order of the reciprocal of the signal bandwidth.

We have thus far in the present paper neglected the possibility of doppler shift in our discussion of radar systems. It is worthwhile to insert a few words on this topic now; we shall limit ourselves, however, to consideration of narrow-band band-pass signals, which are the only ones for which we can meaningfully speak of a doppler "*shift*" of frequency. Suppose, in addition to being delayed, the target return signal may also be shifted in frequency. It turns out then that when additive, white, gaussian noise is present, the ideal receiver should contain a parallel bank of matched filters much like that in the communication receiver of Fig. 5. Each filter is matched to a frequency-shifted version of the transmitted signal, there being a filter for each possible doppler shift. (If there is a continuum of possible doppler shifts, we shall see that the required "continuum" of filters is well approximated by a finite set, in which the frequency shifts are evenly spaced by amounts of the order of $1/T$, the reciprocal of the duration of the transmitted signal.) The (narrow-band) outputs of the bank of matched filters are then all envelope detected and subsequently passed into a device which decides, according to the edicts of hypothesis-testing and estimation theories, whether or not a target is present, and if present, at what delay (range) and doppler shift (velocity).

Now, as in the case of no doppler shift, the detectability of the signal is still governed by (39) and (40). But a modification is required in our previous discussion of the demands of high accuracy, high resolution, and low ambiguity. Generalizing this discussion to the case of nonzero doppler shift, it is clear that we require the following: each target represented at the receiver input should excite *only* the filter in the matched-filter bank which corresponds to the target doppler shift (velocity), and, further, should cause a sharp peak to appear in this filter's output envelope *only* at a time corresponding to the delay of the target, and nowhere else. In order to state this requirement mathematically, let us compute the output envelope at time t of a filter matched to a target return with doppler shift ϕ, in response to a target return with zero doppler shift and zero delay. (We lose no generality in assuming these particular target parameters.) We first note that if the (double-sided) spectrum of the transmitted signal is $S(j2\pi f)$, then for the narrow-band band-pass signal we are considering, the spectrum of the signal after undergoing a doppler shift ϕ is approximately $S[j2\pi(f - \phi)]$ *in the positive-frequency region.*[16] A filter matched to this shifted signal therefore has the approximate transfer function $S^*[j2\pi(f - \phi)]$, again in the positive-frequency region. Then the spectrum of the response of this filter to a non-doppler-shifted, nondelayed signal is approximately $S(j2\pi f)\ S^*[j2\pi(f - \phi)]\ (f > 0)$; the response itself, at time t, is the real part of the complex Fourier transform

$$\chi(t, \phi) = 2 \int_0^\infty S(j2\pi f) S^*[j2\pi(f - \phi)] e^{j2\pi ft}\, df. \tag{43}$$

The envelope of the response is just $|\chi(t, \phi)|$, and it is this which we require to be large at $t = 0$ if $\phi = 0$, and small otherwise. That is, we require that $|\chi(t, \phi)|$ have the general shape shown in Fig. 12: a sharp central peak at the origin, and small values elsewhere [23], [44], [57].

Note that (42) is a special case of (43) for $\phi = 0$; hence, the magnitude of (42) corresponds to the intersection of the $|\chi(t, \phi)|$ surface in Fig. 12 with the plane $\phi = 0$. It follows therefore from our discussion of (42) that the central peak in Fig. 12 cannot be narrower than the order of $1/W$ in the t direction. The use of an uncertainty relation of the form of (35) similarly reveals that the "width" of the peak cannot be less than the order of $1/T$ in the ϕ direction, where T is the effective signal duration [57]. (From this is implied a previous statement, that when a continuum of doppler shifts is possible, a good approximation to the ideal receiver involves the use of matched filters spaced apart in frequency by $1/T$; for, a target return which has doppler shift ϕ will cause responses in filters matched to signals with doppler shifts in an interval at least $1/T$ cps wide centered on ϕ.) More generally, one may show that the cross-sectional "area" of the central peak of the surface in Fig. 12—*i.e.*, the size of the (t, ϕ) region over which the peak has appreciable height—cannot be less than the order of $1/TW$ [44]. Thus, in order to attain a very sharp central peak it is

[16] In the negative-frequency region, the shifted spectrum has the approximate form $S[j2\pi(f + \phi)]$.

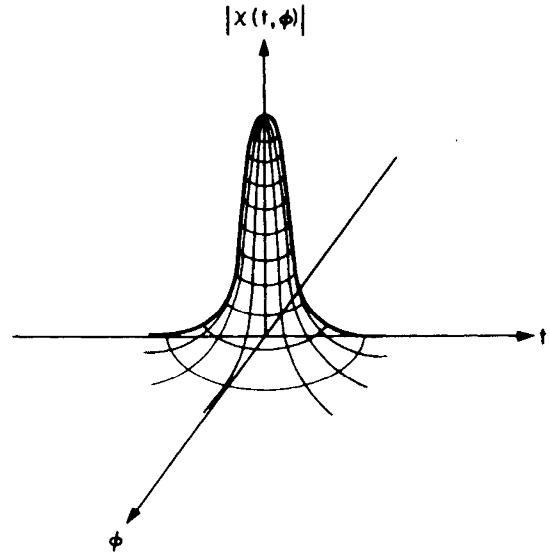

Fig. 12—A desirable $|\chi(t, \phi)|$ function.

necessary—but not sufficient, as we shall see—to make the TW product of the signal very large. Elimination of spurious peaks in $|\chi(t, \phi)|$ away from the (t, ϕ) origin is much more complicated, and has been studied elsewhere [23], [45], [57].

So much for radar signals. The requirements on communication signals are somewhat different, in some ways laxer and in others more stringent. In the simplest case, on-off communication, in which we are concerned with only one signal, the situation is obviously much like radar, except that in general one need not worry about doppler shifts of the transmitted signal. In particular, for gaussian noise, the detectability of the signal, and hence the probability of error, will in this case depend solely on ρ_0 of (39) and (40).

