INFORMATION THEORY AND RADAR:

MUTUAL INFORMATION AND THE DESIGN AND ANALYSIS OF RADAR WAVEFORMS AND SYSTEMS

Thesis by

Mark Robert Bell

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1988

(Submitted March 30, 1988)

ACKNOWLEDGEMENTS

I wish to thank my thesis advisor, Dr. Edward C. Posner, for his insightful guidance, discussions, and reviews of my work throughout the course of this research. His comments and questions have made this work both clearer and more relevant. I have been very fortunate to have an advisor with both the depth and breadth of knowledge that he possesses.

I also wish to thank the professors who make up my thesis committee, Dr. Yaser Abu-Mostafa, Dr. Charles Elachi, Dr. Joel N. Franklin, and Dr. Robert J. McEliece, for both taking the time to be on my committee and offering the excellent courses I have had the priviledge of taking with each of them.

I also thank Hughes Aircraft Company for sponsoring this work through a Howard Hughes Doctoral Fellowship.

Finally, I thank my wife Catherine for her love, patience, and encouragement while I pursued this research. This work is dedicated to my parents, who have given me many years of support and encouragement.

ABSTRACT

This thesis examines the use of information theory in the analysis and design of radar, with a particular emphasis on the information-theoretic design of radar waveforms. First, a brief review of information theory is presented and then the applicability of mutual information to the measurement of radar performance is examined. The idea of the *radar target channel* is introduced. The *Radar/Information Theory Problem* is formulated and solved for a number of radar target channels, providing insight into the problem of designing radar waveforms that maximize the mutual information between the target and the received radar signal. Radar-scattering models are examined in order to obtain usable models for practical waveform design problems. The *target impulse response* is introduced as a method of characterizing the spatial range distribution of radar targets. The target impulse response is used to formulate a new generalization of the matched filter in radar that matches a transmitted-waveform/receiver-filter pair to a target of known impulse response, providing the maximum signal-to-noise ratio at the receiver under a constraint on transmitted energy and the time duration of the waveform. Next, the problem is formulated and solved of designing radar waveforms that maximize the mutual information between the target and the received radar waveform for a target characterized by an impulse response that is a finite-energy random process. The characteristics of waveforms for optimum detection and for obtaining maximum information about a target are compared. Finally, the information content of radar images is examined. It is concluded that the information-theoretic viewpoint can improve the performance of practical radar systems.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF ILLUSTRATIONS ix

LIST OF TABLES xiii

CHAPTER 1. INTRODUCTION 1

1.1 Background 1

1.2 Overview 7

1.3 Chapter 1 References 10

CHAPTER 2. THE INFORMATION-THEORETIC ANALYSIS OF RADAR .. 11

2.1 The Information-theoretic Analysis of Radar Systems 12

2.2 Information Theory and Discrete Random Variables 22

2.3 Information Theory and Continuous Random Variables 37

2.4 Mutual Information and Radar Measurement Performance 45

2.5 Chapter 2 References 52

CHAPTER 3. THE RADAR/INFORMATION THEORY PROBLEM 54

3.1 General Formulation of the Radar/Information Theory Problem 55

3.2 The Radar/Information Theory Problem for Discrete Target Channels 58

3.2.1 Solution for Discrete Memoryless Target Channels 60

3.2.2 The Radar/Information Theory Problem for N Observations of a Fixed Target 76

3.2.3 The Radar/Information Theory Problem for Finite-State Target Channels 79

3.3 The Radar/Information Theory Problem for Continuous Target Channels .. 95

3.3.1 The Continuous Memoryless Target Channel 97

3.4 Conclusions 103

3.5 Chapter 3 References 104

CHAPTER 4. RADAR SCATTERING MODELS 105

4.1 Radar Reflectivity and Radar Cross Section 106

4.2 The Analytic Signal Representation 112

4.3 Polarization, Depolarization, and Scattering 114

4.4 Statistical Models of Radar Backscatter from Land 118

4.5 A Model for Cross Polarization Measurements from Rough Surfaces 126

4.6 Spatial Resolution Characteristics of Radar Waveforms 128

4.7 Chapter 4 References 146

CHAPTER 5. MATCHING A RADAR WAVEFORM/RECEIVER-FILTER PAIR TO A TARGET OF KNOWN IMPULSE RESPONSE 148

5.1 Radar Target Detection by Energy Threshold Test 149

5.2 Matching a Waveform/Receiver-Filter Pair to a Target of Known Impulse Response 156

5.3 Optimum Waveform/Receiver-Filter Pairs for Sphere Detection 170

5.4 Chapter 5 References 182

CHAPTER 6. INFORMATION THEORY AND RADAR WAVEFORM DESIGN 183

6.1 Radar Waveform Design for Maximum Information Transfer Between the Target and the Received Radar Signal 184

6.2 A Numerical Example 213

6.3 Radar Waveform Design and Implementation 228

6.4 Comparison of Waveforms for Optimum Detection and Maximum Information Extraction 233

6.5 Chapter 6 References 235

CHAPTER 7. THE INFORMATION CONTENT OF RADAR IMAGES 237

7.1 Radar Images 238

7.2 Information Theory and Speckle Noise in Radar Images 259

7.3 Chapter 7 Appendix 276

7.3.1 Verification of Eq. (7.48) 276

7.3.2 Verification of Eq. (7.49) 277

7.3.3 Verification of Eq. (7.54) 279

7.4 Chapter 7 References 283

CHAPTER 8. SUMMARY AND CONCLUSIONS 286

LIST OF ILLUSTRATIONS

Figure Page

1.1. Block Diagram of a Generic Radar System 2

2.1. Block Diagram of a Generic Communication System 13

2.2. Block Diagram of a Generic Radar System 16

2.3. Self Information $I(x)$ versus $p(x)$ 23

2.4. Block Diagram of a Simple Communication System 31

2.5. Binary Symmetric Channel 32

2.6. The Binary Entropy Function $\mathcal{H}(\beta)$ 34

2.7. Additive Gaussian Noise Channel 41

2.8. Capacity of the Additive Gaussian Noise Channel 44

2.9. Block Diagram of a Measurement System 45

2.10. Rate Distortion Function of the Gaussian Measurement Example 45

3.1. Block Diagram of Radar/Information Theory Problem 56

3.2. Block Diagram of The Target Channel 57

3.3. Radar/Information Theory Problem for Discrete Target Channels 59

3.4. Diagram of Target Channel for $N = 1$ and$X = \alpha_j$ 61

3.5. Memoryless Binary Target Channel of *Example 3.1* 65

3.6. Memoryless Discrete Target Channel of *Example 3.2* 69

Figure Page

3.7. Memoryless Discrete Target Channel of *Example 3.3* 70

3.8. Finite-State Target Channel of *Example 3.5* 83

3.9. Finite-State Target Channel of *Example 3.6* 84

3.10. Radar/Information Theory Problem for Continuous Target Channels ... 96

3.11. Continuous Target Channel Model 99

4.1. Waveform Spectra $X(f)$ and $R(f)$ 130

4.2. Constructive Interference from Two Scatterers 132

4.2. Two-Frequency Constructive Interference from Two-Scatterers 132

4.3. Single-Frequency Constructive Interference from Two-Scatterers 134

4.4. Radar Measurement of a Rough Surface 134

4.5. Constructive Interference and Waveform Bandwidth 139

4.6. Coordinate System of Scattering Problem 142

4.7. Linear Time-Invariant System Representation of Stationary Target-Scattering Mechanism 142

4.8. Continuous Spectra $X(f)$ and $R(f)$ of Radar Waveform 142

5.1. Probability Density Functions $f(e|H_0)$ and $f(e|H_1)$ 152

5.2. Block Diagram of Radar Waveform/Receiver-Filter Design Problem ... 157

5.3. Linear System Representation of the Relation Between $Q(f)$ and $X(f)$ 163

5.4. The Backscatter Impulse Response $h(t)$ of a Sphere of Radius a 171

Figure Page

5.5. Magnitude-Squared Spectrum $|H(f)|^2$ versus f 173

5.6. Wave shape of $\hat{x}(t)$ for $T = 1$ 177

5.7. Wave shape of $\hat{x}(t)$ for $T = 25$ 177

5.8. Wave shape of $\hat{x}(t)$ for $T = 50$ 178

5.9. Wave shape of $\hat{x}(t)$ for $T = 100$ 178

6.1. Block Diagram of The Radar Target Channel 185

6.2. $E(\alpha)$ as a function of $\alpha = T/W$ 190

6.3. Additive Gaussian Noise Channel 193

6.4. Another Interpretation the of Radar Target Channel 206

6.5. "Water-Filling Interpretation of $|X(f)|^2$ 210

6.6. Example illustrating the resulting $|X(f)|^2$ for a given $\sigma_G^2(f)$ and $P_{nn}(f)$ 211

6.7. $r(f) = P_{nn}(f)\widetilde{T}/2\sigma_g^2(f)$ as a function of f 216

6.8. $|X(f)|^2$ for $T = 10\,\text{ms}$ and $P_x = 1000\,\text{W}$ 217

6.9. $|X(f)|^2$ for $T = 10\,\text{ms}$ and $P_x = 500\,\text{W}$ 218

6.10. $|X(f)|^2$ for $T = 10\,\text{ms}$ and $P_x = 100\,\text{W}$ 220

6.11. $I_{\max}(\mathbf{y}(t); \mathbf{g}(t)|x(t))$ as a function of T and P_x 221

6.12. Spectral Analysis of Bandwidth W Using M Bins of Bandwidth Δf .. 225

7.1. Spatial Sampling in an Imaging Radar System 241

7.2. Discrete Radar Image as an $m \times n$ Array of Resolution Cells 242

Figure Page

7.3. Radar Image of San Francisco Area.....247

7.4. Resolution Cell Neighborhoods on Which $f(q)$ Is Conditioned.....255

7.5. 10×10 Binary Markov Image.....257

7.6. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=4$.....272

7.7. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=8$.....272

7.8. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=10$.....273

7.9. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=20$.....273

7.10. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=50$.....274

7.11. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=100$ 274

LIST OF TABLES

Table Page

5-1. Signal-To-Noise Ratio for $x_n(t) = \sqrt{\mathcal{E}} S_{0n}(\alpha T/2, 2t/T)$ 168

5-2. Eigenvalues $\mu_{\max}$ and $\lambda_{\max}$ for Various T 176

5-3. Signal-to-Noise ratios multiplied by N_0 for Pulsed Sinusoid and Optimal Detection Waveforms for Various T 180

7-1. Radar Operating Characteristics 248

7-2. Minimum Required Mutual Information for Classification into One of J Equiprobable Classes 275

CHAPTER 1

INTRODUCTION

1.1. Background.

In the last fifty years, radar has grown from its infancy as a technology with relatively few applications into a mature technology with a wide range of applications. The acronym *radar* (*ra*dio *d*etection *a*nd *r*anging) indicates the impetus for the initial development of systems that use radiowave or microwave scattering to make measurements of a remote object. Its primary use in the Second World War was detecting and locating enemy aircraft. This provided both advanced warning of attack and information for the direction of anti-aircraft weapons. It was also effectively used by the British to detect and locate German submarines [1.1–1.3].

The modern uses of radar, while still including these early applications, include several additional applications. These include such remote sensing applications as the measurement of water resources, agricultural resources, global ice-coverage, forest conditions, and wind, as well as such radar techniques as ionospheric sounding, geological mapping, radar meteorology, planetary remote sensing, and radar astronomy. Radar has also expanded in its application in navigation. In aviation this can be seen in its use in air traffic control and aircraft navigation radar. In navigation at sea, radar is used aboard ships for collision avoidance and on land for harbor traffic management. Applications in the areas of military surveillance have also expanded to include terrain mapping and target identification. In radar target identification,

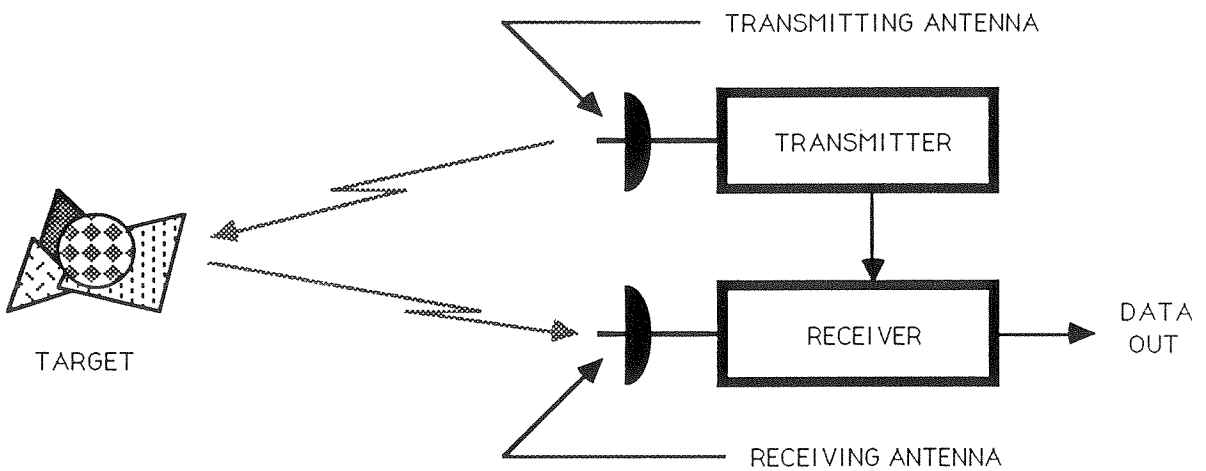


Figure 1.1. Block Diagram of a Generic Radar System.

not only is the target detected and located, but the physical characteristics of the target are determined, or the target is classified by type based on its radar signature [1.4, 1.5].

A *radar*, as considered in this thesis, is any system that uses the scattering of microwave radiation from an object to obtain information about that object. Such objects will be referred to as *targets*. Fig. 1.1 shows a block diagram of a generic radar system. The radar consists of a *transmitter*, which generates the signal to be transmitted, a *transmitting antenna*, which radiates the transmitter waveform as an electromagnetic field, a *receiving antenna*, which intercepts a portion of the

electromagnetic field scattered in the direction of the receiver, and the *receiver*, which detects and processes the signal collected by the receiving antenna. The transmitting and receiving antennas may or may not be the same physical antenna. In this diagram, the "transmitter" corresponds to all signal generators, modulators, and power amplifiers used in the transmission process; the "receiver" corresponds not only to the standard receiver elements such as RF amplifiers, detectors, IF amplifiers, and filters, but also to any signal processing elements that might be used to extract information from the received radar signal (e.g., Doppler filters, range-gating circuitry, SAR processors, etc.). While not radars in the strict sense, instruments that make radio science measurements, or perform "radio sounding," can also be considered radars for the purpose of this thesis. These instruments transmit a known—or at least partially known—waveform through a medium to be characterized, and the received waveform is then analyzed to determine how the propagating medium has distorted it. From these distortions, characteristics of the observed medium are inferred. If the "distortion of the field by the medium" is viewed as functionally equivalent to the "scattering of the field by the target" in the above description of the generic radar system, then the generic radar model is also applicable to such radio science instruments.

When a radar system is making a measurement, the transmitted waveform radiated by the transmitting antenna propagates through space until it impinges on the target. It is then scattered by the target. In general, the field incident on the target is scattered in all directions. That portion of the scattered field that is

intercepted by the receiving antenna is then fed to the receiver. The receiver then processes this received field and the desired information is extracted from it. Early radar work dealt primarily with detecting the presence or absence of the target, but modern radar work concentrates on the extraction of additional information about the target as well.

Note that in Fig 1.1, a link is depicted between the transmitter and the receiver. This represents the fact that in most radar systems, the receiver has at least partial knowledge of the transmitted waveform. A typical reason as to why this knowledge may be only partial is that the phase of the transmitted waveform may not be known. In the case of a *bistatic* radar system—one in which the transmitter and receiver are not collocated—the time of transmission as well as the phase of the carrier may not be known. The general shape and polarization of the transmitted waveform will be known, however, and this knowledge will exhibit itself in the design of the radar receiver. Some systems have very detailed knowledge of the transmitted waveform at the receiver. Systems with quadrature detectors, for example, generally have phase-lock between the transmitter and receiver, with a small portion of the transmitter carrier fed directly to the receiver.

As many of the modern applications of radar deal not only with detecting and locating objects but also with measuring the scattering characteristics of these objects, a method of characterizing measurement performance becomes desirable. In making these measurements of the target-scattering characteristics, the primary goal is to obtain information about the target. So an appropriate choice of per-

formance metric might be one that measures the information obtained about the target by radar observation. Such a metric would allow one not only to determine how much information can be determined about a target but also to design a radar system that determines the maximum amount of information about a target for a given set of design constraints—for example, a fixed bandwidth and a maximum average power.

In addition, such a metric would allow one to determine whether or not a radar is capable of performing its desired task. Standard approaches to the problem of radar performance have been developed in the case of radar target detection, but in the case of target identification or the precision measurement of a target parameter, a metric based on information obtained about the target would be useful. We will now investigate a simple example that illustrates the usefulness of such a metric.

As will be discussed in detail in Chapter 4, the range resolution of a radar system is inversely proportional to the bandwidth of the radar system. As a result, if one is making measurements of an object such as a ship, for example, greater bandwidth is required to measure the ship's fine structure than is required to make rough measurements of the ship's size. Now if we represent the received radar waveforms by discrete samples, the Sampling Theorem requires a greater sampling rate to represent the signal of greater bandwidth—that corresponding to the measurement of fine structure—than it does to represent the signal of lesser bandwidth corresponding to the rough measurement of the ship's size. Intuitively, this makes sense. One would expect the measurement of the ship's fine structure to contain

more information than the rough measurement of the ship's size, as the measurement of the ship's fine structure would include within it a measurement of the ship's size as well as the additional information characterizing the ship's structure. What this example points out is the following: *Different types of radar measurements may require different information acquisition capabilities of a radar system.* In this thesis we will analyze this point quantitatively.

This simple example, while intuitively appealing, ignores many significant points. There is not a direct relationship between the number of samples and the amount of information conveyed. A single sample in the first case does not necessarily provide the same amount of information as a single sample in the second case. Increasing the bandwidth may increase the number of samples required to represent the target, taking full advantage of the information contained in this larger bandwidth, but if the same amount of total energy is available for obtaining the samples in both cases, less energy is available per sample in the larger bandwidth case. As a result, in the presence of noise, the individual samples will not be as reliable as in the larger bandwidth case. This example does, however, bring up the question of the trade-off between the number of samples and the energy per sample in the design of a radar system that maximizes the information obtained about the target. In terms of practical system design, the question could be stated as follows: *How does one distribute the transmitted power in frequency to maximize the information obtained when there is a constraint on the average transmitter power?*

In this thesis, the radar measurement process will be analyzed in terms of

information theory. The measure of information proposed by Shannon [1.6] will be used to examine the ability of a radar system to collect information about a target—that is, the information content of radar measurements will be examined. We will then examine the information-theoretic design of radar systems so that their measurements will yield the maximum amount of information about the target being observed. A particularly interesting result of this analysis is that radar systems designed for optimal target detectability—the most common approach to radar system design today—differ significantly from those that would be designed to collect the maximum amount of information about a target known to be present. This suggests that a different approach to radar waveform design should be used in the case of radars designed for target identification and precision measurement than for target detection.

1.2. Overview.

In Chapter 2 of this thesis, we will briefly review information theory by defining the information-theoretic quantities that will be used in this thesis. This will be done primarily to establish the notation to be used in the remainder of this thesis and to highlight general concepts in information theory. More detailed and in-depth discussions of information theory are found in References [1.7–1.9]. After our brief review, we will look at the relationship between information theory and radar measurement. This will provide the motivation for a more detailed look at the application of information theory to the problem of radar measurement by pointing out the relationship between information acquired and target identification

capabilities or measurement error.

In Chapter 3, we will introduce the *Radar/Information Theory Problem*, which forms the basis for most of the rest of this thesis. The *Radar/Information Theory Problem* addresses the problem of what a radar should transmit in order to obtain the maximum amount of information about the target under observation. We then look at the solution of this problem in the general case.

Before the results of Chapter 3 can be applied to specific radar problems, physical models of the radar measurement process must be established. This is done in Chapter 4. Here we will look at statistical models of electromagnetic scattering and the appropriateness of these models to the radar measurement process. We will begin with a brief survey of scattering models and then will examine in detail those that will be useful in our analysis.

In Chapter 5 we will apply the results of Chapter 4 to the problem of optimum target detection. We will derive a new extension of the matched filter concept for radar developed by North [1.10]. North's matched filter assumed that the target being observed was a point target—one with no significant spatial extent. The results in Chapter 5 take into account the interference patterns arising in the scattered electric field because of the spatial extent of a target distributed in space. The result of our analysis is a design procedure that gives both a realizable waveform and a receiver filter, which together have optimum detection properties in additive Gaussian noise with an arbitrary power spectral density.

In Chapter 6, the information-theoretic design of radar waveforms is considered.

Here we look at the problem of designing radar waveforms that maximize the rate of information transfer about the target to the radar. These results are of particular interest in the case of radar design for target identification and measurement. A design procedure is then developed for radar systems that maximize the rate of information extraction about the target. We then compare the optimum detection waveforms of Chapters 5 to those derived in Chapter 6, which provide the maximum amount of information about the target. This results in a very interesting physical and information-theoretic interpretation of the waveforms for optimal detection and information extraction.

In Chapter 7, we apply mutual information to the analysis of imaging radar. Here we examine the information content of radar images generated by both real and synthetic aperture radars. We also give an information-theoretic interpretation to some well-known results in the processing of radar images.

In Chapter 8, we summarize our results from the previous chapters and examine their implications for the design of radar waveforms and systems.

1.3. Chapter 1 References.

1.1 Alvarez, L. W., *Adventures of a Physicist*, Basic Books, New York, NY, 1987.

1.2 Ridenour, L. N., *Radar System Engineering*, McGraw-Hill, New York, NY, 1948.

1.3 Brookner, E., *Radar Technology*, Artech House, Dedham, MA, 1977.

1.4 Skolnik, M. L., *Introduction to Radar Systems*, 2nd ed., McGraw-Hill, New York, NY, 1980.

1.5 Stimson, G. W., *Introduction to Airborne Radar*, Hughes Aircraft Company, El Segundo, CA, 1983.

1.6 Shannon, C. E., "A Mathematical Theory of Communication," *Bell Sys. Tech. J.* 27, 1948. pp. 379-423, 623-656. Reprinted in C. E. Shannon and W. W. Weaver, *The Mathematical Theory of Communication*, Univ. Ill. Press, Urbana, IL, 1949.

1.7 McEliece, R. J., *The Theory of Information and Coding*, Addison-Wesley, Reading, MA, 1977.

1.8 Blahut, R. E., *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

1.9 Gallager, R. G., *Information Theory and Reliable Communication*, John Wiley and Sons, New York, NY, 1968.

1.10 North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems", RCA Laboratory Report PTR-6C; reprinted in *Proc. IEEE*, 51 (July 1963), pp. 1016-27.

CHAPTER 2

THE INFORMATION-THEORETIC ANALYSIS OF RADAR

In this chapter, we will examine the applicability of information theory to the analysis of radar systems. In Section 2.1, we will consider the rationale for examining radar systems from the viewpoint of information theory. This is done by examining the similarities and differences between communication systems—which have been analyzed for the last forty years using information theory with great success—and radar systems—to which information theory was early considered applicable [2.13], but to which information theory has not been traditionally applied. We will see that a radar system can be seen as a "communication system" of an unusual type, and that it is indeed reasonable to use information theory in the analysis of radar systems. In Sections 2.2 and 2.3, we will briefly review the main points of information theory. The purpose here is twofold. First, it serves to establish the notation to be used throughout this thesis with regard to information, but more importantly it serves to introduce its basic principles to those radar engineers reading this thesis, who may not be familiar with information theory. These sections are by no means a complete introduction to information theory. An excellent introduction and reference can be found in [2.1], on which much of these two sections is based. Having introduced the basic concepts of information theory, we will look at their relevance to the radar measurement problem in Section 2.4. There we will see the relevance

of information theory to radar measurement performance.

2.1. The Information-Theoretic Analysis of Radar Systems.

Information theory, introduced by Shannon in 1948 [2.2], provided a new and powerful framework for the analysis of communication systems. But information theory was more than a new tool for examining previously known results in communication theory. It opened up a whole new realm of previouly unknown results about the communication process, by offering new and fundamental insights into its nature. This in turn spawned the field of error-correcting codes, providing the means by which many of Shannon's results would be realized in practice. Information theory has had a profound impact on the design of today's communication systems and the methods of transferring information from one point to another, or from one time to another in the case of computer memories.

In this thesis, we will examine radar systems using information theory in order to derive some insights which information theory can provide into the design of radar systems. It may not be immediately apparent that information theory provides an appropriate or even desirable framework in which to analyze radar systems. In this section, we will motivate such an approach by examining the similarities and differences of radar systems and communication systems. References [2.3] and [2.4] provide many of the details for the analysis of communications and radar systems, both generally and specifically.

Consider the block diagram of a generic communication system shown in Fig. 2.1. The purpose of the communication system is to transfer information from the source

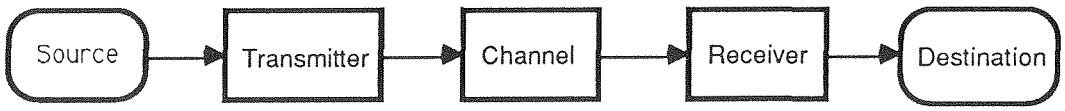


Figure 2.1. Block Diagram of a Communication System.

to the destination. These two terminals are separated by the *channel*. The transfer of information is done using the transmitter and receiver. We will now examine the function of each block in the communication system in Fig. 2.1 and determine how they fit into the overall function of the communication system.

