

# Information Theory and Radar Waveform Design

Mark R. Bell

**Abstract**—The use of information theory to design waveforms for the measurement of extended radar targets exhibiting resonance phenomena is investigated. The target impulse response is introduced to model target scattering behavior. Two radar waveform design problems with constraints on waveform energy and duration are then solved. In the first, a deterministic target impulse response is used to design waveform/receiver-filter pairs for the optimal detection of extended targets in additive noise. In the second, a random target impulse response is used to design waveforms that maximize the mutual information between a target ensemble and the received signal in additive Gaussian noise. The two solutions are contrasted to show the difference between the characteristics of waveforms for extended target detection and information extraction. The optimal target detection solution places as much energy as possible in the largest target scattering mode under the imposed constraints on waveform duration and energy. The optimal information extraction solution distributes the energy among the target scattering modes in order to maximize the mutual information between the target ensemble and the received radar waveform.

**Index Terms**—Mutual information, radar, matched filter, radar waveforms, target impulse response, extended radar targets, target scattering resonance, ultra-wideband radar.

## I. INTRODUCTION

**S**HORTLY after the publication of Shannon's *A Mathematical Theory of Communication* [1] in 1948, Woodward and Davies began investigating the application of information theory to radar [2]–[5]. In 1953, Woodward published the book *Probability and Information Theory with Applications to Radar* [6], in which he presented both an introductory tutorial of information theory from the viewpoint of radar detection as well as a summary of results from his investigations with Davies. From these works, it is apparent that, at its inception, information theory was considered applicable to radar problems, particularly in the area of radar detection. The rationale behind these early investigations of the application of information theory to radar problems was summarized by Woodward [6, p. 62] as follows:

The problem of reception is to gain information from a mixture of signal and unwanted noise, and a considerable literature exists on the subject. Much of it has been concerned with methods of obtaining as large a signal-to-noise ratio as possible on the grounds that noise ultimately limits sensitivity and the less there is of it the better. This is a valid attitude as far as it goes,

but it does not face up to the problem of extracting information. Sometimes it can be misleading, for there is no general theorem that maximum output signal-to-noise ratio insures maximum gain of information.

In their work, Woodward and Davies use information-theoretic ideas to formulate the *a posteriori* radar receiver, which in the case of bandlimited additive white Gaussian noise results in the correlation receiver. They did not, however, extend the rationale noted above to the problem of radar waveform design for the purpose of extracting information about a target.

Since the time of Woodward's and Davies' investigations, a few researchers have considered the relationship between information theory and radar detection and estimation problems, but none of them have considered the use of information theory in radar waveform design. Frost and Shanmugan [7] examined the use of information theory in order to calculate the information content of synthetic aperture radar images. They then used the information content of synthetic aperture radar images generated using noncoherent integration with a variable number of looks in order to determine a measure of radiometric resolution. Wilcox [8] considered the problem of designing waveforms from the radar ambiguity function for narrowband signals. Naparst [9] recently investigated the problem of wideband waveform design and processing to resolve targets in dense target environments. None of these investigations, nor any others that could be found, addressed the issue of using mutual information in the design of radar waveforms and processing.

In this paper, we consider the problems of radar waveform design for optimal target detection (maximum output signal-to-noise ratio) and optimal target information extraction, when the radar targets are modeled as extended radar targets. Extended radar targets are targets of significant physical extent, so they no longer behave as simple point targets which scatter a scaled and attenuated version of the transmitted waveform back to the radar receiver, but instead they exhibit interference and resonance effects in the scattered electric field as a result of the target's physical extent.

The first problem, that of waveform design for the optimal detection of radar targets that exhibit resonance phenomena, involves the design of radar waveforms and receiver-filters that maximize the output signal-to-noise ratio at the receiver-filter output under constraints on transmitted waveform energy and duration. The second problem deals with the design of radar waveforms which maximize the mutual information between an ensemble of extended targets and the receiver-filter output. The waveforms of this second problem will be shown to be optimal, in a certain sense, for characterizing or identifying the target under observation. As we will see, information theory

Manuscript received June 14, 1990; revised December 28, 1992. This work was supported by Hughes Aircraft Company and by the National Science Foundation under Research Initiation Award MIP-9010834.

The author is with the School of Electrical Engineering, Purdue University, West Lafayette, IN 47907.

IEEE Log Number 9211657.

provides a unique insight into the relative characteristics of the two families of waveforms that arise in the solution of these two problems.

The waveforms solved for by maximizing the mutual information between the random target ensemble and the received radar signal will be called information extraction waveforms or estimation waveforms. We call these waveforms estimation waveforms because of the relationship between mutual information and the lower bound on parameter error as expressed through the rate distortion function. We do not explicitly discuss parameter estimation with these waveforms using standard parameter estimation techniques. Rather, the sense in which these are good estimation (or classification) waveforms is discussed in Section II-C.

In Section II, we formulate the problems of waveform design for target detection and target estimation. In formulating these waveform design problems, we will introduce the idea of *target impulse response* for both deterministic targets and random target ensembles. The target impulse response will allow us to represent the scattering characteristics of the radar target using an approach motivated by signal theory rather than electromagnetic scattering theory. We then formulate the two waveform design problems, considering the models that give rise to them and the rationale behind the optimality criteria used in formulating the problems.

In Section III, we summarize the major results obtained for both the detection waveform problem and the estimation waveform problem. In Theorem 1, we present an algorithm for the design of optimal waveform/receiver-filter pairs for detection of targets with known impulse response under constraints on waveform duration and energy. We also give the resulting signal-to-noise ratio obtained using the optimal waveform/receiver-filter pairs. In Theorem 2 we present the main result on estimation waveforms. Theorem 2 describes the spectral characteristics of waveforms that maximize the mutual information between a Gaussian target ensemble and the received radar waveform.

In Section IV, we solve the problem of finding the waveform/receiver-filter pair for optimal extended target detection and prove Theorem 1. In Section V, we consider the problem of designing optimal estimation waveforms and prove Theorem 2. In Section VI, we present examples of the design and performance of waveforms for both the optimal detection and estimation problems. We also compare the characteristics of the optimal detection and estimation waveforms. Finally, in Section VII, we summarize the results of our investigation.

## II. FORMULATION OF PROBLEMS

### A. Target Impulse Response

Radar targets are commonly modeled as *point targets*—targets of infinitesimal physical extent. The resulting simplification is that the reflected radar waveform observed at the receiver is an amplitude-scaled and time-delayed replica of the transmitted waveform. For narrow bandwidth waveforms, the point target model is often valid, but as the waveform bandwidth  $\Delta f$  becomes comparable to  $c/2\Delta z$ , where  $c$  is the

speed of light and  $\Delta z$  is the spatial extent of the radar target in range, the point-target model does not accurately reflect the behavior of radar scatterers. As  $\Delta z$  becomes comparable to  $c/2\Delta f$ , the return must instead be viewed as coming from several—or even a continuum—of points in an extended region of space. As a result, the received radar signal is the sum of multiple delayed versions of the transmitted waveform. Targets exhibiting such scattering behavior are called *extended targets*.

The propagation and scattering of electromagnetic waves, as they typically occur in radar measurement, are linear processes, obeying the principles of superposition and homogeneity. One common method of studying linear processes such as scattering is to view them as linear systems and to study the input/output relationships of their representative linear systems. In addition to being linear, the system may also be time-invariant, as would be true if the target was stationary with respect to the radar. The system impulse response is then a convenient tool for characterizing the input/output relationships of the system.

To apply linear systems analysis to scattering problems, we will first define the input and output quantities to be the electric field present at a pair of points in space. We will assume a fixed, although not necessarily identical, polarization at each point. The input  $e(t)$  is the  $\hat{x}$  polarized electric field at point  $P_1$ . We assume that the plane wave is propagating along the line connecting  $P_1$  and the origin. If the plane wave is incident on the target located at the origin, a scattered electric field will be present at an arbitrarily chosen observation point  $P_2$ . We select an arbitrary polarization  $\hat{x}'$  at  $P_2$  and view as the output of our linear system the electric field  $v(t)$  at point  $P_2$  with polarization  $\hat{x}'$ . Thus, restricting the direction and polarization of the incident plane wave and selecting a point  $P_2$  for measurement of the scattered wave for a fixed polarization, we have that the relationship between  $e(t)$  and  $v(t)$  is that of a linear system. We will also assume that the scatterer is stationary during the period of observation, and that the system relating  $e(t)$  and  $v(t)$  is a linear time-invariant system.

We will designate the impulse response of this system by  $h(t)$ . For general  $e(t)$ , the output  $v(t)$  of the linear system is given by the convolution integral

$$v(t) = \int_{-\infty}^{\infty} h(\tau)e(t-\tau) d\tau.$$

Let the Fourier transforms of  $e(t)$ ,  $v(t)$ , and  $h(t)$  be given by  $E(f)$ ,  $V(f)$ , and  $H(f)$ , respectively. Then

$$V(f) = E(f)H(f).$$

Although we have assumed that the target is stationary with respect to the radar, this approach can be generalized for a wide class of targets in motion as well. For a wide class of target motions with respect to the radar (e.g., radial target motion in a monostatic radar system), the received waveform  $v(t)$  for the target in motion is the same as that for the stationary target, except with a contraction or dilation of the time-axis induced by the Doppler effect. Hence, the analysis can be generalized for target motion in much the same way that the matched filter is generalized for target motion in a standard radar processor.

While a deterministic target impulse response may be useful in the case where the target of interest is known *a priori*, there are many instances where such *a priori* knowledge is not available. In such cases, it may be possible to approximate the target impulse response from known physical characteristics of the target under consideration, or else treat the target impulse response as a finite-energy random process. The former approach may be useful in the design of waveforms for optimal target detection, whereas the latter approach will be used in the problem of waveform design for target information extraction.

The random impulse response  $g(t)$  is a finite-energy random process used to model the scattering characteristics of a random target. The random process  $g(t)$  can be thought of as an ensemble  $\{g(t, \omega)\}$  of functions, where  $\omega \in \Omega$  and  $\Omega$  is the underlying sample space. We will now examine some properties of the random impulse response  $g(t)$ .

The first property which  $g(t)$  must possess is that all of the sample functions  $g(t, \omega)$  must satisfy

$$\int_{-\infty}^{\infty} |g(t, \omega)|^2 dt \leq 1.$$

This follows from conservation of energy and the fact that electromagnetic scattering is a passive process. The next property of  $g(t)$  we will assume is that all of its sample functions are causal impulse responses; that is,  $g(t, \omega) = 0, \forall t < 0, \forall \omega \in \Omega$ . This is a property of all *physical* linear time-invariant systems. In addition, we will also assume that the Fourier transform  $\{G(f, \omega)\}$  of each sample function  $\{g(t, \omega)\}$  exists. A sufficient condition for this is that each sample function  $\{g(t, \omega)\}$  have only finite discontinuities, along with the first condition on the sample functions given above [12]. With these conditions, the Fourier transform  $G(f)$  of  $g(t)$  exists.

Finally, we will assume that  $g(t)$  is a Gaussian random process. This is a reasonable assumption for targets consisting of a large number of scattering centers randomly distributed in space, since both the in-phase and quadrature components of the received signal in such cases are, at least approximately, Gaussian random processes.

Throughout the remainder of this paper, a deterministic target impulse response will be denoted  $h(t)$ , while a random target impulse response will be denoted by the random process  $g(t)$ .

### B. Optimal Detection Waveforms

When detecting radar targets, the presence or absence of a target is generally determined by a threshold test on the energy in the received signal. When a target is present, we expect that there will be greater energy in the received signal than when no target is present.

As is well known from standard detection theory results, in order to obtain the best target detection performance in a radar using either a Neyman-Pearson or Bayes decision rule, the signal-to-noise ratio at the output of the radar receiver should be made as large as possible [13], [14]. It follows then that for optimal detection, we should design our waveform and

receiver to make the signal-to-noise ratio as large as possible under the physical constraints placed on the waveform and receiver.

The problem of interest can be stated as follows. Given a target impulse response  $h(t)$  and stationary additive Gaussian noise  $n(t)$  with power spectral density  $S_{nn}(f)$ , find a transmitted waveform  $x(t)$  with total energy  $E_x$  and a receiver-filter impulse response  $r(t)$  such that the signal-to-noise ratio of the receiver output  $y(t)$  is maximized at time  $t_0$ . In addition, because real radar waveforms are of finite duration, restrict the waveform  $x(t)$  such that it is zero outside the interval  $[-T/2, T/2]$ .

The primary difference between this problem and a standard matched filtering problem [11] is the effect of the target in changing the shape of the transmitted waveform in accordance with its scattering characteristics. The optimal receiver-filter itself should be matched to the waveform scattered by the target, not the transmitted waveform itself. But this alone is not sufficient in order to achieve the maximum attainable signal-to-noise ratio under the transmitted waveform constraints. We must find the waveforms meeting the constraints that, when filtered with the proper receiver-filter matched to the waveform scattered by the target, will yield the greatest signal-to-noise ratio.

### C. Optimal Estimation Waveforms

A radar system may make measurements of a target in order to determine unknown characteristics of the target. Stated differently, we can say that a radar system may make measurements of a target in order to decrease the *a priori* uncertainty about the target. In the analysis of communication channels, information theory provides a method of quantifying the decrease in the *a priori* message uncertainty of the transmitted message by observing the channel output. Such an approach has successfully allowed for the information transmission capabilities of a communication channel to be determined. This being the case, let us examine the measurement process in general in light of information theory in order to determine the information transmitted to an observer by a measurement mechanism. These results then can be applied to the radar measurement problem.

Consider a measurement system in which we have an object to be measured, a measurement mechanism, and an observer. We assume that the random vector  $X$  consists of parameters characterizing the object we wish to measure and that a probabilistic model of the unknown parameters is meaningful. The measurement mechanism maps  $X$  into the random vector  $Y$ , and the observer observes  $Y$ . From this observation, the observer determines the desired description of  $X$ . The measurement mechanism is assumed to have inherent inaccuracies, so its measurements contain errors. This can be modeled by assuming that the measurement mechanism stochastically maps the random vector  $X \in R_X$  to the random vector  $Y \in R_Y$ . We will denote the mutual information between  $X$  and  $Y$  to  $I(X; Y)$ .