It should be expressly noted, however, that the problems of ranging which we encountered in the radar case are not entirely absent in the communication case. For example, in on-off communication, in order to sample the output of the matched filter at the time its signal component passes through its maximum, we must know when this maximum occurs. In many transmission media, however, the transmitted signal is randomly delayed, thus necessitating a ranging operation, in communication parlance called synchronization. We are not interested in this synchronization time *per se*, as in radar ranging, and are therefore not necessarily interested in eliminating large ambiguous peaks in $y_s(t)$ of (41). If such subsidiary peaks are of the same height as the central one, as will occur if the signal is periodic, we will be just as happy to sample one of them, rather than the central peak. On the other hand, we should not like to synchronize on and sample a peak appreciably smaller than the central peak, for then we would lose in signal-to-noise ratio. Thus, roughly, we require that a spurious peak in $y_s(t)$ be either very large or nonexistent.

We may no longer even require the peaks in $y_s(t)$ to be narrow, for we are not interested in determining the exact location of the peak, but only in sampling the matched-filter output at or near this peak. Clearly, an error in the sampling instant will be less disastrous if the peak is broad than if it is narrow. On the other hand, a narrow peak *is* often desirable in a multipath situation. There are often several independent paths between the transmitter and receiver, which represent independent sources of information about the transmitted signal. It behooves us to keep these sources separate, *i.e.*, to be able to resolve one path from another: this is the same as the radar resolvability requirement, illustrated in Fig. 11. From the frequency-domain viewpoint, requiring the matched-filter output pulse width to be small enough to resolve the various paths is the same as requiring the signal bandwidth to be large enough to avoid nonselective fading of the whole frequency band of the signal.

Such are the considerations which must be given to the choice of a signal for an on-off communication system, or to each signal individually in multisignal systems. However, in systems of the latter type, such as in Fig. 5, one must also consider the relationships *between* signals. In the system of Fig. 5, for example, it is not enough to specify that each signal individually have high energy and excite an output in its associated matched filter which consists, say, of a single narrow pulse at $t = 0$; if this were sufficient, we could choose all the signals to be identical, patently a ridiculous choice. We must also require that the various signals be *distinguishable*. More precisely, if the signal component of $x(t)$ in Fig. 5 is, say, $s_i(t)$, then the signal component at the output of the ith filter should be as large as possible at $t = 0$, *and* at the same instant the signal components of the outputs of all other filters should be "as much different" from the ith filter signal output as possible.

The phrase "as much different" needs defining: one wishes to specify M signals so that the over-all probability of error in reception is minimized. Unfortunately, this problem has not been solved in general for the phase-coherent receiver of Fig. 5. For the special case of binary transmission ($M \doteq 2$), however, it turns out that if the two signals are *a priori* equally probable, one should use equal-energy antipodal signals, *i.e.*, $s_1(t) = -s_2(t)$ [20], [52]. For then, on reception of the ith signal with zero delay, the output signal component of its associated matched filter at $t = 0$ is, from (32),

$$y_s(0) = \int_{-\infty}^{\infty} s_i^2(\tau)\, d\tau = E \qquad (i = 1, 2), \tag{44}$$

where E is the signal energy. The kth filter signal output at $t = 0$ is clearly

$$\int_{-\infty}^{\infty} s_k(-\tau) s_i(-\tau)\, d\tau = -E \qquad (i \neq k), \tag{45}$$

which is easily shown to be "as much different" from (44) as possible.[17]

If one considers the band-pass case in which the carrier phases are unknown, we have seen that the optimum system in Fig. 5 is modified by the insertion of envelope detectors between the matched filters and samplers. In this case a reasonable conjecture, which has been established for the binary case [20], [52], is that the M signals be "envelope-orthogonal." That is, if the ith signal is received, the envelopes of the signal-component outputs of all but the ith filter should be zero at the instant $t = 0$. Now, if the spectrum of the ith signal is denoted by $S_i(j2\pi f)$, then the envelope of the signal-component output of the kth filter is:

$$2\left| \int_0^{\infty} S_k^*(j2\pi f) S_i(j2\pi f) e^{j2\pi f t}\, df \right|. \tag{46}$$

This, evaluated at $t = 0$, must therefore be zero for all $k \neq i$. There are several ways of assuring that this be so, the most obvious being the use of signals with non-overlapping bands, so that $S_k^*(j2\pi f)\ S_i(j2\pi f) = 0$.[18] Another way involves the use of signals which are rectangular bursts of sine waves, the sine-wave frequencies of the different signals being spaced apart by integral multiples of $1/T$ cps, where T is the duration of the bursts. A third method is considered elsewhere in this issue [49].

If random multipath propagation is involved and the modulation delays of the paths are known, then we have noted that the samplers in Fig. 5 should in general sample both the matched-filter outputs and their envelopes at several instants, corresponding to the various path delays. We may in this case conjecture, again for the situation of unknown path phase-shifts (*i.e.*, only envelope sampling), that for an optimum set of signals, (46) should vanish for $k \neq i$, but now at values of t corresponding to all path delay *differences*, including zero. For, suppose the ith signal is sent. Then the spectrum

$$\sum_{l=1}^{L} S_i(j2\pi f) e^{-j2\pi f t_l} e^{-j\theta_l} \tag{47}$$

is received, where t_l and θ_l are, respectively, the modulation delay and carrier phase-shift of the lth path [50].

[17] Since the writing of this paper, Dr. A. V. Balakrishnan has informed the author that he has proved the following long-standing conjecture concerning the general case of M equiprobable signals. If the dimensionality of the signal space (roughly $2TW$) is at least $M - 1$, then the signals, envisaged as points in signal space, should be placed at the vertices of an $(M - 1)$-dimensional regular simplex (*i.e.*, a polyhedron, each vertex of which is equally distant from every other vertex); in this situation, (44) holds for all i, and the right-hand side of (45) becomes $-E/(M - 1)$. The problem of a signal space of smaller dimensionality than $M - 1$ has not been solved in general.

[18] If the signals $s_i(t)$ are of finite duration, then their spectra extend over all f, and nonoverlapping bands are therefore not possible. An approximation is achieved by spacing the band centers by amounts large compared to the bandwidths.

The envelope of the output signal component of the kth matched filter is then

$$\left| 2 \sum_{l=1}^{L} e^{-i\theta l} \int_0^\infty S_k^*(j2\pi f) S_i(j2\pi f) e^{j2\pi f(t-t_l)} \, df \right|. \quad (48)$$

Since the output envelopes of the matched filters are to be sampled at $t = t_r$ $(r = 1, \cdots, L)$, it seems reasonable to require that in the absence of noise these samples should all be zero for $k \neq i$. That is, (48) should be zero for all $k \neq i$ at all $t = t_r$. For lack of knowledge of the θ_L's, a sufficient condition for this to occur is that (46) be zero at all $t = t_r - t_l$.