The *source* generates the message to be sent to the destination. The source may represent any of a number of message-generating mechanisms. It could be a person speaking, a thermometer measuring temperature, a camera imaging a planetary surface, or a computer memory with some file to be transferred. A characteristic of all sources in communication systems must be that the output of the source is not known at the destination before the source output is transmitted across the channel. If it were known, there would be no purpose in the communication system, as it would not provide any information, either in the mathematical sense or the informal sense, about the source to the destination. Thus, there is at least some a *priori* uncertainty in the output of the source at the destination.

The *transmitter* maps the output from the source into a suitable form for the channel across which the message is to be transferred. Thus, the transmitter matches the message from the source to the channel. The form of the transmitter then will be a function of both the message source and the channel. For example, if the source is human speech and the channel is free space, a microwave transmitter using wide deviation frequency modulation may be a good match of source to channel, whereas if the source is a computer system with digital data files to be transmitted and the channel is an optical fiber, a pulse modulated laser may be a good choice as a transmitter matching the source to the channel.

The *channel* is determined by the medium across which the information is to be transferred from the source to the destination. It may, for example, represent free space, the earth's atmosphere, a coaxial cable, or an optical fiber when the source and destination are spatially separated. It may also represent a magnetic, optical, or solid-state storage medium when the source and destination are separated in time, such as in a computer program, where data are stored away in memory and then recalled at some later time. A characteristic of almost all physical channels is that when a waveform is being transmitted across the channel, it is distorted by physical processes present within the channel. As a result, the waveform received at the channel output may not be interpreted as corresponding to the message produced by the source. These distortions may be due to any of a number of physical processes. For example, in the case of electromagnetic waveform transmission, thermally excited molecules emit electromagnetic radiation due to charge motion. This

exhibits itself in radio communication systems as additive noise. Also present in some radio channels is multipath fading that is due to the received signals from multiple propagation paths. Any change in the geometry of the transmitter, receiver, or propagation paths can cause the amplitude of the received signal to change. This exhibits itself as a multiplicative noise. In optical communication systems, noise in the solid-state detectors can cause the detector to determine the presence of a pulse from the transmitter when none is present.

The *receiver* observes the output of the channel and makes a decision as to what the transmitted message was. Its function is complicated by noise and distortion present in the channel. After making its estimate of the message from the source, it passes this decision on to the destination. The *destination* represents or displays the point to which the information is to be transmitted. It receives the decision on the received message from the receiver and functions as the end user.

The role of information theory in the design of such a communication system is to determine how the transmitter and receiver will be designed in order to efficiently and reliably transmit the information from the source to the destination, given the characteristics of the source and the channel. That is, given a model of what types of messages the source is likely to produce and a characterization of the distorting or noise properties of the channel, information theory determines whether or not a receiver and transmitter can be designed so as to provide reliable communication. It also provides some guidance as to how to design this transmitter and receiver. The actual implementation of the transmitter and receiver often involves coding,

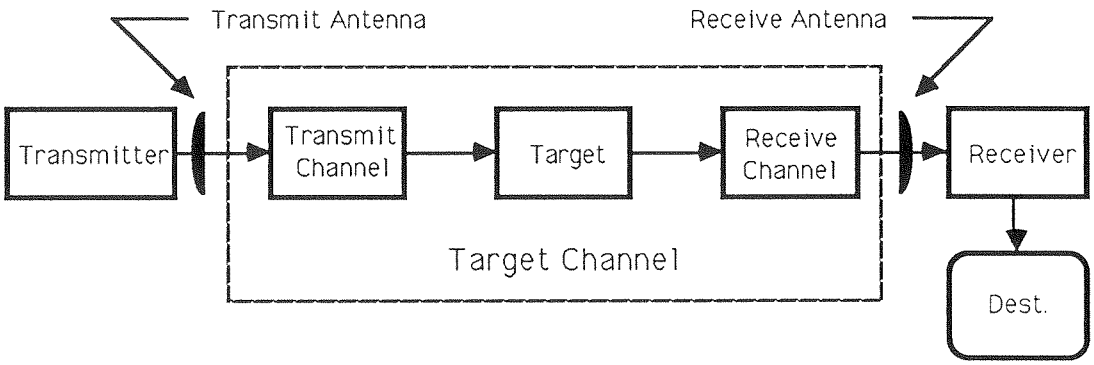


Figure 2.2. Block Diagram of a General Radar System.

and the field of error-correcting codes [2.1] provides the details of the specific coding design. Information theory does, however, provide specific insights into the higher level of system design. For example, in Reference [2.5], Shannon considers the spectral distribution of transmitter power for optimal communications for an additive Gaussian noise channel with a given noise power spectral density.

We now examine a general radar system. Consider the block diagram of the radar system shown in Fig. 2.2. The purpose of this radar system is to determine characteristics of the target. We will now consider the function and effect of each of these blocks in carrying out the purpose of radar.

The *transmitter* of the radar system provides the electromagnetic waves by which the target will be probed. Typical radar systems have transmitters operating at frequencies between 1 GHz and 50 GHz, although lower frequencies are some-

times used for over-the-horizon, long-range detection radars where the increased reflectivity of the ionosphere can be used to good advantage [2.4]. Generally, larger average transmitter powers, although not necessarily greater received powers, can be obtained at lower frequencies, since the physical size of the components is greater, and thus greater heat dissipation can take place. In some applications, however, such as airborne and spaceborne systems, the limitation on average transmitter power is determined by size and weight constraints on the transmitter power supplies [2.4]. The frequency and bandwidth of the transmitted waveform also play a significant role in the measurement capabilities of the radar system. This will be examined in detail in Chapter 4.

We next encounter the *transmit antenna*, whose function it is to radiate the transmitter waveform into space at the target. In the case of the abstract communication channel, we did not explicitly note the antenna or the method of coupling the transmitter to the channel. The reason for this was that it was not necessary in order to understand the function of the communication system. In the case of radar, however, the antenna plays a more pronounced role. The characteristics of the antenna—its beamwidth, polarization, and geometry—significantly affect the outcome of the resulting radar measurements [2.6]. Antenna system design is a significant part of the design of a radar system and in particular determines the angular resolution with which radar measurements can be made.

We next encounter the block labelled as the *target channel*, made up of the *transmit channel*, the *target*, and the *receive channel*. In a real radar system, we

cannot directly separate the effects of these three elements by measurement. Any processing that allows us to separate these effects is dependent on the estimation of the the individual effects based on assumed physical models of the individual processes involved and measurements involving all of the effects acting simultaneously.

The *transmit channel* represents the path across which the transmitted waveform travels as it propagates from the transmit antenna to the target. For free space or the earth's atmosphere, this channel has little effect other than to attenuate the signal. This attenuation is the result of space loss (dispersion of the signal in space) and atmospheric attenuation. Space loss is defined so as not to be a frequency-selective process, so all frequencies in the transmitted waveform are attenuated equally when they reach the target. Atmospheric absorption, on the other hand, is highly frequency-selective, as it is due to the absorption of electromagnetic energy by specific molecules making up the atmosphere. The energy is absorbed at frequencies corresponding to the molecular bonds of these molecules [2.3,pp.23,35]. In some cases, the transmit channel can distort the transmitted electromagnetic field in still more severe ways. For example, plasmas made up of charged particles interact with electromagnetic waves propagating through them. This results in a decrease in velocity of propagation, dispersion (frequency-dependent propagation time differences for different waveform spectral components) of the waveform, and both amplitude and phase scintillation (rapid fluctuations about the mean value) of the wave passing through the plasma [2.7]. Such plasmas exist in the solar system and could exhibit themselves as transmit channel distortions if the radar system

being modeled is used in radar astronomy measurements at high enough frequencies (although these frequencies would be higher than those currently used in radar astronomy [2.14]).

The *target* is the object on which measurements are being made. It is the element in the block diagram of which there is generally the greatest uncertainty, yet about which the most information is desired. In fact, the sole purpose of the radar system is to obtain information about the target to reduce this uncertainty. If there were no uncertainty about the target—about its presence, geometry, physical characteristics, and motion—there would be no need to make radar measurements of it. In a sense, the target plays a parallel role in the radar system to that played by the source in the communication system. The transmitted waveform, having been radiated by the transmit antenna, propagates through space until it reaches the target. Once the transmitted electromagnetic wave is intercepted by the target, it is scattered by the target. This scattering can be viewed as a retransmission by the target of the wave impinging on the target, but the resulting scattered wave is modified as a function of the target's geometry and physical characteristics.

The *receive channel*, as shown in Fig. 2.2, includes the medium between the target and the antenna. As in the case of the channel in the communication system, there will generally be additive thermal noise present. In the case of the radar system, the effect of the additive noise may be particularly severe. This is because of the two-way transmission path from the transmitter to the target and the target to the receiver. After suffering the one-way path loss squared, the received signal from

the target may be very faint. This is especially true if the target is at a considerable distance. In the case of a monostatic (single antenna) radar, for instance, the received signal power is inversely proportional to the fourth power of the range from the antenna to the target.

The *receiver* detects the signal intercepted by the receive antenna and performs any processing required to extract the desired information about the target. This processing may involve filtering out unwanted noise, spectral analysis in order to determine the spectrum of the Doppler shift resulting from target motion, and any radar or target motion compensation that may be required to extract the desired target characteristics from the received signal. So we should and do include any postprocessing of the detected signal in the receiver block as well.

The output of the receiver goes to the *destination*, representing the end user of the information obtained about the target. The output of the receiver that is provided to the destination could take on any of a number of forms. For example, it could be a radar map of the target in a geological mapping system, the position and velocity of the target in an air-traffic control system, the Doppler spectrum of the target in a system used to study surface scattering behavior in remote sensing, or the class of a target in a target-identification system.

If we compare the block diagram of the communication system in Fig. 2.1 to the block diagram of the radar system in Fig. 2.2, we see that several of the elements in the two systems are similar. Both systems have a transmitter that radiates electromagnetic energy, and both systems have a receiver that detects and

processes electromagnetic energy. Both systems also have a channel through which the signal passes as it propagates from transmitter to receiver, but the nature of these channels is quite different.

In the communication system, the information to be received by the destination has its origin at the source. The channel acts primarily as the medium across which the message is to be transferred, and as a result, acts primarily to corrupt the message sent from the source to the destination. In the radar, however, it is the channel, or the *target channel* to be precise, that contains the source of information. In the communication system, then, the transmitter responds to the message generated by the source and transmits a waveform to the receiver. In the radar system, the target responds to the transmitted waveform and scatters a modified waveform to the receiver. The source of uncertainty to be reduced in the scattered waveform is only the result of the target channel itself, since the receiver has knowledge of the transmitted waveform. Thus, the radar system functions by probing the target and measuring its response. So the target itself acts as the source of information or the "message source." The transmitter merely provides the energy in an appropriate form for it to do so.

The target will not, however, respond to differing waveforms of identical energy in identical ways. While this is true of the idealized point targets encountered in theoretical radar analysis, scattering from targets of spatial extent generates interference patterns. These interference patterns can differ significantly for transmitted waveforms made up of different frequency components. Thus, the question arises:

How does the shape or frequency content of the transmitted waveform affect the amount of information obtained about the target by radar measurement? In this thesis, we will examine this question, using information theory. The following two sections provide a brief introduction and review of information theory.

2.2. Information Theory and Discrete Random Variables.

Let X be a discrete random variable (finite or countable) taking on values from a set $R_X = \{x_1, x_2, \ldots\}$. For each $x \in R_X$, let $p(x) = P\{X = x\}$, the probability distribution of X. We wish to obtain a measure of the information obtained by observing X. Equivalently, we wish to obtain a measure of the a *priori* uncertainty in the outcome of X. In order to do this, we will define for each $x \in R_X$ a quantity $I(x)$ called the *self-information* of x, by

$$I(x) = -\log p(x). \tag{2.1}$$

The base of the logarithm is left unspecified and determines the units of $I(x)$. The two most commonly used bases are base-2 and base-e, yielding units of *bits* and *nats*, respectively. In this thesis, base-e or *natural* logarithms will be used almost exclusively when a specific base needs to be specifed. This is done to simplify calculations. Nats can be converted to bits by dividing by a scale factor of $\ln 2$.

Fig. 2.3 shows a graph of $I(x)$ as a function of $p(x)$. For any given $x \in R_X$, $p(x) \in [0, 1]$. Note that as the event $X = x$ becomes less probable, the self-information $I(x)$ increases, and that as the event becomes more probable, the self-information decreases. This then says that the occurrence of an unlikely event—"It

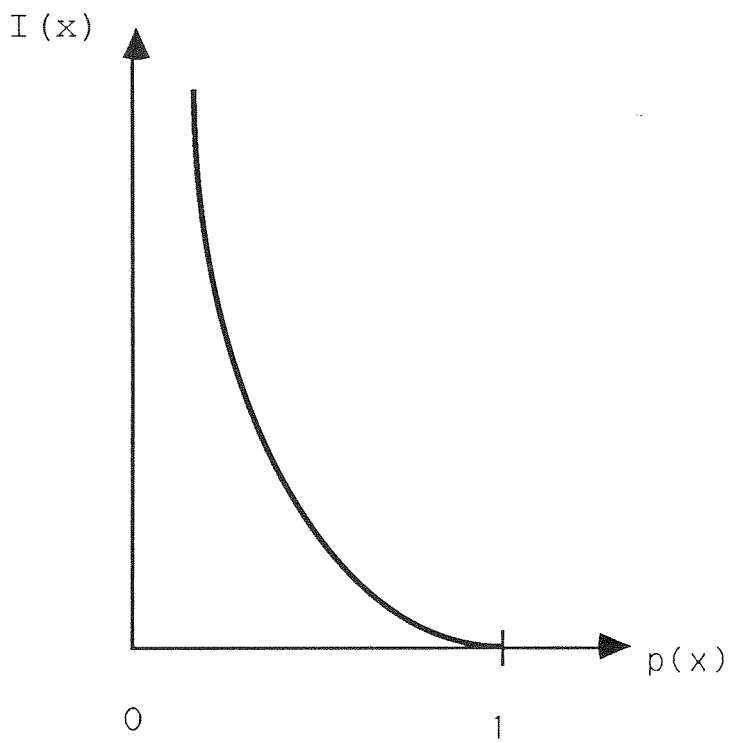


Figure 2.3. Self-Information $I(x)$ versus $p(x)$.

is raining in Los Angeles"—contains more self-information than a likely event—"It is raining somewhere." Intuitively, this is appealing, since if we are told the obvious, we feel we have obtained very little "information," whereas when we are told the not so obvious or the unlikely, we feel we have obtained significant "information." Note as well the relationship between self-information and the certainty of an event's occurrence. When we are quite certain an event will occur, the self-information

provided by its occurrence is small. As we become less certain of its occurrence, the self-information provided by its occurrence becomes larger.

Defining $I(x)$ as we have, we see that $I(X)$ is a new random variable, defined in terms of the random variable X. The expectation of the random variable $I(X)$, denoted by $H(X)$, is called the *entropy* of X, and is given by

$$H(X) = -\sum_{x \in R_X} p(x) \log p(x). \tag{2.2}$$

If $p(x)$ is equal to zero for any $x \in R_X$, $p(x) \log p(x)$ is defined as $p(x) \log p(x) = 0$, the limit as p approaches zero from above:

$$\lim_{p \to 0^+} p \log p = 0.$$

If X and Y are jointly distributed discrete random variables taking on values from $R_X = \{x_1, x_2, \ldots\}$ and $R_Y = \{y_1, y_2, \ldots\}$ respectively, and having joint probability distribution $p(x,y) = P\{X = x, Y = y\}$, then the self-information in the joint occurrence of $X = x$ and $Y = y$ is $I(x,y) = -\log p(x,y)$. The joint entropy $H(X,Y)$ of X and Y is

$$H(X,Y) = -\sum_{x \in R_X} \sum_{y \in R_Y} p(x,y) \log p(x,y). \tag{2.3}$$

We now consider several properties of the entropy function. These properties are easily proved, as in References [2.1, 2.8]:

1. Let $\mathbf{p} = (p(x_1), p(x_2), \ldots)$ be the probability distribution of X. Then $H(X)$ is continuous in $\mathbf{p}$.
2. $H(X) \geq 0$, with equality if and only if all but one of the $p(x_j)$ are equal to zero.
3. For a finite random variable X with $R_X = \{x_1, \ldots, x_r\}$, $H(X) \leq \log r$, with equality if and only if for all $x_j \in R_X$, $p(x_j) = 1/r$.
4. If X and Y are jointly distributed random variables,

$$H(X,Y) \leq H(X) + H(Y),$$

with equality if and only if X and Y are statistically independent.

5. $H(X)$ is a convex $\cap$ function of $\mathbf{p}$.

We have previously noted that the entropy $H(X)$ is the mean value of the self-information $I(X)$, and that $I(x)$ was small for events that occurred with great certainty. It follows then that $H(X)$ must in some sense measure the average uncertainty of the outcome of X. When $H(X)$ is large, there is a greater *a priori* uncertainty in the outcome of the random variable X. We will now consider the five properties of entropy listed above in light of viewing entropy as a measure of *a priori* uncertainty.

Property 1 states that entropy of X is continuous in the probability distribution of X. This can be interpreted as saying that small changes in the probability distribution of X produce small changes in the entropy of X. This is a reasonable property of an uncertainty measure, as it states that very similar distributions have very similar *a priori* uncertainty.

Property 2 states that entropy measures the uncertainty of X as a positive quantity, and that if the outcome of X is certain, the entropy of X is zero. This is an intuitively appealing property of an uncertainty measure.

Property 3 states that if there are a fixed, finite number of outcomes of X, then there is an upper bound on the entropy of X, and this upper bound occurs only when each of the outcomes is equally likely. This is a reasonable property of an uncertainty measure, as intuitively, the case of equiprobable outcomes has the greatest uncertainty—no single outcome is more favorable than another.

Property 4 states that for two random variables, the joint entropy of their outcomes is less than or equal to the sum of the entropies of the individual outcomes. Furthermore, the joint entropy of their outcome is exactly the sum of the entropies of the individual outcomes when the two outcomes are statistically independent. This, too, is a reasonable property of an uncertainty measure. It states that if the two outcomes are statistically independent, then the uncertainty of the outcome of the two jointly is the sum of the individual uncertainties. If, however, there is statistical dependence between the two outcomes, observation of the outcome of one **must** provide some information on the outcome of the other, and in such an

instance the joint uncertainty of the two random variables is strictly less than the sum of the two individual uncertainties.

Property 5, convexity down, while not having the direct intuitive appeal of the previous four properties, is significant in that it greatly simplifies finding distributions that maximize the entropy of X, as well as solving related optimization problems in information theory. This is a fortunate benefit of adopting entropy as an uncertainty measure.

As can be seen from these five properties of entropy, entropy is a very reasonable measure of the *a priori* uncertainty in the outcome of a discrete random variable. In fact, although several measures of uncertainty that have some of the above properties have been proposed [2.9], only the entropy function (or actually $\alpha H(X)$, where α is any positive real number) satisfies all of these properties [2.10]. Also, the entropy function measures the average code length needed to specify a source [2.1,2.3], but we shall not pursue this point of view here.

Consider again the jointly distributed discrete random variables X and Y, with joint distribution $p(x,y)$. Let $p(x)$ be the (marginal) distribution of X and $p(y)$ be the distribution of Y. Then the conditional probability distribution of Y conditioned on $X = x$ is

$$p(y|x) = \frac{p(x,y)}{p(x)}, \tag{2.4}$$

and the conditional probability distribution of X conditioned on $Y = y$ is

$$p(x|y) = \frac{p(x,y)}{p(y)}. \tag{2.5}$$

Given that $X = x$, the conditional entropy of Y, i.e., the entropy of Y conditioned on $X = x$, is

$$H(Y|X = x) = -\sum_{y \in R_Y} p(y|x) \log p(y|x). \tag{2.6}$$

Similarly, the entropy of X conditioned on $Y = y$ is

$$H(X|Y = y) = -\sum_{x \in R_X} p(x|y) \log p(x|y). \tag{2.7}$$

We define the conditional entropy $H(Y|X)$ by averaging over all $x \in R_X$. This yields

$$\begin{aligned} H(Y|X) &= -\sum_{x \in R_X} p(x) \sum_{y \in R_Y} p(y|x) \log p(y|x) \\ &= -\sum_{x \in R_X} \sum_{y \in R_Y} p(x,y) \log p(y|x). \end{aligned} \tag{2.8}$$

Similarly,

$$\begin{aligned} H(X|Y) &= -\sum_{y \in R_Y} p(y) \sum_{x \in R_X} p(x|y) \log p(x|y) \\ &= -\sum_{x \in R_X} \sum_{y \in R_Y} p(x,y) \log p(x|y). \end{aligned} \tag{2.9}$$

We now define a quantity of central importance in information theory, known as *mutual information*. Consider again the jointly distributed discrete random variables X and Y. The mutual information between X and Y, denoted $I(X;Y)$, is defined as

$$I(X;Y) = H(X) - H(X|Y). \tag{2.10}$$

The mutual information represents the difference between the *a priori* uncertainty in X and the uncertainty in X after observing Y. So $I(X;Y)$ is a measure of the

amount of information Y provides about X. It is interesting to note that

$$\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= -\sum_{x\in R_X} p(x)\,\log p(x) + \sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log p(x|y) \\
&= \sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log\frac{p(x|y)}{p(x)} \\
&= \sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log\frac{p(y|x)}{p(y)} \\
&= -\sum_{y\in R_Y} p(x)\,\log p(y) + \sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log p(y|x) \\
&= H(Y) - H(Y|X).
\end{aligned}$$

Thus, there is a symmetry in X and Y exhibited by $I(X;Y)$, since $I(X;Y)$ is not only equal to $H(X) - H(X|Y)$, but also to $H(Y) - H(Y|X)$. Thus, we have

$$I(X;Y) = I(Y;X) = \sum_{x\in R_X}\sum_{y\in R_Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}, \tag{2.11}$$

and we see not only that $I(X;Y)$ is the information that observation of Y provides about X, but also that $I(X;Y)$ is the information that observation of X provides about Y; hence, the name *mutual information* given to $I(X;Y)$.

The mutual information $I(X;Y)$ has several interesting properties. We note some of them [2.1]:

1. $I(X;Y) \geq 0$,with equality if and only if X and Y are statistically independent.
2. $I(X;Y) = I(Y;X)$, as was previously shown.

3. $I(X;Y)$ is a convex $\cap$ function of the input probabilities $p(x)$.

4. $I(X;Y)$ is a convex $\cup$ function of the conditional probabilities $p(y|x)$.

We can generalize the definitions of entropy and mutual information to include not only discrete random variables, but discrete random vectors as well. Let $\mathrm{X} = (X_1, \ldots, X_m)$ be an m-dimensional random vector of random variables $X_1, \ldots, X_m$, with X taking on values from the set R_{X}. Let $p(\mathrm{x}) = P\{\mathrm{X} = \mathrm{x}\} = P\{X_1 = x_{j_1}, \ldots, X_m = x_{j_m},\}$ be the probability distribution on X. Then the entropy of the random variable X is

$$H(\mathrm{X}) = -\sum_{\mathrm{x} \in R_{\mathrm{X}}} p(\mathrm{x}) \log p(\mathrm{x}). \tag{2.12}$$

Let $\mathrm{Y} = (Y_1, \ldots, Y_n)$ be an n-dimensional random vector of random variables $Y_1, \ldots, Y_n$, with Y taking on values from the set R_{Y}. Let $p(\mathrm{y}) = P\{\mathrm{Y} = \mathrm{y}\} = P\{Y_1 = y_{j_1}, \ldots, Y_n = y_{j_n},\}$ be the probability distribution on Y. Let X and Y be jointly distributed with joint distribution $p(\mathrm{x},\mathrm{y}) = P\{\mathrm{X} = \mathrm{x}, \mathrm{Y} = \mathrm{y}\}$. Let $p(\mathrm{x}|\mathrm{y}) = P\{\mathrm{X} = \mathrm{x}|\mathrm{Y} = \mathrm{y}\}$ and $p(\mathrm{y}|\mathrm{x}) = P\{\mathrm{Y} = \mathrm{y}|\mathrm{X} = \mathrm{x}\}$. Then the joint entropy of X and Y is

$$H(\mathrm{X},\mathrm{Y}) = -\sum_{x \in R_{\mathrm{X}}} \sum_{y \in R_{\mathrm{Y}}} p(\mathrm{x},\mathrm{y}) \log p(\mathrm{x},\mathrm{y}). \tag{2.13}$$

The entropy of Y conditioned on X is

$$\begin{aligned} H(\mathrm{Y}|\mathrm{X}) &= -\sum_{x \in R_{\mathrm{X}}} p(\mathrm{x}) \sum_{y \in R_{\mathrm{Y}}} p(\mathrm{y}|\mathrm{x}) \log p(\mathrm{y}|\mathrm{x}) \\ &= -\sum_{x \in R_{\mathrm{X}}} \sum_{y \in R_{\mathrm{Y}}} p(\mathrm{x},\mathrm{y}) \log p(\mathrm{y}|\mathrm{x}). \end{aligned} \tag{2.14}$$

Figure 2.4. Block Diagram of a Simple Communication System.