The mutual information  $I(X; Y)$  between two random vectors  $X$  and  $Y$  tells us the quantity of information observation

of  $Y$  provides about  $X$ ; that is,  $I(X; Y)$  is the amount of information that the measurement  $Y$  provides about the object parameter vector  $X$ . The greater this mutual information is, the greater the quantity of information describing the object we obtain from our measurement and the greater the reduction in the *a priori* uncertainty as a result of this measurement. Intuitively, we might expect that the greater the mutual information between a measurement and the quantity being measured, the more accurately we can classify or estimate the parameters characterizing the entity we are trying to measure. Berger alludes to this idea when he makes the statement: "Rate distortion theory provides knowledge about how the frequency of faulty categorization will vary with the number and quality of observations" [15, p. 9], since the rate distortion function relates the average distortion or error to the minimum mutual information required to achieve that error.

We shall justify the idea that the greater the mutual information between the parameter we are measuring and the measurement, the better our ability to classify or estimate the parameters describing the object, in two ways. We will determine the maximum number of equiprobable classes into which we can assign  $X$  by observation of  $Y$ ; then we relate  $I(X; Y)$  to the average measurement error through use of the *rate distortion function*. This is done in the following two propositions. We also cite a recent result relating to this problem [16].

Consider the problem of putting  $X$  into one of  $N$  equiprobable classes based on observation of  $Y$ . That is, assume that  $R_X$  has been partitioned into  $N$  equiprobable subsets, and we wish to assign  $X$  to its proper subset based on observing the  $Y$  generated by the measurement process.

**Proposition 1:** For any decision rule assigning  $X$  to a subset of a partition based on observation of  $Y$ , and for all possible equiprobable partitions of  $R_X$ , the maximum number of partitions  $N$  for which this can be done with an arbitrarily small probability of error is

$$N = \lfloor e^{I(X; Y)} \rfloor. \quad (1)$$

To see that this is true, we note that, given  $I(X; Y) = I_0$  nats, we can calculate the associated  $N$ , which we will designate  $N_0$ , as

$$N_0 = \lfloor e^{I_0} \rfloor.$$

Then

$$N_0 \leq e^{I_0} < N_0 + 1.$$

Since the logarithm is a monotonically increasing function of its argument for all positive real numbers,

$$\ln N_0 \leq I_0 < \ln(N_0 + 1).$$

By Shannon's Theorem for the noisy channel, it is not possible to classify  $X$  into one of  $N_0 + 1$  equiprobable classes. This cannot be done without the channel's transferring  $\ln(N_0 + 1)$  nats of information. But the measurement mechanism cannot possibly do so, because  $I_0 < \ln(N_0 + 1)$ .

We now consider the relationship between mutual information and measurement error. When examining the accuracy

of measurements, it is common to talk about the accuracy in terms of some error criterion, for example, mean-squared error of relative mean squared error. It would be useful if we could relate the mutual information  $I(X; Y)$  to the relevant measurement error criterion. Rate distortion theory provides a framework for doing this.

In making a measurement, we are trying to obtain a description of the object parameter vector  $X$  from the measurement vector  $Y$ . Because of inaccuracies in the measurement process, we cannot generally obtain  $X$  perfectly, so there is an error associated with a given parameter vector  $x$  and a given measurement vector  $y$ . Let us designate this error or *distortion* as  $d(x, y)$ . We assume that this distortion is a nonnegative function defined for all pairs of  $x \in R_X$  and  $y \in R_Y$ . The mean distortion  $\delta$  is the expectation of  $d(x, y)$ . Thus,

$$\delta = E\{d(x, y)\}.$$

The *rate distortion function*  $R(D)$  for a single measurement is defined as

$$R(D) = \min \{I(X; Y) : \delta \leq D\}. \quad (2)$$

The minimization is over all measurement mechanisms that satisfy the condition that the fidelity criterion  $\delta$  is less than or equal to  $D$ . The minimization may also be constrained to measurement mechanisms which satisfy a specified set of conditions (e.g., the condition that a single measurement can use at most  $E_0$  joules of energy). The rate distortion function  $R(D)$  gives the minimum possible rate at which information must be transferred by a measurement mechanism in order to have an average error or distortion  $\delta$  less than or equal to  $D$ .

It is well known that  $R(D)$  is a nonincreasing function of  $D$ . So the smaller the average error  $D$ , the larger is the minimum required information rate  $R(D)$  required of the measurement mechanism in order to achieve this average error  $D$ . We summarize these ideas in the following proposition.

**Proposition 2:** Let  $D$  be the largest allowable mean error between the object parameter vector  $X$  and the measurement vector  $Y$ . Then the minimum possible value of  $I(X; Y)$  for which  $D$  can be achieved is a nonincreasing function of  $D$ .

Intuitively, this makes sense. It says that if greater accuracy is required in the measurements, the measurement mechanism must provide more information about the object being measured.

In general, the greater the mutual information between the parameters we wish to measure and the measurements themselves, the more we can say about the object being measured. In the case where we examined the number of equiprobable classes to which we could assign  $X$  based on observation of  $Y$ , we saw that the larger  $I(X; Y)$ , the larger the number of classes. In the case of the rate distortion function, we saw that the more precise we wanted our measurements to be, the greater the minimum rate of information transfer by the measurement mechanism.

A recent result by Kanaya and Nakagawa [16] relates mutual information to Bayes risk in statistical decision problems in a mathematically rigorous manner. For a random parameter  $\theta$  taking on values from a finite parameter set, they define

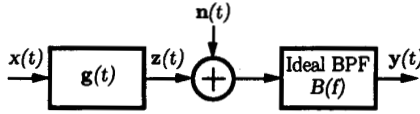


Fig. 1. Block diagram of the radar target channel.

a function  $R(P, L)$  that determines the minimum required mutual information between  $\theta$  and a measurement such that the Bayes risk is less than or equal to a value  $L$  given that the random parameter  $\theta$  has probability distribution  $P$ . This function is closely related to the rate distortion function. They then prove asymptotic results that show that for a sufficiently large number of independent experiments, the probability that the average Bayes loss is greater than  $L$  goes to zero as the number of experiments grows large, if the measurements associated with each experiment exceed  $R(P, L)$ . This result, perhaps of limited practical importance in the design of statistical decision procedures, does point directly to the fact that the greater the mutual information between the parameter and its measurement, the better the expected performance in the best Bayes risk decision procedure.

In applying Propositions 1 and 2 to the measurement mechanism of radar, we see that if we design radar systems in such a way as to maximize the mutual information between the target parameters of interest and their measurements, then the better we can expect our system performance to be, at least if we measure system performance in terms of target classification ability or average measurement error. This being the case, we will consider optimal information extraction waveforms to be those waveforms that maximize the mutual information between the observed target ensemble and the output of the radar receiver. It is this class of waveforms, under the imposed duration and energy constraints, that we are interested in finding.

Consider the radar target channel model shown in Fig. 1. Here,  $x(t)$ , a finite-energy deterministic waveform with energy  $E_x$  and of duration  $T$  is transmitted by the transmitter in order to make a measurement of the radar target. We will assume that  $x(t)$  is confined to the symmetric time interval  $[-T/2, T/2]$ . Thus,

$$E_x = \int_{-T/2}^{T/2} |x(t)|^2 dt. \quad (3)$$

Since the energy constraint in most real radar systems is not on the total energy in the transmitted waveform, but rather on the average power of the waveform, we will be interested in the average power  $P_x$ , which satisfies the relation  $E_x = P_x T$ . We also assume that  $x(t)$  is confined to a frequency interval  $\mathcal{W} = [f_0, f_0 + W]$ . While strictly speaking, we cannot have an  $x(t)$  with finite support whose Fourier transform has finite support, we assume that  $\mathcal{W}$  is selected so that only negligible energy resides outside the frequency interval  $\mathcal{W}$ .

After transmission, the radar waveform  $x(t)$  is scattered by the target, which has a scattering characteristic modeled by the random impulse response  $g(t)$ . The resulting scattered signal  $z(t)$  received at the receiver is a finite-energy random process,

and is given by the convolution integral

$$z(t) = \int_{-\infty}^{\infty} g(\tau)x(t-\tau) d\tau.$$

The random process  $z(t)$  is received at the receiver in the presence of the zero-mean additive Gaussian noise process  $n(t)$ . This noise process is assumed to be stationary and ergodic, and to have one-sided power spectral density  $P_{nn}(f) = 2S_{nn}(f)$  for  $f > 0$ . In addition,  $n(t)$  is assumed to be statistically independent of both the transmitted waveform  $x(t)$  and the target impulse response  $g(t)$ .

The waveform received at the receiver is shown in Fig. 1 to be  $z(t) + n(t)$  filtered by the ideal linear time-invariant bandpass filter  $B(f)$ , passing only frequencies in the band  $\mathcal{W}$ . The explicit inclusion of the filter  $B(f)$  is just a statement of the fact that we assume that the transmitted signal has no significant energy outside the frequency interval  $\mathcal{W}$ . Thus, neither does  $z(t)$ , since it is the response of a linear time-invariant system to the transmitted signal.

For a given sample function  $g(t, \omega_0)$  with the Fourier transform  $G(f, \omega_0)$ , the resulting spectrum of the scattered signal  $z(t)$  is given by  $Z(f, \omega_0) = X(f)G(f, \omega_0)$ . The magnitude squared of this spectrum is  $|Z(f, \omega_0)|^2 = |X(f)|^2 |G(f, \omega_0)|^2$ . Taking the expectation with respect to  $G(f)$ , the mean-square spectrum of  $z(t)$  is

$$E|Z(f)|^2 = |X(f)|^2 E\{|G(f)|^2\}.$$

Now,

$$E\{|G(f)|^2\} = |\mu_G(f)|^2 + \sigma_G^2(f),$$

where  $\mu_G(f)$  is the mean of  $G(f)$  and  $\sigma_G^2(f)$  is the variance of  $G(f)$ ; that is,

$$\mu_G(f) = E\{G(f)\},$$

and

$$\sigma_G^2(f) = E\{|G(f) - \mu_G(f)|^2\}.$$

We are interested primarily in  $\sigma_G^2(f)$  for the Gaussian target model, as the signal component of  $z(t)$  corresponding to the mean  $\mu_G(f)$  is known since  $x(t)$  is known. It thus tells us nothing about the target. In most cases,  $\mu_G(f) = 0$ , since there is a random delay  $d$  in  $g(t)$  because of the target's random position in space. This corresponds to a random phase factor of  $\exp\{-i2\pi f d\}$ , which has expectation zero for a wide class of distributions on  $d$ . We will thus assume that  $\mu_G(f) = 0$ .

Similarly, if we define

$$\mu_Z(f) = E\{Z(f)\}$$

and

$$\sigma_Z^2(f) = E\{|Z(f) - \mu_Z(f)|^2\},$$

then

$$E|Z(f)|^2 = |\mu_Z(f)|^2 + \sigma_Z^2(f).$$

Referring again to Fig. 1, we will assume that the radar receiver observes  $y(t)$  for a period  $\tilde{T}$  in order to obtain

information about the target. The duration of observation  $\tilde{T}$  must be long enough to allow the receiver to capture all but a negligible portion of the energy in the sum of the scattered signal  $z(t)$ . We know that the duration of the transmitted waveform is  $T$ , and we know that  $z(t)$  must be at least this long, since the convolution of two waveforms of finite duration  $T_1$  and  $T_2$  produces a waveform of duration  $T_1 + T_2$ . So if  $T_g$  is the duration of  $g(t)$ , then the duration of  $z(t)$  is  $T + T_g$ .

The received  $y(t)$  consists of the scattered signal  $z(t)$  and the additive Gaussian noise  $n(t)$  passed through the ideal bandpass filter  $B(f)$ , passing the frequency interval  $\mathcal{W}$ . The impulse response  $h_W(t)$  of this filter is

$$h_W(t) = W \frac{\sin \pi W t}{\pi W t} \cos(f_0 + W/2)t.$$

The duration of this pulse is infinite, but, as is well known, most of the energy is concentrated in an interval of duration  $1/W$ . Thus, it is reasonable to assume the impulse response duration  $T_W$  of the ideal passband filter at the receiver to be  $T_W \approx 1/W$ .

It is reasonable to assume that the bandwidth over which most radar targets exhibit significant scattering of electromagnetic waves is much larger than the bandwidth of the radar system used in order to make the scattering measurements [27, Section 27.6]. Hence it is reasonable to assume

$$T_g \ll T_W.$$

Also, for most radar signals, the duration of the transmitted signal is much larger than  $1/W$ . This is often necessary to provide enough signal energy for reliable target detection. An example of this is the linear FM or "chirp" signal commonly encountered in radar systems. This allows a range resolution equivalent to a much narrower pulse than that actually transmitted. Long transmission time, or "time-on-target" is also common in radar target recognition problems, where the longer observation time allows better frequency resolution in the measured Doppler spectrum. For such signals, the actual duration  $T$  of transmission is much larger than the  $1/W$ . For such systems,  $T \gg T_W \approx 1/W$ . So in summary,  $\tilde{T} = T + T_g + T_W \approx T + T_W \approx T + 1/W$ , and for systems that satisfy the condition  $T \gg 1/W$ ,  $\tilde{T} \approx T$ .

The problem of interest can now be stated as follows. Given a Gaussian target ensemble with random impulse response  $g(t)$  having spectral variance  $\sigma_G^2(f)$ , find the waveforms  $x(t)$  confined to the symmetric time interval  $[-T/2, T/2]$  and having all but a negligible fraction of their energy confined in (one-sided) frequency to  $\mathcal{W} = [f_0, f_0 + W]$  that maximize the mutual information  $I(y(t); g(t) | x(t))$  in additive Gaussian noise with one-sided power spectral density  $P_{nn}(f)$ .

### III. SUMMARY OF RESULTS

We now present the main results of our investigation as two theorems. Theorem 1 summarizes the main results on the design of waveform/receiver-filter pairs for optimal target detection. Theorem 2 summarizes the main results for the design of optimal estimation waveforms.

#### A. Results on Detection Waveforms

The main results arising from the solution of the waveform/receiver-filter design problem for optimal detection can be summarized in the following design algorithm, which explicitly states how to calculate the optimal waveform/receiver-filter pair, and gives an expression for the resulting signal-to-noise ratio.