An extension of this argument to the case of unknown modulation delays considered in connection with Fig. 6 leads similarly to a requirement that, for all $k \neq i$, (46) vanish over the whole interval $-A \leq t \leq A$, where $A = t_b - t_a$, t_a and t_b being the parameters referred to in Fig. 6. This may not be possible with physical signals, and some approximation must then be sought. Here is another unsolved problem.

A final desirable property of both radar and communication signals which is worth mentioning is one arising from the use of a peak-power limited transmitter. For such a transmitter, operation at rated average power often points to the use of a constant-amplitude signal, *i.e.*, one in which only the phase is modulated. A signal of this sort is also demanded by certain microwave devices.

Having thus closed parentheses on a rather lengthy detour into the problem of signal specification, let us recall the question on which we opened them. We had decided that the constraints of practical realizability had limited us to the consideration of filters which can, in fact, be built. Looking at any particular class of such filters—for example, those with less than 1000 lumped elements, with coils having Q's less than 200—we ask, "Which members of this class are matched to desirable signals?" We have gotten some idea of what constitutes a desirable signal. Let us now briefly examine a few proposed classes of filters and see to what extent this question has been answered for these classes.

VI. Some Forms of Matched Filters

It is not intended to give herein an exhaustive treatment of all solutions obtained to the problem of matched-filter realization; indeed, such a treatment would be neither possible nor desirable. We shall, rather, concentrate on three classes of solutions which seem to have attracted the greatest attention. Even in consideration of these we shall be brief, for details are adequately given elsewhere, in many cases in this issue.

A. Tapped-Delay-Line Filters

Let us first consider matched filters for the class of signals generatable as the impulse response of a filter of the form of Fig. 13. The spectrum of a signal of this class has the form

$$S(j2\pi f) = F(j2\pi f) \sum_{i=0}^{n} G_i(j2\pi f) e^{-j2\pi f \Delta_i}, \quad (49)$$

which is, of course, the transfer function of the filter of Fig. 13. It is immediately clear that a filter matched to this signal may be constructed in the form of Fig. 14, for the transfer function of the filter shown there is

$$\begin{aligned} H(j2\pi f) &= F^*(j2\pi f) \sum_{i=0}^{n} G_i^*(j2\pi f) e^{-j2\pi f(\Delta_n - \Delta_i)} \\ &= S^*(j2\pi f) e^{-j2\pi f \Delta_n}. \end{aligned} \quad (50)$$

The filters of Figs. 13 and 14 are thus candidates for the filter pair appearing in Fig. 7.

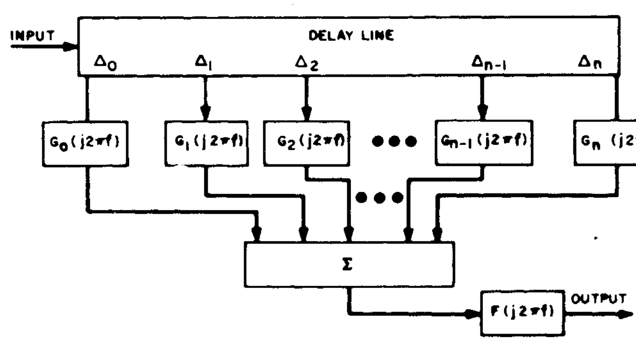

Fig. 13—A tapped-delay-line signal generator.

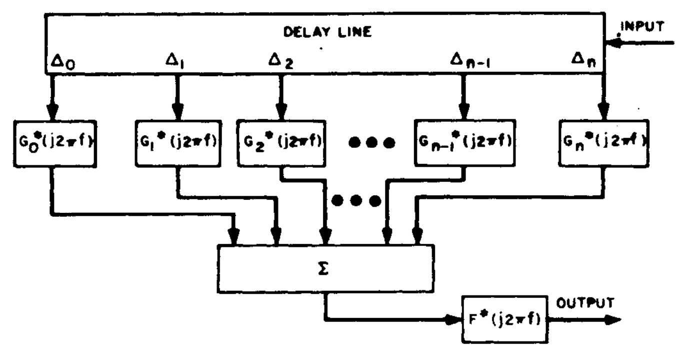

Fig. 14—A tapped-delay-line matched filter.

If $F(j2\pi f)$ and $G_i(j2\pi f)$ are assigned phase functions which are uniformly zero, then $F^*(j2\pi f) = F(j2\pi f)$ and $G_i^*(j2\pi f) = G_i(j2\pi f)$, and the two filters of Figs. 13 and 14 become identical except for the end of the delay line which is taken as the input. [The same identity of Figs. 13 and 14 is obviously also obtained, except for an unimportant discrepancy in delay, if $F(j2\pi f)$ and $G_i(j2\pi f)$ have linear phase functions, the slopes of all the latter being equal.] The advantages of having a single filter which can perform the tasks both of signal generation and signal processing are obvious, especially in situations such as radar, where the transmitter and receiver are physically at the same location.

Having defined a generic form of matched filter, we are still left with the problem of adjusting its characteristics [$F(j2\pi f)$, $G_i(j2\pi f)$ and Δ_i, all i] to correspond to a desirable signal. A possibility which immediately comes to mind is to set

$$\left.\begin{aligned} F(j2\pi f) &= \begin{cases} \dfrac{1}{2W}, & |f| \leq W \\ 0, & |f| > W \end{cases} \\ G_i(j2\pi f) &= a_i \\ \Delta_i &= \frac{i}{2W} \end{aligned}\right\}. \tag{51}$$