The entropy of X conditioned on Y is

$$\begin{aligned} H(\mathrm{X}|\mathrm{Y}) &= -\sum_{y\in R_Y} p(\mathrm{y}) \sum_{x\in R_X} p(\mathrm{x}|\mathrm{y}) \log p(\mathrm{x}|\mathrm{y}) \\ &= -\sum_{x\in R_X} \sum_{y\in R_Y} p(\mathrm{x},\mathrm{y}) \log p(\mathrm{x}|\mathrm{y}). \end{aligned} \tag{2.15}$$

The mutual information between the random vectors X and Y is

$$\begin{aligned} I(\mathrm{X};\mathrm{Y}) &= H(\mathrm{X}) - H(\mathrm{X}|\mathrm{Y}) \\ &= H(\mathrm{Y}) - H(\mathrm{Y}|\mathrm{X}) \\ &= \sum_{x\in R_X} \sum_{y\in R_Y} p(\mathrm{x},\mathrm{y}) \log \frac{p(\mathrm{x},\mathrm{y})}{p(\mathrm{x})p(\mathrm{y})}. \end{aligned} \tag{2.16}$$

We will now introduce the concept of the communication channel in the discrete case. Consider again the discrete random variables X and Y. We have previously looked at them abstractly as jointly distributed random variables, but now we will examine how they might arise in a typical discrete communication system. Consider the communication system depicted in Fig. 2.4. Here we have a transmitter that transmits a message X consisting of one of m symbols from an alphabet R_X =

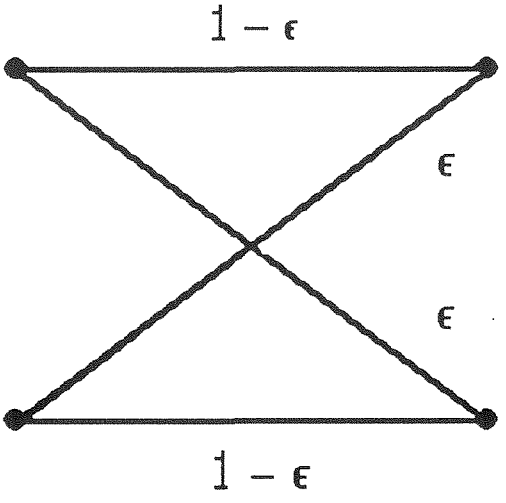


Figure 2.5. Binary Symmetric Channel

$\{x_1, \ldots, x_m\}$. The message is the random variable X with distribution $p(x)$. The transmitted message X proceeds through the channel, which stochastically maps to a discrete random variable Y at the channel output. Y takes on one of n symbols from the alphabet $R_Y = \{y_1, \ldots, y_n\}$. The stochastic mapping of the channel is governed by the conditional probability distribution

$$p(y|x) = P\{Y = y|X = x\}. \tag{2.17}$$

The resulting channel output Y thus has the marginal probability distribution

$$p(y) = \sum_{x \in R_X} p(x)p(y|x).$$

As a concrete example of a discrete communication channel, consider the binary symmetric channel (BSC) shown in Fig. 2.5. Here the input and output alphabets are identical; that is, $R_X = R_Y = \{0,1\}$. We will assume the input probability distribution is

$$p(x) = \begin{cases} p, & x = 0; \\ 1-p, & x = 1. \end{cases} \tag{2.18}$$

The conditional probability distribution that governs the channel behavior is

$$p(y|x) = \begin{cases} 1-\epsilon, & x = 0, y = 0; \\ \epsilon, & x = 0, y = 1; \\ \epsilon, & x = 1, y = 0; \\ 1-\epsilon, & x = 1, y = 1. \end{cases} \tag{2.19}$$

If we assume that the goal of the BSC is to reproduce faithfully the input symbol at its output, then it can be seen that the probability of error for the BSC is ϵ, and the probability of correct transmission is $1-\epsilon$. Of course, we would intuitively expect that the smaller ϵ, the better the channel. Let us verify this by calculating the mutual information between X and Y for the BSC. From Eq.s (2.2) and (2.9) we have

$$H(X) = -p\log p - (1-p)\log(1-p),$$

and

$$\begin{aligned} H(X|Y) &= -p(1-\epsilon)\log(1-\epsilon) - p\epsilon\log\epsilon \\ &\quad - (1-p)(1-\epsilon)\log(1-\epsilon) - (1-p)\epsilon\log\epsilon \\ &= -\epsilon\log\epsilon - (1-\epsilon)\log(1-\epsilon). \end{aligned}$$

Thus, from Eq. (2.10) we have (noting that $I(X;Y) \geq 0$)

$$I(X;Y) = \max[\,0\,, -p\log p - (1-p)\log(1-p) + \epsilon\log\epsilon + (1-\epsilon)log(1-\epsilon)].$$

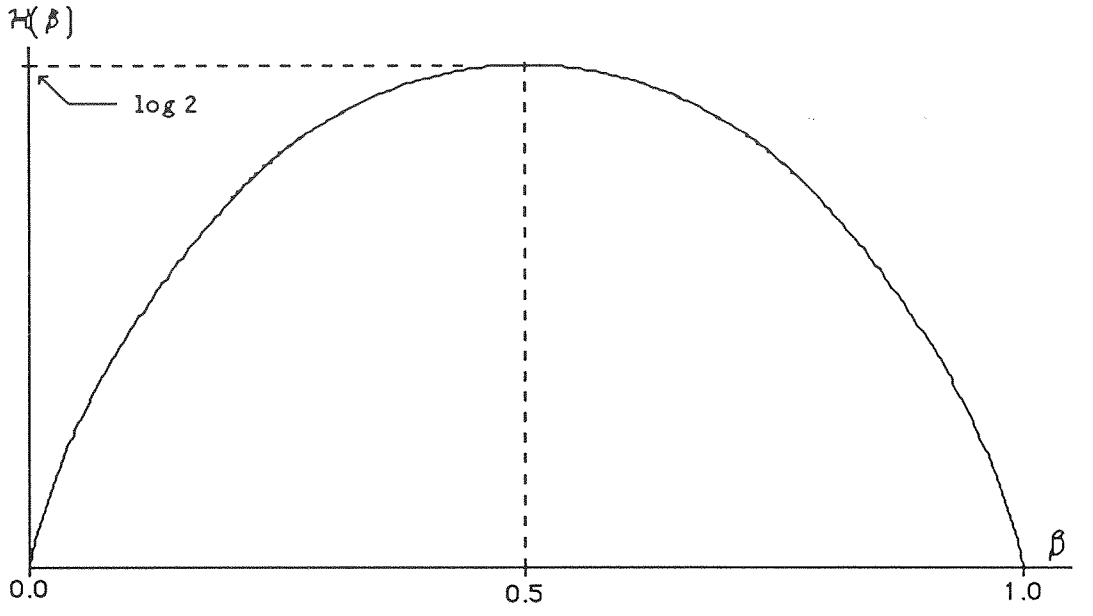


Figure 2.6. The Binary Entropy Function $\mathcal{H}(\beta)$.

If for all $\beta \in [0,1]$, we define (recalling as above $0 \log 0 = 0$)

$$\mathcal{H}(\beta) = -\beta \log \beta - (1-\beta)\log(1-\beta), \tag{2.20}$$

then

$$I(X;Y) = \max[\,0\,, \mathcal{H}(p) - \mathcal{H}(\epsilon)], \tag{2.21}$$

where $\mathcal{H}(\beta)$ is called the *binary entropy function.*

A plot of $\mathcal{H}(\beta)$ versus β is shown in Fig. 2.6. We note that $\mathcal{H}(\beta)$ is symmetric about $\beta = 0.5$, that it is a convex $\cap$ function on $\beta \in [0,1]$, and that it takes on its maximum value of $\log 2$ at the point $\beta = 0.5$. In order to examine the behavior of

$I(X;Y)$ as given in Eq. (2.21), we note that for all p and ϵ such that $\mathcal{H}(p) > \mathcal{H}(\epsilon)$, $I(X;Y)$ is just the difference $\mathcal{H}(p) - \mathcal{H}(\epsilon)$, and for all other p and ϵ, $I(X;Y)$ is zero. From Eq. (2.21) we see that $I(X;Y)$ is maximized when $p = 0.5$. This can be seen by noting that Eq. (2.21) is maximized when $\mathcal{H}(p)$ is maximized, and $\mathcal{H}(p)$ is maximized when $p = 0.5$. So the BSC provides the maximum mutual information between input and output when the input symbols 0 and 1 are equiprobable. We also see that $I(X;Y)$ is maximized when $\mathcal{H}(\epsilon) = 0$, which occurs when either $\epsilon = 0$ or $\epsilon = 1$. That this is true for $\epsilon = 0$ makes intuitive sense, since this is the case in which no errors whatsoever occur. That this is true for the case of $\epsilon = 1$ may at first seem surprising, since this is the case of the channel's *always* making an error. But if a binary symmetric channel *always* makes an error, $X = 0$ *always* produces $Y = 1$ and $X = 1$ *always* produces $Y = 0$. Such a channel is really a very reliable channel. The receiver simply assumes that when $Y = 1$ is received, $X = 0$ was sent, and that when $Y = 0$ is received, $X = 1$ was sent.

Of particular interest is the fact that for all p and ϵ such that $\mathcal{H}(p) \geq \mathcal{H}(\epsilon)$, $I(X;Y) = 0$, and the channel is of no use whatsoever. In such cases, the channel error rate ϵ satifies the following inequality:

$$p \leq \epsilon \leq 1 - p \qquad \text{for all } p \text{ and } \epsilon \text{ such that} \qquad \mathcal{H}(p) \geq \mathcal{H}(\epsilon).$$

Such channels are of no use in conveying information because in effect there is greater uncertainty in the channel's performance than there is in what message was sent. If one chooses the most likely symbol X to be transmitted as the transmitted symbol, the probability of error is no greater than that obtained by any decision

rule based on actual observation of the channel output.

In the above example, we noted that for any given binary symmetric channel and its associated $\mathcal{H}(\epsilon)$, when $p = 0.5$, the mutual information $I(X;Y)$ was maximized. In general, for any channel with mutual information $I(X;Y) = H(X) - H(X|Y)$, there will be a maximum value of $I(X;Y)$ over all probability distributions $p(x)$. This quantity is known as the *capacity* of the channel. We will denote the channel capacity by $\mathcal{C}$. Thus, the channel capacity is defined as

$$\mathcal{C} = \max_{p(x)}\{I(X;Y)\}. \tag{2.22}$$

The channel capacity is the largest rate at which information can be transferred across the channel. Shannon [2.2] showed more importantly not only that the capacity $\mathcal{C}$ of a channel was the maximum rate at which information could be sent across the channel, but also that, with proper encoding, information can be sent across the channel at any rate less than $\mathcal{C}$ with arbitrarily small error. The capacity of a channel sets an upper bound on the rate at which highly reliable information can be transferred across the channel.

2.3. Information Theory and Continuous Random Variables.

In the previous section, we examined entropy and mutual information for discrete random variables and vectors, but usually, and particularly in this thesis, we are interested in the case of continuous random variables and vectors. We will now obtain expressions for the mutual information between two continuous random variables.

Consider a continuous random variable X defined on the real line $\mathbf{R}$ with probability density function (PDF) $f(x)$. The *differential entropy* of X is defined by

$$h(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)\,dx. \tag{2.23}$$

The differential entropy is *not* the limit of the Riemann sum obtained by discretizing the real line into intervals of size Δx [2.1]. For such a case we would have a Riemann sum of the form

$$\begin{aligned}\lim_{\Delta x\to 0} H(X) &= \lim_{\Delta x\to 0}\{-\sum_k f(x_k)\log f(x_k)\Delta x\}\\ &= \lim_{\Delta x\to 0}\{-\sum_k f(x_k)\log f(x_k)\} + \lim_{\Delta x\to 0}\{-\log\Delta x\}\\ &= -\int_{-\infty}^{\infty} f(x)\log f(x)\,dx - \lim_{\Delta x\to 0}\{\log\Delta x\}.\end{aligned} \tag{2.24}$$

But

$$-\lim_{\Delta x\to 0}\{\log\Delta x\} = \infty,$$

so the limit as $\Delta x \to 0$ of the Riemann sum will in general diverge, and this limit will be of little interest or use to us. Note, however, that the first term of the limit in Eq. 2.24 is what we have defined as the differential entropy of X.

Assume we have another continuous random variable Y defined on $\mathbf{R}$ and that X and Y are jointly distributed with joint PDF $f(x,y)$. Assume also that $f(x)$ is the marginal pdf of X, $f(y)$ is the marginal pdf of y, $f(x|y)$ is the PDF of X conditioned on Y, and $f(y|x)$ is the PDF of Y conditioned on X. We can then define the joint differential entropy

$$h(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(x,y)\,dx\,dy, \tag{2.25}$$

and the conditional differential entropies

$$h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(x|y)\,dx\,dy, \tag{2.26}$$

and

$$h(Y|X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(y|x)\,dx\,dy. \tag{2.27}$$

Consider again the limiting process of the Riemann sum approximation to an entropy, but this time using the conditional entropy $H(X|Y)$. In the limit as $\Delta x \to 0$, we obtain

$$\begin{aligned}
\lim_{\Delta x\to 0}\lim_{\Delta y\to 0} H(X|Y) &= \lim_{\Delta x\to 0}\lim_{\Delta x\to 0}\{-\sum_j\sum_k f(x_j,y_k)\Delta x\Delta y\log f(x_j|y_k)\Delta x\} \\
&= \lim_{\Delta x\to 0}\lim_{\Delta x\to 0}\{-\sum_j\sum_k f(x_j,y_k)\Delta x\Delta y\log p(x_j|y_k) \\
&\qquad -\lim_{\Delta x\to 0}\lim_{\Delta x\to 0}\{\sum_j\sum_k f(x_j,y_k)\Delta x\Delta y\log\Delta x\} \\
&= -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\log f(x|y)\,dx\,dy - \lim_{\Delta x\to 0}\log\Delta x.
\end{aligned} \tag{2.28}$$

Again, we note that the second term of the Riemann sum diverges as $\Delta x \to 0$, so this approximation is of little use to us either. We do, however, note that the first

term that results in the limit is what we have defined as the differential entropy of X conditioned on Y.

We considered evaluating the entropy of a continuous random variable by discretizing it into intervals of length Δx and forming a Riemann sum representing the entropy of the discretized random variable, and then letting Δx go to zero. This was not useful because the Riemann sums diverge as Δx goes to zero. In both Eqs. (2.28) and (2.29), the problem was a term of the form $-\log \Delta x$. If, however, we do the discretization of both $H(X)$ and $H(X|Y)$, form the mutual information $I(X;Y)$, and then take the limit as $\Delta x \to 0$, we obtain

$$\begin{aligned} \lim_{\Delta x \to 0} I(X;Y) &= \lim_{\Delta x \to 0} \{H(X) - H(X|Y)\} \\ &= h(X) - \lim_{\Delta x \to 0} \log \Delta x - [h(X|Y) - \lim_{\Delta x \to 0} \log \Delta x] \\ &= h(X) - h(X|Y) + \lim_{\Delta x \to 0} [\log \Delta x - \log \Delta x] \\ &= h(X) - h(X|Y). \end{aligned} \tag{2.29}$$

Thus, in the limit, as $\Delta x \to 0$, we have that the mutual information is well defined and is given by

$$I(X;Y) = h(X) - h(X|Y). \tag{2.30}$$

So the mutual information between two jointly distributed continuous random variables X and Y is given by Eq. (2.30). A more detailed proof of this relationship is given in Reference [2.1]. The above argument, however, illustrates the process by which Eq. (2.30) is obtained.

The mutual information $I(X;Y)$ between the continuous random variables X and Y has all four properties of mutual information mentioned previously for the

mutual information between discrete random variables. On the other hand, the differential entropy of continuous random variables does *not* have many of the properties of the entropy of discrete random variables. We will not consider these in detail, as we will refer to differential entropy only when calculating calculating mutual information

We can generalize the definitions of differential entropy and mutual information for continuous random variables to include continuous random vectors as well. Let $\mathbf{X} = (X_1, \ldots, X_m)$ be an m-dimensional vector taking on values from the m-dimensional vector space $\mathbf{R}^m$ defined over the field of real numbers $\mathbf{R}$. Let $f(\mathbf{x})$ be the PDF of the random vector $\mathbf{X}$. The differential entropy of the random variable $\mathbf{X}$ is given by

$$\begin{aligned} h(\mathbf{X}) &= -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) \log f(\mathbf{x})\, dx_1 \cdots dx_m \\ &= -\int_{\mathbf{R}^m} f(\mathbf{x}) \log f(\mathbf{x})\, d\mathbf{x}. \end{aligned} \tag{2.31}$$

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be an n-dimensional random vector taking on values from the n-dimensional vector space $\mathbf{R}^m$ defined over the field of real numbers $\mathbf{R}$. Let $f(\mathbf{y})$ be the PDF of $\mathbf{Y}$. Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly distributed with joint PDF $f(\mathbf{x}, \mathbf{y})$. Let $f(\mathbf{x}|\mathbf{y})$ be the PDF of $\mathbf{X}$ conditioned on $\mathbf{Y} = \mathbf{y}$ and $f(\mathbf{y}|\mathbf{x})$ be the PDF of $\mathbf{Y}$ conditioned on $\mathbf{X} = \mathbf{x}$. Then the joint differential entropy of $\mathbf{X}$ and $\mathbf{Y}$ is

$$h(\mathbf{X}, \mathbf{Y}) = -\int_{\mathbf{R}^n} \int_{\mathbf{R}^m} f(\mathbf{x}, \mathbf{y}) \log f(\mathbf{x}, \mathbf{y})\, d\mathbf{x}\, d\mathbf{y}. \tag{2.32}$$

The differential entropy of $\mathbf{X}$ conditioned on $\mathbf{Y}$ is

$$h(\mathbf{X}|\mathbf{Y}) = -\int_{\mathbf{R}^n} \int_{\mathbf{R}^m} f(\mathbf{x}, \mathbf{y}) \log f(\mathbf{x}|\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}. \tag{2.33}$$

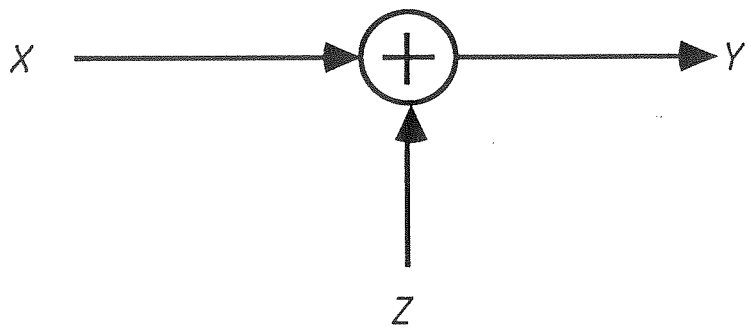


Figure 2.7. Additive Gaussian Noise Channel.

The differential entropy of Y conditioned on X is

$$h(\mathrm{Y}|\mathrm{X}) = -\int_{\mathrm{R}^n}\int_{\mathrm{R}^m} f(\mathrm{x},\mathrm{y}) \log f(\mathrm{y}|\mathrm{x})\, d\mathrm{x}\, d\mathrm{y}. \tag{2.34}$$

The mutual information between X and Y is

$$\begin{aligned} I(\mathrm{X};\mathrm{Y}) &= h(\mathrm{X}) - h(\mathrm{X}|\mathrm{Y}) \\ &= h(\mathrm{Y}) - h(\mathrm{Y}|\mathrm{X}) \\ &= \int_{\mathrm{R}^n}\int_{\mathrm{R}^m} f(\mathrm{x},\mathrm{y}) \log \frac{f(\mathrm{x},\mathrm{y})}{f(\mathrm{x})f(\mathrm{x})}\, d\mathrm{x}\, d\mathrm{y}. \end{aligned} \tag{2.35}$$

We will now consider an example of a continuous communication channel that will illustrate many of the principles of information theory for continuous random variables. Consider the continuous communication channel shown in Fig. 2.7. Here we have an additive noise channel in which the channel input is a Gaussian random variable X, with zero mean and variance σ_X^2. As X proceeds through the channel,

another Gaussian random variable Z with zero mean and variance σ_Z^2 is added to it. The input X represents a message that we wish to transmit across the channel, and Z represents some additive contaminating noise. The resulting output of the channel is a random variable Y, which is the sum of X and Z. We will assume that X and Z are statistically independent. It is well known that the sum of two independent Gaussian random variables is a Gaussian random variable whose mean is the sum of the means of the two component random variables and whose variance is the sum of the variances of the component random variables [2.11]. Thus, Y is a Gaussian random variable with zero mean and variance $\sigma_Y^2 = \sigma_X^2 + \sigma_Z^2$.

Consider a Gaussian random variable S with zero mean and variance σ^2. The PDF $f(s)$ of S is

$$f(s) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-s^2}{2\sigma^2}}.$$

Then the differential entropy of S is

$$\begin{aligned} h(S) &= -\int_{-\infty}^{\infty} f(s) \log f(s)\, ds \\ &= -\int_{-\infty}^{\infty} f(s) \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-s^2}{2\sigma^2}} \right] ds \\ &= \int_{-\infty}^{\infty} f(s) \left[\log \sqrt{2\pi}\sigma + \frac{s^2}{2\sigma^2} \right] ds \\ &= \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{2} \\ &= \frac{1}{2} \log(2\pi e\sigma^2). \end{aligned}$$

Thus, since Y is a zero-mean Gaussian random variable with variance $\sigma_Z^2 + \sigma_X^2$, it

has differential entropy

$$h(Y) = \frac{1}{2}\log\left(2\pi e(\sigma_Z^2 + \sigma_X^2)\right).$$

It is straightforward to show analytically that $h(Y|X) = h(Z)$ [2.8], so

$$h(Y|X) = \frac{1}{2}\log(2\pi e\sigma_Z^2).$$

The mutual information $I(X;Y)$ between X and Y is thus

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= \frac{1}{2}\log\left(2\pi e(\sigma_Z^2 + \sigma_X^2)\right) - \frac{1}{2}\log(2\pi e\sigma_Z^2) \\ &= \frac{1}{2}\log\left(\frac{\sigma_Z^2 + \sigma_X^2}{\sigma_Z^2}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right). \end{aligned}$$

The capacity of the additive Gaussian noise channel with noise variance σ_Z^2 and input X with variance σ_X^2 is found by maximizing over all input density functions with variance σ_X^2. Thus, if $\mathbf{F}$ is the set of all PDF's with variance σ_X^2, the capacity of the additive Gaussian noise channel is

$$C = \max_{f(x)\in\mathbf{F}} I(X;Y).$$

It can be shown [2.1] that the maximum is attained by any Gaussian random variable, regardless of its mean value, with variance σ_X^2. So the $I(X;Y)$ calculated above for the additive Gaussian noise channel achieves capacity. Thus, the capacity C of the additive Gaussian noise channel with signal variance σ_X^2 and noise variance σ_Z^2 is

$$\begin{aligned} C &= \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right) \\ &= \frac{1}{2}\log(1 + r), \end{aligned}$$

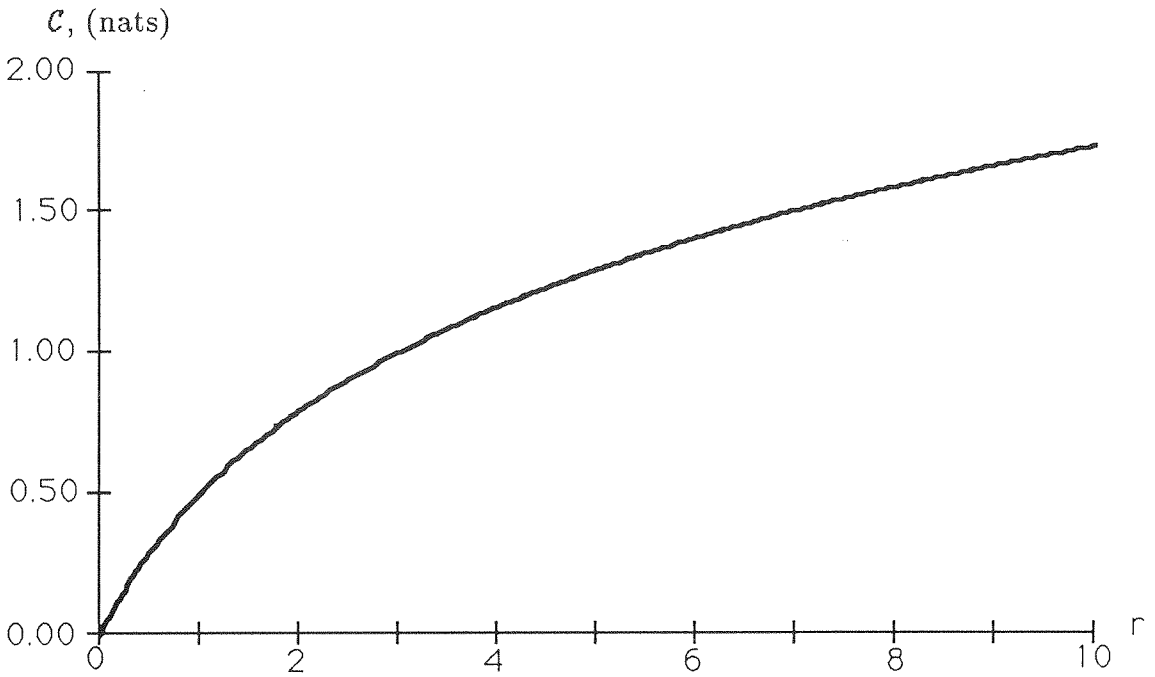


Figure 2.8. Capacity C of the Additive Gaussian Noise Channel versus $r = \sigma_X^2/\sigma_Z^2$.

where $r = \sigma_X^2/\sigma_Z^2$, and can be interpreted as the signal-to-noise ratio. In a real transmitter, we would be sending zero-mean signals since our real constraint is power, and then the variance would be equal to the power (mean-square value). A plot of C versus r is shown in Fig. 2.8.

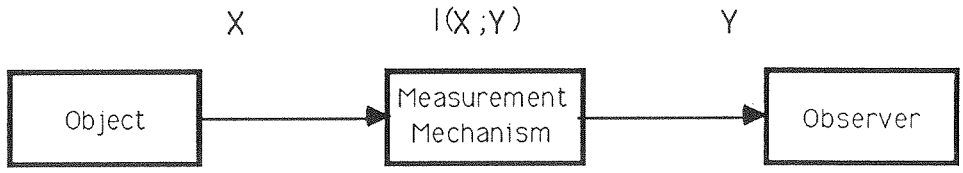


Figure 2.9. Block Diagram of a Measurement System.

2.4. Mutual Information and Radar Measurement Performance.

In Sections 2.2 and 2.3, we reviewed the basic definitions and principles of information theory, concentrating on the expressions for the mutual information between two random variables or vectors. We will now examine the usefulness of mutual information in describing radar performance.

As noted in Section 2.1, a radar system is a measurement system that makes measurements of a target in order to obtain information about the target, decreasing the a *priori* uncertainty about the target. This being the case, let us examine the measurement process in general in light of information theory.