**Theorem 1:** A waveform/receiver-filter pair maximizing the signal-to-noise ratio at the output of the receiver-filter can be designed using the following algorithm.

a) Compute

$$L(t) = \int_{-\infty}^{\infty} \frac{|H(f)|^2}{S_{nn}(f)} e^{i2\pi f t} df.$$

Here  $S_{nn}(f)$  is the two-sided power spectral density of the noise  $n(t)$ , and  $h(t)$  is the impulse response of the target.

b) Solve for an eigenfunction  $\hat{x}(t)$  corresponding to the maximum eigenvalue  $\lambda_{\max}$  of the integral equation

$$\lambda_{\max} \hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau) L(t - \tau) d\tau.$$

Scale  $\hat{x}(t)$  so that it has energy  $E_x$ .

c) Compute the spectrum  $\hat{X}(f)$  corresponding to the optimal waveform  $\hat{x}(t)$ :

$$\hat{X}(t) = \int_{-\infty}^{\infty} \hat{x}(t) e^{-i2\pi f t} dt.$$

d) Implement a receiver-filter of the form

$$R(f) = \frac{K \hat{X}(f) H(f) e^{-i2\pi f t_0}}{S_{nn}(f)},$$

where  $K$  is a complex constant.

e) The resulting signal-to-noise ratio for this design, which is the maximum obtainable under the specified constraints, is

$$\left( \frac{S}{N} \right)_{t_0} = \lambda_{\max} E_x.$$

The proof of this result is given in Section IV.

#### B. Results on Estimation Waveforms

The main results arising from the solution of the optimal estimation waveform design problem can be summarized in the following theorem.

**Theorem 2:** If  $x(t)$  is a finite-energy waveform with energy  $E_x$  confined to the symmetric time interval  $[-T/2, T/2]$ , and having all but a negligible fraction of its energy confined to the frequency interval  $\mathcal{W} = [f_0, f_0 + W]$ , the mutual information  $I(y(t); g(t) | x(t))$  between  $y(t)$  and  $g(t)$  in additive Gaussian noise with one-sided power spectral density  $P_{nn}(f)$  is maximized by an  $x(t)$  with a magnitude-squared

spectrum

$$\begin{aligned} |X(f)|^2 &= \max \left[ 0, A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right] \\ &= \max [0, A - r(f)], \end{aligned} \quad (4)$$

where  $r(f) = P_{nn}(f)\tilde{T}/2\sigma_G^2(f)$ , and  $A$  is found by solving the equation

$$E_x = \int_{\mathcal{W}} \max \left[ 0, A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right] df. \quad (5)$$

The resulting maximum value  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  of  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is

$$\begin{aligned} I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) \\ &= \tilde{T} \int_{\mathcal{W}} \max \left[ 0, \ln A - \ln \left( \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right) \right] df \\ &= \tilde{T} \int_{\mathcal{W}} \max [0, \ln A - \ln r(f)] df. \end{aligned} \quad (6)$$

The proof of this result is given in Section V.

#### IV. WAVEFORMS FOR DETECTION OF EXTENDED TARGETS

We now show that the design procedure of Theorem 1 gives a waveform/receiver-filter pair that produces the largest possible signal-to-noise ratio for an extended target. The receiver output is given by

$$\mathbf{y}(t) = \mathbf{y}_s(t) + \mathbf{y}_n(t).$$

Here,  $\mathbf{y}_s(t)$  is the signal component in the receiver output, and  $\mathbf{y}_n(t)$  is the noise component in the receiver output. These two components are given by

$$\begin{aligned} \mathbf{y}_s(t) &= \mathbf{r}(t) * \mathbf{x}(t) * \mathbf{h}(t) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{x}(\tau) \mathbf{h}(p - \tau) \mathbf{r}(t - p) d\tau dp, \end{aligned} \quad (7)$$

and

$$\mathbf{y}_n(t) = \mathbf{r}(t) * \mathbf{n}(t) = \int_{-\infty}^{\infty} \mathbf{n}(p) \mathbf{r}(t - p) dp. \quad (8)$$

We are interested in finding a waveform/receiver-filter pair that will maximize the output signal-to-noise ratio at time  $t_0$ . This signal-to-noise ratio is defined as

$$\left( \frac{S}{N} \right)_{t_0} \stackrel{\text{def}}{=} \frac{|y_s(t_0)|^2}{E|\mathbf{y}_n(t_0)|^2}. \quad (9)$$

Define the Fourier transforms  $X(f)$ ,  $H(f)$ ,  $V(f)$ , and  $R(f)$ , of  $x(t)$ ,  $h(t)$ ,  $v(t)$ , and  $r(t)$ , respectively. We then have the following lemma, the proof of which is quite standard.

**Lemma 1:** A necessary condition for the maximization of the signal-to-noise ratio is

$$R(f) = \frac{K \overline{X(f)H(f)} e^{-i2\pi f t_0}}{S_{nn}(f)}, \quad (10)$$

and with  $R(f)$  satisfying the relation of (10), the resulting signal-to-noise ratio is

$$\left( \frac{S}{N} \right)_{t_0} = \int_{-\infty}^{\infty} \frac{|X(f)H(f)|^2}{S_{nn}(f)} df. \quad (11)$$

Although (10) is a necessary and sufficient condition to achieve the maximum signal-to-noise ratio for a fixed  $x(t)$ , it is not sufficient to provide a waveform/receiver-filter pair  $(x(t), r(t))$  that achieves the maximum possible signal-to-noise ratio for all waveforms satisfying the imposed constraints. In order to do this,  $r(t)$  must satisfy (10), but in addition, we must find a signal  $x(t)$  that maximizes (11) under the constraints

$$\begin{aligned} \int_{-\infty}^{\infty} |X(f)|^2 df &= E_x, \\ x(t) &= 0, \quad \text{for all } t \notin [-T/2, T/2]. \end{aligned} \quad (12)$$

Here,  $E_x$  is the total energy available for the transmitted waveform  $x(t)$ .

**Lemma 2:** The function  $\hat{x}(t)$  that maximizes the signal-to-noise ratio at the receiver-filter output is a solution to the Fredholm equation

$$\lambda_{\max} \hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau) L(t - \tau) d\tau, \quad (13)$$

where  $\lambda_{\max}$  is the maximum eigenvalue of (13),  $\hat{x}(t)$  is a corresponding eigenfunction scaled to have energy  $E_x$ , and the kernel  $L(t)$  is given by

$$L(t) = \int_{-\infty}^{\infty} \frac{|H(f)|^2}{S_{nn}(f)} e^{i2\pi f t} df.$$

The resulting signal-to-noise ratio is

$$\left( \frac{S}{N} \right)_{t_0} = \lambda_{\max} E_x.$$

*Proof:* We can rewrite (11) as

$$\begin{aligned} \left( \frac{S}{N} \right)_{t_0} &= \int_{-\infty}^{\infty} \left| X(f) \left[ \frac{H(f)}{\sqrt{S_{nn}(f)}} \right] \right|^2 df \\ &= \int_{-\infty}^{\infty} |X(f)B(f)|^2 df, \end{aligned} \quad (14)$$

where

$$B(f) = \frac{H(f)}{\sqrt{S_{nn}(f)}}.$$

Maximizing (14) is equivalent to maximizing an integral of the form

$$\begin{aligned} \int_{-\infty}^{\infty} |q(t)|^2 dt &= \int_{-\infty}^{\infty} |Q(f)|^2 df \\ &= \int_{-\infty}^{\infty} |X(f)B(f)|^2 df, \end{aligned} \quad (15)$$

where  $q(t)$  can be viewed as the output of a linear time-invariant system with input  $x(t)$  and impulse response  $b(t)$ . Here,  $Q(f)$  and  $B(f)$  are the Fourier transforms of  $q(t)$  and  $b(t)$ , respectively.

We must maximize (15) with respect to all  $X(f)$  that satisfy the constraints of (12). We can rewrite (15) for the energy in  $q(t)$ , noting that both  $x(t)$  and  $q(t)$  are real, as

$$\int_{-\infty}^{\infty} q^2(t) dt = \int_{-\infty}^{\infty} \left[ \int_{-T/2}^{T/2} x(\tau) e^{-i2\pi f\tau} d\tau \right] \cdot \left[ \int_{-T/2}^{T/2} x(p) e^{i2\pi fp} dp \right] |B(f)|^2 df. \quad (16)$$

Now define  $L(t)$  as the inverse Fourier transform of  $|B(f)|^2$ ; that is,

$$L(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |B(f)|^2 e^{i2\pi ft} df. \quad (17)$$

Then from (16) and (17), we have, changing the order of integration,

$$\int_{-\infty}^{\infty} q^2(t) dt = \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} x(\tau) x(p) L(p - \tau) d\tau dp. \quad (18)$$

We wish to find the function  $\hat{x}(t)$  on  $[-T/2, T/2]$ , which maximizes (18). It can be shown [14, p. 125] that  $\hat{x}(t)$  satisfies the integral equation

$$\lambda \hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau) L(t - \tau) d\tau, \quad (19)$$

where  $\lambda$  is the maximum eigenvalue of (19). So  $\hat{x}(t)$  is an eigenfunction corresponding to  $\lambda$ , the maximum eigenvalue of (19), and having energy  $E_x$ . From (18) and (19),

$$\begin{aligned} \int_{-\infty}^{\infty} q^2(t) dt &= \int_{-T/2}^{T/2} \hat{x}(t) \int_{-T/2}^{T/2} \hat{x}(p) L(p - \tau) dp d\tau \\ &= \int_{-T/2}^{T/2} \hat{x}(\tau) \lambda \hat{x}(\tau) d\tau \\ &= \lambda \int_{-T/2}^{T/2} \hat{x}(\tau) \hat{x}(\tau) d\tau \\ &= \lambda E_x. \end{aligned}$$

Substituting  $B(f) = H(f)/\sqrt{S_{nn}(f)}$  and then defining  $L(t)$  as in (17) yields

$$L(t) = \int_{-\infty}^{\infty} \frac{|H(f)|^2}{S_{nn}(f)} e^{i2\pi ft} df.$$

Then the waveform  $\hat{x}(t)$  time limited to the interval  $[-T/2, T/2]$ , which maximizes the signal-to-noise ratio at the receiver output, is given by the solution to

$$\lambda_{\max} \hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau) L(t - \tau) d\tau, \quad (20)$$

where  $\lambda_{\max}$  is the maximum eigenvalue of (20) and  $\hat{x}(t)$  is a corresponding eigenfunction scaled to have energy  $E_x$ . It

follows then from (11) that the resulting signal-to-noise ratio is

$$\left( \frac{S}{N} \right)_{t_0} = \lambda_{\max} E_x. \quad \square$$

Since the waveform  $x(t)$  is designed to take on nonzero values only on the interval  $[-T/2, T/2]$ , it can take on nonzero values for  $t < 0$ , and so the waveform  $x(t)$  is not necessarily causal. However, since  $x(t) = 0$  for all  $t < -T/2$ , we can obtain a causal waveform  $\hat{x}(t) = \hat{x}(t - T/2)$ , which will also yield the optimal response at the receiver output, except with delay  $T/2$ . To see that this waveform also maximizes the signal-to-noise ratio, we note that  $\hat{X}(f) = \hat{X}(f) e^{-i\pi fT}$ . But from (11), we see that the phase term  $e^{-i\pi fT}$  does not affect the resulting signal-to-noise ratio. We do, however, note from (10) that the response occurs at time  $t_0 + T/2$  instead of time  $t_0$ . So an optimal waveform  $\hat{x}(t)$  that is causal, and thus physically realizable, exists. In addition, the target impulse response is causal for all real physical targets. This being the case, the resulting receiver-filter also has causal impulse response  $r(t)$ , and thus is also a realizable filter. So from the optimal waveform/receiver-filter solution, we can find a waveform/receiver-filter pair that is physically realizable.

## V. MUTUAL INFORMATION IN REFLECTED WAVEFORMS AND WAVEFORMS FOR ESTIMATION

We now consider the problem of finding waveforms that maximize the mutual information between the target ensemble and the received radar waveform. We will prove Theorem 2 of Section III-B. We are interested in finding waveforms  $x(t)$  that maximize the mutual information  $I(\mathbf{g}(t); \mathbf{y}(t) | x(t))$  between the random target impulse response and the received radar waveform. The waveform  $x(t)$  is deterministic. It is explicitly denoted in  $I(\mathbf{g}(t); \mathbf{y}(t) | x(t))$  because the mutual information is a function of  $x(t)$ , and we are interested in finding those functions  $x(t)$  that maximize  $I(\mathbf{g}(t); \mathbf{y}(t) | x(t))$  under constraints on their energy and bandwidth. In order to find the functions  $x(t)$  that maximize  $I(\mathbf{g}(t); \mathbf{y}(t) | x(t))$ , we will first find  $I(\mathbf{z}(t); \mathbf{y}(t) | x(t))$  and those functions  $x(t)$  that maximize it. We will then show that the functions  $x(t)$  that maximize  $I(\mathbf{z}(t); \mathbf{y}(t) | x(t))$  also maximize  $I(\mathbf{g}(t); \mathbf{y}(t) | x(t))$ , and that for these  $x(t)$ ,  $I(\mathbf{g}(t); \mathbf{y}(t) | x(t)) = I(\mathbf{z}(t); \mathbf{y}(t) | x(t))$ .