For, the signal which corresponds to such a choice has, from (49), the spectrum

$$S(j2\pi f) = \begin{cases} \dfrac{1}{2W} \displaystyle\sum_{i=0}^{n} a_i e^{-j2\pi f(i/2W)}, & |f| \leq W, \\ 0, & |f| > W, \end{cases} \tag{52}$$

and the signal itself therefore has the form

$$s(t) = \sum_{i=0}^{n} a_i \frac{\sin \pi(2Wt - i)}{\pi(2Wt - i)}. \tag{53}$$

It is well known [43] that *any* signal limited to the band $|f| \leq W$ can be represented in a form similar to (53), but with the summation running over all values of i; the a_i's are in fact just the values of $s(t)$ at $t = i/2W$. In (53) we therefore have a band-limited low-pass signal for which

$$s\left(\frac{i}{2W}\right) = \begin{cases} a_i, & i = 0, \cdots, n \\ 0, & \text{other integral values of } i \end{cases}. \tag{54}$$

Although $s(t)$ is not *uniformly* zero outside of $0 \leq t \leq n/2W$, it is seen from (53) and (54) that, at least for large n, the duration of $s(t)$ is effectively $T = n/2W$. The time-bandwidth product of the signal, which we have found to be a very important parameter, is thus approximately $TW = n/2$. Notice that this is proportional to the number of taps on the delay line and the number of multipliers, a_i, which in this case justifies the use of the TW product as a measure of the complexity of the filter.

It would seem that we have here, in one swoop, solved the problem of signal specification and matched-filter design, for we now have means available for obtaining both any desired band-limited signal and the filter matched to it.[19] There are two drawbacks which mar this hopeful outlook, however, one practical and the other theoretical. The first is that, at least at present, we do not know how to choose the a_i's so that $s(t)$ is desirable in the senses, say, of our discussions of (43) and (46). More basic is the fact that truly band-limited signals are not physically realizable; that is, the transfer function $F(j2\pi f)$ of (51) cannot be achieved, even theoretically. Using a realizable approximation to $F(j2\pi f)$ complicates the choice of the a_i's: for even if we were aware of how to choose the a_i's for the ideal $F(j2\pi f)$, it would not be clear how the use of an approximation would then affect the "desirability" of $s(t)$. This is not to say that the solution embodied in (51) should be discarded, but only that it must be further investigated to render it practically realizable.

Another possible choice of characteristics for the filters of Figs. 13 and 14 involves letting the pass bands of the filters $G_i(j2\pi f)$ be nonoverlapping, or essentially so [27]. [$F(j2\pi f)$ may here be considered to be unity, since limitation of the signal bandwidth is accomplished by the G_i's.] The purpose here is to afford independent control of various frequency bands of $S(j2\pi f)$ [cf. (49)], to the end of satisfying whatever requirements have been placed on the signal spectrum by constraints on (43) and (46). If we are considering only one signal which is not subject to doppler shift, then we are concerned only with constraints on (42), a special case of (43). In this case, control of various frequency ranges of $S(j2\pi f)$ is a direct method of achieving the energy-spectrum shaping mentioned in connection with (42); further, the use of a long enough delay line in conjunction with enough filters, G_i, will result in the desired large TW product.[20] For this special case, then, the use of a tapped delay line with nonoverlapping filters yields a possible desirable solution to our problem. More generally, however, when there are doppler shifts and/or many signals, we again must profess ignorance of how to select the G_i's and Δ_i's properly. Again further investigation is called for.

A third choice of characteristics for the filters of Figs. 13 and 14, *viz.*,

$$\left.\begin{aligned} F(j2\pi f) &= \frac{\sin \pi f \Delta}{\pi f} e^{-j\pi f \Delta} \\ G_i(j2\pi f) &= b_i = \pm 1 \\ \Delta_i &= i\,\Delta \end{aligned}\right\}, \tag{55}$$

has received considerable attention [12], [27], [44]. Note that the impulse response of $F(j2\pi f)$ is a rectangular pulse of unit height, and width Δ, starting at $t = 0$.[21] The impulse response of the filter of Fig. 13, *i.e.*, the matched signal, therefore has the form of Fig. 15—a low-pass sequence of positive and negative pulses, shown

[19] We have explicitly given only the low-pass case, in which, incidentally, the F and G's have the desirable zero phase functions. The band-pass case is obtainable by replacing the low-pass $F(j2\pi f)$ of (51) with its band-pass equivalent, letting the a_i's be complex (*i.e.*, contain phase shifts), and letting $\Delta_i = i/W$, where W is the band-pass band-width.

[20] If, in particular, the impulse responses of the filters G_i all have an effective duration of Δ, and the tap spacings in Fig. 13 are $\Delta_i - \Delta_{i-1} = \Delta$, then a "stepped-frequency" signal is obtained; that is, $s(t)$ is a continuous succession of nonoverlapping "pulses" of different frequencies. If, in addition, the bands of the G_i's are adjacent and have widths of the order of $1/\Delta$, the TW product for $s(t)$ and its matched filter is of the order of $n\Delta(n/\Delta) = n^2$. Here is a degenerate case in which we have generated a signal whose TW product is of the order of n^2 with a filter comprising a number of parameters proportional to n; that is, in this case TW is not a good measure of the filter's complexity, being much too large. The degeneracy involved here is not without its disastrous effects, however, as we shall see later when considering "chirp" signals.

[21] Needless to say, the $F(j2\pi f)$ can be eliminated in the signal-generator of Fig. 13, and the resulting filter driven by such a rectangular pulse, instead of an impulse. The $F(j2\pi f)$ is still required in the matched filter, however [40].

for $n = 10$. An equivalent band-pass case is obtained merely by using a band-pass equivalent of the $F(j2\pi f)$ in (55); the resulting signal is a sequence of sine-wave pulses of some frequency f_0, each pulse having either 0° or 180° phase. The search for a desirable signal is now simply the search for a desirable sequence, $b_0, b_1, \cdots, b_n$, of +1's and −1's.

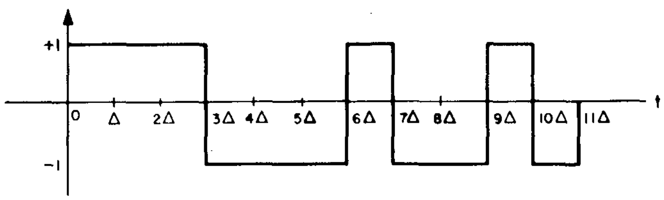

Fig. 15—A signal corresponding to (55).