Consider the measurement system shown in Fig. 2.9. Here we have an object to be measured, a measurement mechanism, and an observer. We assume that the random vector $\mathbf{X}$ consists of parameters characterizing the object we wish to measure, the measurement mechanism maps them into the random vector $\mathbf{Y}$, and the

observer observes Y and from this observation determines the desired information about X. The measurement mechanism, like all real measurement mechanisms, is assumed to have inherent inaccuracies, and thus the measurements obtained with it contain error. This can be modeled by assuming that the measurement mechanism stochastically maps the random vector X to the random vector Y. We will assume that the mutual information between X and Y is $I(\mathrm{X};\mathrm{Y})$.

As an example of such a situation, consider a situation in which the object being measured is a rectangular box whose dimensions X_1, X_2, and X_3 are unknown (we could assume, for example, that the box was randomly selected from a large collection of boxes of various sizes). The parameter vector X would be $\mathrm{X} = (X_1, X_2, X_3)$. Assume that the measurement mechanism consists of some method of measuring these dimensions—perhaps a person with a crudely marked ruler. The measurements Y_1, Y_2, and Y_3 obtained as measurements of X_1, X_2, and X_3 would form the measurement vector Y; that is, $\mathrm{Y} = (Y_1, Y_2, Y_3)$. Since X and Y are jointly distributed random vectors, they have a mutual information $I(\mathrm{X};\mathrm{Y})$ between them.

The mutual information $I(\mathrm{X};\mathrm{Y})$ between two random vectors X and Y tells us how much information observation of Y provides about X; that is, $I(\mathrm{X};\mathrm{Y})$ is the amount of information that the measurement Y provides about the object parameter vector X. The significance of this is that the greater this mutual information, the more information we are obtaining about the object by our measurement. We will examine this in two ways. We will first examine this by determining the maxi-

mum number of equiprobable classes into which we can assign X by observation of Y, and then we will relate $I(\mathrm{X};\mathrm{Y})$ to average measurement error through use of the *rate distortion function*.

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 partition based on observing the Y generated by the measurement process. For any decision rule assigning X to a partition based on observation of X, 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

$$N = \left\lfloor e^{I(\mathrm{X};\mathrm{Y})} \right\rfloor . \tag{2.36}$$

Here $\lfloor \alpha \rfloor$ is the largest integer less than or equal to α.

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

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

Let $\mathcal{N} = e^{I_0}$; then

$$N_0 \leq \mathcal{N} < N_0 + 1.$$

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

$$\ln N_0 \leq \ln \mathcal{N} < \ln (N_0 + 1) ,$$

or equivalently,

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

That sufficient information is being conveyed across the measurement channel to classify $\mathbf{X}$ into one of N_0 equiprobable classes, given observation of $\mathbf{Y}$, follows from the fact that only $\ln N_0$ nats are required to do so, while I_0 nats are currently being transferred. That it is not possible to classify $\mathbf{X}$ into one of $N_0 + 1$ equiprobable classes follows from the fact that by Shannon's Theorem for the noisy channel, this cannot be done without the channel's transferring $\ln(N_0 + 1)$ nats of information. But the measurement mechanism cannot possibly do so, since $I_0 < \ln(N_0 + 1)$.

When examining the accuracy of measurements, it is common to talk about the accuracy in terms of some error criterion, for example, mean-square error or relative mean square error. It would be useful if we could relate the mutual information $I(\mathbf{X}; \mathbf{Y})$ to measure of measurement error. Fortunately, the framework for doing this in our measurement problem has already been developed in information theory. It is called *Rate Distortion Theory*, and was developed by Shannon in the last chapter of [2.2] and in [2.12]. References [2.1] and [2.3] discuss rate distortion theory and its application in communications. We will now look at the application of rate distortion theory to our measurement problem.

Referring again to Fig. 2.9, we have a measurement system by which a measurement mechanism conveys a quantity of information $I(\mathbf{X}; \mathbf{Y})$ about an object to an observer. In the measurement process, we are trying to determine the characteristics of the object parameter vector $\mathbf{X}$ from the measurement vector $\mathbf{Y}$. Because of

inaccuracies in the measurement process, we cannot generally obtain $\mathbf{X}$ perfectly, so there is an error associated with a given parameter vector $\mathbf{x}$ and a given measurement vector $\mathbf{y}$. Let us designate this error or *distortion* as $d(\mathbf{x}, \mathbf{y})$. We assume that this distortion is a non-negative function defined for all pairs of $\mathbf{x} \in R_{\mathbf{X}}$ and $\mathbf{y} \in R_{\mathbf{Y}}$. This being the case, the average distortion or error D, also known as the *fidelity criterion*, is the expectation of $d(\mathbf{x}, \mathbf{y})$. Thus,

$$D = E\{d(\mathbf{x}, \mathbf{y})\}. \tag{2.37}$$

The *rate distortion function*, $R(\delta)$, is defined as

$$R(\delta) = \min\{I(\mathbf{X}, \mathbf{Y}) : D \leq \delta\}. \tag{3.38}$$

The minimization is over all measurement mechanisms that satisfy the condition that the the fidelity criterion D is less than or equal to δ. The rate distortion function $R(\delta)$ gives the minimum possible rate at which information must be transferred by a measurement mechanism in order to have an average error or distortion D less than or equal to δ.

It is significant to note that for $\delta_1 < \delta_2$, $R(\delta_1) \geq R(\delta_2)$. Thus, $R(\delta)$ is a non-increasing function of δ. So the smaller the average error δ, the larger is the minimum required information rate $R(\delta)$ required of the measurement mechanism. This intuitively makes sense, as it says that if greater accuracy is required in the measurements, the measurement mechanism must provide more information about the object being measured.

We will now look at a simple example of the use of the rate distortion function in a measurement problem. Suppose we wish to measure a scalar parameter X, which

is known to be a zero-mean, Gaussian random variable with variance σ_X^2. Assume that the measurement mechanism is known to provide a measurement $Y = X + Z$, where Z is a zero-mean, Gaussian random variable with variance σ_Z^2. Assume X and Z to be statistically independent. Then the measurement problem is described by the additive Gaussian noise channel of Fig. 2.7. Now assume our error or distortion measure is the mean-square error between X and Y. Then the error or distortion measure is

$$d(x, y) = (x - y)^2 ,$$

and

$$D = E\left\{(x - y)^2\right\} = E\left\{Z^2\right\} = \sigma_Z^2 .$$

From the example in Section 2.3 we have

$$I(X;Y) = \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_Z^2}\right).$$

The resulting rate distortion function is [2.8]

$$R(\delta) = \frac{1}{2}\log\left(\frac{\sigma_x^2}{\delta}\right).$$

A plot of $R(\delta)$ is shown in Fig. 2.10.

In general, we see that the greater the mutual information between the parameters we wish to measure and the measurements themselves, the more we can say about the target. 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 $I(\mathrm{X};\mathrm{Y})$, the larger the number of classes. In the case of the rate distortion

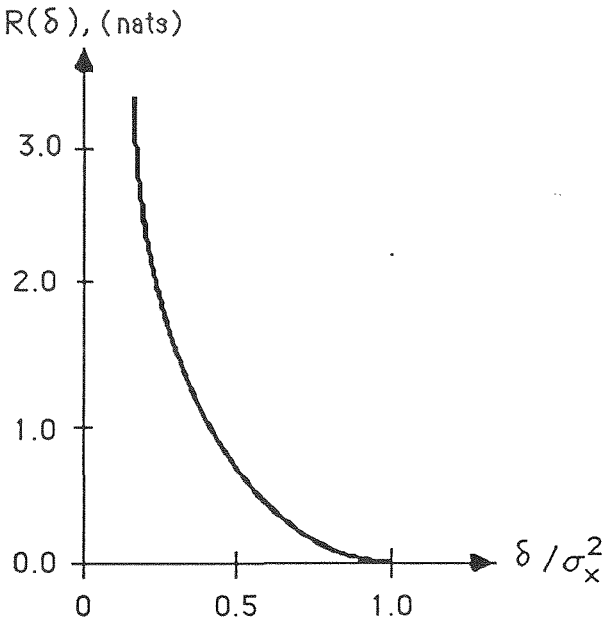


Figure 2.10. Rate Distortion Function of the Gaussian Measurement Example.

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.

In applying these results to the measurement mechanism known as 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, the better we can expect our system performance to be. In the next chapter, we will look at the problem of the design of radar systems. This will be done using the formalism of *The Radar/Information Theory Problem.*

2.5. Chapter 2 References.

2.1 McEliece, R. J., *The Theory of Information and Coding*, Addison-Wesley, Reading, MA, 1977.

2.2 Shannon, C. E., "A Mathematical Theory of Communication," *Bell Sys. Tech. J.* 27, 1948. 379-423, 623-656. Reprinted in C. E. Shannon and W. W. Weaver, *The Mathematical Theory of Communication*, Univ. Ill. Press, Urbana, IL, 1949.

2.3 Pierce, J. R. and E. C. Posner, *Introduction to Communication Science and Systems*, Plenum, New York, NY, 1980.

2.4 Stimson, G. W., *Introduction to Airborne Radar*, Hughes Aircraft Company, El Segundo, CA, 1983.

2.5 Shannon, C. E., "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, 11-21, 1949.

2.6 Ullaby, F. T., R. K. Moore, and A. K. Fung, *Microwave Remote Sensing*, vol. 2, Addison-Wesley, Reading, MA, 1982.

2.7 Posner, E. C. and R. Stevens, "Deep Space Communications," *Encyclopedia of Physical Science and Technology*, Academic Press, New York, NY, 1987. pp. 165-185.

2.8 Blahut, R. E., *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA, 1987.

2.9 Simon, J. C., *Patterns and Operators: The Foundations of Data Representation*, McGraw-Hill, New York, NY, 1986.

2.10 Khinchin, A. I. *Mathematical Foundations of Information Theory,* Dover, New York, NY, 1957.

2.11 Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics,* MacMillan, London, England, 1970.

2.12 Shannon, C. E., "Coding Theorems for a Discrete Source with a Fidelity Criterion," *Information and Decision Processes,* R. E. Machol, ed., McGraw-Hill, New York, NY, 1960.

2.13 Woodward, P. M., *Probability and Information Theory with Applications to Radar*, Pergamon Press, London, England, 1953.

2.14 Ostro, S. J., "Planetary Radar Astronomy,"*Encyclopedia of Physical Science and Technology,* Academic Press, New York, NY, 1987. pp. 612-634.

CHAPTER 3

THE RADAR/INFORMATION THEORY PROBLEM

In the previous chapter, we examined the use of information theory to characterize the performance of radar systems. In particular, we examined how the mutual information between a vector of target parameters and the measurement obtained by the radar characterized the radar measurement performance. As a result of this analysis, we concluded that the greater this mutual information between target and measurement, the better the radar's ability to reduce uncertainty about the target—the purpose of making the radar measurement in the first place. This leads us to the question: *How do we design a radar system in order to maximize the mutual information between the target and the radar measurement of the target under a given set of design constraints?*

In this chapter, we will begin to address this question by studying a problem we call *The Radar/Information Theory Problem.* The Radar/Information Theory Problem allows us to formulate a mathematical framework that will provide insight into the problem of radar waveform design for maximum mutual information between the target and the measurement obtained at the radar receiver.

3.1. General Formulation of the Radar/ Information Theory Problem.

Consider the situation shown in Fig. 3.1. Here we have a radar system making measurements on a target. The transmitter transmits an N-tuple $\mathbf{X} \in \mathbf{A}^{(N)}$ of waveforms, where $\mathbf{A}^{(N)}$ is an ensemble of N-tuples of waveforms, with probability distribution $\Pi(\mathbf{x})$, and $\mathbf{A} = \{\alpha_1, \ldots, \alpha_m\}$ is the set of possible transmitted waveforms. The transmitted N-tuple $\mathbf{X}$ proceeds through the target channel $\mathbf{C}$, which stochastically maps it into an N-tuple $\mathbf{Y} \in \mathbf{B}^{(N)}$. The set $\mathbf{B}$ is the set of possible waveforms that can be output by the target channel in response to a transmitted waveform from $\mathbf{A}$. The set of possible received waveforms $\mathbf{B}$ may be either a finite, countable, or uncountable set.

The *target channel* $\mathbf{C}$ that stochastically maps $\mathbf{X}$ to $\mathbf{Y}$ can be viewed as being made up of two mechanisms. The first of these represents the scattering characteristics of the physical target that we wish to measure. We will assume that these target characteristics we are interested in measuring are characterized by a random parameter vector $\mathbf{Z} \in \Gamma^{(N)}$. In general, Γ may be a finite, countable, or uncountable set. In addition to the target-scattering mechanism characterized by $\mathbf{Z}$, in general there will also be random noise present in the target channel, which corrupts the measurement of the waveform scattered by the target. Such noise may be independent of the target, such as thermal noise in the radar receiver, or may be due to the physical target itself, as in the case of speckle noise in synthetic aperture radar (SAR) images of terrain. This noise mechanism in the target channel model includes all measurement noise processes present in the radar system, whether they are due

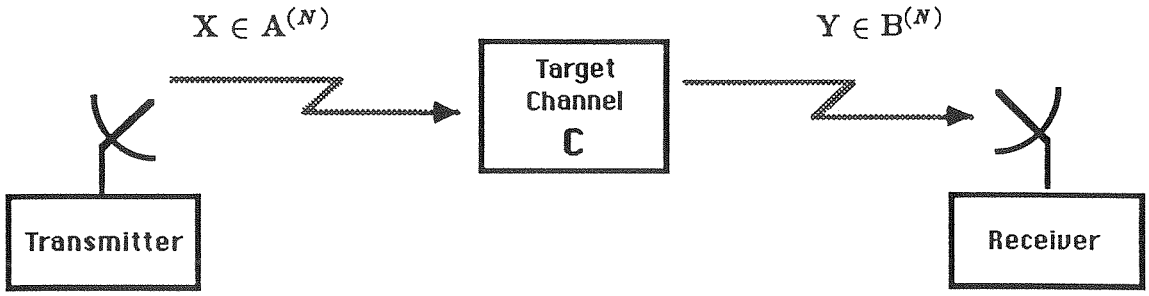


Figure 3.1. Block Diagram of Radar/Information Theory Problem.

to thermal noise in the receiver, spurious reflections or "clutter" from the physical environment in which the desired physical target is located, target "self-noise" such as speckle noise in SAR that is due to target scattering effects which we are not interested in, or to other system noise or error such as quantization noise in the radar receiver itself when the receiver is not considered available for modification.

In Fig. 3.2 we have a block diagram showing these two mechanisms in our target channel model. Here we assume we have a source that generates the random target parameter vector $\mathbf{Z} \in \Gamma^{(N)}$. The random vector $\mathbf{Z}$ then determines the behavior of the *scattering mechanism*, which is itself deterministic for a fixed $\mathbf{Z} = \mathbf{z}$. The resulting scattered signal then proceeds into the block labelled *noise mechanism*, where any random errors corrupt it. The resulting signal then leaves the noise channel and proceeds to the receiver.

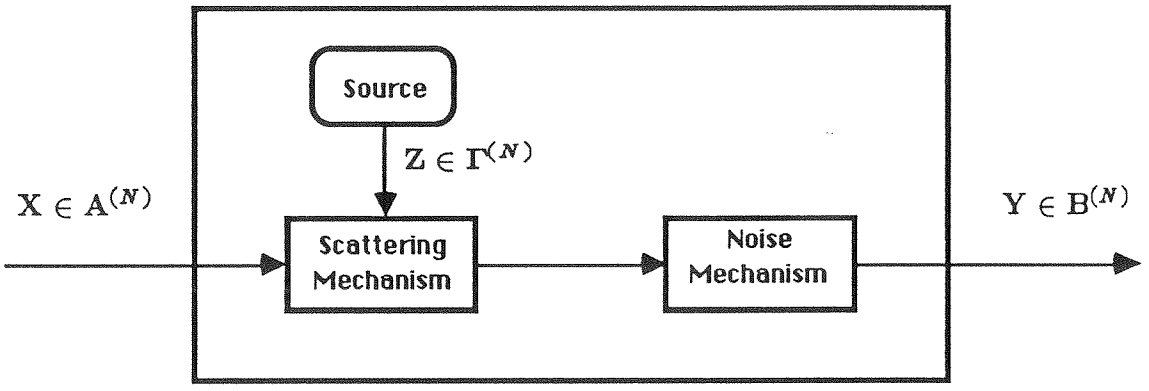


Figure 3.2. Block Diagram of The Target Channel.

From the model of Fig. 3.2, we see that the target channel output N-tuple Y is a function of the transmitted N-tuple X, the target-scattering mechanism characterized by Z, and by random noise effects in the target channel. Thus, the radar transmits X, receives Y, and from Y determines information about the target parameters Z. The *Radar/Information Theory Problem*, which we are interested in investigating, can be stated as follows: Assume that the transmitter transmits an N-tuple X through the target channel C and that the receiver as a result receives the N-tuple Y. Assume that the receiver has knowledge of the N-tuple of transmitted waveforms X. Then which distribution or distributions $\Pi(\mathrm{x})$ maximize the mutual information $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$, and what is the resulting maximum value $R(\mathrm{Y};\mathrm{Z})$ of $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$?

In this chapter, we will examine the Radar/Information Theory Problem for several different types of target models, including discrete and continuous output target models and target models that are memoryless as well as those with memory. The results of this chapter will then, in subsequent chapters, be extended to the analysis of practical radar systems and the design of radar waveforms for maximum information extraction.

3.2. The Radar/Information Theory Problem for Discrete Target Channels.

Consider the situation shown in Fig. 3.3, in which a transmitter transmits an N-tuple $\mathbf{X} \in \mathrm{A}^{(N)}$, where $\mathrm{A}^{(N)}$ is an ensemble of N-tuples of waveforms with probability distribution $\Pi(\mathbf{x})$, and $\mathrm{A} = \{\alpha_1, \ldots, \alpha_m\}$ is the set of possible transmitted waveforms. The transmitted N-tuple $\mathbf{X}$, representing the transmission of N successive waveforms by the radar, proceeds through the target channel C, which perturbs $\mathbf{X}$ in such a way that an N-tuple $\mathbf{Y} \in \mathrm{B}^{(N)}$ is received at the receiver. Here B is the set of possible received waveforms $\mathrm{B} = \{\beta_1, \ldots, \beta_n\}$. The received waveform set B is a discrete set with n members. Such a situation might arise when the radar receiver observes the received signal and discretizes or classifies it into one of n levels or classes. The channel C perturbs the transmitted N-tuple X in such a way that if $\mathbf{x}$ is a particular transmitted N-tuple, the N-tuple $\mathbf{y} \in \mathrm{B}^{(N)}$ will be received with probability $p(\mathbf{y}|\mathbf{x})$.

It is assumed that the transmitted N-tuple $\mathbf{X}$ is known at the receiver, which observes the channel output $\mathbf{Y}$. Let the random vector $\mathbf{Z}$ be the parameter vector

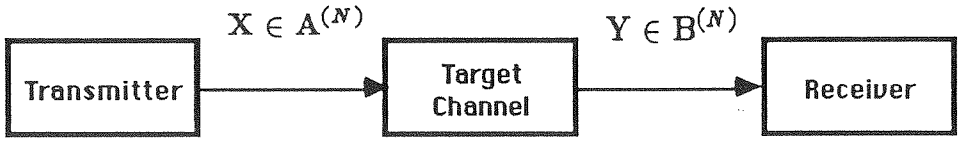


Figure 3.3. The Radar/Information Theory Problem for Discrete Target Channels.

specifying the random behavior of the scattering mechanism in the target channel C. So for a particular occurrence $\mathbf{Z} = \mathbf{z}$, if the noise mechanism of the target channel was not present, the channel output would be known with certainty if $\mathbf{z}$ was known. In analyzing the *Radar/Information Theory Problem for Discrete Target Channels*, we wish to find the input distribution $\Pi(\mathbf{x})$ that maximizes the mutual information between the channel output $\mathbf{Y}$ and the target parameter vector $\mathbf{Z}$, conditioned on the transmitted waveform N-tuple $\mathbf{X}$. This conditional information is designated $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$. In addition to finding the maximizing distribution $\Pi(\mathbf{x})$, we wish to find the maximum value of $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$ corresponding to this maximizing distribution. This maximum value of $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$ will be designated $R(\mathbf{Y};\mathbf{Z})$, and represents the maximum amount of information about $\mathbf{Z}$ that can be determined by transmitting an N-tuple $\mathbf{X}$ and observing the resulting N-tuple $\mathbf{Y}$.

We will now look at the solution of this problem for two specific types of

discrete target channel models. First, we will look at the solution for memoryless target channels, and then we will examine the solution for discrete target channels with a simple memory mechanism. In each case the maximizing distribution $\Pi(\mathbf{x})$ will be determined, as well as the maximum information rate $R(\mathbf{X};\mathbf{Z})$.

3.2.1 Solution for Discrete Memoryless Target Channels.

Consider the Radar/Information theory problem as outlined in Fig. 3.3, but with a memoryless target channel, by which we mean the propagation of each individual waveform in the transmitted N-tuple propagates through the target channel in a manner independent of the others. For any transmitted $\mathbf{x} \in \mathbf{A}^{(N)}$ and $\mathbf{y} \in \mathbf{B}^{(N)}$, we have

$$\begin{aligned} &\mathbf{x} = (x_1, \ldots, x_N), \quad x_l \in \mathbf{A}, \quad \text{for all } l = 1, \ldots, N, \\ &\mathbf{y} = (y_1, \ldots, y_N), \quad y_l \in \mathbf{B}, \quad \text{for all } l = 1, \ldots, N. \end{aligned} \tag{3.1}$$

Since the channel is memoryless, the channel transition probabilities are given by

$$p(\mathbf{y}|\mathbf{x}) = \prod_{l=1}^{N} p(y_l|x_l). \tag{3.2}$$

In order to analyze the problem for the discrete memoryless target channel, we will first examine the case where $N = 1$ (a single use of the target channel). The case for general N will then be examined as an extension of the case for $N = 1$. In analyzing the memoryless channel for $N = 1$, it will be advantageous to use the following notation: If $x \in \mathbf{A}$ and $y \in \mathbf{B}$, then

$$p_{jk} = \Pr\{y = \beta_k | x = \alpha_j\} = p(\beta_k|\alpha_j). \tag{3.3}$$

For the case where $N = 1$, a single waveform from the ensemble $\mathbf{A}$ is transmitted through the target channel and is perturbed by the target channel such that

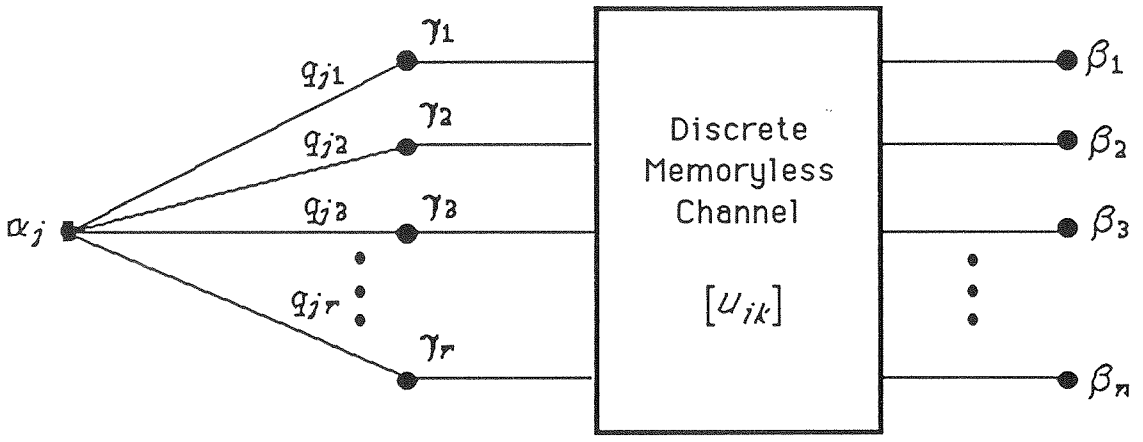


Figure 3.4. Diagram of Target Channel for $N = 1$ and $X = \alpha_j$.

an element of the set $\mathbf{B}$ is observed at the output of the channel. If the waveform $x = \alpha_j$ is transmitted, the probability of receiving $y = \beta_k$ at the channel output is p_{jk}.