Consider the small frequency interval  $\mathcal{F}_k = [f_k, f_k + \Delta f]$  of bandwidth  $\Delta f$  sufficiently small such that for all  $f \in \mathcal{F}_k$ ,  $X(f) \approx X(f_k)$ ,  $Z(f) \approx Z(f_k)$ , and  $Y(f) \approx Y(f_k)$ . Let  $\hat{x}_k(t)$  correspond to the component of  $x(t)$  with frequency components in  $\mathcal{F}_k$ ,  $\hat{z}_k(t)$  correspond to the component of  $z(t)$  with frequency components in  $\mathcal{F}_k$ , and  $\hat{\mathbf{y}}_k(t)$  correspond to the component of  $\mathbf{y}(t)$  with frequency components in  $\mathcal{F}_k$ . Then, over the time interval  $\mathcal{T} = [t_0, t_0 + \bar{T}]$ , the mutual information between  $\hat{\mathbf{y}}_k(t)$  and  $\hat{z}_k(t)$ , given that  $x(t)$  is transmitted, is [17, pp. 192–196]

$$\begin{aligned} &I(\hat{\mathbf{y}}_k(t); \hat{z}_k(t) | x(t)) \\ &= \bar{T} \Delta f \ln \left[ 1 + \frac{2|X(f_k)|^2 \sigma_G^2(f_k)}{P_{nn}(f_k) \bar{T}} \right]. \end{aligned}$$



If we consider two disjoint frequency intervals  $\mathcal{F}_j$  and  $\mathcal{F}_k$ , with  $\hat{\mathbf{y}}_j(t)$ ,  $\hat{\mathbf{z}}_j(t)$ , and  $\hat{\mathbf{n}}_j(t)$  the components in  $\mathcal{F}_j$  and  $\hat{\mathbf{y}}_k(t)$ ,  $\hat{\mathbf{z}}_k(t)$ , and  $\hat{\mathbf{n}}_k(t)$  the components in  $\mathcal{F}_k$  of  $\mathbf{y}(t)$ ,  $\mathbf{z}(t)$ , and  $\mathbf{n}(t)$ , respectively, then  $\hat{\mathbf{y}}_j(t)$  is statistically independent of  $\hat{\mathbf{y}}_k(t)$ ,  $\hat{\mathbf{z}}_j(t)$  is statistically independent of  $\hat{\mathbf{z}}_k(t)$ , and  $\hat{\mathbf{n}}_j(t)$  is statistically independent of  $\hat{\mathbf{n}}_k(t)$ . We can see that this is true by noting that each of the pairs of independent processes is made up of two Gaussian random processes with disjoint power spectral densities, and such processes are known to be statistically independent [18, p. 353]. Since the processes are statistically independent, the mutual information between  $[\hat{\mathbf{y}}_j(t), \hat{\mathbf{y}}_k(t)]$  and  $[\hat{\mathbf{z}}_j(t), \hat{\mathbf{z}}_k(t)]$  given that  $x(t)$  is transmitted is equal to the sum of the mutual information between  $\hat{\mathbf{y}}_j(t)$  and  $\hat{\mathbf{z}}_j(t)$  given that  $x(t)$  was transmitted and the mutual information between  $\hat{\mathbf{y}}_k(t)$  and  $\hat{\mathbf{z}}_k(t)$  given that  $x(t)$  is transmitted:

$$\begin{aligned} I([\hat{\mathbf{y}}_j(t), \hat{\mathbf{y}}_k(t)]; [\hat{\mathbf{z}}_j(t), \hat{\mathbf{z}}_k(t)] | x(t)) \\ = I(\hat{\mathbf{y}}_j(t); \hat{\mathbf{z}}_j(t) | x(t)) + I(\hat{\mathbf{y}}_k(t); \hat{\mathbf{z}}_k(t) | x(t)). \end{aligned}$$

If we now consider the frequency interval  $\mathcal{W} = [f_0, f_0 + W]$ , partition it into a large number of disjoint intervals of bandwidth  $\Delta f$ , and then let the number of intervals increase as  $\Delta f \rightarrow 0$ , in the limit we obtain an integral for the mutual information  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$ , where we assume that  $x(t)$ ,  $\mathbf{y}(t)$ , and  $\mathbf{z}(t)$  are confined to the frequency interval  $\mathcal{W}$ . This limit is

$$I(\mathbf{y}(t); \mathbf{z}(t) | x(t)) = \tilde{T} \int_{\mathcal{W}} \ln \left[ 1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f)\tilde{T}} \right] df.$$

**Lemma 3:** The magnitude-squared spectrum  $|X(f)|^2$  that maximizes  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$  under the average power constraint

$$\int_{\mathcal{W}} |X(f)|^2 df = E_x. \quad (21)$$

is

$$\begin{aligned} |X(f)|^2 &= \max \left[ 0, A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right] \\ &= \max [0, A - r(f)]. \end{aligned} \quad (22)$$

Here, the value of the constant  $A$  is determined by solving the following equation for  $A$ :

$$E_x = \int_{\mathcal{W}} \max \left[ 0, A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right] df. \quad (23)$$

The resulting maximum value of  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$  is

$$\begin{aligned} I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)) \\ = \tilde{T} \int_{\mathcal{W}} \max \left[ 0, \ln A - \ln \left( \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right) \right] df \\ = \tilde{T} \int_{\mathcal{W}} \max [0, \ln A - \ln r(f)] df. \end{aligned} \quad (24)$$

Here

$$r(f) = \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)}.$$

*Proof:* Using the Lagrange multiplier technique [19, pp. 357–359], we form the objective function

$$\begin{aligned} \Phi(|X(\cdot)|^2) &= \tilde{T} \int_{\mathcal{W}} \ln \left[ 1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f)\tilde{T}} \right] df \\ &\quad - \lambda \left[ \int_{\mathcal{W}} |X(f)|^2 df - E_x \right]. \end{aligned} \quad (25)$$

This is equivalent to maximizing  $\phi(|X(f)|^2)$  with respect to  $|X(f)|^2$ , for each  $f \in \mathcal{W}$ , where

$$\phi(|X(f)|^2) = \tilde{T} \ln \left[ 1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f)\tilde{T}} \right] - \lambda |X(f)|^2, \quad (26)$$

and  $\lambda$  is the Lagrange multiplier, to be determined from the constraint of (21). So we obtain an  $|X(f)|^2$  that maximizes (25) when we solve for an  $|X(f)|^2$  that maximizes (26). Thus, the  $|X(f)|^2$  that maximizes  $\Phi(|X(f)|^2)$  is

$$|X(f)|^2 = A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)}. \quad (27)$$

Here,  $A = \tilde{T}/\lambda = \text{constant}$ .

Substituting the expression for  $|X(f)|^2$  of (27) into the constraint of (21), we obtain

$$\begin{aligned} \int_{\mathcal{W}} |X(f)|^2 df &= \int_{\mathcal{W}} \left[ A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right] df \\ &= WA - \int_{\mathcal{W}} \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} df \\ &= E_x. \end{aligned} \quad (28)$$

Solving for the constant  $A$ , we have

$$A = \frac{1}{W} \left[ E_x + \int_{\mathcal{W}} \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} df \right]. \quad (29)$$

So the  $|X(f)|^2$  that maximizes  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$  is given by

$$|X(f)|^2 = \frac{1}{W} \left[ E_x + \int_{\mathcal{W}} \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} df \right] - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)}.$$

If we define  $r(f)$  as

$$r(f) \stackrel{\text{def}}{=} \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)},$$

then we can write  $|X(f)|^2$  as

$$|X(f)|^2 = A - r(f).$$

The maximum value of  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$ , which this  $|X(f)|^2$  achieves, is

$$\begin{aligned} I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)) \\ = \tilde{T} \int_{\mathcal{W}} \ln \left[ 1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f)\tilde{T}} \right] df \\ = \tilde{T} W \ln A - \tilde{T} \int_{\mathcal{W}} \ln r(f) df \text{ (nats)}. \end{aligned} \quad (30)$$

Since  $|X(f)|^2$  is the magnitude squared of the transmitted signal spectrum, it must be real and nonnegative for all  $f \in \mathcal{W}$  (we assume it to be zero for all  $f \notin \mathcal{W}$ ). Yet from (27),

$$|X(f)|^2 = A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)}.$$

It may be possible that for a given  $E_x$ ,  $\sigma_G^2(f)$ , and  $\mathcal{W}$ , one obtains a value of  $A$  such that for some  $f \in \mathcal{W}$ ,  $A < P_{nn}(f)\tilde{T}/2\sigma_G^2(f)$ . This would result in an invalid  $|X(f)|^2$ , as  $|X(f)|^2$  would be negative for such  $f$ . In order to obtain the  $|X(f)|^2$  that maximizes  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$ , we must actually solve for the value of  $A$  that satisfies (23). In most cases, the solution of this equation will have to be done numerically. However, for any (positive) value of  $E_x$ ,  $A$  can be bounded as follows:

$$\max_{f \in \mathcal{W}} \left\{ \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right\} \leq A \leq \frac{E_x}{W} - \min_{f \in \mathcal{W}} \left\{ \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right\}.$$

Once  $A$  has been solved for using (23), we have that  $I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t))$  is given by (24). The magnitude-squared spectrum  $|X(f)|^2$  that achieves  $I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t))$  is given by (22).  $\square$

The following two lemmas will be used in proving Theorem 2.

**Lemma 4:** Let  $\mathbf{a}(t)$  and  $\mathbf{b}(t)$  be finite-energy random processes and let  $D$  be a reversible transformation of  $\mathbf{a}(t)$  to a finite-energy random process  $\mathbf{c}(r)$  (where  $r$  is a new independent variable, but  $r$  could equal  $t$ ). Then

$$I(\mathbf{a}(t); \mathbf{b}(t)) = I(\mathbf{c}(r); \mathbf{b}(t)). \quad (31)$$

**Proof:** If  $D$  is a reversible transformation between  $\mathbf{a}(t)$  and  $\mathbf{c}(r)$ , that is,  $\mathbf{c}(r) = D\{\mathbf{a}(t)\}$ , then [1, pp. 90–91]

$$h(\mathbf{c}(r)) = h(\mathbf{a}(t)) + \mathcal{K}(D),$$

and

$$h(\mathbf{c}(r) | \mathbf{b}(t)) = h(\mathbf{a}(t) | \mathbf{b}(t)) + \mathcal{K}(D).$$

Here,  $\mathcal{K}(D)$  is a function only of the transformation  $D$ , not of the specific processes  $\mathbf{a}(t)$  and  $\mathbf{b}(t)$ . Thus,

$$\begin{aligned} I(\mathbf{c}(r); \mathbf{b}(t)) &= h(\mathbf{c}(r)) - h(\mathbf{c}(r) | \mathbf{b}(t)) \\ &= h(\mathbf{a}(t)) + \mathcal{K}(D) - h(\mathbf{a}(t) | \mathbf{b}(t)) - \mathcal{K}(D) \\ &= h(\mathbf{a}(t)) - h(\mathbf{a}(t) | \mathbf{b}(t)) \\ &= I(\mathbf{a}(t); \mathbf{b}(t)). \end{aligned} \quad \square$$

**Lemma 5:** Let  $\mathbf{a}(t)$  and  $\mathbf{b}(t)$  be finite-energy random processes with Fourier transforms  $\mathbf{A}(f)$  and  $\mathbf{B}(f)$ , respectively. Then if  $I(\mathbf{a}(t); \mathbf{b}(t))$  is the mutual information between  $\mathbf{a}(t)$  and  $\mathbf{b}(t)$  and  $I(\mathbf{A}(f); \mathbf{B}(f))$  is the mutual information between  $\mathbf{A}(f)$  and  $\mathbf{B}(f)$ , we have

$$I(\mathbf{a}(t); \mathbf{b}(t)) = I(\mathbf{A}(f); \mathbf{B}(f)). \quad (32)$$

**Proof:** We choose as our reversible transform  $D$  the Fourier transform, and apply it to both  $\mathbf{a}(t)$  and  $\mathbf{b}(t)$ , yielding

$$D\{\mathbf{a}(t)\} = \int_{-\infty}^{\infty} \mathbf{a}(t)e^{-i2\pi ft} dt = \mathbf{A}(f),$$

and

$$D\{\mathbf{b}(t)\} = \int_{-\infty}^{\infty} \mathbf{b}(t)e^{-i2\pi ft} dt = \mathbf{B}(f).$$

Applying Lemma 4 first to the transformation  $D\{\mathbf{a}(t)\} = \mathbf{A}(f)$ , we get

$$I(\mathbf{A}(f); \mathbf{b}(t)) = I(\mathbf{a}(t); \mathbf{b}(t)).$$

Applying Lemma 4 again, this time to the transformation  $D\{\mathbf{b}(t)\} = \mathbf{B}(f)$ , we have

$$I(\mathbf{A}(f); \mathbf{B}(f)) = I(\mathbf{A}(f); \mathbf{b}(t)).$$

So it follows that

$$I(\mathbf{a}(t); \mathbf{b}(t)) = I(\mathbf{A}(f); \mathbf{B}(f)). \quad \square$$

With Lemmas 4 and 5, we now prove Theorem 2.

**Proof of Theorem 2:** Define the two-sided set of frequencies

$$\hat{\mathcal{W}}_2 \stackrel{\text{def}}{=} \{f: |f| \in \mathcal{W}, |X(f)|^2 \neq 0\}.$$

Recall that  $\mathcal{W}$  is a one-sided set of frequencies, containing only positive frequencies. If  $X(f)$  has frequencies limited to  $\hat{\mathcal{W}}_2$ , then so does  $Z(f)$ , since

$$Z(f) = X(f)G(f).$$

So for  $f \in \hat{\mathcal{W}}_2$ , we can determine  $G(f)$  from  $Z(f)$ . For  $f \notin \hat{\mathcal{W}}_2$ , we cannot determine  $G(f)$  from  $Z(f)$ , since  $X(f) = 0$  and thus  $1/X(f)$  is indeterminate.