Notice that $s(t)$ is a constant-amplitude signal, a property we have already seen to be important in some situations. For such signals, for the band-pass case implicit in (43), we easily may show that

$$|\chi(0, \phi)| = T\left|\frac{\sin \pi\phi T}{\pi\phi T}\right|, \tag{56}$$

where T is the signal duration, equal to $(n + 1)\ \Delta$ in Fig. 15. This is the cross section of the $|\chi(t, \phi)|$ surface in the ϕ direction; it has the shape shown in Fig. 16. In view of our discussion of the properties of $|\chi(t, \phi)|$, this shape seems reasonably acceptable: a central peak of the "minimum" possible "width," with no serious spurious peaks.

The cross section of $|\chi(t, \phi)|$ in the t direction—*i.e.*, $|\chi(t, 0)|$—can be made to have a desirable shape by proper choice of the sequence $b_0, b_1, \cdots, b_n$. Classes of sequences suitable from this point of view have, in fact, been found. The members of one such class [1], [47] have the desirable property that the central peak has a width of the order of Δ [the reciprocal of the "bandwidth" of $s(t)$] and is $(n + 1)$ times as high as any subsidiary peak. The sequence of Fig. 15, of length $n + 1 = 11$, is in fact a member of this class and has the $|\chi(t, 0)|$ shown in Fig. 17. Other members of the class with lengths 1, 2, 3, 4, 5, 7, and 13 have been found. Unfortunately, no odd-length sequences of length greater than 13 exist; it is a strong conjecture that no longer even-length sequences exist either.

If we for the moment redefine $|\chi(t, 0)|$ as the response envelope of the matched filter to the signal made periodic with period $(n + 1)\ \Delta$, we find that at least for the odd-length sequences we have just discussed, $|\chi(t, 0)|$ retains a desirable (albeit necessarily periodic) shape; for example, the new, "periodic" $|\chi(t, 0)|$ function for the sequence in Fig. 15 is shown in Fig. 18. This property is of interest in radar, where the signal is often repeated periodically or quasi-periodically.

Another class of binary sequences, $b_0, b_1, \cdots, b_n$, which have periodic $|\chi(t, 0)|$ functions of the desirable

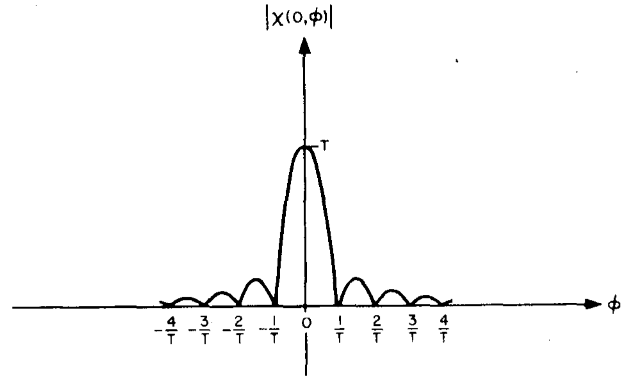

Fig. 16—The $|\chi(0, \phi)|$ function for the signal in Fig. 15.

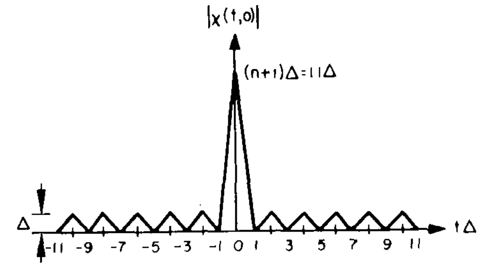

Fig. 17—The $|\chi(t, 0)|$ function for the signal in Fig. 15.

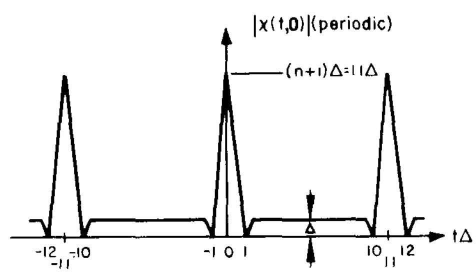

Fig. 18—The periodic $|\chi(t, 0)|$ function for the signal in Fig. 15.

form shown in Fig. 18 are the so-called linear maximal-length shift-register sequences. These have lengths $2^p - 1$, where p is an integer; the number of distinct sequences for any p is $\Phi(2^p - 1)/p$. where $\Phi(x)$, the Euler phi-function, is the number of integers less than x which are prime to x [59]. The class of linear maximal-length shift-register sequences has been studied in great detail [10], [16], [17], [42], [44], [46], [59], [61]. In general, their aperiodic $|\chi(t, 0)|$ functions leave something to be desired [26].

Other classes of binary sequences with two-level periodic $|\chi(t, 0)|$ functions such as shown in Fig. 18 have been investigated [18], [60].

So far we have discussed the desirability of certain binary sequences for use in (55) only from the point of view of the form of the $|\chi(t, \phi)|$ function along the coordinate axes. No general solution has been given for the form of the function off the axes, but certain conjectures can be made, which have been roughly corroborated by trying special cases. It is felt [44], at least for long linear shift-register sequences, that the central peak of $|\chi(t, \phi)|$ indeed has the minimum cross-sectional "area" of $1/TW$, and has subsidiary peaks in the plane of height no greater than the order of $1/\sqrt{n+1}$ of the height of the central peak [$(n+1)$ is the length of the sequence].

No general conclusions have been reached concerning the *joint* desirability, from the viewpoint of (46), of several of these binary codes.

The actual synthesis of matched filters with the characteristics of (55) is considered in detail elsewhere in this issue [24].

As a final example of a choice of characteristics of the filters of Figs. 13 and 14, let us consider

$$\left.\begin{aligned} F(j2\pi f) &= \frac{\sin \pi f a}{\pi f} e^{-i\pi f a} \\ G_i(j2\pi f) &= 1, \quad \text{all } i \\ \Delta_i &= i\,\Delta, \qquad \Delta > a \end{aligned}\right\}. \tag{57}$$

The impulse response of the signal-generating filter in this case is a sequence of $(n+1)$ rectangular pulses of width a, starting at times $t = i\Delta$. (Such a signal, of course, is more easily generated by other means.) The ambiguity function for the band-pass analog of this signal, which is often used in radar systems, is rather undesirable, having pronounced peaks periodically both in the t and ϕ directions [44], [57], but the signal has the virtue of simplicity. The filter which is matched to the signal is often called a pulse integrator, since its effect is to add up the received signal pulses; alternatively, it is called a comb filter, since its transmission function, $|S(j2\pi f)|^2$, has the form of a comb with roughly (Δ/a) teeth of "width" $1/(n+1)\,\Delta$, spaced by $1/\Delta$ cps [14], [54].