Referring to Fig. 3.3, which depicts the general form of the target channel, we now construct a discrete target channel model for the case of $N = 1$. Assume that a particular $x = \alpha_j$ is transmitted. Then we can view Z as being generated by a source that produces letters from the ensemble $\Gamma = \{\gamma_1, \ldots, \gamma_r\}$ with probability distribution $Q_{\alpha_j} = (q_{j1}, \ldots, q_{jr})$. Here

$$q_{ji} = \Pr\{Z = \gamma_i | X = \alpha_j\} = q(\gamma_i | \alpha_j).$$

This source is then followed by a discrete memoryless channel (DMC), which represents the noise mechanism of Fig. 3.3. This DMC has channel transition probabilities given by the conditional distribution $u(y|z)$. In our analysis, it will be convenient to use the following notation: If $z \in \Gamma$ and $y \in \mathbf{B}$, then

$$u_{ik} = \Pr\{Y = \beta_k | Z = \gamma_i\} = u(\beta_k | \gamma_i). \tag{3.4}$$

Fig. 3.4 displays a diagram of this discrete target channel for the case where $x = \alpha_j$ is transmitted. Note that p_{jk} can be obtained from q_{ji} and u_{ik} by the relationship

$$p_{jk} = \sum_{i=1}^{r} q_{ji} u_{ik}. \tag{3.5}$$

Now let $Z \in \Gamma$ be the random target parameter that characterizes the target channel's behavior. Given that $X = \alpha_j$ is transmitted, the mutual information between Y and Z conditioned on $X = \alpha_j$ is given by

$$I(Y; Z|X = \alpha_j) = H(Y|X = \alpha_j) - H(Y|X = \alpha_j, Z). \tag{3.6}$$

Evaluating the conditional entropies $H(Y|X = \alpha_j)$ and $H(Y|X = \alpha_j, Z)$, we have

$$H(Y|X = \alpha_j) = -\sum_{k=1}^{n} p_{jk} \log p_{jk}, \tag{3.7}$$

and

$$\begin{aligned} H(Y|X = \alpha_j, Z) &= -\sum_{i=1}^{r} q_{ji} \sum_{k=1}^{n} u_{ik} \log u_{ik} \\ &= -\sum_{i=1}^{r} \sum_{k=1}^{n} q_{ji} u_{ik} \log u_{ik}. \end{aligned} \tag{3.8}$$

Thus, we have

$$I(Y; Z|X = \alpha_j) = -\sum_{k=1}^{n} p_{jk} \log p_{jk} + \sum_{i=1}^{r} \sum_{k=1}^{n} q_{ji} u_{ik} \log u_{ik}. \tag{3.9}$$

Taking the expectation of Eq. (3.9) over the ensemble A, we have

$$I(Y;Z|X) = \sum_{x \in \mathbf{A}} \Pi(x) I(Y;Z|X=x). \tag{3.10}$$

Defining $\pi_j = \Pi(\alpha_j)$, this can be rewritten as

$$I(Y;Z|X) = \sum_{j=1}^{m} \pi_j I(Y;Z|X=\alpha_j). \tag{3.11}$$

We wish to find the distribution $\Pi = \{\pi_1, \ldots, \pi_m\}$ that maximizes $I(Y;Z|X)$. We must maximize Eq. (3.11) over the set $\{\pi_1, \ldots, \pi_m\} \subset \mathrm{R}^m$ under the constraints

$$\begin{cases} \pi_j \geq 0, \quad \text{for } j = 1, \ldots, m; \\ \sum_{j=1}^{m} \pi_j = 1. \end{cases} \tag{3.12}$$

In order to do this, we define $I_{\max}^{(1)}$ as

$$I_{\max}^{(1)} \stackrel{\text{def}}{=} \max_j \{I(Y;Z|X=\alpha_j)\}. \tag{3.13}$$

Define $\hat{\mathbf{A}}^{(1)} \subseteq \mathbf{A}$ as follows:

$$\hat{\mathbf{A}}^{(1)} \stackrel{\text{def}}{=} \left\{\alpha_j \in \mathbf{A} : I(Y;Z|X=\alpha_j) = I_{\max}^{(1)}\right\}. \tag{3.14}$$

Then $I(Y;Z|X)$ is maximized by any distribution $\Pi = \{\pi_1, \ldots, \pi_m\}$ that satisfies the following conditions:

$$\begin{cases} \pi_j = 0, \quad \text{for all } j \text{ such that } \alpha_j \notin \hat{\mathbf{A}}^{(1)}; \\ \pi_j \geq 0, \quad \text{for all } j \text{ such that } \alpha_j \in \hat{\mathbf{A}}^{(1)}; \\ \sum_{j=0}^{m} \pi_j = 1. \end{cases} \tag{3.15}$$

The conditions on the maximizing distribution given by Eq. (3.15) have the simple interpretation of assigning all of the probability to those α_j that provide maximum information about Z by observation of Y. None of the probability is assigned to the set $\{\mathbf{A} - \hat{\mathbf{A}}^{(1)}\}$, as these waveforms do not provide the maximum amount of information about the target. We note that the maximum is always achieved by at least one of the α_j, and thus $\hat{\mathbf{A}}^{(1)}$ always has at least one element. The maximum value $R(Y;Z)$ of $I(Y;Z|X)$, obtained with any distribution that satisfies Eq. (3.15), is

$$R(Y;Z) = I_{\max}^{(1)} = \max_j \{I(Y;Z|X=\alpha_j)\}. \tag{3.16}$$

This is, of course, as expected.

Note that the problem has a deterministic solution as well. If we transmit any waveform x in $\hat{\mathbf{A}}^{(1)}$ with certainty, we will achieve the maximum mutual information $R(Y;Z)$. This is because the resulting probability distribution in this case—probability one of transmitting the selected waveform in $\hat{\mathbf{A}}^{(1)}$ and probability zero of transmitting all others—satisfies the requirements for a maximizing distribution as given in Eq. (3.15).

This basic analysis will be applied to the analysis of more complicated channels subsequently, but first we will examine a few examples of the discrete memoryless target channel for the case of $N = 1$.

Example 3.1: Consider the discrete memoryless target channel in Fig. 3.5. This target channel is made up of two binary discrete channels in cascade. The first represents the effects of uncertainty in the scattering mechanism and the

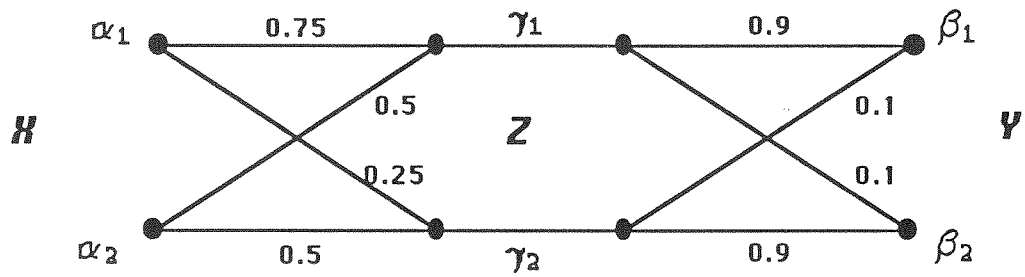


Figure 3.5. Memoryless Binary Target Channel of *Example 3.1*.

second represents observation noise. In this discrete target channel model, we have $\mathbf{A} = \{\alpha_1, \alpha_2\}$, $\Gamma = \{\gamma_1, \gamma_2\}$, and $\mathbf{B} = \{\beta_1, \beta_2\}$. The discrete memoryless channels making up the target channel have transition probabilities

$$q_{11} = \Pr\{Z = \gamma_1, X = \alpha_1\} = 0.75,$$

$$q_{12} = \Pr\{Z = \gamma_2, X = \alpha_1\} = 0.25,$$

$$q_{21} = \Pr\{Z = \gamma_1, X = \alpha_2\} = 0.50,$$

$$q_{22} = \Pr\{Z = \gamma_2, X = \alpha_2\} = 0.50,$$

$$u_{11} = \Pr\{Y = \beta_1, Z = \gamma_1\} = 0.90,$$

$$u_{12} = \Pr\{Y = \beta_2, Z = \gamma_1\} = 0.10,$$

$$u_{21} = \Pr\{Y = \beta_1, Z = \gamma_2\} = 0.10,$$

$$u_{22} = \Pr\{Y = \beta_2, Z = \gamma_2\} = 0.90.$$

It is then straightforward to calculate the overall target channel transition probabilities, which are

$$p_{11} = \Pr\{Y = \beta_1, X = \alpha_1\} = 0.7,$$

$$p_{12} = \Pr\{Y = \beta_2, X = \alpha_1\} = 0.3,$$

$$p_{21} = \Pr\{Y = \beta_1, X = \alpha_2\} = 0.5,$$

$$p_{22} = \Pr\{Y = \beta_2, X = \alpha_2\} = 0.5.$$

Applying Eqs. (3.7) and (3.8) and using base-2 logarithms, we have

$$H(Y|X = \alpha_1) = -0.7 \log_2 0.7 - 0.3 \log_2 0.3 = 0.8813 \text{ bits},$$

$$H(Y|X = \alpha_2) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ bit},$$

and

$$\begin{aligned} H(Y|X = \alpha_1, Z) &= 0.75\,[-0.9 \log 0.9 - 0.1 \log 0.1] \\ &\quad + 0.25\,[-0.9 \log 0.9 - 0.1 \log 0.1] \\ &= 0.4690 \text{ bits}, \end{aligned}$$

$$\begin{aligned} H(Y|X = \alpha_2, Z) &= 0.5\,[-0.9 \log 0.9 - 0.1 \log 0.1] \\ &\quad + 0.5\,[-0.9 \log 0.9 - 0.1 \log 0.1] \\ &= 0.4690 \text{ bits}. \end{aligned}$$

From Eq. (3.6), we have

$$I(Y; Z|X = \alpha_1) = 0.4123 \text{ bits},$$

$$I(Y; Z|X = \alpha_2) = 0.5310 \text{ bits}.$$

Thus, we have

$$I_{\max}^{(1)} = \max\{I(Y; Z|X = \alpha_1), I(Y; Z|X = \alpha_2)\} = 0.5310 \text{ bits}.$$

It follows then that

$$\hat{\mathrm{A}}^{(1)} = \{\alpha_2\}.$$

The maximizing distribution Π of X is thus obtained by assigning all of the probability to α_2; that is,

$$\Pi = (\pi_1, \pi_2) = (0, 1).$$

So $I(Y; Z|X)$ is maximized by transmitting α_2 with certainty, and $R(Y; Z)$, the resulting maximum value of $I(X; Y|X)$, is

$$R(Y; Z) = I_{\max}^{(1)} = 0.5310 \text{ bits}.$$

Example 3.2: Consider the discrete memoryless target channel shown in Fig. 3.6. This target channel is again made up of two cascaded discrete memoryless channels. The first, with three inputs and two outputs, represents the target-scattering process. The second, a noiseless binary symmetric channel, represents the observation noise—in this case no observation noise is present. Here we have $\mathrm{A} = \{\alpha_1, \alpha_2, \alpha_3\}$, $\Gamma = \{\gamma_1, \gamma_2\}$, and $\mathrm{B} = \{\beta_1, \beta_2\}$. The discrete

memoryless channels making up the target channel have transition probabilities

$$q_{11} = \Pr\{Z = \gamma_1, X = \alpha_1\} = 0.50,$$

$$q_{12} = \Pr\{Z = \gamma_2, X = \alpha_1\} = 0.50,$$

$$q_{21} = \Pr\{Z = \gamma_1, X = \alpha_2\} = 0.75,$$

$$q_{22} = \Pr\{Z = \gamma_2, X = \alpha_2\} = 0.25,$$

$$q_{31} = \Pr\{Z = \gamma_1, X = \alpha_3\} = 0.50,$$

$$q_{32} = \Pr\{Z = \gamma_2, X = \alpha_3\} = 0.50,$$

$$u_{11} = \Pr\{Y = \beta_1, Z = \gamma_1\} = 1,$$

$$u_{12} = \Pr\{Y = \beta_2, Z = \gamma_1\} = 0,$$

$$u_{21} = \Pr\{Y = \beta_1, Z = \gamma_2\} = 0.10,$$

$$u_{22} = \Pr\{Y = \beta_2, Z = \gamma_2\} = 1.$$

The overall target channel transition probabilities are thus

$$p_{11} = \Pr\{Y = \beta_1, X = \alpha_1\} = 0.50,$$

$$p_{12} = \Pr\{Y = \beta_2, X = \alpha_1\} = 0.50,$$

$$p_{21} = \Pr\{Y = \beta_1, X = \alpha_2\} = 0.75,$$

$$p_{22} = \Pr\{Y = \beta_2, X = \alpha_2\} = 0.25,$$

$$p_{31} = \Pr\{Y = \beta_1, X = \alpha_3\} = 0.50,$$

$$p_{32} = \Pr\{Y = \beta_2, X = \alpha_3\} = 0.50.$$

Applying Eqs. (3.7) and (3.8) and using base-2 logarithms, we have

$$H(Y|X = \alpha_1) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ bit},$$

$$H(Y|X = \alpha_2) = -0.75 \log_2 0.75 - 0.25 \log_2 0.25 = 0.8113 \text{ bits},$$

$$H(Y|X = \alpha_3) = -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1 \text{ bit},$$

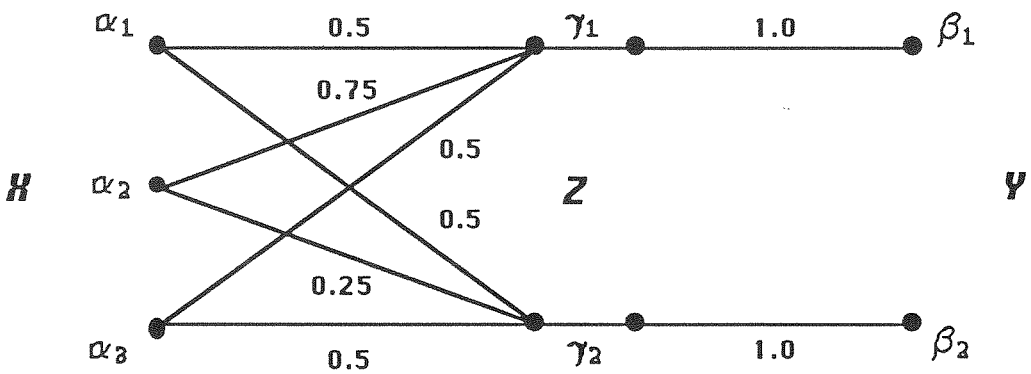


Figure 3.6. Memoryless Discrete Target Channel of *Example 3.2*.

and

$$H(Y|X=\alpha_1, Z) = 0 \text{ bits},$$

$$H(Y|X=\alpha_2, Z) = 0 \text{ bits}.$$

It follows from Eq. (3.6) that

$$I(Y;Z|X=\alpha_1) = 1 \text{ bit},$$

$$I(Y;Z|X=\alpha_2) = 0.8113 \text{ bits},$$

$$I(Y;Z|X=\alpha_3) = 1 \text{ bit}.$$

Thus,

$$I_{\max}^{(1)} = \max\{I(Y;Z|X=\alpha_1), I(Y;Z|X=\alpha_2), I(Y;Z|X=\alpha_3)\} = 1 \text{ bit}.$$

So we have

$$\hat{\mathbf{A}}^{(1)} = \{\alpha_1, \alpha_3\}.$$

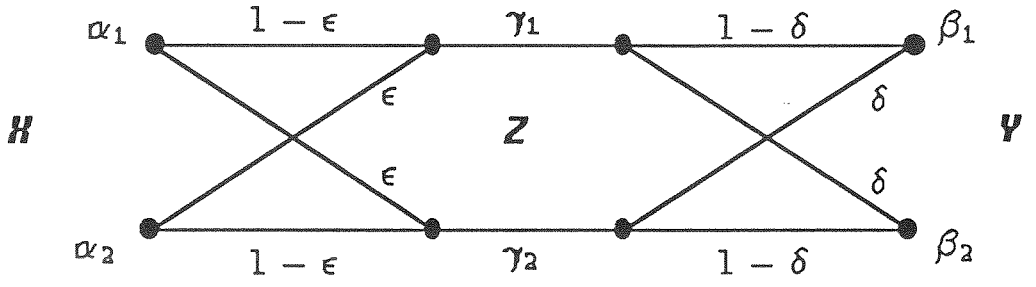


Figure 3.7. Memoryless Discrete Target Channel of *Example 3.3*.

The maximizing distribution Π of X is thus any distribution of the form

$$\Pi = (\pi_1, \pi_2, \pi_3) = (p, 0, 1-p), \qquad \text{for all } p \in [0,1].$$

The maximum $R(Y;Z)$ of $I(Y;Z|X)$ obtained by these distributions is

$$R(Y;Z) = I_{\max}^{(1)} = 1 \text{ bit}.$$

Example 3.3: Consider the discrete memoryless target channel in Fig. 3.7. This target channel is made up of two binary symmetric channels in cascade, the first representing the scattering mechanism and the second representing the observation noise. In this model, we have $\mathbf{A} = \{\alpha_1, \alpha_2\}$, $\Gamma = \{\gamma_1, \gamma_2\}$, and $\mathbf{B} = \{\beta_1, \beta_2\}$. The binary symmetric channels making up the target channel

have transition probabilities

$$q_{11} = \Pr\{Z = \gamma_1, X = \alpha_1\} = 1 - \epsilon,$$

$$q_{12} = \Pr\{Z = \gamma_2, X = \alpha_1\} = \epsilon,$$

$$q_{21} = \Pr\{Z = \gamma_1, X = \alpha_2\} = \epsilon,$$

$$q_{22} = \Pr\{Z = \gamma_2, X = \alpha_2\} = 1 - \epsilon,$$

$$u_{11} = \Pr\{Y = \beta_1, Z = \gamma_1\} = 1 - \delta,$$

$$u_{12} = \Pr\{Y = \beta_2, Z = \gamma_1\} = \delta,$$

$$u_{21} = \Pr\{Y = \beta_1, Z = \gamma_2\} = \delta,$$

$$u_{22} = \Pr\{Y = \beta_2, Z = \gamma_2\} = 1 - \delta.$$

The overall target channel transition probabilities are thus

$$p_{11} = \Pr\{Y = \beta_1, X = \alpha_1\} = 1 - (\epsilon + \delta - 2\epsilon\delta),$$

$$p_{12} = \Pr\{Y = \beta_2, X = \alpha_1\} = \epsilon + \delta - 2\epsilon\delta,$$

$$p_{21} = \Pr\{Y = \beta_1, X = \alpha_2\} = \epsilon + \delta - 2\epsilon\delta,$$

$$p_{22} = \Pr\{Y = \beta_2, X = \alpha_2\} = 1 - (\epsilon + \delta - 2\epsilon\delta).$$

Applying Eqs. (3.7) and (3.8) , we have

$$H(Y|X = \alpha_1) = \mathcal{H}(\epsilon + \delta - 2\epsilon\delta),$$

$$H(Y|X = \alpha_2) = \mathcal{H}(\epsilon + \delta - 2\epsilon\delta),$$

and

$$H(Y|X = \alpha_1) = \mathcal{H}(\delta),$$

$$H(Y|X = \alpha_2) = \mathcal{H}(\delta).$$

From Eq. (3.6), we have

$$I(Y; Z|X = \alpha_1) = I(Y; Z|X = \alpha_2) = \mathcal{H}(\epsilon + \delta - 2\epsilon\delta) - \mathcal{H}(\delta).$$

Thus, we have

$$I_{\max}^{(1)} = \max\{I(Y;Z|X=\alpha_1), I(Y;Z|X=\alpha_2)\} = \mathcal{H}(\epsilon+\delta-2\epsilon\delta) - \mathcal{H}(\delta),$$

where $\mathcal{H}(\epsilon)$ is the binary entropy function defined in Chapter 2. It follows that

$$\hat{\mathbf{A}}^{(1)} = \{\alpha_1, \alpha_2\}.$$

So any valid distribution $\Pi = (\pi_1, \pi_2)$ maximizes $I(Y;Z|X)$, and the maximum value $R(Y;Z)$, which is obtained by all distributions of X, is

$$R(Y;Z) = \mathcal{H}(\epsilon+\delta-2\epsilon\delta) - \mathcal{H}(\delta).$$

As can be seen from these three single-observation examples, cases can arise in which only one, several, or all of the possible transmitted waveforms in $\mathbf{A}$ achieve the maximum information transfer between the target and the receiver.

We will now look at the solution for the general case of N observations of the memoryless target channel. Assume that an N-tuple $\mathbf{X} \in \mathbf{A}^{(N)}$ is transmitted across the memoryless target channel, and the N-tuple $\mathbf{Y} \in \mathbf{B}^{(N)}$ is received at the channel output. Given that a particular N-tuple $\mathbf{x}$ is transmitted, the probability that a specific N-tuple $\mathbf{y}$ will be received is $p(\mathbf{y}|\mathbf{x})$. Since the target channel is memoryless, it perturbs each waveform in the N-tuple independently. Thus, if $p(y|x)$ is the probability of receiving $y \in \mathbf{B}$, given that $x \in \mathbf{A}$ was transmitted, then as previously noted in Eq. (3.2),

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} p(y_i|x_i).$$

Since the single-channel probabilities occur in the analysis of N uses of the memoryless channel, we will again use the notation

$$p_{jk} = \Pr\{y = \beta_k | x = \alpha_j\} = p(\beta_k | \alpha_j).$$

In transmitting an N-tuple across a memoryless channel, the independence of each channel use means that we are using the channel N distinct times incorporating the results of these N uses to form the N-tuple $\mathbf{Y}$. Thus each of the N distinct uses can be modeled as a single use as shown in Fig. 3.4. This means that the whole process of the target channel's perturbing the transmitted N-tuple of waveforms, such that the resulting received N-tuple is obtained, can be modeled as a source that produces an N-tuple $\mathbf{Z}$ of letters from the ensemble $\Gamma^{(N)}$. If a particular N-tuple $\mathbf{x} = (\alpha_{j_1}, \ldots, \alpha_{j_N})$ is transmitted, the probability of a particular $\mathbf{z} = (\gamma_{k_1}, \ldots, \gamma_{k_N})$ by this source is

$$q(\mathbf{z}|\mathbf{x}) = q_{j_1 k_1} q_{j_2 k_2} \cdots q_{j_N k_N} = \prod_{i=1}^{N} q_{j_i k_i}, \tag{3.17}$$

where the single-use channel transition probabilities $p_{j_i k_i}$ corresponding to $x_\tau = \alpha_j$ have distribution Q_{α_j} as in the case of the single channel use.

Once a particular N-tuple $\mathbf{z} \in \Gamma^{(N)}$ is generated, it is mapped to the N-tuple $\mathbf{y} \in \mathrm{B}^{(N)}$ at the channel output by N successive operations of the DMC representing the observation noise process. If a particular N-tuple $\mathbf{z} = \{\gamma_{i_1}, \ldots, \gamma_{i_N}\}$ is generated by the target source, the probability that a particular $\mathbf{y} = \{\beta_{k_1}, \ldots, \beta_{k_N}\}$ is received by the receiver is

$$u(\mathbf{y}|\mathbf{z}) = u_{i_1 k_1} u_{i_2 k_2} \cdots u_{i_2 k_2} = \prod_{i=1}^{N} u_{i_i k_i}, \tag{3.18}$$

where the channel transition probabilities $u_{i_l k_l}$ are as defined in Eq. (3.4) for the single channel use. The overall target channel transition distribution $p(\mathrm{y}|\mathrm{x})$ can be written in terms of $q(\mathrm{z}|\mathrm{x})$ and $u(\mathrm{y}|\mathrm{z})$ as

$$p(\mathrm{y}|\mathrm{x}) = \sum_{\mathrm{z}\in\Gamma^{(N)}} u(\mathrm{y}|\mathrm{z})q(\mathrm{z}|\mathrm{x}). \tag{3.19}$$

Given that $\mathrm{x} = (\alpha_{j_1}, \ldots, \alpha_{j_N})$ is transmitted, we have that the mutual information between Y and Z conditioned on $\mathrm{X} = \mathrm{x}$ is

$$I(\mathrm{Y};\mathrm{Z}|\mathrm{X}=\mathrm{x}) = H(\mathrm{Z}|\mathrm{X}=\mathrm{x}) - H(\mathrm{Z}|\mathrm{X}=\mathrm{x},\mathrm{Y}). \tag{3.20}$$

By the independence of the individual target channel uses, we have that

$$I(\mathrm{Y};\mathrm{Z}|\mathrm{X}=\mathrm{x}) = \sum_{l=1}^{N} I(Y_l; Z_l|X_l = \alpha_{j_l}). \tag{3.21}$$

Taking the expectation of Eq. (3.21) with respect to X, we have

$$\begin{aligned} I(\mathrm{Y};\mathrm{Z}|\mathrm{X}) &= \sum_{\mathrm{z}\in\Gamma^{(N)}} \Pi(\mathrm{x}) I(\mathrm{Y};\mathrm{Z}|\mathrm{X}=\mathrm{x}) \\ &= \sum_{\mathrm{z}\in\Gamma^{(N)}} \Pi(\mathrm{x}) \left\{ \sum_{r=1}^{N} I(Y_r; Z_r|X_r = \alpha_{j_r}) \right\}. \end{aligned} \tag{3.22}$$

In order to find the distribution of X that maximizes $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$, define

$$I_{\max}^{(1)} \stackrel{\text{def}}{=} \max_j \{I(Y;Z|X=\alpha_j)\}. \tag{3.23}$$

Then we can once again define $\hat{\mathrm{A}} = \{\alpha_j \in \mathrm{A} : I(Y;Z|X=\alpha_j) = I_{\max}^{(1)}\}$, and the set $\hat{\mathrm{A}}^{(N)} \subseteq \mathrm{A}^{(N)}$ as the N-fold Cartesian product of $\hat{\mathrm{A}}^{(1)}$. We see from Eq. (3.22) that $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$ is maximized by any distribution $\Pi(\mathrm{x})$ that satisfies the following

conditions:

$$\begin{cases} \Pi(\mathbf{x}) = 0, & \text{for all } \mathbf{x} \notin \hat{\mathbf{A}}^{(N)}. \\ \Pi(\mathbf{x}) \geq 0, & \text{for all } \mathbf{x} \in \hat{\mathbf{A}}^{(N)}. \\ \\ \sum\limits_{\mathbf{x} \in \hat{\mathbf{A}}^{(N)}} \Pi(\mathbf{x}) = 1. \end{cases} \tag{3.24}$$

Any distribution $\Pi(\mathbf{x})$ that satisfies the conditions of Eq. (3.24) achieves the maximum value $R(\mathbf{Y};\mathbf{Z})$ of $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$, and this maximum is given by (see Eq. (3.22))

$$R(\mathbf{Y};\mathbf{Z}) = N I_{\max}^{(1)} = N \max_j \{I(Y;Z|X=\alpha_j)\}. \tag{3.25}$$

This is as expected.

Note that $R(\mathbf{Y};\mathbf{Z})$, the maximum amount of information in the case when an N-tuple of waveforms $\mathbf{x}$ is transmitted, is equal to N times $R(Y;Z)$, the maximum amount of information obtained when a single waveform is transmitted (see Eq. (3.16)). This is not surprising, as $R(Y;Z)$ corresponds to the entropy of a single letter from the ensemble Γ obtained from the channel model source, while $R(\mathbf{Y};\mathbf{Z})$ in the case of the memoryless target channel corresponds to joint entropy of N independent uses of this source. Since the entropy of multiple independent events taken jointly is equal to the sum of the entropies of the events individually, we would expect that $R(\mathbf{Y};\mathbf{Z}) = NR(Y;Z)$.