Define

$$\hat{G}(f) = \begin{cases} G(f), & \text{for } f \in \hat{\mathcal{W}}_2; \\ 0, & \text{elsewhere;} \end{cases} \quad (33)$$

$$\hat{Z}(f) = \begin{cases} Z(f), & \text{for } f \in \hat{\mathcal{W}}_2; \\ 0, & \text{elsewhere;} \end{cases} \quad (34)$$

$$\hat{\mathbf{g}}(t) = \int_{-\infty}^{\infty} \hat{G}(f)e^{i2\pi ft} df; \quad (35)$$

$$\hat{\mathbf{z}}(t) = \int_{-\infty}^{\infty} \hat{Z}(f)e^{i2\pi ft} df. \quad (36)$$

Then, from Lemma 5, we have

$$I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)) = I_{\max}(\mathbf{Y}(f); \mathbf{Z}(f) | x(t)). \quad (37)$$

For  $f \notin \hat{\mathcal{W}}_2$ ,  $Z(f) = 0$ , since  $Z(f) = X(f)G(f)$  and  $X(f) = 0$  for  $f \notin \hat{\mathcal{W}}_2$ . Thus,  $\hat{Z}(f) = Z(f)$  from the definition of  $\hat{Z}(f)$  in (34). So from Lemma 4,

$$I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)) = I_{\max}(\mathbf{Y}(f); \hat{\mathbf{Z}}(f) | x(t)). \quad (38)$$

Now for all  $f \in \hat{\mathcal{W}}_2$ ,  $X(f) \neq 0$ , so

$$\hat{G}(f) = \frac{Z(f)}{X(f)}.$$

Thus for all  $f \in \hat{\mathcal{W}}_2$ , there is a reversible transformation  $D$  that maps  $\hat{Z}(f)$  to  $\hat{G}(f)$ , given by

$$D[\hat{Z}(f)] = \begin{cases} \frac{\hat{Z}(f)}{X(f)}, & \text{for } f \in \hat{\mathcal{W}}_2; \\ 0, & \text{elsewhere.} \end{cases}$$

But as we know from Lemma 4, mutual information is invariant under reversible transformations, so

$$I(Y(f); \hat{Z}(f) | x(t)) = I(Y(f); \hat{G}(f) | x(t)).$$

This result, along with (38), yields

$$I(\mathbf{y}(t); \mathbf{z}(t) | x(t)) = I(Y(f); \hat{G}(f) | x(t)). \quad (39)$$

Note that  $\hat{G}(f)$ , defined by (33), is equal to  $G(f)$  for  $f \in \hat{\mathcal{W}}_2$  and is zero elsewhere. Thus the quantity of information obtained about  $\hat{G}(f)$  by observation of  $Y(f)$  on  $\hat{\mathcal{W}}_2$  is equal to the quantity of information obtained about  $G(f)$  on  $\hat{\mathcal{W}}_2$ . It follows immediately then that for observation on the set of all frequencies observed,

$$I(Y(f); G(f) | x(t)) \geq I(Y(f); \hat{G}(f) | x(t)). \quad (40)$$

But for  $f \notin \hat{\mathcal{W}}_2$ ,  $Y(f)$  provides no additional information about  $G(f)$ , because  $Y(f)$  is a function only of the noise  $\mathbf{n}(t)$ . Since  $\mathbf{n}(t)$  and  $\mathbf{g}(t)$  are statistically independent, the mutual information between the components of these processes with frequency components  $f \notin \hat{\mathcal{W}}_2$  is zero, since the mutual information between statistically independent random processes is zero. So the inequality of (40) is actually an equality:

$$I(Y(f); G(f) | x(t)) = I(Y(f); \hat{G}(f) | x(t)). \quad (41)$$

From Lemma 5 we have

$$I(Y(f); G(f) | x(t)) = I(\mathbf{y}(t); \mathbf{g}(t) | x(t)). \quad (42)$$

Thus, from (39), (41), and (42), we have

$$I(\mathbf{y}(t); \mathbf{g}(t) | x(t)) = I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)).$$

We have shown that for the class of functions  $x(t)$  that maximize  $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$ , we have  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t)) = I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$ . We have not yet shown that there is not some other waveform  $\tilde{x}(t)$  confined to the time interval  $[-T/2, T/2]$  satisfying the energy constraint of (3) resulting in a larger mutual information between  $\mathbf{g}(t)$  and  $\mathbf{y}(t)$ .

In order to show that there exists no  $\tilde{x}(t)$  resulting in a larger mutual information, we redraw the target channel model of Fig. 1 as shown in Fig. 2. Here we view both  $\mathbf{g}(t)$  and  $x(t)$  as inputs. The target impulse response  $\mathbf{g}(t)$  is observed by illuminating the target, resulting in the scattered waveform shown as the output of "Channel 1." "Channel 2" then accounts for the additive noise process  $\mathbf{n}(t)$  and the observation of the received waveform. From the Data Processing Theorem of

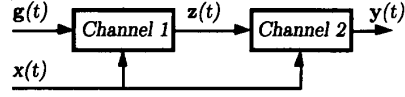


Fig. 2. Another interpretation of the radar target channel.

information theory [20, p. 31], we have that, for any  $x(t)$  transmitted,

$$I(\mathbf{y}(t); \mathbf{g}(t) | x(t)) \leq I(\mathbf{y}(t); \mathbf{z}(t) | x(t)). \quad (43)$$

In order to show that there is no  $\tilde{x}(t)$  for which

$$I(\mathbf{y}(t); \mathbf{g}(t) | x(t)) > I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)),$$

we will assume that such an  $\tilde{x}(t)$  exists. Then, from (43), it must be that

$$I(\mathbf{y}(t); \mathbf{z}(t) | \tilde{x}(t)) > I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)).$$

But this is a contradiction, since  $I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t))$  is the maximum value that the mutual information between  $\mathbf{y}(t)$  and  $\mathbf{z}(t)$  can achieve for any valid  $x(t)$ . Thus, for the class of  $x(t)$  with magnitude-squared spectrum  $|X(f)|^2$  given by (23),  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is maximized, and the maximum value  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is

$$I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) = I_{\max}(\mathbf{y}(t); \mathbf{z}(t) | x(t)). \quad \square$$

Note the behavior of the magnitude-square spectrum

$$|X(f)|^2 = \max \left[ 0, A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right],$$

which maximizes  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ . If the variance  $\sigma_G^2(f)$  of  $G(f)$  is held constant for  $f \in \mathcal{W}$ ,  $|X(f)|^2$  gets larger as  $P_{nn}(f)$  gets smaller, and  $|X(f)|^2$  gets smaller as  $P_{nn}(f)$  gets larger, becoming zero for  $P_{nn}(f) \geq 2A\sigma_G^2(f)/\tilde{T}$ . Similarly, if  $P_{nn}(f)$  is constant for all  $f \in \mathcal{W}$ , as would be the case for additive white Gaussian noise,  $|X(f)|^2$  gets larger as  $\sigma_G^2(f)$  gets larger and  $|X(f)|^2$  gets smaller as  $\sigma_G^2(f)$  gets smaller, with  $|X(f)|^2 \approx A$  for  $\sigma_G^2(f) \gg P_{nn}(f)\tilde{T}/2A$  and  $|X(f)|^2 = 0$  for  $\sigma_G^2(f) \leq P_{nn}(f)\tilde{T}/2A$ . In order to interpret this behavior physically, recall that  $\sigma_G^2(f)$  is the variance of the frequency spectrum  $G(f)$ . We see that frequencies  $f \in \mathcal{W}$  with large  $\sigma_G^2(f)$  provide greater information about the target than those with small  $\sigma_G^2(f)$ . This is not surprising, since for frequencies with small  $\sigma_G^2(f)$ , there is less uncertainty about the target response at that frequency in the first place. In fact, for those frequencies at which  $\sigma_G^2(f) = 0$ , there is no uncertainty at all in the outcome of  $\sigma_G^2(f)$ , and thus, there is no point in making any measurement at these frequencies.

Note that  $A = A(E_x, \sigma_G^2(f), P_{nn}(f))$  is a function of the transmitted energy  $E_x$ , the target spectral variance  $\sigma_G^2(f)$ , and of the noise power spectral density  $P_{nn}(f)$ . The fact that  $|X(f)|^2 = 0$  for all  $f$  such that  $\sigma_G^2(f) \leq P_{nn}(f)\tilde{T}/2A$  can then be interpreted as saying that a greater return in mutual information can be obtained by using the energy at another frequency or set of frequencies.

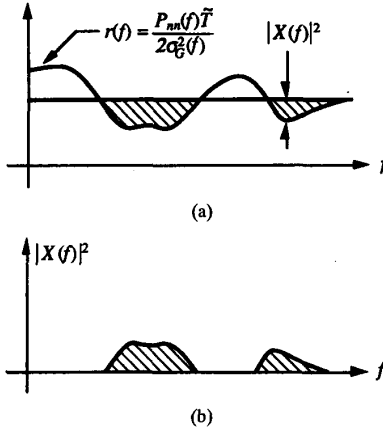


Fig. 3. (a) "Water-filling" interpretation of the magnitude-squared spectrum  $|X(f)|^2$  that maximizes the mutual information  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ . (b) Magnitude-squared spectrum  $|X(f)|^2$  that maximizes  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ . Note the relationship to the shaded area in (a).

An interesting interpretation of the relationship between  $|X(f)|^2$ ,  $A$ ,  $P_{nn}(f)$ , and  $\sigma_G^2(f)$  is shown in Fig. 3. Comparing (5) and Fig. 3, we see that the total energy  $E_x$  corresponds to the shaded area in Fig. 3(a). The difference between the line of value  $A$  forming the upper boundary of the shaded region and the curve forming the lower boundary of the shaded region is  $|X(f)|^2$ . This difference is displayed in Fig. 3(b).

This interpretation of Fig. 3, called the "water-filling" interpretation, arises in many problems dealing with the spectral distribution of power and energy in information theory [21, p. 389].

To further illustrate the behavior of  $|X(f)|^2$  as a function of the target spectral variance  $\sigma_G^2(f)$  and the noise power spectral density, consider the example of Fig. 4. In Fig. 4(a) we have the spectral variance  $\sigma_G^2(f)$ . In Fig. 4(b) we have the power spectral density  $P_{nn}(f)$ . In Fig. 4(c) we have  $r(f) = P_{nn}(f)T/2\sigma_G^2(f)$ , a function of both the power spectral density  $P_{nn}(f)$  of the noise and of the spectral variance  $\sigma_G^2(f)$ . In Fig. 4(d) we have the resulting magnitude-squared spectrum  $|X(f)|^2$  for the waveforms  $x(t)$  that maximize  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ . Note that because of the assumed bandwidth constraint,  $\mathcal{W} = [f_0, f_0 + W]$ ,  $|X(f)|^2 = 0$  for all  $f \notin \mathcal{W}$ . In the next section, we will examine a more realistic and detailed example numerically, in order to illustrate these results more clearly.

We have assumed that the random impulse response  $\mathbf{g}(t)$  is a Gaussian random process. As a result, the scattered signal  $\mathbf{z}(t)$  is a Gaussian random process. The received signal  $\mathbf{y}(t)$  is also a Gaussian random process, since the noise in the channel is additive Gaussian noise. Thus, for a given  $\sigma_G^2(f)$ , we are solving for the mutual information in the case of an additive Gaussian noise channel with a Gaussian input. As is well known, in the case of the additive Gaussian noise channel, for a channel input with a given variance  $\sigma^2$ , the mutual information between the channel input and the channel output is maximized when the input is Gaussian. Then by assuming that  $\mathbf{g}(t)$  is a Gaussian random process, we have selected a Gaussian input process for an additive Gaussian noise channel in our problem. By solving for the maximum mutual information

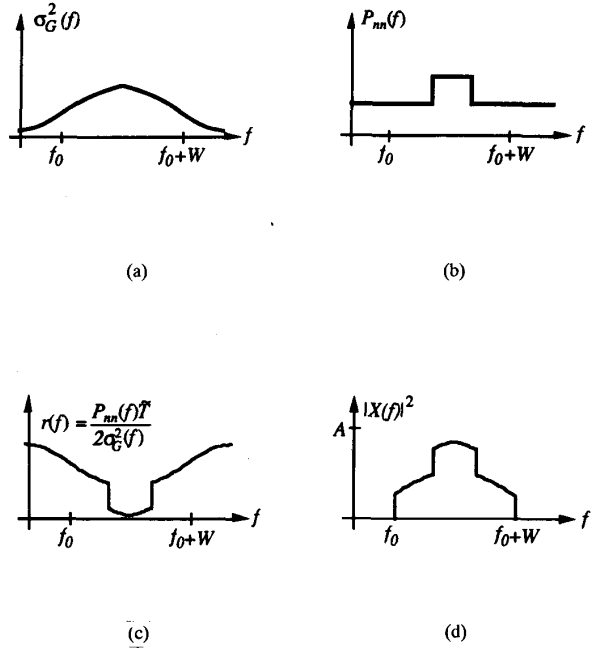


Fig. 4. Example illustrating the resulting  $|X(f)|^2$  for a given  $\sigma_G^2(f)$  and  $P_{nn}(f)$ . (a) Example  $\sigma_G^2(f)$ . (b) Example  $P_{nn}(f)$ . (c) Resulting  $r(f) = P_{nn}(f)T/2\sigma_G^2(f)$ . (d) Resulting  $|X(f)|^2$ .

$I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ , we have derived an upper bound on the maximum, achievable mutual information between  $\mathbf{y}(t)$  and  $\mathbf{g}(t)$  for any  $\mathbf{g}(t)$  with spectral variance  $\sigma_G^2(f)$  under the imposed bandwidth and energy constraints, whether  $\mathbf{g}(t)$  is Gaussian or not. In the case when  $\mathbf{g}(t)$  is Gaussian, as we have assumed, this upper bound is achieved.

## VI. EXAMPLES AND COMPARISONS

### A. Detection Waveform Examples

We first consider an example that illustrates the use of the design procedure and shows the effect that the transmission of various waveforms with identical energy can have on the output signal-to-noise ratio. Assume that the stationary additive Gaussian noise is white, with power spectral density  $S_{nn}(f) = N_0/2$ , and assume that the target impulse response  $h(t)$  has Fourier transform  $H(f)$  given by

$$H(f) = \begin{cases} k, & \text{for } |f| \leq W, \\ 0, & \text{for } |f| > W. \end{cases}$$

Here  $k$  is a constant, which for convenience is taken to be  $\sqrt{N_0/2}$  so that

$$|B(f)|^2 = \frac{|H(f)|^2}{S_{nn}(f)} = \begin{cases} 1, & \text{for } |f| \leq W, \\ 0, & \text{for } |f| > W. \end{cases}$$

TABLE I  
SIGNAL-TO-NOISE RATIO FOR  $x_n(t) = \sqrt{\epsilon} S_{0n}(\alpha T/2, 2t/T)$

$n$	$2WT = 2.55$	$2WT = 5.10$
0	$0.966E_x$	$1.000E_x$
1	$0.912E_x$	$0.999E_x$
2	$0.519E_x$	$0.997E_x$
3	$0.110E_x$	$0.961E_x$
4	$0.009E_x$	$0.748E_x$
5	$0.004E_x$	$0.321E_x$
6	—	$0.061E_x$
7	—	$0.006E_x$

This being the case, we have

$$L(t) - \int_{-W}^W e^{i2\pi ft} df = 2W \frac{\sin 2\pi Wt}{2\pi Wt} = \frac{\alpha \sin \alpha t}{\pi \alpha t}.$$

Here  $\alpha = 2\pi W$ . So the optimal  $\hat{x}(t)$  is a solution to the integral equation

$$\lambda_n x_n(t) = \int_{-T/2}^{T/2} x_n(t) 2W \frac{\sin 2\pi W(t - \tau)}{2\pi W(t - \tau)} d\tau,$$

which is known [13] to have a countable number of solutions for  $n = 0, 1, 2, \dots$ . The solutions to this equation are known as the *angular prolate spheroidal functions*, and are designated  $S_{0n}(\alpha T/2, 2t/T)$ . The associated eigenvalues  $\lambda_n$  can be written in terms of the *radial prolate spheroidal functions*, which are designated as  $R_{0n}^{(1)}(\alpha T/2, 1)$ . The eigenvalues and their associated eigenfunctions are given by  $\lambda_n = 2WTR_{0n}^{(1)}(\alpha T/2, 1)$ , and  $x_n(t) = S_{0n}(\alpha T/2, 2t/T)$ , for  $n = 0, 1, 2, \dots$ .