B. Cascaded All-Pass Filters

Another way to get a signal-generating and matched filter pair is through the use of cascaded elementary all-pass networks [27], [48]. An elementary all-pass network is defined here as one which has a pole-zero configuration like that in Fig. 19. Such a network clearly has a transfer function which has constant magnitude at all frequencies, f, on the imaginary axis.

Now, suppose that many (say, n) elementary all-pass networks are cascaded, as in Fig. 20, to form an over-all network with the uniform pole-zero pattern of Fig. 21. The over-all transfer function, $G(j2\pi f)$, will have a constant magnitude; its phase will be approximately linear over some bandwidth W, as shown by the upper curve of Fig. 22.[22] If some of the elementary networks of Fig. 20 are now grouped together into a network, A, with transfer function $G_A(j2\pi f)$, and the rest into a network, B, with transfer function $G_B(j2\pi f)$, both $G_A(j2\pi f)$ and $G_B(j2\pi f)$ still have constant magnitudes, but neither necessarily will have a linear phase function; however, the two phase

[22] We are again considering the "low-pass" case for convenience. The "band-pass" analog is obvious.

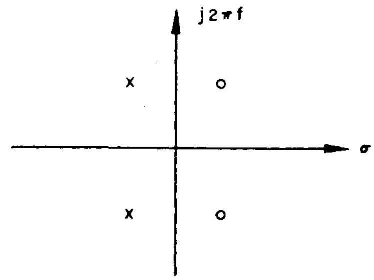

Fig. 19—The pole-zero configuration of an elementary all-pass network.

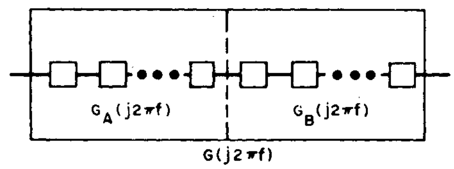

Fig. 20—A cascade of elementary all-pass networks.

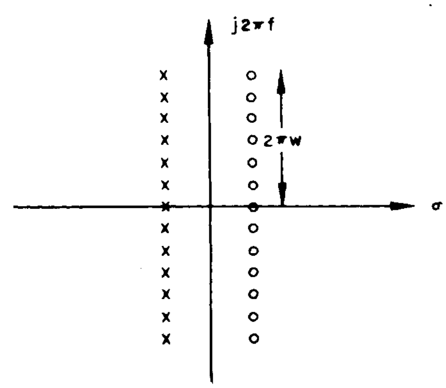

Fig. 21—The pole-zero configuration for the over-all network of Fig. 20.

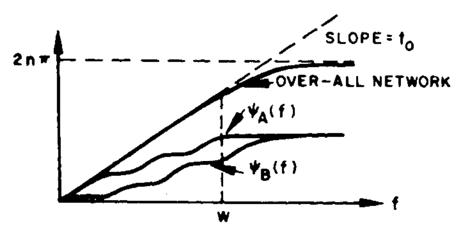

Fig. 22—Phase functions corresponding to the networks in Fig. 20.

functions must add to approximately a linear phase within the band $0 \leq f \leq W$ (see Fig. 22). *Within this band*, the relationship between the A and B transfer functions is

$$e^{-i\psi_A(f)}e^{-i\psi_B(f)} = e^{-i2\pi f t_0}, \tag{58}$$

where t_0 is the slope of the over-all phase function, and ψ_A and ψ_B are the phase functions of the A and B networks, respectively.

Suppose network A is driven by a pulse of bandwidth W, such as might be supplied by the pulse-forming network, $F(j2\pi f)$, of (57) ($a \approx 1/W$). Then the output signal of network A will have the spectrum $S(j2\pi f) = F(j2\pi f)e^{-i\psi_A(f)}$. From (58), therefore, over the bandwidth of $F(j2\pi f)$,

$$\begin{aligned} F^*(j2\pi f)e^{-i\psi_B(f)} &= F^*(j2\pi f)e^{i\psi_A(f)}e^{-j2\pi f t_0} \\ &= S^*(j2\pi f)e^{-j2\pi f t_0}. \end{aligned} \tag{59}$$

We hence have, at least approximately, a filter pair $F(j2\pi f)\, G_A(j2\pi f)$ and $F^*(j2\pi f)\, G_B(j2\pi f)$ which are matched.

The function of the filter $F(j2\pi f)$ is merely that of band limitation; therefore F should be made as simple as possible, preferably with zero or linear phase. The problems of choosing $G_A(j2\pi f)$ so as to obtain a desirable signal, and of actually synthesizing the filters, are considered elsewhere in this issue [48].

C. Chirp Filters

Let us now consider, as a final class of signals, signals of the form

$$s(t) = A(t) \cos(\psi_0 + 2\pi f_0 t + 2\pi k t^2), \quad 0 \leq t \leq T, \tag{60}$$

where $A(t)$ is some slowly varying envelope. Such signals, whose frequency—more correctly, whose phase derivative—varies linearly with time, have appropriately been called "chirp" signals.

In general, the spectrum of a chirp signal has a very complicated form. For the case of a constant-amplitude pulse, $A(t)$ = constant, an exact expression has been given [4]; this indicates, as one might expect, that in practical cases the spectrum has approximately constant amplitude and linearly increasing group time delay of slope k over the band $f_0 \leq f \leq f_0 + kT$, and is approximately zero elsewhere. The corresponding matched filter then has approximately constant amplitude and linearly *decreasing* group time delay of slope $-k$ over the same band.