Examples of the use of the memoryless channel for general N can be obtained by making N-fold extensions of *Example 3.1, Example 3.2, and Example 3.3*. We simply form the set $\hat{\mathbf{A}}^{(N)}$ from the set $\hat{\mathbf{A}}$ in each of these examples, construct the distributions such that all of the probability in the ensemble $\mathbf{A}^{(N)}$ is contained in

$\hat{\mathrm{A}}^{(N)}$ as specified in Eq. (3.24), and calculate $R(\mathrm{Y};\mathrm{Z})$ as N times the result for a single use of the channel.

3.2.2 The Radar/Information Theory Problem for N Observations of a Fixed Target.

We now consider the case in which multiple observations are made of a fixed target. In this case we have a target parameter Z, which is a discrete random variable (vector) taking on values from the ensemble Γ. We can view Z as being generated by a source that has probability distribution $q(z)$ as shown in Fig. 3.2. Now, however, we assume that once Z is generated, it remains fixed while N observations are made by the radar system. These N observations are made by transmitting an N-tuple of waveforms X and receiving a corresponding N-tuple Y of measurements. So we assume that $\mathrm{X} \in \mathrm{A}^{(N)}$, with probability density $\Pi(\mathrm{x})$. Here again, $\mathrm{A} = \{\alpha_1, \ldots, \alpha_m\}$ is the set of possible transmitted waveforms.

When we transmit the N-tuple of waveforms $\mathrm{X} = \{X_1, \ldots, X_N\}$, we receive an N-tuple $\mathrm{Y} = \{Y_1, \ldots, Y_N\} \in \mathrm{B}^{(N)}$ at the output of the target channel (the output Y_k in response to the input X_k, for $k = 1, \ldots, N$). The output N-tuple Y is thus dependent on the transmitted N-tuple X, the target parameter Z, and the noise present in each observation. The probability distribution governing the output y, given that the channel parameter $Z = z$ and the transmitted waveform $\mathrm{X} = \mathrm{x}$ is

$$u(\mathrm{y}|z,\mathrm{x}) = \Pr\{\mathrm{Y} = \mathrm{y}|Z = z, \mathrm{X} = \mathrm{x}\}. \tag{3.26}$$

If we define the conditional density $p(\mathrm{y}|\mathrm{x})$ as

$$p(\mathrm{y}|\mathrm{x}) = \Pr\{\mathrm{Y} = \mathrm{y}|\mathrm{X} = \mathrm{x}\}, \tag{3.27}$$

then we can write $p(\mathbf{y}|\mathbf{x})$ in terms of $u(\mathbf{y}|z,\mathbf{x})$ and $q(z)$ as

$$p(\mathbf{y}|\mathbf{x}) = \sum_{z\in\Gamma} q(z)u(\mathbf{y}|z,\mathbf{x}). \tag{3.28}$$

The mutual information $I(\mathbf{Y};Z|\mathbf{X}=\mathbf{x})$ in which we are interested is given by

$$\begin{aligned} I(\mathbf{Y};Z|\mathbf{X}=\mathbf{x}) &= H(\mathbf{Y}|\mathbf{X}=\mathbf{x}) - H(\mathbf{Y}|Z,\mathbf{X}=\mathbf{x}) \\ &= H(Z|\mathbf{X}=\mathbf{x}) - H(Z|\mathbf{Y},\mathbf{X}=\mathbf{x}). \end{aligned} \tag{3.29}$$

Using the form given on the first line of Eq. (3.29), the conditional entropies of interest are

$$H(\mathbf{Y}|\mathbf{X}=\mathbf{x}) = -\sum_{\mathbf{y}\in\mathrm{B}^{(N)}} p(\mathbf{y}|\mathbf{x})\log p(\mathbf{y}|\mathbf{x}), \tag{3.30}$$

and

$$\begin{aligned} H(\mathbf{Y}|Z,\mathbf{X}=\mathbf{x}) &= -\sum_{z\in\Gamma} q(z) \sum_{\mathbf{y}\in\mathrm{B}^{(N)}} u(\mathbf{y}|z,\mathbf{x})\log u(\mathbf{y}|z,\mathbf{x}) \\ &= -\sum_{z\in\Gamma}\sum_{\mathbf{y}\in\mathrm{B}^{(N)}} q(z)u(\mathbf{y}|z,\mathbf{x})\log u(\mathbf{y}|z,\mathbf{x}). \end{aligned} \tag{3.31}$$

Thus, from Eqs. (3.29) and (3.30),

$$I(\mathbf{Y};Z|\mathbf{X}=\mathbf{x}) = \sum_{\mathbf{y}\in\mathrm{B}^{(N)}} \left\{ -p(\mathbf{y}|\mathbf{x})\log p(\mathbf{y}|\mathbf{x}) + \sum_{z\in\Gamma} q(z)u(\mathbf{y}|z,\mathbf{x})\log u(\mathbf{y}|z,\mathbf{x}) \right\}. \tag{3.32}$$

We can now write

$$I(\mathbf{Y};Z|\mathbf{X}) = \sum_{\mathbf{x}\in\mathrm{A}^{(N)}} \Pi(\mathbf{x})\left\{I(\mathbf{Y};Z|\mathbf{X}=\mathbf{x})\right\}. \tag{3.33}$$

In order to maximize $I(\mathbf{Y};Z|\mathbf{X})$, we define

$$I_{\max}^{(N)} \stackrel{\text{def}}{=} \max_{\mathbf{x}\in\mathrm{A}^{(N)}} \left\{I(\mathbf{Y};Z|\mathbf{X}=\mathbf{x})\right\}, \tag{3.34}$$

and $\hat{\mathbf{A}}^{(N)} \subseteq \mathbf{A}^{(N)}$ as

$$\hat{\mathbf{A}}^{(N)} \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbf{A}^{(N)} : I(\mathbf{Y}; Z|\mathbf{X} = \mathbf{x}) = I_{\max}^{(N)} \right\}. \tag{3.35}$$

Then $I(\mathbf{Y}; Z|\mathbf{X})$ is maximized by any distribution $\Pi(\mathbf{x})$ that satisfies the following conditions:

$$\begin{cases} \Pi(\mathbf{x}) = 0, & \text{for all } \mathbf{x} \notin \hat{\mathbf{A}}^{(N)}; \\ \Pi(\mathbf{x}) \geq 0, & \text{for all } \mathbf{x} \in \hat{\mathbf{A}}^{(N)}; \\ \\ \sum\limits_{\mathbf{x} \in \hat{\mathbf{A}}^{(N)}} \Pi(\mathbf{x}) = 1. \end{cases} \tag{3.36}$$

Any distribution $\Pi(\mathbf{x})$ that satisfies the conditions of Eq. (3.36) achieves the maximum value $R(\mathbf{Y}; Z)$ of $I(\mathbf{Y}; Z|\mathbf{X})$, which is

$$R(\mathbf{Y}; Z) = I_{\max}^{(N)} = \max_{\mathbf{x} \in \mathbf{A}^{(N)}} \{ I(\mathbf{Y}; Z|\mathbf{X} = \mathbf{x}) \}. \tag{3.37}$$

From Eq. (3.29) and the fact that mutual information is a non-negative quantity, we note that this $R(\mathbf{Y}; Z)$ is bounded by

$$0 \leq R(\mathbf{Y}; Z) \leq \min \{ H(\mathbf{Y}|\mathbf{X} = \mathbf{x}), H(Z|\mathbf{X} = \mathbf{x}) \}.$$

But since $H(Z|\mathbf{X} = \mathbf{x}) = H(Z)$—that is, $H(Z|\mathbf{X} = \mathbf{x})$ is not independent of the transmitted X—we have that no matter how large N grows for a fixed target,

$$R(\mathbf{Y}; Z) \leq H(Z). \tag{3.38}$$

This makes sense, as it states that if there is an initial uncertainty $H(Z)$ in the fixed target, the most information we can obtain about it by making multiple measurements of it is $H(Z)$. After this much uncertainty has been obtained about the fixed target, there is no more uncertainty to be resolved.

3.2.3 The Radar/Information Theory Problem for Finite-State Target Channels.

In Section 3.2.1, we examined memoryless target channels where, although the channel was used N times, the channel acted independently from use to use. While this may be an accurate model of the target channel in some instances, in many others it may not be. In order to illustrate this, we have the following example:

Example 3.4: Consider the radio sounding system that has a single waveform $\mathrm{A} = \{\alpha_1\}$ and which transmits an N-tuple $\mathbf{x} = (x_1, \ldots, x_N) = (\alpha_1, \ldots, \alpha_1)$ across the target channel, with each of the N waveforms α_1 transmitted T seconds apart. Suppose that the target channel across which the N-tuple is being transmitted is a slowly fading channel, with the typical fade having a period of T_f. We will assume $T_f \gg T$. Such a channel could arise in radio sounding studies investigating shortwave propagation. Now assume that the receiver receives the waveform at the output of the target channel and decides between outputs from $\mathrm{B} = \{\beta_1, \beta_2\}$, based on whether the energy in the received waveform is greater than or less than some threshhold E_0. Assume that E_0 is selected such that, on the average, $\Pr\{Y = \beta_1\} = \Pr\{Y = \beta_2\} = 1/2$ for each of the transmitted waveforms. Since $T_f \gg T$, by looking at the output of the receiver for the last few waveforms received, we can make a pretty good guess as to what the next output will be. By just guessing the next receiver output to be equal to the last receiver output observed, we would expect to be right significantly more often than half the time.

This simple example illustrates that the memoryless target channel does not

accurately model all target channels encountered in practice. As we will see in the analysis of practical problems forthcoming in this thesis, proper formulation of the target channel model for a given problem often allows the use of memoryless target models where it might not at first be apparent that the memoryless model is applicable. Still, it will be useful to have a target channel model that takes into account target channels with memory. We will now analyze the Radar/Information Theory Problem for target channels with a simple memory mechanism. In order to do this, we will formulate a target channel model which we will call the *Finite-State Target Channel*(FSTC). In this model, the target channel transition probabilities will be controlled by a finite-state Markov process or Markov chain.

The *Finite-State Target Channel* (FSTC) maps the input sequence of waveforms $x_1, \ldots, x_N$ represented by the N-tuple $\mathbf{x} = (x_1, \ldots, x_N)$ into the output sequence

$y_1, \ldots, y_N$ represented by the sequence $\mathbf{y} = (y_1, \ldots, y_N)$, where $\mathbf{x} \in \mathrm{A}^{(N)}$ and $\mathbf{y} \in \mathrm{B}^{(N)}$. This is done by the following mechanism.

The channel has a finite set of states $\mathcal{S} = \{\sigma_1, \ldots, \sigma_Q\}$. At any given time, the channel is in one of the states in $\mathcal{S}$, and with each channel use the channel changes to another state, with all elements of S (including the current state) being possible candidates. Given that the channel is currently in a state $\sigma \in \mathcal{S}$, if any input $x \in \mathrm{A}$ is input into the channel, the probability that $y \in \mathrm{B}$ is received at the output is given by $p^{(\sigma)}(y|x)$.

The state transitions of the channel are assumed to be independent of both

channel inputs and states prior to the current state at the time of transition. The sequence of states $\mathbf{s} = (s_1, \ldots, s_N)$ that characterize the channel during the time that $\mathbf{x}$ is being transmitted across the channel is assumed a subsequence of a finite-state ergodic Markov chain $(\ldots, s_{-1}, s_0, s_1, \ldots)$. For all $\sigma_a, \sigma_b \in \mathbf{S}$, the state transition probabilities of the Markov chain are given by

$$w(\sigma_b|\sigma_a) = \Pr\{s_{t+1} = \sigma_b | s_t = \sigma_a\}. \tag{3.39}$$

We also define the joint probability

$$w(\sigma_a, \sigma_b) = \Pr\{s_t = \sigma_a, s_{t+1} = \sigma_b\}. \tag{3.40}$$

We note that Eqs. (3.39) and (3.40) are valid for all integer values of the time index t, since the ergodicity of the Markov chain implies that the Markov chain is stationary. Eqs. (3.39) and (3.40) are related by the expression

$$w(\sigma_a, \sigma_b) = w(\sigma_b|\sigma_a) w(\sigma_a), \tag{3.41}$$

where

$$w(\sigma_a) = \Pr\{s_t = \sigma_a\}. \tag{3.42}$$

In the FSTC, the overall target channel transition probabilities when the channel is in state σ is $p^{(\sigma)}(y|x)$. Inside the FSTC, we assume that a source generates the target parameter Z, and that the value of Z can be statistically dependent on both the transmitted waveform X and the current channel state σ_t. The probability that a particular z is generated by the target source when a particular x is transmitted and the channel is in state σ_t is $q^{(\sigma_t)}(z|x)$. The probability that a particular z will

be mapped to a particular Y at the channel output when the target channel is in state σ_i is $u^{(\sigma_i)}(y|z)$. $p^{(\sigma)}(y|x)$ is related to $q^{(\sigma_i)}(z|x)$ and $u^{(\sigma_i)}(y|z)$ by the relation

$$p^{(\sigma)}(y|x) = \sum_{z \in \Gamma} u^{(\sigma_i)}(y|z) q^{(\sigma_i)}(z|x). \tag{3.43}$$

In order to illustrate the FSTC, we will look at a pair of examples.

Example 3.5: Consider the FSTC that takes inputs from $\mathrm{A} = \{\alpha_1, \alpha_2\}$ and maps them to outputs from $\mathrm{B} = \{\beta_1, \beta_2\}$. Assume that the channel has states $\mathrm{S} = \{\sigma_1, \sigma_2\}$ and that the state transition probabilities are

$$w(\sigma_1|\sigma_1) = 0.99, \qquad w(\sigma_2|\sigma_1) = 0.01,$$

$$w(\sigma_1|\sigma_2) = 0.2, \qquad w(\sigma_2|\sigma_2) = 0.8.$$

Assume also that the channel transition probabilities as a function of the current state are

$$p^{(\sigma_1)}(\beta_1|\alpha_1) = 1.0, \qquad p^{(\sigma_1)}(\beta_2|\alpha_1) = 0.0,$$

$$p^{(\sigma_1)}(\beta_1|\alpha_2) = 0.0, \qquad p^{(\sigma_1)}(\beta_2|\alpha_2) = 1.0,$$

$$p^{(\sigma_2)}(\beta_1|\alpha_1) = 0.8, \qquad p^{(\sigma_1)}(\beta_2|\alpha_1) = 0.2,$$

$$p^{(\sigma_1)}(\beta_1|\alpha_2) = 0.2, \qquad p^{(\sigma_1)}(\beta_2|\alpha_2) = 0.8.$$

This channel is described graphically by the state diagram and channel transition diagrams in Fig. 3.8. In a communications problem, such a channel would be useful in modeling a burst error channel [3.1, pp.97-99].

Recall *Example 3.4*, the fading channel. Here we noted that observation of the current received waveform gave a pretty good indication of what the next received waveform would be. Let us now construct a model of such a channel using the FSTC.

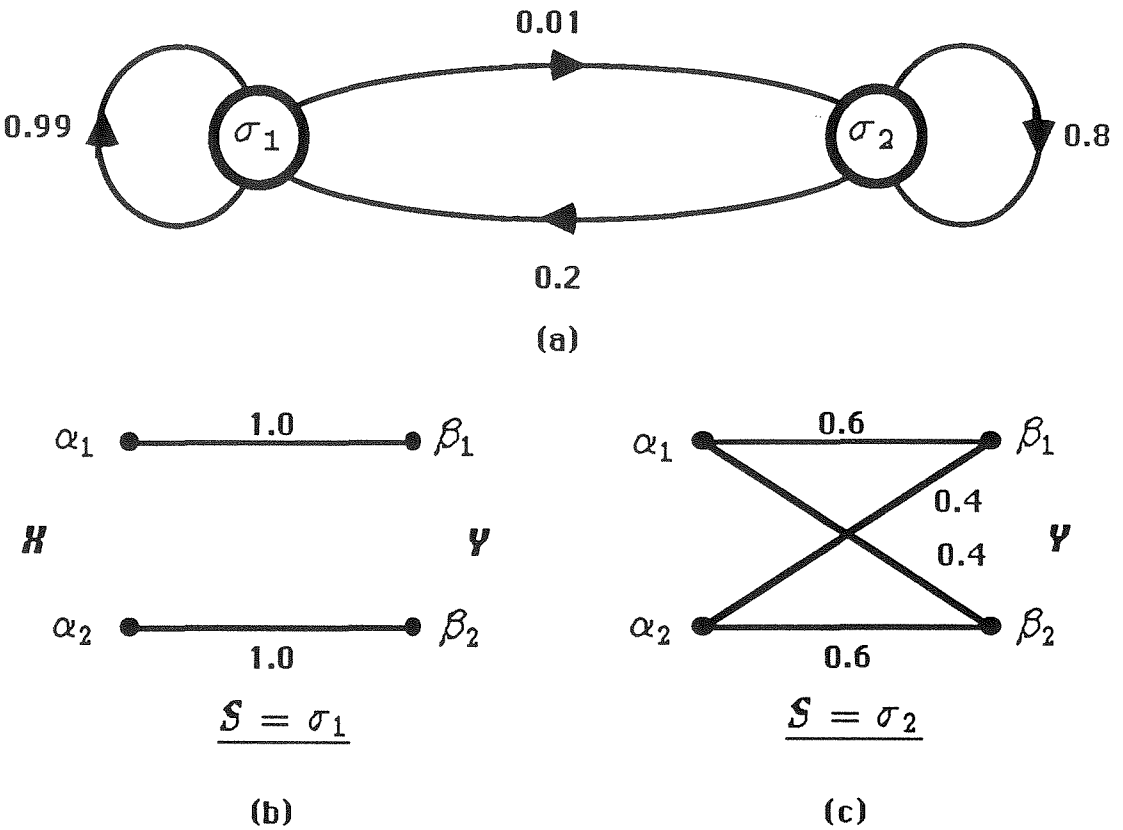


Figure 3.8. Finite-State Target Channel of *Example 3.5*. (a) State Diagram of Channel; (b) Channel Transition Diagram for $S = \sigma_1$; (c) Channel Transition Diagram for $S = \sigma_2$.

Example 3.6: Consider the fading channel described in *Example 3.4*. Here the period of a typical fade was T_f and a waveform $x = \alpha_1$ was transmitted every T seconds. We also assumed that $T_f \gg T$. As a result of the fade, half the time $y = \beta_1$ was received at the output and half the time $y = \beta_2$ was received at the output. Thus, a reasonable model of this fading channel using an FSTC

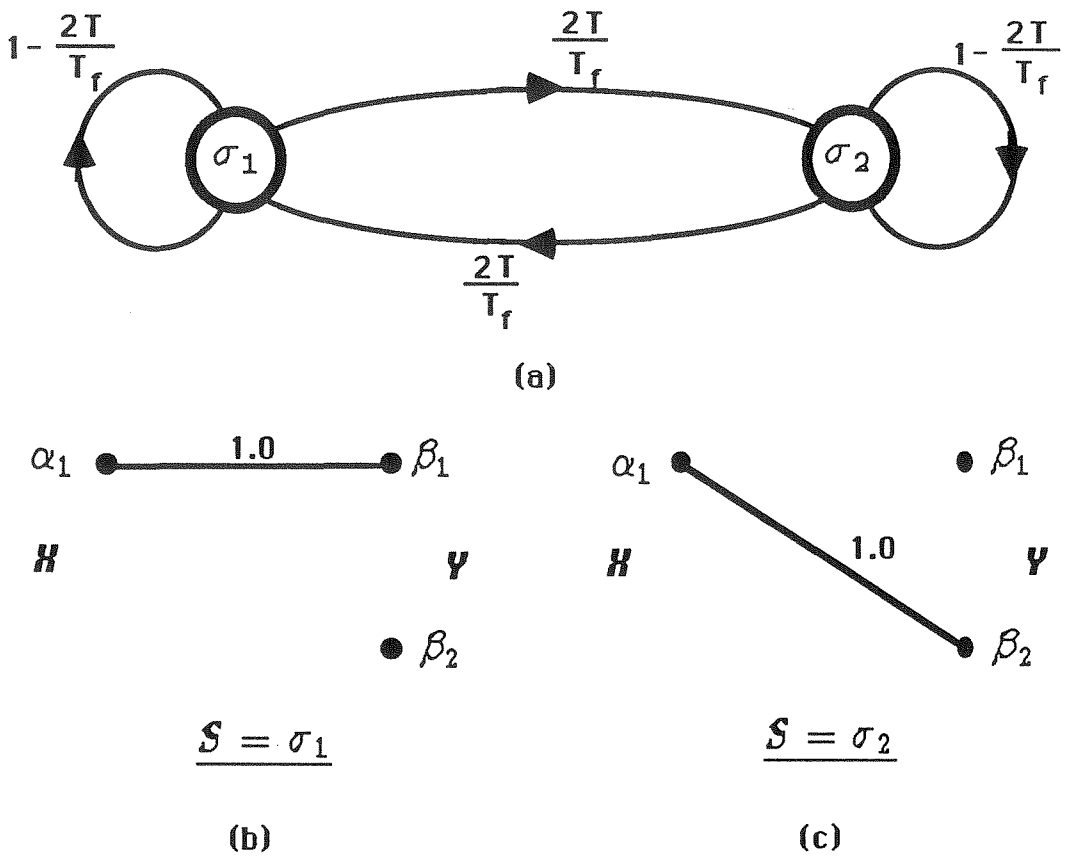


Figure 3.9. Finite-State Target Channel of *Example 3.6*. (a) State Diagram of Channel; (b) Channel Transition Diagram for $\mathcal{S} = \sigma_1$; (c) Channel Transition Diagram for $\mathcal{S} = \sigma_2$.

with $\mathrm{S} = \{\sigma_1, \sigma_2\}$ is as follows: The transmit and receive waveform sets are

$$\mathrm{A} = \{\alpha_1\}, \qquad \mathrm{B} = \{\beta_1, \beta_2\}.$$

The state transition probabilities are

$$w(\sigma_1|\sigma_1) = 1 - \frac{2T}{T_f}, \qquad w(\sigma_2|\sigma_1) = \frac{2T}{T_f},$$

$$w(\sigma_2|\sigma_2) = 1 - \frac{2T}{T_f}, \qquad w(\sigma_1|\sigma_2) = \frac{2T}{T_f}.$$

The channel transition probabilities are

$$p^{(\sigma_1)}(\beta_1|\alpha_1) = 1.0, \qquad p^{(\sigma_1)}(\beta_2|\alpha_1) = 0.0,$$

$$p^{(\sigma_2)}(\beta_1|\alpha_1) = 0.0, \qquad p^{(\sigma_2)}(\beta_2|\alpha_1) = 1.0.$$

The state diagram and channel transition diagrams of this channel are shown in Fig. 3.9.

We will now look at the solution of the Radar/Information Theory Problem for the FSTC. We are interested primarily in finding the distribution of $\mathbf{X}$ that maximizes $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$ for the general case of N uses, as well as determine the maximum value $R(\mathbf{Y};\mathbf{Z})$. In the process we will look at several related problems.

The FSTC for $N = 1$:

Although the case of a single channel use of the FSTC does not illustrate the statistical dependence between multiple uses of the channel, it is worth examining because it illustrates several important features of the FSTC in as simple a context as possible. Once the FSTC has been analyzed for the case of $N = 1$, the extension to general N is straightforward, although not as trivial as it was in the case of the memoryless channel. For the case of $N = 1$, we are primarily interested in finding $I(Y;Z|X)$ and the distribution $\Pi(x)$ of X that maximizes $I(Y;Z|X)$. Before doing this, however, we will first find $I(Y;Z|X,S = \sigma_i)$ and $I(Y;Z|X,S)$. In practical applications, these two quantities are of little interest, since S, the state of the Markov chain characterizing the FSTC, is not known *a priori*. Still, determination of $I(Y;Z|X,S = \sigma_i)$ and $I(Y;Z|X,S)$ gives insight into the FSTC and is intermediate

to obtaining the desired solution, so we will solve for these conditional mutual informations.

Determination of $I(Y;Z|X,S=\sigma_t)$ for $N=1$:

Consider the situation where $X=\alpha_j$, $\alpha_j \in \mathrm{A}$, and $S=\sigma_t$, $\sigma_t \in \mathrm{S}$. The mutual information $I(Y;Z|X,S=\sigma_t)$ is given by

$$I(Y;Z|X=\alpha_j,S=\sigma_t)=H(Y|X=\alpha_j,S=\sigma_t)-H(Y|X=\alpha_j,Z,S=\sigma_t). \tag{3.44}$$

Evaluating these conditional entropies, we have

$$H(Y|X=\alpha_j,S=\sigma_t)=-\sum_{k=1}^{n} p_{jk}^{(\sigma_t)} \log p_{jk}^{(\sigma_t)} \tag{3.45}$$

and

$$\begin{aligned} H(Y|X=\alpha_j,Z,S=\sigma_t) &= -\sum_{i=1}^{r} q_{ji}^{(\sigma_t)} \sum_{k=1}^{n} u_{ik}^{(\sigma_t)} \log u_{ik}^{(\sigma_t)} \\ &= -\sum_{i=1}^{r}\sum_{k=1}^{n} q_{ji}^{(\sigma_t)} u_{ik}^{(\sigma_t)} \log u_{ik}^{(\sigma_t)}. \end{aligned} \tag{3.46}$$

Combining Eqs. (3.44), (3.45) and (3.46), we have

$$\begin{aligned} I(Y;Z|X=\alpha_j,S=\sigma_t) = &-\sum_{k=1}^{n} p_{jk}^{(\sigma_t)} \log p_{jk}^{(\sigma_t)} \\ &+\sum_{i=1}^{r}\sum_{k=1}^{n} q_{ji}^{(\sigma_t)} u_{ik}^{(\sigma_t)} \log u_{ik}^{(\sigma_t)}. \end{aligned} \tag{3.47}$$

Taking the expectation of Eq. (3.47) with respect to X, we have

$$\begin{aligned} I(Y;Z|X,S=\sigma_t) &= \sum_{x\in \mathrm{A}} \Pi(x) I(Y;Z|X=x,S=\sigma_t) \\ &= \sum_{j=1}^{m} \pi_j \left\{ I(Y;Z|X=\alpha_j,S=\sigma_t) \right\}. \end{aligned} \tag{3.48}$$

We now solve for the distributions $\Pi(x) = (\pi_1, \ldots, \pi_m)$ that maximize Eq. (3.48). Define

$$I_{\max}^{(\sigma_i)} \stackrel{\text{def}}{=} \max_j \{I(Y; Z|X = \alpha_j, S = \sigma_i)\}, \tag{3.49}$$

and define $\hat{\mathbf{A}}^{(\sigma_i)} \subseteq \mathbf{A}$ as

$$\hat{\mathbf{A}}^{(\sigma_i)} \stackrel{\text{def}}{=} \max_j \left\{\alpha_j \in \mathbf{A} : I(Y; Z|X = \alpha_j, S = \sigma_i) = I_{\max}^{(\sigma_i)}\right\}. \tag{3.50}$$

Then a distribution $\Pi(x) = \{\pi_1, \ldots, \pi_m\}$ maximizes $I(Y; Z|X, S = \sigma_i)$ if and only if it satisfies the following constraints:

$$\begin{cases} \pi_j = 0, & \text{for all } \alpha_j \notin \hat{\mathbf{A}}^{(\sigma_i)}, \\ \pi_j \geq 0, & \text{for all } \alpha_j \in \hat{\mathbf{A}}^{(\sigma_i)}, \\ \\ \displaystyle\sum_{j:\alpha_j \in \hat{\mathbf{A}}^{(\sigma_i)}} \pi_j = 1. \end{cases} \tag{3.51}$$

Any distribution satisfying Eq.(3.51) achieves the maximum value $R^{(\sigma_i)}(Y; Z)$ of $I(Y; Z|X, S = \sigma_i)$, given by

$$R^{(\sigma_i)}(Y; Z) = I_{\max}^{(\sigma_i)} = \max_j \{I(Y; Z|X = \alpha_j, S = \sigma_i)\}. \tag{3.52}$$

This is, as before, the expected result.