The sequence of eigenvectors  $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_n, \dots$  is a positive decreasing sequence in  $n$ . So the largest eigenvector occurs when  $n = 0$ . Thus, the solution  $\hat{x}(t)$  with energy  $E_x$  is

$$\hat{x}(t) = \sqrt{E_x} S_{00}(\alpha T/2, 2t/T),$$

and its associated eigenvalue is

$$\lambda_{\max} = 2WTR_{00}^{(1)}(\alpha T/2, 1).$$

The signal-to-noise ratio in this case is

$$\left(\frac{S}{N}\right)_{t_0} = \lambda_{\max} E_x = 2WTE_x R_{00}^{(1)}(\alpha T/2, 1).$$

In order to demonstrate the effect of the wave shape of the transmitted waveform on the output signal-to-noise ratio, consider the effect of transmitting the waveforms  $x_n(t) = \sqrt{E_x} S_{0n}(\alpha T/2, 2t/T)$ , for  $n = 0, 1, 2, \dots$ . All of these waveforms have transmitted energy  $E_x$ , but the resulting signal-to-noise ratios are  $\lambda_n E_x$ . As noted previously,  $\{\lambda_n\}$  is a positive decreasing sequence of  $n$ , and thus  $\lambda_0 E_x > \lambda_1 E_x > \lambda_2 E_x > \dots$ . So we see that the output signal-to-noise ratio is definitely a function of the transmitted waveform. In Table I, we show the resulting signal-to-noise ratio for the cases of  $2WT = 2.55$  and  $2WT = 5.10$ . (The eigenvalues were obtained from [13, p. 194].) As we can see, the signal-to-noise ratios drop off very quickly for  $n > 2WT$ . So we see that the wave-shape or spectral content of the transmitted waveform plays a significant role in the resulting signal-to-noise ratio.

Physically, we can interpret these results by noting that the maximum signal-to-noise ratio occurs when the target mode with the largest eigenvalue is excited by the transmitted waveform. In order to obtain the largest response possible from the target, we put as much of the transmitted energy as is possible into exciting this mode. This maximizes the signal-to-noise ratio and thus provides the best possible target detection performance under the imposed constraints.

Table I shows that it is possible to have more than one waveform with almost maximum response. For example, in the case where  $2WT = 5.10$ , the waveforms corresponding to  $n = 1, 2, 3$  give an output signal-to-noise ratio almost as large as the optimal waveform ( $n = 0$ ). In fact, the  $n = 3$  waveform produces a signal-to-noise ratio only 0.18 dB below that of the optimal waveform. Thus, not only can any one of these four waveforms be used with a resulting output signal-to-noise ratio comparable to the optimal, but any linear combination of these four waveforms can be used as well, yielding a family of waveforms with nearly optimal detection properties.

We now consider the problem of designing a waveform/receiver-filter pair that is optimal for detecting a perfectly conducting metal sphere of radius  $a$  in the presence of stationary white noise. We will compare the output signal-to-noise ratio of the resulting waveform to that of a pulse-modulated sinusoid and compare their target detection capabilities.

In order to apply the waveform/receiver-filter pair design procedure to a perfectly conducting sphere of radius  $a$ , we must find its impulse response  $h(t)$ . We assume a monostatic radar system with identical linear transmit and receive antenna polarizations. Thus, we are interested in the backscatter impulse response. The backscatter impulse response of a perfectly conducting sphere, when both transmit and receive antennas have identical linear polarization, has been calculated by Ken-nough and Moffatt, using the physical optics approximation [22]. The physical optics approximation is used here in order to simplify the calculation of  $h(t)$  and provide an analytically tractable solution. The physical optics approximation to the impulse response accurately predicts high frequency scattering behavior and corresponds to the early-time component of the target impulse response, which for many real targets contains most of the total scattered energy [23]. For the mathematical basis of the more general treatment, see [24].

This impulse response can be written as

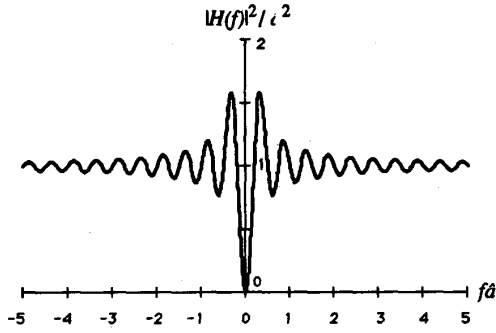
$$h(t) = -\frac{a}{c} \delta(t) + \frac{1}{2} [u(t) - u(t - 2a/c)].$$

Here,  $a$  is the radius of the sphere,  $c$  is the velocity of light,  $\delta(t)$  is the Dirac delta function, and  $u(t)$  is the unit step function, defined as

$$u(t) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{for } t \geq 0; \\ 0, & \text{for } t < 0. \end{cases}$$

Applying the waveform/receiver-filter design procedure of Theorem 1, we must first calculate the Fourier transform of  $h(t)$ . Doing so, we find that

$$H(f) = -\hat{a} + \hat{a} e^{-i2\pi f \hat{a}} \left[ \frac{\sin 2\pi f \hat{a}}{2\pi f \hat{a}} \right], \quad (44)$$

Fig. 5. Magnitude-squared spectrum  $|H(f)|^2$  versus  $f$ .

where  $\hat{a} = a/c$ . The magnitude-squared spectrum  $|H(f)|^2$  is thus

$$|H(f)|^2 = \hat{a}^2 \left[ 1 + \left( \frac{\sin 2\pi f \hat{a}}{2\pi f \hat{a}} \right)^2 - 2 \left( \frac{\sin 2\pi f \hat{a}}{2\pi f \hat{a}} \right) \cos 2\pi f \hat{a} \right]. \quad (45)$$

A plot of  $|H(f)|^2$  appears in Fig. 5.

Next, we find the inverse Fourier transform of the function  $|H(f)|^2/S_{nn}(f)$ . For white noise,  $S_{nn}(f) = N_0/2$ . Thus, we have

$$\frac{|H(f)|^2}{N_0/2} = \frac{2\hat{a}^2}{N_0} + \frac{2\hat{a}^2}{N_0} \left( \frac{\sin 2\pi f \hat{a}}{2\pi f \hat{a}} \right)^2 - \frac{4\hat{a}^2}{N_0} \left( \frac{\sin 2\pi f \hat{a}}{2\pi f \hat{a}} \right) \cos 2\pi f \hat{a}. \quad (46)$$

Calculating the inverse Fourier transform of (46), we obtain  $L(t)$  as specified in step 1 of our design procedure. Define the two functions

$$V_\alpha(t) = \begin{cases} 1 - |t|/\alpha, & \text{for } |t| \leq \alpha; \\ 0, & \text{elsewhere} \end{cases}$$

and

$$B_\alpha(t) = \begin{cases} 1, & \text{for } |t| \leq \alpha; \\ 0, & \text{elsewhere.} \end{cases}$$

Then we have

$$\begin{aligned} L(t) &= \int_{-\infty}^{\infty} \frac{|H(f)|^2}{N_0/2} e^{i2\pi ft} dt \\ &= \frac{2\hat{a}^2}{N_0} \delta(t) + \frac{\hat{a}}{N_0} V_{2\hat{a}}(t) - \frac{\hat{a}}{N_0} B_{\hat{a}}(t) \\ &\quad * [\delta(t - \hat{a}) + \delta(t + \hat{a})] \\ &= \frac{2\hat{a}^2}{N_0} \delta(t) + \frac{\hat{a}}{N_0} V_{2\hat{a}}(t) - \frac{\hat{a}}{N_0} B_{2\hat{a}}(t). \end{aligned} \quad (47)$$

Next, we solve the integral equation

$$\lambda_{\max} \hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau) L(t - \tau) d\tau$$

for an eigenfunction  $\hat{x}(t)$  corresponding to the maximum eigenvalue  $\lambda_{\max}$ . For  $L(t)$  as given in (47), we must solve

$$\begin{aligned} \lambda_{\max} \hat{x}(t) &= \int_{-T/2}^{T/2} \left[ \frac{2\hat{a}^2}{N_0} \delta(t) + \frac{\hat{a}}{N_0} V_{2\hat{a}}(t) - \frac{\hat{a}}{N_0} B_{2\hat{a}}(t) \right] \hat{x}(\tau) d\tau \\ &= \frac{2\hat{a}^2}{N_0} \int_{-T/2}^{T/2} \delta(t - \tau) \hat{x}(\tau) d\tau \\ &\quad + \frac{\hat{a}}{N_0} \int_{-T/2}^{T/2} [V_{2\hat{a}}(t - \tau) - B_{2\hat{a}}(t - \tau)] \hat{x}(\tau) d\tau. \end{aligned} \quad (48)$$

From (48), we see that if  $\hat{x}(t)$  is an eigenfunction corresponding to the maximum eigenvalue  $\lambda_{\max}$ , it must also be an eigenvector of the integral equation

$$\mu_{\max} \hat{x}(t) = \frac{\hat{a}}{N_0} \int_{-T/2}^{T/2} [V_{2\hat{a}}(t - \tau) - B_{2\hat{a}}(t - \tau)] \hat{x}(\tau) d\tau \quad (49)$$

corresponding to the maximum eigenvalue  $\mu_{\max}$ . Note that  $\lambda_{\max} = 2\hat{a}^2/N_0 + \mu_{\max}$ .

For convenience in our analysis, we will assume that  $\hat{x}(t)$  has unit energy (i.e.,  $E_x = 1$ ). In addition, for computational convenience, we assume a normalized value of  $\hat{a} = 1$ . Although such a value of  $\hat{a}$  does not correspond to values typically encountered in practice, these results can be applied to more typical values by scaling both the amplitude of the received signal and the time axis linearly in the length unit.

Solving the integral of equation (49) numerically for  $T = 1, 25, 50, 100, 250$ , and  $500$ , we obtain the eigenvalues  $\mu_{\max}$  of (49) and thus the eigenvalues  $\lambda_{\max}$  of (48), allowing us to determine the resulting signal-to-noise ratio in each of these cases.

For the purpose of comparison with more typical radar waveforms, we will consider the response of the target to a pulse modulated sinusoid of duration  $T$  with unit energy. The receiver-filter will be a matched filter matched to the transmitted waveform—the form of receiver-filter normally used in radar detection problems.

Such a waveform can be expressed as

$$x(t) = \beta B_{T/2}(t) \cos 2\pi f_0 t. \quad (50)$$

Here, for fixed  $T$ ,  $\beta$  is a normalizing constant such that the waveform has unit energy.

In order to obtain the most favorable result when transmitting a waveform as specified in (50), we must select the carrier frequency in order to take advantage of the resonance of the spherical scatterer as expressed by  $|H(f)|^2$ . From Fig. 5, we see that  $|H(f)|^2$  has its peak value at a frequency between  $0.25/\hat{a}$  and  $0.5/\hat{a}$ . It was determined numerically that  $|H(f)|^2$  takes on a maximum value of  $1.5862\hat{a}^2$  at a frequency of  $f = 0.3251/\hat{a}$ .

Let  $X(f)$  be the Fourier transform of the transmitted waveform  $x(t)$  as given in (50). Then the matched filter matched to this waveform that gives the maximum signal-to-noise ratio at time  $t_0$  is specified by the transfer function

$$R(f) = \frac{k \overline{X(f)} e^{-i2\pi f t_0}}{S_{nn}(f)}, \quad (51)$$

TABLE II  
SIGNAL-TO-NOISE RATIOS MULTIPLIED BY  $N_0$  FOR PULSED SINUSOID  
AND OPTIMAL DETECTION WAVEFORMS FOR VARIOUS  $T$

$T$	Pulsed Sinusoid	Optimal	Improvement
1	1.1454	2.1737	2.78 dB
25	2.7917	3.9108	1.46 dB
50	2.8183	3.8682	1.38 dB
100	2.8354	3.7477	1.21 dB
250	2.8615	3.7464	1.17 dB
500	2.8645	3.4967	0.87 dB

where  $k$  is any nonzero constant. The signal-to-noise ratio at time  $t_0$  is given by

$$\begin{aligned} \left(\frac{S}{N}\right)_{t_0} &= \frac{|y_s(t_0)|^2}{E|y_n(t_0)|^2} \\ &= \frac{\left|\int_{-\infty}^{\infty} X(f)H(f)R(f)e^{i2\pi ft_0} df\right|^2}{\int_{-\infty}^{\infty} |R(f)|^2 S_{nn}(f) df}. \end{aligned}$$

For white noise with power spectral density  $S_{nn}(f) = N_0/2$  and  $R(f)$  as given in (51), noting that  $H(f)$  is a conjugate-symmetric function of  $f$ , this simplifies to

$$\left(\frac{S}{N}\right)_{t_0} = \left(\frac{4}{N_0}\right) \frac{\left|\int_{-\infty}^{\infty} |X(f)|^2 \operatorname{Re}\{H(f)\} df\right|^2}{\int_{-\infty}^{\infty} |X(f)|^2 df}. \quad (52)$$

From (44), we have that when  $\hat{a} = 1$  as in our example,

$$\operatorname{Re}\{H(f)\} = \left(\frac{\sin 2\pi f}{2\pi f}\right) \cos 2\pi f - 1. \quad (53)$$

For  $x(t)$  as given in (50),  $X(f)$  is given by

$$X(f) = \frac{\beta T}{2} \left[ \frac{\sin \pi(f - f_0)T}{\pi(f - f_0)T} + \frac{\sin \pi(f + f_0)T}{\pi(f + f_0)T} \right],$$

and  $|X(f)|^2$  is given by

$$\begin{aligned} |X(f)|^2 &= \frac{\beta^2 T^2}{4} \left[ \left( \frac{\sin \pi(f - f_0)T}{\pi(f - f_0)T} \right)^2 \right. \\ &\quad + \left( \frac{\sin \pi(f + f_0)T}{\pi(f + f_0)T} \right)^2 \\ &\quad \left. + 2 \left( \frac{\sin \pi(f - f_0)T}{\pi(f - f_0)T} \right) \left( \frac{\sin \pi(f + f_0)T}{\pi(f + f_0)T} \right) \right]. \quad (54) \end{aligned}$$

Using (52), (53), and (54), we can solve for the signal-to-noise ratio that results when  $x(t)$  is the unit energy pulsed sinusoid given in (50). This is done for  $T = 1, 25, 50, 100, 250$ , and  $500$ . The values of  $\beta$  that provides a unit-energy waveform for each of these  $T$  are calculated numerically. Table II shows the resulting signal-to-noise ratio for these unit-energy pulsed sinusoids with their associated matched filters as well as the signal-to-noise ratio that results when an optimal waveform/receiver-filter pair is matched to the sphere being detected. In addition, we note the improvement (in decibels) in the output signal-to-noise ratio for the optimal waveform/receiver-filter pair over the pulsed sinusoid and its associated matched filter.