Chirp signals and the synthesis of their associated matched filters have been studied in great detail [3], [4], [6], [8], [22], [32], [44], [57]. The chief virtue of this class of signals is their ease of generation: a simple frequency-modulated oscillator will do. Similarly, chirp matched filters are relatively easy to synthesize. On the other hand, from the point of view of the ambiguity function, $|\chi(t, \phi)|$—*i.e.*, of radars in which doppler shift is important—a chirp signal is rather undesirable. The area in the (t, ϕ) plane over which the ambiguity function for a chirp signal with a gaussian envelope is large is shown in Fig. 23 [44], [57]. Along the t and ϕ axes the "width" of the $|\chi(t, \phi)|$ surface is satisfactorily small: $1/W$ and $1/T$, respectively. But the surface, instead of being concentrated around the origin as desired, is spread out along the line $\phi = kt$; in fact, the area of concentration is very much greater than the minimum value, $1/TW$. This is equivalent to saying that it is very hard to determine whether the chirp signal has been given a time delay t or a frequency shift kt, a fact which is obvious from an inspection of the signal waveform.

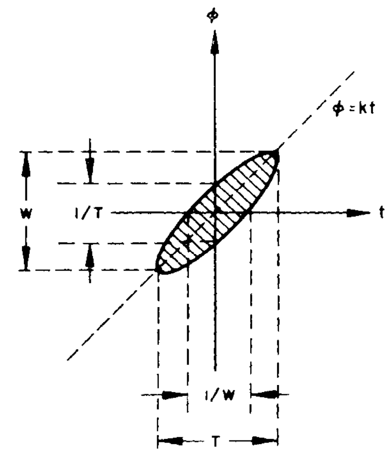

Fig. 23—The ambiguous "area" for a chirp signal.

VII. Conclusion

It must be admitted that, in regard to matched filters, the proverbial "state of the art" is somewhat less than satisfactory. We have seen that there are two basic problems to be solved simultaneously: the specification of a desirable signal or set of them, and the synthesis of their associated matched filters. In regard to the former, we have but the barest knowledge of the freedom with which we can constrain the "desirability" functions (43) and/or (46) and still expect physical signals; we know even less how to solve for the signals once both desirable and allowable constraints are set. It was partially for these reasons that we attacked a different problem: of a set of physical signals corresponding to a class of filters which we can build, which are the most desirable? In even this we were stopped, except for some special cases.

Nor can we assume that we are very sophisticated in the area of actually constructing filters which we say, on paper, we can build. It has become apparent that the TW product for a signal is a most important parameter; in general, the larger the product, the better the signal. But at the present time the synthesis of filters with TW products greater that several hundred may be deemed exceptional.

Clearly, much more effort is needed in this field. One hopes that the present issue of the Transactions will stimulate just such an effort.

VIII. Bibliography

[1] Barker, R. H., "Group synchronizing of binary digital systems," in "Communication Theory," W. Jackson, Ed., Academic Press, New York, N. Y.; 1953.

[2] Brennan, D. G., "On the maximum signal-to-noise ratio realizable from several noisy signals," Proc. IRE, vol. 43, p. 1530; October, 1955.

[3] Cauer, W., German Patent No. 892, 772; December 19, 1950.

[4] Cook, C. E., "Modification of pulse compression waveforms," *Proc. 1958 Natl. Electronics Conf.*, pp. 1058–1067.

[5] Cutrona, L. J., *et al.*, "Optical data processing and filtering systems," this issue, p. 386.

[6] Darlington, S., U. S. Patent No. 2, 678, 997; May 18, 1954.

[7] Davenport, W. B., Jr., and Root, W. L., "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill Book Co., Inc., New York, N. Y.; 1958.

[8] Dicke, R. H., U. S. Patent No. 2, 624, 876; January 6, 1953.

[9] Dwork, B. M., "Detection of a pulse superimposed on fluctuation noise," PROC. IRE, vol. 38, pp. 771–774; July, 1950.

[10] Elspas, B., "The theory of autonomous linear sequential networks," IRE TRANS. ON CIRCUIT THEORY, vol. CT-6, pp. 45–60; March, 1959.

[11] Fano, R. M., "Communication in the presence of additive Gaussian noise," in "Communication Theory," W. Jackson, Ed., Academic Press, New York, N. Y., 1953.

[12] Fano, R. M., "On Matched-Filter Detection in the Presence of Multipath Propagation," unpublished paper, M. I. T., Cambridge, Mass.; 1956.

[13] Gabor, D., "Theory of communication," *J. IEE*, vol. 93, pt. III, pp. 429–457; November, 1946.

[14] Galejs, J., "Enhancement of pulse train signals by comb filters," IRE TRANS. ON INFORMATION THEORY, vol. IT-4, pp. 114–125; September, 1958.

[15] George, T. S., "Fluctuations of ground clutter return in airborne radar equipment," *J. IEE*, vol. 99, pt. IV, pp. 92–99; April, 1952.

[16] Golomb, S. W., "Sequences with Randomness Properties," Glenn L. Martin Co., Baltimore, Md., Internal Report; June 14, 1955.

[17] Golomb, S. W., "Sequences with the Cycle-and-Add Property," Jet Propulsion Lab., C. I. T., Pasadena, Calif., Section Rept. 8–573; December 19, 1957.

[18] Golomb, S. W., and Welch, L. R., "Nonlinear Shift-Register Sequences," Jet Propulsion Lab., C. I. T., Pasadena, Calif., Memo. 20–149; October 25, 1957.

[19] Guillemin, E. A., "Synthesis of Passive Networks," John Wiley and Sons, Inc., New York, N. Y.; 1957.

[20] Helstrom, C. W., "The resolution of signals in white, gaussian noise," PROC. IRE, vol. 43, pp. 1111–1118; September, 1955.

[21] Kailath, T., "Correlation detection of signals perturbed by a random channel," this issue, p. 381.

[22] Krönert, R., "Impulsverdichtung," *Nachrichtentech.*, vol. 7, pp. 148–152 and 162, April, 1957; pp. 305–308, July, 1957.

[23] Lerner, R. M., "Signals with uniform ambiguity functions," 1958 IRE NATIONAL CONVENTION RECORD, pt. 4, pp. 27–36.

[24] Lerner, R. M., "A matched filter detection system for doppler-shifted signals," this issue, p. 373.

[25] Levin, M. J., "Optimum estimation of impulse response in the presence of noise," 1959 IRE NATIONAL CONVENTION RECORD, pt. 4, pp. 174–181.

[26] Lytle, D. W., "Experimental Study of Tapped Delay-Line Filters," Stanford Electronics Labs., Stanford Univ., Stanford, Calif., Tech. Rept. 361–3; July 30, 1956.