Determination of $I(Y;Z|X,S)$ for $N=1$:

From Eq. (3.48), we have

$$I(Y;Z|X,S=\sigma_t)=\sum_{j=1}^{m}\pi_j\left[I(Y;Z|X=\alpha_j,S=\sigma_t)\right].$$

Let $w_t=w(\sigma_t)=\Pr\{S=\sigma_t\}$. Then taking the expectation with respect to the state S, we have

$$\begin{aligned}I(Y;Z|X,S)&=\sum_{t=1}^{Q}w_t\left\{\sum_{j=1}^{m}\pi_j\left[I(Y;Z|X=\alpha_j,S=\sigma_t)\right]\right\}\\&=\sum_{j=1}^{m}\pi_j\left\{\sum_{t=1}^{Q}w_t\left[I(Y;Z|X=\alpha_j,S=\sigma_t)\right]\right\}.\end{aligned}\tag{3.53}$$

In order to maximize this quantity, we proceed as follows: Define

$$I_{\max}^{(S)}\stackrel{\text{def}}{=}\max_j\left\{\sum_{t=1}^{Q}w_t\left[I(Y;Z|X=\alpha_j,S=\sigma_t)\right]\right\},\tag{3.54}$$

and $\hat{\mathrm{A}}^{(S)}\subseteq\mathrm{A}$ such that

$$\hat{\mathrm{A}}^{(S)}\stackrel{\text{def}}{=}\left\{\alpha_j\in\mathrm{A}:I(Y;Z|X=\alpha_j,S)=I_{\max}^{(S)}\right\},\tag{3.55}$$

where

$$I(Y;Z|X=\alpha_j,S)\stackrel{\text{def}}{=}\sum_{t=1}^{Q}w_t\left[I(Y;Z|X=\alpha_j,S=\sigma_t)\right].\tag{3.56}$$

Then a distribution $\Pi(x)=\{\pi_1,\ldots,\pi_m\}$ maximizes $I(Y;Z|X,S)$ if and only if it satisfies the following conditions:

$$\begin{cases}\pi_j=0, & \text{for all } \alpha_j\notin\hat{\mathrm{A}}^{(S)},\\ \pi_j\geq 0, & \text{for all } \alpha_j\in\hat{\mathrm{A}}^{(S)},\\ \displaystyle\sum_{j:\alpha_j\in\hat{\mathrm{A}}^{(S)}}\pi_j=1.\end{cases}\tag{3.57}$$

Any distribution $\Pi(x)$ that satisfies the conditions of Eq.(3.57) achieves the maximum $R^{(S)}(Y;Z)$ of $I(Y;Z|X,S)$, given by

$$R^{(S)}(Y;Z) = I^{(S)}_{\max} = \max_j \left\{ \sum_{t=1}^{Q} w_t \left[I(Y;Z|X=\alpha_j, S=\sigma_t) \right] \right\}. \tag{3.58}$$

Determination of $I(Y;Z|X)$ for $N=1$:

For the case of $N = 1$, we are most interested in finding $I(Y;Z|X)$. This is because the mutual information between Y and Z is conditioned only on X, which is known to us *a priori*, whereas $I(Y;Z|X,S=\sigma_t)$ and $I(Y;Z|X,S)$ are conditioned on the state of the channel, which is not generally known to us *a priori*. We now look into the problem of determining $I(Y;Z|X)$, finding the maximizing distribution $\Pi(x)$ of X, and finding the resulting maximum $R(Y;Z)$ of $I(Y;Z|X)$.

If q_{ji}, u_{ik}, and p_{jk} are defined as

$$\begin{aligned} q_{ji} &\stackrel{\text{def}}{=} \Pr\{Z=\gamma_i|X=\alpha_j\}, \\ u_{ik} &\stackrel{\text{def}}{=} \Pr\{Y=\beta_k|Z=\gamma_i\}, \\ p_{jk} &\stackrel{\text{def}}{=} \Pr\{Y=\beta_k|X=\alpha_j\}, \end{aligned} \tag{3.59}$$

and

$$\begin{aligned} q_{ji}^{(\sigma_t)} &\stackrel{\text{def}}{=} \Pr\{Z=\gamma_i|X=\alpha_j, S=\sigma_t\}, \\ u_{ik}^{(\sigma_t)} &\stackrel{\text{def}}{=} \Pr\{Y=\beta_k|Z=\gamma_i, S=\sigma_t\}, \\ p_{jk}^{(\sigma_t)} &\stackrel{\text{def}}{=} \Pr\{Y=\beta_k|X=\alpha_j, S=\sigma_t\}, \end{aligned} \tag{3.60}$$

then we can express q_{ji}, u_{ik}, and p_{jk} in terms of $q_{ji}^{(\sigma_t)}$, $u_{ik}^{(\sigma_t)}$, and $p_{jk}^{(\sigma_t)}$, respectively,

and the current state probabilities w_t as

$$\begin{aligned} q_{ji} &= \sum_{t=1}^{Q} w_t q_{ji}^{(\sigma_t)}, \\ u_{ik} &= \sum_{t=1}^{Q} w_t u_{ik}^{(\sigma_t)}, \\ p_{jk} &= \sum_{t=1}^{Q} w_t p_{jk}^{(\sigma_t)}. \end{aligned} \tag{3.61}$$

Once again, assuming $X = \alpha_j$, we can model the FSTC as a new source with source alphabet $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ followed by a discrete channel representing the observation noise. This new source produces the random variable $Z \in \Gamma = \{\gamma_1, \ldots, \gamma_n\}$ with probability distribution $Q_{\alpha_j} = (q_{j1}, \ldots, q_{jn})$. Now

$$I(Y; Z|X = \alpha_j) = H(Y|X = \alpha_j) - H(Y|X = \alpha_j, Z). \tag{3.62}$$

The conditional entropies $H(Y|X = \alpha_j)$ and $H(Y|X = \alpha_j, Z)$ are given by

$$\begin{aligned} H(Y|X = \alpha_j) &= -\sum_{k=1}^{n} p_{jk} \log p_{jk} \\ &= -\sum_{k=1}^{n} \left[\sum_{t=1}^{Q} w_t p_{jk}^{(\sigma_t)} \right] \log \left[\sum_{t=1}^{Q} w_t p_{jk}^{(\sigma_t)} \right], \end{aligned} \tag{3.63}$$

and

$$\begin{aligned} H(Y|X = \alpha_j, Z) &= -\sum_{i=1}^{r} \sum_{k=1}^{n} q_{ji} u_{ik} \log u_{ik} \\ &= -\sum_{i=1}^{r} \sum_{k=1}^{n} \left\{ \left[\sum_{t=1}^{Q} w_t q_{ji}^{(\sigma_t)} \right] \left[\sum_{t=1}^{Q} w_t u_{ik}^{(\sigma_t)} \right] \log \left[\sum_{t=1}^{Q} w_t u_{ik}^{(\sigma_t)} \right] \right\} \end{aligned} \tag{3.64}$$

Thus, we have

$$\begin{aligned} I(Y; Z|X = \alpha_j) = \sum_{k=1}^{n} \Bigg\{ & \left[\sum_{t=1}^{Q} w_t p_{jk}^{(\sigma_t)} \right] \log \left[\sum_{t=1}^{Q} w_t p_{jk}^{(\sigma_t)} \right] \\ & + \sum_{i=1}^{r} \left\{ \left[\sum_{t=1}^{Q} w_t q_{ji}^{(\sigma_t)} \right] \left[\sum_{t=1}^{Q} w_t u_{ik}^{(\sigma_t)} \right] \log \left[\sum_{t=1}^{Q} w_t u_{ik}^{(\sigma_t)} \right] \right\} \Bigg\}. \end{aligned} \tag{3.65}$$

Taking the expectation with respect to X yields

$$I(Y;Z|X) = \sum_{j=1}^{m} \pi_j \left\{ I(Y;Z|X=\alpha_j) \right\}. \tag{3.66}$$

In order to maximize $I(Y;Z|X)$ with respect to the input distribution of X, we proceed as follows: Define

$$I_{\max}^{(1)} \stackrel{\text{def}}{=} \max_j \left\{ I(Y;Z|X=\alpha_j) \right\}, \tag{3.67}$$

and $\hat{\mathbf{A}}^{(1)} \subseteq \mathbf{A}$ by

$$\hat{\mathbf{A}}^{(1)} \stackrel{\text{def}}{=} \left\{ \alpha_j \in \mathbf{A} : I(Y;Z|X=\alpha_j) = I_{\max}^{(1)} \right\}. \tag{3.68}$$

Thus, we see that a distribution $\Pi(x) = (\pi_1, \ldots, \pi_m)$ maximizes $I(Y;Z|X)$ if and only if it satisfies the following conditions:

$$\begin{cases} \pi_j = 0, & \text{for all } j : \alpha_j \notin \hat{\mathbf{A}}^{(1)}, \\ \pi_j \geq 0, & \text{for all } j : \alpha_j \in \hat{\mathbf{A}}^{(1)}, \\ \displaystyle\sum_{j:\alpha_j \in \hat{\mathbf{A}}^{(1)}} \pi_j = 1. \end{cases} \tag{3.69}$$

The maximum value $R(Y;Z)$ of $I(Y;Z|X)$ achieved by the distribution $\Pi(x)$ satisfying Eq. (3.69) is

$$R(Y;Z) = I_{\max}^{(1)}, \tag{3.70}$$

where $I_{\max}^{(1)}$ is as defined in Eq. (3.67). This is , of course, as expected.

Determination of $I(Y;Z|X)$ for General N:

It is now assumed that the N-tuple $\mathbf{X} \in \mathbf{A}^{(N)}$ is transmitted across the FSTC and the N-tuple $\mathbf{Y} \in \mathbf{B}^{(N)}$ is received at the channel output. Given that a particular N-tuple $\mathbf{x}$ is transmitted, the probability that a particular N-tuple $\mathbf{y}$ is received will be denoted by $p(\mathbf{y}|\mathbf{x})$.

Let $\mathbf{s} = \{s_1, \ldots, s_N\}$ be the particular sequence of states that occurs during the N uses of the channel in which $\mathbf{x}$ is mapped to $\mathbf{y}$. Then we have $\mathbf{s} \in \mathcal{S}^{(N)}$. The probability of a given set of states $\mathbf{s}$ occurring is given by

$$w(\mathbf{s}) = \Pr\{\mathbf{S} = \mathbf{s}\} = w(s_1, \ldots, s_N). \tag{3.71}$$

Because the state sequence of the channel is governed by an ergodic Markov chain, we have

$$w(\mathbf{s}) = w(s_1, \ldots, s_N) = w(s_1)w(s_2|s_1)q(s_3|s_2)\cdots w(s_N|s_{n-1}), \tag{3.72}$$

where $w(s_k|s_{k-1})$ and $w(s_1)$ are defined in Eqs. (3.41) and (3.42), respectively. Knowing this, we can now calculate $q(\mathbf{z}|\mathbf{x})$, $u(\mathbf{y}|\mathbf{z})$, and $p(\mathbf{y}|\mathbf{x})$.

Given that the channel has state sequence $\mathbf{s} = \{s_1, \ldots, s_N\}$, the conditional probabilities associated with the particular N-tuples $\mathbf{x}$, $\mathbf{z}$, and $\mathbf{y}$ are

$$\begin{aligned} q^{(\mathbf{s})}(\mathbf{z}|\mathbf{x}) &= q^{(s_1)}(z_1|x_1)pq^{(s_2)}(z_2|x_2)\cdots q^{(s_N)}(z_N|x_N), \\ u^{(\mathbf{s})}(\mathbf{y}|\mathbf{z}) &= p^{(s_1)}(y_1|z_1)p^{(s_2)}(y_2|z_2)\cdots p^{(s_N)}(y_N|z_N), \\ p^{(\mathbf{s})}(\mathbf{y}|\mathbf{x}) &= p^{(s_1)}(y_1|x_1)p^{(s_2)}(y_2|x_2)\cdots p^{(s_N)}(y_N|x_N), \end{aligned} \tag{3.73}$$

where $\mathbf{x} = (x_1, \ldots, x_N)$, $\mathbf{z} = (z_1, \ldots, z_N)$, and $\mathbf{y} = (y_1, \ldots, y_N)$. Thus, we have

that $q(z|x)$, $u(y|z)$, and $p(y|x)$ are given by

$$\begin{aligned} q(z|x) &= \sum_{s\in\mathcal{S}^{(N)}} w(s)q^{(s)}(z|x), \\ u(y|z) &= \sum_{s\in\mathcal{S}^{(N)}} w(s)u^{(s)}(y|z), \\ p(y|x) &= \sum_{s\in\mathcal{S}^{(N)}} w(s)p^{(s)}(y|x). \end{aligned} \tag{3.74}$$

The mutual information between Y and Z given that $X = x$ is

$$I(Y;Z|X=x) = H(Y|X=x) - H(Y|X=x,Z). \tag{3.75}$$

The conditional entropies $H(Y|X=x)$ and $H(Y|X=x,Z)$ are given by

$$\begin{aligned} H(Y|X=x) &= -\sum_{y\in\mathcal{B}^{(N)}} p(y|x)\log p(y|x) \\ &= -\sum_{y\in\mathcal{B}^{(N)}} \left\{ \left[\sum_{s\in\mathcal{S}^{(N)}} w(s)p^{(s)}(y|x)\right] \log \left[\sum_{s\in\mathcal{S}^{(N)}} w(s)p^{(s)}(y|x)\right] \right\}, \end{aligned} \tag{3.76}$$

and

$$\begin{aligned} H(Y|X=x,Z) &= -\sum_{z\in\Gamma^{(N)}}\sum_{y\in\mathcal{B}^{(N)}} q(z|y)u(y|z)\log u(y|z) \\ &= -\sum_{z\in\Gamma^{(N)}}\sum_{y\in\mathcal{B}^{(N)}} \left\{ \begin{array}{c} \left[\sum_{s\in\mathcal{S}^{(N)}} w(s)q^{(s)}(z|x)\right]\left[\sum_{s\in\mathcal{S}^{(N)}} w(s)u^{(s)}(y|z)\right] \\ \cdot\log\left[\sum_{s\in\mathcal{S}^{(N)}} w(s)u^{(s)}(y|z)\right] \end{array} \right\}. \end{aligned} \tag{3.78}$$

Thus, we have

$$\begin{aligned} I(Y;Z|X=x) = -\sum_{y\in\mathcal{B}^{(N)}} \Bigg\{ &-\left[\sum_{s\in\mathcal{S}^{(N)}} w(s)p^{(s)}(y|x)\right] \log \left[\sum_{s\in\mathcal{S}^{(N)}} w(s)p^{(s)}(y|x)\right] \\ &+ \sum_{z\in\Gamma^{(N)}} \left\{ \begin{array}{c} \left[\sum_{s\in\mathcal{S}^{(N)}} w(s)q^{(s)}(z|x)\right]\left[\sum_{s\in\mathcal{S}^{(N)}} w(s)u^{(s)}(y|z)\right] \\ \cdot\log\left[\sum_{s\in\mathcal{S}^{(N)}} w(s)u^{(s)}(y|z)\right] \end{array} \right\} \Bigg\}. \end{aligned} \tag{3.79}$$

Upon taking the expectation of Eq. (3.79) with respect to X, we have

$$I(\mathrm{Y};\mathrm{Z}|\mathrm{X}) = \sum_{\mathrm{x}\in \mathrm{A}^{(N)}} \Pi(\mathrm{x})\,\{I(\mathrm{Y};\mathrm{Z}|\mathrm{X}\}\,. \tag{3.80}$$

In order to maximize Eq. (3.80) with respect to the distribution $\Pi(\mathrm{x})$, we proceed as follows: Define

$$I_{\max}^{(N)} \stackrel{\mathrm{def}}{=} \max_{\mathrm{x}\in \mathrm{A}^{(N)}} \{I(\mathrm{Y};\mathrm{Z}|\mathrm{X}=\mathrm{x})\}\,, \tag{3.81}$$

and $\hat{\mathrm{A}}^{(N)} \subseteq \mathrm{A}^{(N)}$ by

$$\hat{\mathrm{A}}^{(N)} \stackrel{\mathrm{def}}{=} \left\{\mathrm{x}\in \mathrm{A}^{(N)} : I(\mathrm{Y};\mathrm{Z}|\mathrm{X}=\mathrm{x}) = I_{\max}^{(N)}\right\}. \tag{3.82}$$

Then a distribution $\Pi(\mathrm{x})$ will maximize $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$ if and only if it satisfies the following conditions:

$$\begin{cases} \Pi(\mathrm{x}) = 0, & \text{for all } \mathrm{x}\notin \hat{\mathrm{A}}^{(N)}; \\ \Pi(\mathrm{x}) \geq 0, & \text{for all } \mathrm{x}\in \hat{\mathrm{A}}^{(N)}; \\ \displaystyle\sum_{\mathrm{x}\in \hat{\mathrm{A}}^{(N)}} \Pi(\mathrm{x}) = 1, \end{cases} \tag{3.83}$$

and the maximum value $R(\mathrm{Y};\mathrm{Z})$ of $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$ obtained by any distribution satisfying Eq. (3.83) is

$$R(\mathrm{Y};\mathrm{Z}) = I_{\max}^{(N)}, \tag{3.84}$$

where $I_{\max}^{(N)}$ is as defined in Eq. (3.81).

3.3. The Radar/Information Theory Problem for Continuous Target Channels.

In Section 3.2, we examined the Radar/Information Theory Problem for discrete target channels, in which, for each channel use, one of m waveforms was transmitted by the transmitter and one of n waveforms was received at the receiver. In dealing with real radar systems, it is reasonable to assume that one of a finite number of waveforms is transmitted, but it is often not convenient to assume that one of a finite number of waveforms is received. For example, if the radar receiver is measuring the power scattered from the target, the received power can generally take on a continuum of values. Although one could model the system—at least approximately—by quantizing the received signal, this procedure can complicate the analysis of the radar system. In addition, information is generally lost in the quantization process. Thus, it is more appropriate to assume that the received waveform can take on a continuum of values. In this section, we will examine target channels that have a continuous target channel output. We will refer to such channels as *continuous target channels*, although the transmitted waveform is still assumed to be a member of a finite set.

Consider the situation shown in Fig. 3.10, in which a transmitter transmits an N-tuple $\mathbf{X} \in \mathrm{A}^{(N)}$, where $\mathrm{A}^{(N)}$ is an ensemble of N-tuples of waveforms with probability distribution $\Pi(\mathbf{x})$, and A is the set of possible transmitted waveforms $\mathrm{A} = \{\alpha_1, \ldots, \alpha_m\}$. The transmitted N-tuple X proceeds through the target channel C, which perturbs X in such a way that an N-tuple $\mathbf{Y} \in \mathrm{R}^{(N)}$ is received at the receiver, where R is the set of real numbers. The channel C perturbs the transmitted

Figure 3.10. The Radar/Information Theory Problem for Continuous Target Channels.

N-tuple in such a way that if $\mathbf{x} \in \mathrm{A}^{(N)}$ is a particular N-tuple transmitted, the N-tuple $\mathbf{y} \in \mathrm{R}^{(N)}$ is received with conditional PDF $f(\mathbf{y}|\mathbf{x})$.

It is assumed that the transmitted N-tuple $\mathbf{X}$ is known at the receiver. Let $\mathbf{Z}$ be a continuous random vector that characterizes the parameters of the target channel we wish to measure. This random parameter vector $\mathbf{Z}$ characterizes all uncertainty of the parameters to be measured in the target channel. Additional uncertainty, however, is present because of the random noise in the radar system. We wish to find the distribution $\Pi(\mathbf{x})$ that maximizes the mutual information $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$. In addition we wish to find the resulting maximum value $R(\mathbf{Y};\mathbf{Z})$ of $I(\mathbf{Y};\mathbf{Z}|\mathbf{X})$. We will now look at the solution of the radar/information theory problem for continuous target channels without memory.

3.3.1 Continuous Memoryless Target Channel.

Consider the Radar/Information theory problem as outlined in Fig. 3.10, but with a memoryless continuous target channel, by which we mean that the propagation of each individual waveform in the transmitted N-tuple propagates through the target channel in a manner independent of the others. For any transmitted $\mathbf{x} \in \mathbf{A}^{(N)}$ and $\mathbf{y} \in \mathbf{R}^{(N)}$ we have

$$\begin{aligned} \mathbf{x} &= (x_1, \ldots, x_N), \quad x_k \in \mathbf{A}, \quad \text{for all } k = 1, \ldots, N, \\ \mathbf{y} &= (y_1, \ldots, y_N), \quad y_k \in \mathbf{R}, \quad \text{for all } k = 1, \ldots, N. \end{aligned} \tag{3.85}$$

Since the channel is memoryless, the channel transition PDF of $\mathbf{y}$ conditioned on $\mathbf{x}$ is given by

$$f(\mathbf{y}|\mathbf{x}) = \prod_{r=1}^{N} f(y_r|x_r). \tag{3.86}$$

In order to analyze the problem for the continuous memoryless target channel, we will first examine the case where $N = 1$ (a single use of the target channel). The case for general N will then be an easy extension of the case for $N = 1$.

For the case where $N = 1$, a single waveform from the ensemble $\mathbf{A}$ is transmitted through the target channel and is perturbed by the target channel such that an element of the set $\mathbf{R}$, a real number, is observed at the output of the channel.

In the case of the discrete target channel, where calculations dealt with (absolute, not differential) entropy, we modeled the transmitter and target channel together as a new source with a noiseless channel connecting Y and Z. In the case of the continuous channel, however, a noiseless channel would have infinite capacity. Thus, we cannot directly extend the model we used in the discrete case, as using a

channel model that has infinite capacity is neither intuitively appealing, physically realistic, or mathematically tractable.

In order to develop a model for the continuous target channel, we look to the physical measurement process that takes place in the measurement of target characteristics by a radar. The transmitter transmits a waveform X that illuminates the target in order to measure some characteristic of the target represented by Z. The value of this random variable Z can be dependent on the transmitted waveform X. For example, if Z represented the mean reflectivity of the target, this could be a function of the frequency spectrum of the transmitted waveform. The electromagnetic wave incident on the target is then scattered by the target. A portion of the scattered wave is detected and processed by the receiver, resulting in the measurement Y. Thus, Y is dependent on Z. Y is, of course, also dependent on X, both through the dependence of Y on Z and the fact that receiver processing is a function of the transmitted waveform. Taking these mechanisms into account in the radar measurement process, we have the model shown in Fig. 3.11 for the case of $N = 1$. As can be seen, this model is characterized by two PDFs, $f(z|x)$ and $f(y|z,x)$. In our analysis, we will need the PDF's $f(y|z,x)$ and $f(y|x)$. This latter PDF can be obtained from the two PDF's in our model by

$$f(y|x) = \int_{-\infty}^{\infty} f(y|z,x) f(z|x)\, dz.$$

From the definition of mutual information for continuous random variables as given in Eq (2.35), we have that $I(Y;Z|X)$ is given by both

$$I(Y;Z|X) = h(Y|X) - h(Y|Z,X), \tag{3.87}$$

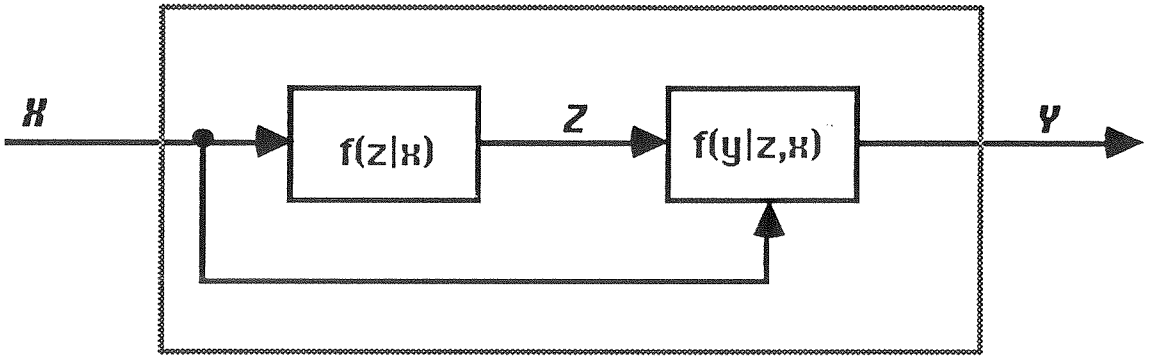


Figure 3.11. Continuous Target Channel Model.

and

$$I(Y;Z|X) = h(Z|X) - h(Z|Y,X). \tag{3.88}$$

Here, $h(Y|X)$, $h(Y|Z,X)$, $h(Z|X)$, and $h(Z|Y,X)$ are conditional differential entropies, given by

$$h(Y|X) = \sum_{x \in \mathbf{A}} \Pi(x) \left\{ - \int_{-\infty}^{\infty} f(y|x) \log f(y|x)\, dy \right\}, \tag{3.89}$$

$$h(Y|Z,X) = \sum_{x \in \mathbf{A}} \Pi(x) \left\{ - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y,z|x) \log f(y|z,x)\, dy\, dz \right\}, \tag{3.90}$$

$$h(Z|X) = \sum_{x \in \mathbf{A}} \Pi(x) \left\{ - \int_{-\infty}^{\infty} f(z|x) \log f(z|x)\, dz \right\}, \tag{3.91}$$

$$h(Y|Z,X) = \sum_{x \in \mathbf{A}} \Pi(x) \left\{ - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y,z|x) \log f(y|z,x)\, dy\, dz \right\}. \tag{3.92}$$

In the derivation that follows, we will use the form in Eq. (3.87), as it will be more straightforward in the applications we will consider. However, our results could also be derived using Eq. (3.88), and in some applications the latter may be more straigtforward.