There is a significant improvement in the resulting signal-to-noise ratio when the optimal waveform/receiver-filter pair is used over that which occurs when a more typical ad hoc procedure is used. This is shown in Table II. For the range of  $T$  examined, the optimal waveform/receiver-filter pair provides approximately 1.2–2.8 dB of gain over the pulsed sinusoid with its associated matched filter. Considering that the received power is inversely proportional to the range to the fourth power, such gains would correspond to an increased detection range of 7–17%. In typical aircraft detection radar systems where the target is assumed to be a point target, the gain could be even greater, since the carrier frequency of the pulsed sinusoid would not be specifically selected to match the resonance of the sphere. Thus, in radar target detection problems where knowledge of the target impulse response is known *a priori*, a significant gain in detection signal-to-noise ratio can be achieved using the waveform/receiver-filter design procedure described in Theorem 1.

### B. Estimation Waveform Examples

We now consider an example illustrating Theorem 2 and the results of Section V. In doing so, we will examine the characteristics of the optimal transmitted signal's spectrum and the amount of information obtained.

We will assume that a radar system is observing a target at a range of 10 km. We will assume that the radar is a monostatic radar with an antenna having an effective area  $A_e = 3 \text{ m}^2$ , an RF bandwidth of 10 MHz, a transmitter frequency centered at 1 GHz, and we will assume that the antenna is pointed directly at the target under observation. This gives us a frequency interval  $\mathcal{W}$  of

$$\mathcal{W} = [f_0, f_0 + W] = [0.995 \text{ GHz}, 1.005 \text{ GHz}].$$

We will consider this radar system with average power constraints ranging from 1 W to 1000 W and observation times ranging from 10  $\mu\text{s}$  to 100 ms.

We assume that the target under observation has a finite-energy, Gaussian impulse response  $g(t)$  with spectral variance  $\sigma_G^2(f)$  given by

$$\sigma_G^2(f) = B \exp\{-\alpha(f - f_p)^2\}.$$

Here,  $B$  and  $\alpha$  are constants that respectively characterize the magnitude of the spectral variance  $\sigma_G^2(f)$  and the rate at which it decreases as  $|f - f_p|$  increases.

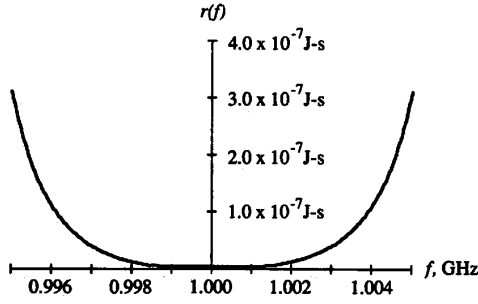
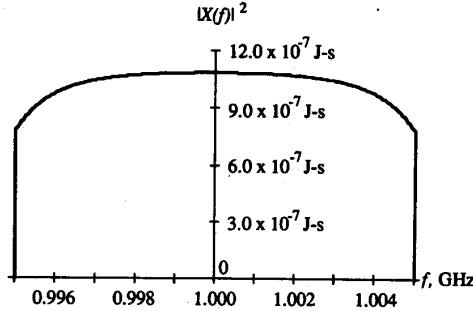
We will assume in our example that

$$\alpha = 10^{-13} \text{ s}^2,$$

a value illustrating well the effect of the transmitted waveform's spectral characteristics for the 10 MHz system bandwidth being considered. We will use a value of  $B$  which results in a spectral variance  $\sigma_G^2(f_p)$  that corresponds to a variation of  $1 \text{ m}^2$  in the target radar cross section at a frequency of  $f_p = 1 \text{ GHz}$ . This value is

$$B = 7.9577 \times 10^{-16}.$$

We will assume that the additive Gaussian noise present at the radar receiver is thermal noise that is white over the

Fig. 6.  $r(f) = P_{nn}(f)\tilde{T}/2\sigma_G^2(f)$  as a function of  $f$ .Fig. 7.  $|X(f)|^2$  for  $T = 10$  ms and  $P_x = 1000$  W.

frequency interval  $\mathcal{W}$  and that its effective noise temperature is  $T_s = 300$  K. Hence, the resulting one-sided noise power spectral density is [26, p. 29]

$$P_{nn}(f) = N_0 = kT_s = (1.381 \times 10^{-23} \text{ J/K})(300 \text{ K}) \\ = 4.1430 \times 10^{-21} \text{ J}.$$

Here,  $k = 1.381 \times 10^{-23} \text{ J/K}$  is Boltzmann's constant.

By definition,  $r(f)$  is given by

$$r(f) = \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)}.$$

For  $\tilde{T} = T = 10$  ms, this becomes

$$r(f) = (2.7835 \times 10^{-8} \text{ J} - s) \exp[\alpha(f - f_p)^2].$$

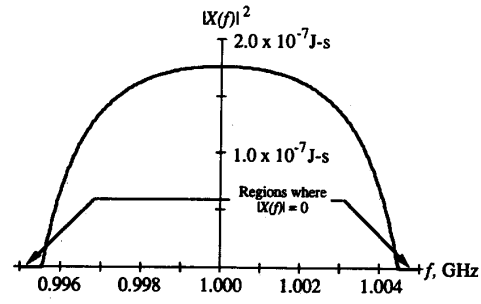
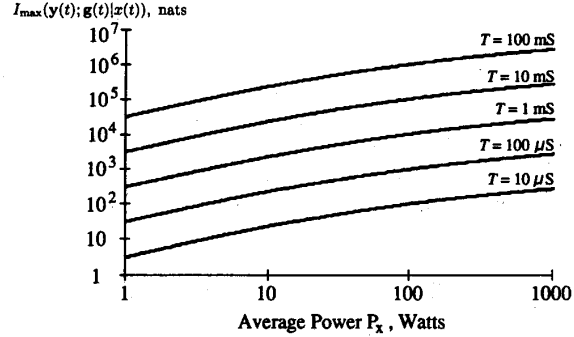
A plot of this  $r(f)$  is shown in Fig. 6.

For  $\tilde{T} = T = 10$  ms, we consider the two cases of  $P_x = 1000$  W and  $P_x = 100$  W. In both cases we must solve

$$E_x = P_x T = \int_{\mathcal{W}} \max[0, A - r(f)] df$$

with respect to  $A$  and then find  $|X(f)|^2 = \max[0, A - r(f)]$ . Recall that this formula gives  $|X(f)|^2$  for positive frequencies only, but that  $|X(f)|^2$  is an even function, so  $|X(f)|^2 = |X(-f)|^2$  for  $f < 0$ . Plots of  $|X(f)|^2$  for positive  $f$  in the cases of  $P_x = 1000$  W and  $P_x = 100$  W are given in Figs. 7 and 8, respectively. From (24), the mutual information  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is given by the integral

$$I(\mathbf{y}(t); \mathbf{g}(t) | x(t)) = \tilde{T} \int_{\mathcal{W}} \max[0, \ln A - \ln r(f)] df.$$

Fig. 8.  $|X(f)|^2$  for  $T = 10$  ms and  $P_x = 100$  W.Fig. 9.  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  as a function of  $T$  and  $P_x$ .

For  $P_x = 1000$  W and  $P_x = 100$  W, the resulting values of  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  are  $2.3815 \times 10^5$  nats and  $2.2152 \times 10^5$  nats, respectively.

In the case of  $P_x = 100$  W,  $\hat{\mathcal{W}} \subset \mathcal{W}$  is given by

$$\hat{\mathcal{W}} = \{f \in \mathcal{W} : A \geq r(f)\} \\ = [0.995698 \text{ GHz}, 1.004302 \text{ GHz}]. \quad (55)$$

So, in this case, only 8.604 MHz of the available 10 MHz of RF bandwidth should be used by the radar system. This is because more information is obtained by concentrating the energy in  $\hat{\mathcal{W}}$  and providing a greater signal-to-noise ratio in  $\hat{\mathcal{W}}$  than by distributing the energy across  $\mathcal{W}$ . The latter would provide a greater number of degrees of freedom to be measured, but they would be measured less reliably. The  $|X(f)|^2$  of (55) optimizes this tradeoff between the measured number of degrees of freedom in the measurement and the reliability of the individual degrees of freedom in the measurement—the optimization being done so as to maximize the mutual information  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ .

We will now display the results of the numerical solution of (23) and (24) for the mutual information  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  as a function of both  $T$  and the average available power  $P_x$ . This numerical solution was carried out for values of  $T$  equal to 10  $\mu$ s, 100  $\mu$ s, 1 ms, 10 ms, and 100 ms. For each of these values, the average power  $P_x$  varies over the range of from 1 W to 1000 W. All integrations were numerically carried out using the Gaussian Quadrature Method [25, pp. 322–326]. The resulting maximum values of  $I(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  are plotted in Fig. 9.



Two interesting points are illustrated by Fig. 9. The first is that the mutual information  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is proportional to  $T$ . Actually, this proportionality is only approximate, but the approximation is very good for  $T \gg 1/W$ , in which case it is reasonable to assume that  $\tilde{T} = T$ . This is true for all  $T$  considered in Fig. 9, since  $1/W = 0.1 \mu\text{s}$ . This proportionality is reflected by noting that each time  $T$  is increased by a factor of 10, the mutual information increases by a constant increment on the logarithmic scale of  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$ . Examining the expression for  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  of (6), we have

$$\begin{aligned} I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) &= \tilde{T} \int_W \max \left[ 0, \ln A - \ln \left( \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)} \right) \right] df \\ &= \tilde{T} \int_W \max \left[ 0, \ln \left( \frac{2A\sigma_G^2(f)}{P_{nn}(f)\tilde{T}} \right) \right] df. \end{aligned}$$

But for  $T \gg 1/W$ , it is reasonable to assume that  $\tilde{T} = T$ . This gives

$$\begin{aligned} I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) &= T \int_W \max \left[ 0, \ln \left( \frac{2A\sigma_G^2(f)}{P_{nn}(f)T} \right) \right] df. \end{aligned}$$

As (5) shows,  $A$  is proportional to  $\tilde{T}$ , and so assuming  $\tilde{T} = T$ ,  $A$  is proportional to  $T$ . If we define

$$a \stackrel{\text{def}}{=} \frac{A}{T},$$

then we can write  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  as

$$\begin{aligned} I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) &= T \int_W \max \left[ 0, \ln \left( \frac{2a\sigma_G^2(f)}{P_{nn}(f)} \right) \right] df. \end{aligned}$$

We can thus see analytically that  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is proportional to  $T$ . In fact, we can write the rate at which information is transferred to the receiver in the radar measurement process as

$$\begin{aligned} R_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) &= \int_W \max \left[ 0, \ln \left( \frac{2a\sigma_G^2(f)}{P_{nn}(f)} \right) \right] df, \end{aligned}$$

which is not a function of  $T$ . We then have

$$I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)) = TR_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t)).$$

The fact that  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is proportional to  $T$  has an interesting interpretation in terms of radar target-recognition problems. Examining the relationship between mutual information and radar-measurement performance, we noted in (1) that, if  $I(X; Y)$  is the mutual information between a set of parameters  $X$  to be measured and their measurement  $Y$ , the maximum number of equiprobable classes  $N$  into which  $X$  can be assigned with statistical reliability by observation of  $Y$  is

$$N = \lfloor e^{I(X; Y)} \rfloor.$$

Applying this result to our problem, we have that given an  $x(t)$  that achieves  $I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))$  is transmitted, the largest number of equiprobable classes into which  $\mathbf{g}(t)$  can be assigned with statistical reliability by observation of  $\mathbf{y}(t)$  is

$$\begin{aligned} N &= \lfloor e^{I_{\max}(\mathbf{y}(t); \mathbf{g}(t) | x(t))} \rfloor \\ &= \lfloor e^{TR(\mathbf{y}(t); \mathbf{g}(t) | x(t))} \rfloor. \end{aligned} \quad (56)$$

This number grows exponentially in  $T$ , the duration of the transmitted signal.  $T$  is often referred to in radar target-recognition problems as the "time-on-target." In radar target-recognition problems, it is well known that all other things being equal, the longer the "time-on-target," the better the performance of the target recognition system. As a result, within constraints imposed by other system requirements such as searching for new targets and tracking targets that have already been detected, the "time-on-target" in radar systems that perform target recognition is generally made as large as possible. This is reflected quantitatively in (56), which shows that the maximum number of equiprobable classes into which  $\mathbf{g}(t)$  can be reliably classified by observation of  $\mathbf{y}(t)$  increases exponentially in  $T$ .

Let us examine this result in terms of a practical methodology often used in radar target-recognition problems. One common method of classifying radar targets in target-recognition problems is by examining the characteristics of the Doppler spectrum of the target by performing spectral analysis on the signal reflected by the target. Such a technique is used, for example, in identifying jet aircraft on the basis of the jet engine modulation (JEM) phenomenon of the scattered signal [27], [28]. Assume that the frequency interval over which this is done has bandwidth  $W$ . Then, using classical methods of spectral analysis [29], the frequency resolution  $\Delta f$  of the measured spectrum is inversely proportional to  $T$ . Thus, the number of frequency bins of bandwidth  $\Delta f$  that span the interval of bandwidth  $W$  is proportional to  $T$ . Call this number of frequency bins  $M$ . Assume that because of noise in the received signal, the energy in each frequency bin can be distinguished to only one of  $Q$  levels. Then the total number of distinguishable spectra  $\tilde{N}$  is

$$\tilde{N} = Q^M. \quad (57)$$

If we now increase  $T$ , holding the power constant, the number of frequency bins  $M$  increases proportional to  $T$ , since the frequency resolution  $\Delta f$  is inversely proportional to  $T$ . We can thus write the number of bins  $M$  as

$$M(T) = mT, \quad (58)$$

where  $m$  is a constant of proportionality.