[27] Lytle, D. W., "On the Properties of Matched Filters," Stanford Electronics Labs., Stanord Univ., Stanford, Calif., Tech. Rept. 17; June 10, 1957.

[28] Marcum, J. I., "A Statistical Theory of Target Detection by Pulsed Radar," Rand Corporation, Santa Monica, Calif., Repts. RM-753 and RM-754; July, 1948, and December, 1947.

[29] Middleton, D., and Van Meter, D., "Detection and extraction of signals in noise from the point of view of statistical decision theory," *J. Soc. Indus. Appl. Math.*, vol. 3, pp. 192–253, December, 1955; vol. 4, pp. 86–119, June, 1956.

[30] Muller, F. A., "Communication in the Presence of Additive Gaussian Noise," Res. Lab. of Electronics, M. I. T., Cambridge, Mass., Tech. Rept. 244; May 27, 1953.

[31] North, D. O., "Analysis of the Factors which Determine Signal/Noise Discrimination in Radar," RCA Laboratories, Princeton, N. J., Rept. PTR-6C; June, 1943.

[32] O'Meara, T. R., "The Synthesis of 'Band-Pass', All-Pass Time-Delay Networks with Graphical Approximation Techniques, "Hughes Res. Labs, Culver City, Calif., Res. Rept. 114; June, 1959.

[33] Peterson, W. W., Birdsall, T. G., and Fox, W. C., "The theory of signal detectability," IRE TRANS. ON INFORMATION THEORY, PGIT-4, pp. 171–212; September, 1954.

[34] Pierce, J. N., "Theoretical diversity improvement in frequency-shift keying," PROC. IRE, vol. 46, pp. 903–910; May, 1958.

[35] Price, R., "Error Probabilities for the Ideal Detection of Signals Perturbed by Scatter and Noise," Lincoln Lab., M. I. T., Lexington, Mass., Group Rept. 34–40; October 3, 1955.

[36] Price, R., "Optimum detection of random signals in noise, with application to scatter-multipath communication," IRE TRANS. ON INFORMATION THEORY, vol. IT-2, pp. 125–135; December, 1956.

[37] Price, R., and Green, P. E., Jr., "A communication technique for multipath channels," PROC. IRE, vol. 46, pp. 555–570; March, 1958.

[38] Reich, E., and Swerling, P., "The detection of a sine wave in gaussian noise," *J. Appl. Phys.*, vol. 24, pp. 289–296; March, 1953.

[39] Reiger, S., "Error probabilities of binary data transmission in the presence of random noise," 1953 IRE CONVENTION RECORD, pt. 9, pp. 72–79.

[40] Rochefort, J. S., "Matched filters for detecting pulsed signals in noise," 1954 IRE CONVENTION RECORD, pt. 4, pp. 30–34.

[41] Root, W. L., and Pitcher, T. S., "Some remarks on statistical detection," IRE TRANS. ON INFORMATION THEORY, vol. IT-1, pp. 33–38; December, 1955.

[42] Scott, B. L., and Welch, L. R., "An Investigation of Iterative Boolean Sequences," Jet Propulsion Lab., C. I. T., Pasadena, Calif., Section Rept. 8–543; November 1, 1955.

[43] Shannon, C. E., 'Communication in the presence of noise," PROC. IRE, vol. 37, pp. 10–21; January, 1949.

[44] Siebert, W. McC., "A radar detection philosophy," IRE TRANS. ON INFORMATION THEORY, vol. IT-2, pp. 204–221; September, 1956.

[45] Siebert, W. McC., "Studies of Woodward's Uncertainty Function," Res. Lab. of Electronics, M. I. T., Cambridge, Mass., Quart. Prog. Rept.; April 15, 1958.

[46] Sloan, R. W., and Marsh, R. W., "The Structure of Irreducible Polynomials Mod 2 Under a Cubic Transformation"; July 7, 1953 (private communication).

[47] Storer, J. E. and Turyn, R., "Optimum finite code groups," PROC. IRE, vol. 46, p. 1649; September, 1958.

[48] Sussman, S., "A matched-filter communication system for multipath channels," this issue, p. 367.

[49] Titsworth, R. C., "Coherent detection by quasi-orthogonal square-wave pulse functions," this issue, p. 410.

[50] Turin, G. L., "Communication through noisy, random-multipath channels," 1956 IRE CONVENTION RECORD, pt. 4, pp. 154–166.

[51] Turin, G. L., "On the estimation in the presence of noise of the impulse response of a random, linear filter," IRE TRANS. ON INFORMATION THEORY, vol. IT-3, pp. 5–10; March, 1957.

[52] Turin, G. L., "Error probabilities for binary symmetric ideal reception through nonselective slow fading and noise," PROC. IRE, vol. 46, pp. 1603–1619; September, 1958.

[53] Turin, G. L., "Some computations of error rates for selectively fading multipath channels," *Proc. 1959 Natl. Electronics Conf.*, pp. 431–440.

[54] Urkowitz, H., "Analysis and synthesis of delay line periodic filters," IRE TRANS. ON CIRCUIT THEORY, vol. CT-4, pp. 41–53; June, 1957.

[55] Van Vleck, J. and Middleton, D., "A theoretical comparison of the visual, aural, and meter reception of pulsed signals in the presence of noise," *J. Appl. Phys.*, vol. 17, pp. 940–971; November, 1946.

[56] Woodward, P. M. and Davies, I. L., "Information theory and inverse probability in telecommunications," *Proc. IEE*, vol. 99, pt. III, pp. 37–44; March, 1952.

[57] Woodward, P. M., "Probability and Information Theory, with Applications to Radar," McGraw-Hill Book Co., Inc., New York, N. Y.; 1953.

[58] Zadeh, L. A. and Ragazzini, J. R., "Optimum filters for the detection of signals in noise," PROC. IRE, vol. 40, pp. 1123–1131; October, 1952.

[59] Zierler, N., "Several binary-sequence generators," *Proc. Am. Math. Soc.*, vol. 7, pp. 675–681; August, 1956.

[60] Zierler, N., "Legendre Sequences," Lincoln Lab., M. I. T., Lexington, Mass., Group Rept. 34–71; May 2, 1958.

[61] Zierler, N., "Linear recurring sequences," *J. Soc. Indus. Appl. Math.*, vol. 7, pp. 31–48; March, 1959.