In increasing  $T$ , both the signal energy and the noise energy increase proportional to  $T$ , so the signal-to-noise ratio within a frequency bin remains constant. Thus, there are still  $Q$  distinguishable signal levels in each bin. This being the case, from (57) and (58), we have

$$\tilde{N} = Q^{M(T)} = Q^{mT}. \quad (59)$$

Hence, the number of discernible frequency spectra also increases exponentially with  $T$ . This is not to say that there is a

direct equivalence between the discernible frequency spectra and the equiprobable classes of  $g(t)$  to which (56) refers, but this heuristic example does show that the concept of the number of classes into which a target recognition system can classify targets increasing exponentially with "time-on-target" is not foreign to radar target-recognition problems. So the behavior of (56) is intuitively satisfying.

It is important to note that when a waveform  $x(t)$  that achieves  $I_{\max}(y(t); g(t) | x(t))$  is transmitted, the  $N$  equiprobable classes referred to in (56) are not under the control of the radar but are a function of the target ensemble. Actually, (56) states that the probability space  $\Omega$  can be partitioned into  $N$  subsets  $\Omega_1, \Omega_2, \dots, \Omega_N$ , where

$$\Pr\{\omega \in \Omega_k\} = \frac{1}{N}, \quad \text{for } k = 1, \dots, N,$$

where  $N$  is given by (56). These  $N$  subsets  $\Omega_1, \Omega_2, \dots, \Omega_N$ , which form a partition of  $\Omega$ , correspond to a set of  $N$  classes into which  $g(t)$  can be reliably classified by observation of  $g(t)$ . These  $N$  classes may not, however, correspond to classes that are of interest to the user of the radar system. Generally, the user will have knowledge of  $x(t)$  and will wish to classify the radar target into one of  $V$  classes  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_V$  based on observation of  $y(t)$ .

These classes may be linked to the physics of the problem, such as the case where the classes  $\mathcal{A}_k$  describe relative target size, or the classes may be less connected with the physics of the problem, such as in the case where only two target classes  $\mathcal{A}_1$  and  $\mathcal{A}_2$  are of interest: whether the aircraft being observed is friendly or hostile. In general, the problem of assigning a target to one of the classes in  $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_V\}$  based on observation of  $y(t)$  can be viewed indirectly as finding a mapping from  $\mathcal{C}: \Omega \rightarrow \mathcal{A}$ , such that for  $k = 1, \dots, N$ , each  $\Omega_k$  is mapped to one of the  $\mathcal{A}_j$  with a reasonable probability of error in determining the proper target class. As  $N$  becomes larger and the partition of  $\Omega$  becomes finer, we would expect the performance of the best mapping  $\mathcal{C}$  to improve for fixed  $V$ . Because  $N$  is an exponentially increasing function of  $I(y(t); g(t) | x(t))$ , we would expect the probability of correct classification into one of the classes in  $\mathcal{A}$  to improve as  $I(y(t); g(t) | x(t))$  becomes larger.

Returning again to the results in Fig. 9, we note the second important point it conveys: in our example, a very large amount of information is contained in the radar measurements. For example, for  $T = 100 \mu\text{s}$  and  $P_x = 10\text{W}$ ,  $I(y(t); g(t) | x(t))$  is approximately 100 nats, which equals approximately 144 bits. The corresponding  $N$  calculated from (56) to  $2.69 \times 10^{43}$ . Thus, we can conclude that a significant amount of information can be obtained about the target in the radar measurement process. In order to put this information to use, signal processing algorithms must be developed. The form these take will generally be highly dependent on the specific purpose for which the radar measurements will be used.

### C. Comparison of Detection and Estimation Waveforms

We now compare the characteristics of the optimal detection and estimation waveforms. This will be done by considering their magnitude squared spectra  $|X(f)|^2$ . If we interpret

$|H(f)|^2$  as "target response" in the detection waveform design problem and  $\sigma_G^2(f)$  as "target response" in the estimation waveform design problem, we see that  $|X(f)|^2$  tends to get larger at frequencies in which the "target response" gets larger, and smaller at those frequencies at which the power spectral density of the noise gets larger. As  $T$  becomes large such that  $L(t)$  in Theorem 1 is approximately 0 for  $|t| > T/2$ , we have

$$|X(f)|^2 \approx \alpha \frac{|H(f)|^2}{S_{nn}(f)}$$

for the optimal detection waveforms, where  $\alpha$  is a constant. In the case of optimal estimation waveforms, from (4), using the two-sided power spectral density  $S_{nn}(f) = P_{nn}(|f|)/2$ , we have

$$|X(f)|^2 = \max \left[ 0, A - \frac{S_{nn}(f)\tilde{T}}{\sigma_G^2(f)} \right].$$

The power spectral density of the noise enters into the two solutions in two quite different ways. Hence the form of the magnitude-squared spectrum of the two waveforms is quite different. While the waveform design for optimal target detection put as much energy as possible into the mode of the target that gave the largest response when weighted with respect to the noise, the waveform design for optimal estimation distributes the available energy in order to maximize the information obtained about the target. This is in agreement with the intuitive idea that lead to the investigation of the information extraction capabilities of radar waveforms in the first place. Because the optimal detection waveforms given by Theorem 1 concentrate on placing as much energy as possible into the largest scattering mode under the imposed waveform duration constraints, they ignore the smaller scattering modes. It is possible that these smaller modes contain a significant quantity of information useful for describing the target, perhaps information useful in differentiating between two very similar targets whose largest scattering modes may be very similar. In Theorem 2, we have looked at how to distribute this energy to maximize the mutual information in the case where the target ensemble could be modeled as a finite-energy Gaussian target impulse response. However, even when this model does not directly apply, the results can serve as a qualitative guide to the proper distribution of energy among target scattering modes for target detection and information extraction radar waveforms.

## VII. SUMMARY

We have shown that in the case of extended radar targets, the resonance phenomenon that occurs when the transmitted radar waveform is scattered by the target can be exploited to provide a larger signal-to-noise ratio at the output of the radar receiver than would result if we simply used an arbitrary waveform  $x(t)$  of energy  $E_x$  with a receiver-filter matched to  $x(t)$  and the noise environment.

Physically, we can interpret this result by noting that the maximum signal-to-noise ratio occurs when the mode of the target with the largest eigenvalue is excited by the transmitted waveform. In order to obtain the largest response possible

from the target, we put as much of the transmitted energy as is possible into exciting this mode. This gives us the largest possible signal-to-noise ratio and thus the best possible target detection performance under the imposed constraints. We note however that other scattering modes of the target may be useful for identifying or characterizing the target, so putting as much energy as possible into the mode with the largest eigenvalue will not generally be the best method of obtaining the maximum amount of information useful in describing the target. When our task is target identification or information extraction, we may thus wish to distribute the transmitted energy among the different target modes in a different manner. Theorem 2 describes how to distribute the energy in such a way that the mutual information between the target ensemble and the received waveform is maximized. As we saw in Section II-C, the greater this mutual information, the better we would expect the radar's classification and estimation capabilities to be.

The direct application of the results in Theorem 1 arise in radar problems where one is attempting to detect the presence of a target when one has *a priori* knowledge of its scattering characteristics—that knowledge being represented by the target impulse response  $h(t)$ . Such a situation might arise if we are trying to detect a high priority target providing a very small radar return. In such a situation, the additional received energy obtained by taking advantage of the target's resonant behavior may provide a significant improvement in detection performance.

As a practical matter, the optimal waveform solutions may not be easy to implement in real radar systems, since for reasons of power efficiency, these systems often operate with transmitter amplifiers in saturation. Because saturated amplifiers typically require that the waveform have (approximately) constant envelope, many of the optimal solutions obtained using the waveform/receiver-filter pair design procedure of Section VI-A cannot be used in radar systems that operate with a saturated transmitter. Still, the basic idea of putting as much of the energy as possible into the target's largest scattering mode under the design constraints on the waveform (which might be expanded to include a constant envelope constraint) is useful to keep in mind. In particular, the pulsed sinusoid used for comparison in Section III-A had its carrier frequency selected to give the largest possible response from the sphere. The selection of  $f_0$  in this example is clearly better than ignoring the target resonance phenomenon and arbitrarily selecting the carrier frequency.

Finally, note that the shape of a radar signal, and not just its energy alone, can have a significant effect on extended target detection performance. This fact is often overlooked in the radar engineering community, largely as a result of the fact that when one uses a matched filter to detect a waveform in stationary additive noise, detection performance does not depend on the shape of the waveform being detected—as long as the receiver-filter being used is matched to the waveform to be detected and the additive noise's power spectral density. As a result, if we are considering a point target, then  $h(t) = \alpha\delta(t - t_0)$ , and the shape of the transmitted waveform would have no effect on the output signal-to-noise ratio. However, as we have seen in Section IV, this is not true of extended targets,

where  $h(t) \neq \alpha\delta(t - t_0)$ . It is advantageous to keep this fact in mind when considering extended targets and not extrapolate the point-target results to the extended target case, as is often done. For a point target, the waveform/receiver-filter design procedure produces a solution where  $x(t)$  is any arbitrary waveform  $x(t)$  with energy  $E_x$  and  $r(t)$  is its associated matched filter.

In the case of the design of a waveform  $x(t)$  that optimizes the detection of a target of known impulse response  $h(t)$ , we noted that the solution corresponds to an eigenfunction with energy  $E_x$  corresponding to the largest eigenvalue of the integral equation of (22). We noted that these results could be interpreted, in the case of additive white Gaussian noise, as putting as much of the transmitted energy as possible into the largest mode of the target under the time and bandwidth constraints on the transmitted waveform. The result was that we obtained the largest possible signal-to-noise ratio, and thus the optimal detection performance, under the constraints on the transmitted waveform. We also noted, however, that the other eigenfunctions, corresponding to different modes of the target, could contain significant information about the target. So if we wished to extract information about the target, it might be advantageous to distribute the available energy among the various modes. This illustrates that optimal waveforms for detection and estimation of extended radar targets can be very different in nature. Viewing the characteristics of these waveforms in terms of the distribution of energy among target scattering modes gives both physical and information-theoretic insights into the design and performance of these waveforms. As the technology for adaptive waveform radars becomes more widespread, these insights should be valuable in effectively utilizing this technology.

#### ACKNOWLEDGMENT

I would like to thank Prof. E. C. Posner of the California Institute of Technology and an anonymous reviewer for their comments and suggestions on improving the presentation of this material.

#### REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948; reprinted in C. E. Shannon and W. W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.
- [2] P. M. Woodward and I. L. Davies, "A theory of radar information," *Phil. Mag.*, vol. 41, pp. 1101–1117, Oct. 1951.
- [3] P. M. Woodward, "Information theory and the design of radar receivers," *Proc. IRE*, vol. 39, pp. 1521–1524, Dec. 1951.
- [4] P. M. Woodward and I. L. Davies, "Information theory and inverse probability in telecommunications," *Proc. IEE*, vol. 99, Part III, pp. 37–44, Mar. 1952.
- [5] I. L. Davies, "On determining the presence of signals in noise," *Proc. IEE*, vol. 99, Part III, pp. 45–51, Mar. 1952.
- [6] P. M. Woodward, *Probability and Information Theory with Applications to Radar*. London, England: Pergamon, 1953.
- [7] V. S. Frost and K. S. Shanmugan, "The information content of synthetic aperture radar images of terrain," *IEEE Trans. Aerospace Electron. Syst.*, vol. AES-19, no. 5, pp. 768–774, Sept. 1983.
- [8] C. H. Wilcox, "The synthesis problem for radar ambiguity functions," Math. Res. Center, U.S. Army, Univ. Wisconsin, Madison, WI, MRC Tech. Summary Rep. 157, Apr. 1960; reprinted in R. E. Blahut, W. M. Miller, and C. H. Wilcox, *Radar and Sonar*, Part I. New York: Springer-Verlag, 1991.

- [9] H. Naparst, "Dense target signal processing," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 2, pp. 317-327, Mar. 1991.
- [10] S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communications Electronics*, 2nd ed. New York: Wiley, 1984.
- [11] D. O. North, "An analysis of the factors which determine signal-noise discrimination in pulsed carrier systems," RCA Lab., Rep. PTR-6C; reprinted in *Proc. IEEE*, vol. 51, pp. 1016-1027, July 1963.
- [12] R. N. Bracewell, *The Fourier Transform and Its Applications*. New York: McGraw-Hill, 1986.
- [13] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part 1. New York: Wiley, 1968.
- [14] C. W. Helstrom, *Statistical Theory of Signal Detection*. New York: Pergamon, 1968.
- [15] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [16] F. Kanaya and K. Nakagawa, "On the practical implication of mutual information for statistical decision making," *IEEE Trans. Inform. Theory*, vol. 37, no. 4, pp. 1151-1156, July 1991.
- [17] M. R. Bell, "Information theory and radar: Mutual information and the design and analysis of radar waveforms and systems," Ph.D. dissertation, California Inst. Technol., Pasadena, 1988.
- [18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1965.
- [19] F. B. Hildebrand, *Advanced Calculus for Applications*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1976.
- [20] R. J. McEliece, *The Theory of Information and Coding*. Reading, MA: Addison-Wesley, 1977.
- [21] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [22] E. M. Kennaugh and D. L. Moffatt, "Transient and impulse response approximations," *Proc. IEEE*, vol. 51, no. 8, pp. 893-901, Aug. 1965.
- [23] M. L. Van Blaricum, "A view of the early-time component in impulsive scattering," in *Ultra-Wideband Radar: Proc. First Los Alamos Symp.*, B. Noel, Ed. Boca Raton, FL: CRC Press, 1991, pp. 191-202.
- [24] C. H. Wilcox, "Sonar and radar echo structure," in *Radar and Sonar*, Part I, R. E. Blahut, W. M. Miller, and C. H. Wilcox, New York: Springer-Verlag, 1991.
- [25] R. W. Hamming, *Numerical Methods for Scientists and Engineers*, 2nd ed. New York: Dover, 1973.
- [26] J. R. Pierce and E. C. Posner, *Introduction to Communication Science and Systems*. New York: Plenum, 1980.
- [27] M. I. Skolnik, *Radar Handbook*. New York: McGraw-Hill, 1970.
- [28] M. R. Bell and R. A. Grubbs, "Modeling of jet engine modulated radar signal returns for target identification," *IEEE Trans. Aerospace Electron. Syst.*, vol. 28, no. 1, Jan. 1993.
- [29] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. New York: Dover, 1958.