From Eqs. (3.87), (3.89), and (3.90), we have

$$I(Y;Z|X) = \sum_{x \in \mathbf{A}} \Pi(x) I(Y;Z|X=x), \tag{3.93}$$

where

$$I(Y;Z|X=x) = -\int_{-\infty}^{\infty} f(y|x) \log f(y|x)\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y,z|x) \log f(y|z,x)\, dy\, dz. \tag{3.94}$$

In order to find the distribution $\Pi(x)$ of X that maximizes $I(Y;Z|X)$, we proceed as follows: Define

$$I_{\max}^{(1)} \stackrel{\text{def}}{=} \max_{x \in \mathbf{A}} \{I(Y;Z|X=x)\}. \tag{3.95}$$

Define $\hat{\mathbf{A}}^{(1)} \subseteq \mathbf{A}$ as

$$\hat{\mathbf{A}}^{(1)} = \left\{x \in \mathbf{A} : I(Y;Z|X=x) = I_{\max}^{(1)}\right\}. \tag{3.96}$$

Then $I(Y;Z|X)$ is maximized by a distribution $\Pi(x)$ if and only if it satisfies the following conditions:

$$\begin{cases} \pi_j = 0, & \text{for all } j \text{ such that } \alpha_j \notin \hat{\mathbf{A}}^{(1)}; \\ \pi_j \geq 0, & \text{for all } j \text{ such that } \alpha_j \in \hat{\mathbf{A}}^{(1)}; \\ \sum_{j=0}^{m} \pi_j = 1. \end{cases} \tag{3.97}$$

The maximum value $R(Y;Z)$ of $I(Y;Z|X)$ obtained by any such distribution is

$$R(Y;Z) = I_{\max}^{(1)}. \tag{3.98}$$

This is as expected.

We now examine the general case of N uses of the memoryless target channel. Assume that an N-tuple $\mathbf{X} \in \mathbf{A}^{(N)}$ is transmitted across the memoryless target channel, and the N-tuple $\mathbf{Y} \in \mathbf{R}^{(N)}$ is received at the channel output. Given that a particular N-tuple $\mathbf{x}$ is transmitted, the PDF of the N-tuple $\mathbf{y}$ being received is $f(\mathbf{y}|\mathbf{x})$. Since the target channel is memoryless, it perturbs each waveform in the N-tuple independently. Thus, if $f(y|x)$ is the PDF of $y \in \mathbf{R}$, given $x \in \mathbf{A}$ was transmitted, then as previously noted in Eq. (3.86),

$$f(\mathbf{y}|\mathbf{x}) = \prod_{r=1}^{N} f(y_r|x_r).$$

In transmitting an N-tuple across the continuous memoryless channel, the independence of each channel use means that we are using the channel N distinct times and are incorporating the results of these N uses to form the N-tuple $\mathbf{Y}$. Thus, each of the single uses can be modeled as shown in Fig. 3.11. If the transmitter transmits an N-tuple $\mathbf{x} = (x_1, \ldots, x_N)$, the PDF of $\mathbf{Z} = (z_1, \ldots, z_N)$ conditioned on $\mathbf{x}$ is

$$f(\mathbf{z}|\mathbf{x}) = \prod_{r=1}^{N} f(z_r|x_r). \tag{3.99}$$

So the N-tuple $\mathbf{Z}$ characterizes the random behavior we wish to measure for the N channel uses, with the r-th element Z_r representing the r-th of N uses.

Given that a specific $\mathbf{x}$ is transmitted, the mutual information between $\mathbf{Y}$ and $\mathbf{Z}$ conditioned on $\mathbf{X} = \mathbf{x}$ is

$$I(\mathbf{Y};\mathbf{Z}|\mathbf{X} = \mathbf{x}) = h(\mathbf{Y}|\mathbf{X} = \mathbf{x}) - h(\mathbf{Y}|\mathbf{X} = \mathbf{x}, \mathbf{Z}). \tag{3.100}$$

Taking the expectation of Eq. (3.100) with respect to $\mathbf{X}$, we have

$$I(\mathrm{Y};\mathrm{Z}|\mathrm{X}) = \sum_{\mathrm{Y}\in\mathrm{A}^{(N)}} \Pi(\mathrm{x}) I(\mathrm{Y};\mathrm{Z}|\mathrm{X}=\mathrm{x}). \tag{3.101}$$

In order to find the distributions of $\mathbf{X}$ that maximize $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$, we proceed as follows: Define

$$I_{\max}^{(1)} \stackrel{\text{def}}{=} \max_{x\in\mathrm{A}} \{I(Y;Z|X=x)\}. \tag{3.102}$$

Define $\hat{\mathrm{A}}^{(1)} \subseteq \mathrm{A}$ as

$$\hat{\mathrm{A}}^{(1)} = \left\{x \in \mathrm{A} : I(Y;Z|X=x) = I_{\max}^{(1)}\right\}. \tag{3.103}$$

Define $\hat{\mathrm{A}}^{(N)}$ as the N-fold Cartesian product of $\hat{\mathrm{A}}^{(1)}$. Then from Eq. (3.101), we see that a distribution $\Pi(\mathrm{x})$ maximizes $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$ if and only if it satisfies the following conditions:

$$\begin{cases} \Pi(\mathrm{x}) = 0, & \text{for all } \mathrm{x} \notin \hat{\mathrm{A}}^{(N)}; \\ \Pi(\mathrm{x}) \geq 0, & \text{for all } \mathrm{x} \in \hat{\mathrm{A}}^{(N)}; \\ \displaystyle\sum_{\mathrm{x}\in\hat{\mathrm{A}}^{(N)}} \Pi(\mathrm{x}) = 1. \end{cases} \tag{3.104}$$

The maximum value $R(\mathrm{Y};\mathrm{Z})$ of $I(\mathrm{Y};\mathrm{Z}|\mathrm{X})$ achieved by distributions that satisfy Eq. (3.104) is

$$R(\mathrm{Y};\mathrm{Z}) = N I_{\max}^{(1)} = N \max_{x\in\mathrm{A}} \{I(Y;Z|X=x)\}. \tag{3.105}$$

This is what we would expect by analogy to the discrete memoryless target channel.

The extension of the memoryless continuous target channel to a finite-state continuous target channel, with memory controlled by a Markov chain with a finite number of states, is straightforward and is done in a manner analogous to that of the discrete case. We assume that the conditional PDF's that characterize the statistical

behavior of the channel are a function of the Markov chain state $s \in \mathcal{S}$, just as was done with the conditional probabilities in the discrete case. Thus, these distributions would be written as $f^{(s)}(y|x)$, $f^{(s)}(y,z|x)$, $f^{(s)}(y|z,x)$. We would then use these distributions to calculate the mutual information $I^{(s)}(Y;Z|X)$ and then calculate $I(Y;Z|X)$ by averaging over s as was done in the discrete case. Determination of distributions that maximize $I(Y;Z|X)$ and the resulting maximum values is then done in an identical manner to that done in the discrete case. Since we will not be using the finite-state continuous target channel in subsequent work, we will not go through this straightforward derivation here.

3.4. Conclusions.

In this chapter we have formulated and investigated *the Radar/Information Theory Problem* for both discrete and continuous target channels. This provides a basis for the design of radar waveforms for obtaining the maximum amount of information about a target, and also provides some of the tools for analyzing the information content of radar images. Both of these problems will be examined in the remainder of this thesis. Before the results in this chapter can be applied to these practical radar problems, we must obtain models of the radar targets of interest. In the next chapter we will review and develop statistical scattering models of radar targets, with the goal of applying the results of this chapter to these models. The results of the Radar/Information Theory Problem in this chapter and the models of the next chapter will form the basis of the information-theoretic design of radar waveforms and analysis of practical radar systems.

3.5 Chapter 3 References.

3.1 Gallager, R. G., *Information Theory and Reliable Communication*, John Wiley and Sons, New York, NY, 1968.

CHAPTER 4

RADAR SCATTERING MODELS

The Radar/Information Theory Problem, examined as the framework for the information-theoretic analysis of radar systems in the previous chapter, is dependent on statistical models of radar scattering and noise behavior. This is reflected in the PDFs that arise in the model. These PDF's are the sole representation of target behavior in the target channel model. Thus, if the results of the previous chapter are to be used in the analysis of practical radar systems, statistical models of the radar targets and target channels for these systems must be developed.

When considering measurements of a radar target, we realize that the electromagnetic wave scattered by the target, besides being a function of the transmitted waveform, is also a function of both the target's spatial and material characteristics. We assume that the target is made up of either multiple scattering centers or a continuum of scatterers distributed in space. Then, when the transmitted waveform is reflected from the target, it is reflected by many points on the target that are spatially separated. If these points have different material composition, the fraction of the incident energy reflected will be a function of the material composition of the points. In addition, the scattered waves from the various points on the target will add constructively and destructively at different points in space, generating an in-

terference pattern in space. This interference pattern will be a function of both the spatial distribution of the target scatterers and the transmitted waveform's wave shape or spectral content. This fact will be of considerable importance in our work dealing with waveform design in Chapters 5 and 6, so we will also be interested in models that characterize the scattering characteristics of spatially extended targets.

In order to facilitate the development of our radar-scattering models, a brief review of radar cross section, the analytic signal representation, and polarization and depolarization in scattering will be presented.

4.1. Radar Reflectivity and Radar Cross Section.

In this section we will briefly review radar reflectivity and introduce the parameter known as *radar cross section*. In the process, other parameters of interest in specifying radar systems will also be introduced. More detailed and specialized discussions of radar reflectivity can be found in References [4.1-4.7].

If an electromagnetic wave is incident on a boundary at which two materials of differing characteristic impedances η_1 and η_2 ($\eta = \sqrt{\mu/\epsilon}$, where μ is the permeability of the material and ϵ is its dielectric constant), then the electromagnetic wave will be at least partially reflected [4.8]. Thus, an electromagnetic wave transmitted by a radar through free space or an atmosphere will generally be at least partially reflected or scattered when it impinges upon a target, such as the planetary surface below or an aircraft flying in the atmosphere. So when a radar is making measurements on an object, the magnitude of the reflected signal is a function of the material making up the object. The magnitude of the reflected wave is also a

function of several other characteristics of the object, such as its size, orientation with respect to the radar, the frequency spectrum of the waveform illuminating the target, and the transmitted electric field polarization.

The *radar cross section* σ of a scatterer is a parameter that measures the amount of power a target scatters toward a receiver. This will be a function of the physical orientation of the transmitter, target, and receiver. In this chapter we will concentrate primarily on *monostatic radar systems*, in which the transmitter and receiver antennas are collocated (and often the same physical antenna), but these results can be extended in a straightforward manner to the case of *bistatic radar systems*, in which the transmitter and receiver antennas are in different locations. In dealing with monostatic radar systems, we are interested in the signal scattered back at the radar, and we refer to the radar cross section in this case as the *backscatter cross section*.

The *radar cross section* σ is a parameter defined to be proportional to the received echo power from the radar target. The received power for a monostatic radar system is related to the radar cross section σ by the relation

$$P_R = \frac{P_T G^2 \lambda^2}{(4\pi)^3 R^4} \sigma. \tag{4.1}$$

Here

$\sigma \stackrel{\text{def}}{=}$ radar cross section in units of area,

$P_R =$ received power,

$P_T =$ transmitted power,

$G =$ antenna gain (the same for both transmit and receive),

$\lambda =$ wavelength of transmitted signal in units of length,

$R =$ radar-to-target range in units of length.

From Eq. (4.1), it can be seen that the radar cross section σ is the cross-sectional area required of an isotropic scatterer in order to receive power P_R given P_T, G, λ, and R. This can be derived from the Friis Transmission Equation [4.9, Sections 1.7 and 3.10].

In considering the reflectivity of complex radar targets, we are often examining the combined return from a large number of individual scatterers. When considering such targets, the concepts of coherent and incoherent scattering can be useful. If the phase of a wave is constant or varies in a deterministic manner, the wave is said to be *coherent*. If the phase of the wave is randomly distributed over the interval $[0, 2\pi]$, it is said to be incoherent. It is a well-known fact that the total power resulting from the sum of multiple incoherent waves is equal to the sum of the powers of each of the waves individually. The power resulting from the sum of multiple coherent waves must be obtained by summing the electric field that is due to each of the individual waves and then by determining the total power from the resulting electric field. This is because the resulting power for the multiple coherent

waves can be much greater than than the sum of the power of the individual waves, or it can be much less—the resulting power can even be zero. The resulting power in the coherent case depends on whether the individual waves add constructively or destructively.

Mathematically, if a target is composed of N individual scatterers with radar cross sections σ_i, the total cross section that results from all of the individual scatterers is

$$\sigma = \left| \sum_{j=1}^{N} \sqrt{\sigma_j}\, e^{i\psi_j} \right|^2 , \tag{4.2}$$

where the ψ_j are the phase angles that are due to the phase delay corresponding to the range of the j-th scatterer from the target. If the j-th scatterer is at distance R_j from the target, assumed constant over the target, the phase ψ_j is given by

$$\psi_j = \frac{4\pi R_j}{\lambda} \pmod{2\pi}. \tag{4.3}$$

If the ψ_j are independently and uniformly distributed on the interval $[0, 2\pi]$, corresponding to the case of incoherent scattering, then one has

$$\begin{aligned} \sigma_{IC} &= E\left| \sum_{j=1}^{N} \sqrt{\sigma_j}\, e^{i\psi_j} \right|^2 = E\left\{ \left[\sum_{j=1}^{N} \sqrt{\sigma_j}\, e^{i\psi_j} \right] \left[\overline{\sum_{k=1}^{N} \sqrt{\sigma_k}\, e^{i\psi_k}} \right] \right\} \\ &= \sum_{j=1}^{N} \sum_{k=1}^{N} \sqrt{\sigma_j}\sqrt{\sigma_k}\, E\left\{ e^{i(\psi_j - \psi_k)} \right\} = \sum_{j=1}^{N} \sigma_j , \end{aligned} \tag{4.4}$$

since for $j \neq k$

$$E\left\{ e^{i(\psi_j - \psi_k)} \right\} = E\left\{ e^{i\psi_j} \right\} E\left\{ e^{-i\psi_k} \right\} = 0,$$

because

$$E\left\{ e^{i\psi_j} \right\} = E\left\{ e^{i\psi_k} \right\} = 0.$$

But for $j = k$,

$$E\left\{e^{i(\psi_j - \psi_k)}\right\} = E\left\{e^{i0}\right\} = E\{1\} = 1.$$

So from Eq. (4.4) we see that in non-coherent scattering, the power from the individual scatterers add, or equivalently, the total radar cross section is the sum of the individual radar cross sections. We are assuming here that there is no significant shielding or multiple reflections between scatterers.

In general, the field scattered from a radar target has both a constant coherent component as well as a random, incoherent component. For example, if the target under consideration is land being illuminated by a scatterometer, there may be a non-coherent component in the radar return because of the roughness of the terrain. There may also be a coherent component that is due to a large specular return from a large scatterer, for example, a large metal roof on a building which is smooth compared to the scale of the wavelength of the radar signal. If the surface being measured is very rough and there are no specular components, as in the case of a perfectly diffuse surface, there may be only an incoherent scattered field. This property is, of course, a (slowly varying) function of the wavelength of the transmitted radar signal.

In talking about the roughness of a scattering surface or radar target, the Rayleigh roughness criterion is often used to classify a surface as rough or smooth. The Rayleigh roughness criterion can be described as follows. Assume that a radar illuminates a target at a grazing angle θ with respect to the normal of a rough planar surface, as shown in Fig. 4.1. If the difference in height (in the direction

normal to the surface) between two neighboring points on the surface is Δh, then the phase difference between two waves reflected from these two points is

$$\Delta\Phi = \frac{4\pi\Delta h \sin\theta}{\lambda}. \tag{4.5}$$

Here

$\Delta h =$ difference in height between the two points,

$\lambda =$ wavelength of electromagnetic wave,

$\theta =$ incidence or grazing angle of wave with respect to local surface tilt.

To give a qualitative description of whether a surface is rough or smooth, the Rayleigh roughness criterion uses Eq. (4.5). A surface is considered rough if

$$\Delta h \sin\theta > \frac{\lambda}{8}, \tag{4.6}$$

and smooth if

$$\Delta h \sin\theta < \frac{\lambda}{8}. \tag{4.7}$$

It is clear that this is an imprecise criterion for surface roughness, but for $\Delta h \sin\theta \ll \lambda/8$, it is safe to assume that the surface is smooth, and for $\Delta h \sin\theta \gg \lambda/8$, it is safe to assume that the surface is rough.

4.2. The Analytic Signal Representation.

Transmitted radar signals are in general the result of the amplitude and phase modulation of a sinusoidal carrier. As such, they can be represented as

$$s(t) = a(t)\cos[2\pi f_0 t + \phi(t)], \tag{4.8}$$

where $s(t)$ is the transmitted radar signal, $a(t)$ is the envelope of the amplitude modulation, $\phi(t)$ is the phase modulation function, and f_0 is the transmitter carrier frequency. It is straightforward to represent standard radar modulation formats (pulse-amplitude modulation, linear and non-linear FM, bi-phase modulation) in terms of Eq. (4.8). In most radar systems of interest, even such wideband systems as direct-sequence spread-spectrum radars, the envelope $a(t)$ and phase $\phi(t)$ are slowly varying functions of time in contrast with $\cos 2\pi f_0 t$. Thus, these waveforms are "narrowband" waveforms in the sense that their Fourier transform $S(f)$ is centered and concentrated around the carrier frequency f_0. Thus, although their bandwidth can be very large, it is for all practical purposes much less than f_0.

Since the waveform $s(t)$ is real, it follows that the Fourier transform $S(f)$ of $s(t)$ is conjugate-symmetric in f. That is,

$$S(f) = \overline{S(-f)}. \tag{4.9}$$

Here the bar over the right-hand side of the equation represents complex conjugation.

In dealing with radar waveforms, it is often convenient to work with a complex

analytic function $\psi(t)$ such that

$$s(t) = \text{Re}\left\{\psi(t)\right\}, \tag{4.10}$$

and

$$\text{Im}\left\{\psi(t)\right\} = \check{s}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau}\, d\tau. \tag{4.11}$$

So $\check{s}(t)$ is the Hilbert transform of $s(t)$. Thus, we have that the analytic signal representation $\psi(t)$ of $s(t)$ is given by

$$\psi(t) = s(t) + i\check{s}(t) = s(t) + \frac{i}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau}\, d\tau. \tag{4.12}$$

If $s(t)$ is a narrowband waveform, that is, if

$$S(f) = 0, \qquad f_0 - f_c \le |f| \le f_0 + f_c, \quad 0 < f_c < f_0, \tag{4.13}$$

then it can be shown [4.10] that

$$\psi(t) = a(t)\exp\left\{i\left[2\pi f_0 t + \phi(t)\right]\right\}. \tag{4.14}$$

Although the narrowband condition of Eq. (4.13) may not be exactly satisfied for a given radar signal $s(t)$, it is approximately satisfied for practical radar systems, even for spread-spectrum radar systems.

We also note that

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(t)e^{-i2\pi ft}\, dt = \int_{-\infty}^{\infty} \left[s(t) + \frac{i}{\pi t} * s(t)\right] e^{-i2\pi ft}\, dt = S(f)\left[1 + Q(f)\right].$$

Here,

$$Q(f) = \begin{cases} 1, & \text{for } f \ge 0; \\ 0, & \text{for } f < 0. \end{cases}$$

So we have

$$\Psi(f) = \begin{cases} 2S(f), & \text{for } f \geq 0; \\ 0, & \text{for } f < 0, \end{cases} \tag{4.15}$$

and the energy in the signal $s(t)$ is given by

$$\mathcal{E} = \int_{-\infty}^{\infty} s(t)s(t)\,dt, \tag{4.16}$$

or

$$\mathcal{E} = \frac{1}{2}\int_{-\infty}^{\infty} \psi(t)\overline{\psi(t)}\,dt. \tag{4.17}$$

These relationships are useful both in representing radar-signals and in calculating their energy or power.

4.3. Polarization, Depolarization, and Scattering.

Most monostatic radars both transmit and receive electromagnetic waves of the same single polarization. When radar measurement of a target is made with such a radar, the information obtained about the target by the radar is limited to the scattering properties of the target for the particular polarization and target orientation with which the measurement was made. Measurements made with additional polarizations and different polarizations on transmit and receive can provide additional information about the scattering characteristics of the target. This information may be useful in classification or identification of the radar target. In addition, some polarizations will provide a larger backscattered field than others, making polarization diversity useful in target detection as well. Reference [4.11] provides a detailed study of polarization in radar scattering problems. Reference [4.2, Ch.7] presents

data from several studies in which the radar returns from land were measured by airborne radars at several polarizations. These results show that polarization diversity is indeed valuable in obtaining information about terrain in radar remote sensing, so we will now briefly examine polarization in radar scattering.

The polarization of an electromagnetic plane wave can be described in terms of two orthogonal linearly polarized components. Let z be the direction of propagation of a plane wave and let x and y be two mutually orthogonal directions, both of which are orthogonal to the direction of propagation z. Assume that the wave is transmitted by a radar, scattered by a target, and that the scattered field is then received by the radar. Let E_x^t, E_y^t, E_x^r, and E_y^r be the components of the transmitted electric field strength in the x-direction, transmitted electric field strength in the y-direction, received electric field strength in the x-direction, and received electric field strength in the y-direction, respectively. Then the received electric field components E_x^r and E_y^r can be expressed in terms of the transmitted electric field components E_x^t and E_y^t by the relation

$$\begin{pmatrix} E_x^r \\ E_y^r \end{pmatrix} = \begin{pmatrix} a_{xx} & a_{yx} \\ a_{xy} & a_{yy} \end{pmatrix} \begin{pmatrix} E_x^t \\ E_y^t \end{pmatrix}. \tag{4.18}$$

The 2×2 matrix relating the transmitted and received fields in Eq. (4.18) is known as the *polarization scattering matrix* and is a function of the scatterer. The elements of this matrix are complex numbers relating the amplitude and phase of the respective fields. If

$$\mathrm{E}^t = \begin{pmatrix} E_x^t \\ E_y^t \end{pmatrix}, \quad \mathrm{E}^r = \begin{pmatrix} E_x^r \\ E_y^r \end{pmatrix}, \quad \text{and} \quad \mathrm{A} = \begin{pmatrix} a_{xx} & a_{yx} \\ a_{xy} & a_{yy} \end{pmatrix},$$

then Eq.(4.18) can be written vectorially as

$$\mathbf{E}^r = \mathbf{A}\mathbf{E}^i. \tag{4.19}$$

The energy scattered by the target depends on the transmitter frequency, target orientation, and position of the target with respect to the radar, and so the elements of **A** are in general also a function of these parameters. In the case of a monostatic radar, the matrix **A** is a symmetric matrix, with $a_{xy} = a_{yx}$ [4.12]. This can be most easily seen by using reciprocity, that is, by interchanging the roles of the collocated transmit and receive antennas of the monostatic radar system.

As was previously noted, polarization diversity can provide significant information about the surfaces being measured in radar remote sensing. This is due in part to the *depolarization* of the illuminating wave when it is scattered wave from the surface being measured. When an electromagnetic wave is scattered by an object or surface the backscattered wave may not have the same polarization as the wave incident on the object or surface. This phenomenon is known as *depolarization*. In specular reflection, the sign of the polarization is reversed, but this does not count as depolarization, since no energy is being transferred to an orthogonal polarization.

Generally, the rougher the surface or object scattering an electromagnetic wave, the greater the depolarization. This is because multiple reflection is one of the major mechanisms from surfaces. For very large, smooth targets, there is usually very little depolarization, but most natural surfaces have both a polarized backscatter component and a depolarized backscatter component. Reference [4.2, Ch.3] discusses several models of depolarization mechanisms in scattering processes.

Since depolarization of the transmitted wave generally occurs in the scattering process, we will now introduce a quantity known as the *cross-polarization cross section*, which is denoted σ_{xy}. It is defined in the same manner as the radar cross section in Eq. (4.1), except that it is assumed that the transmit antenna has linear polarization x and the receive antenna has linear polarization y.

Typically, in dealing with backscattering from surfaces, the directions x and y are taken to be either in the plane of incidence (vertically polarized) or perpendicular to the plane of incidence (horizontally polarized) [4.1, pp.168-9]. Radar backscattering from the surface is then characterized by the four backscattering cross sections σ_{VV}, σ_{HH}, σ_{HV}, and σ_{VH}. These four backscatter cross sections are often used to characterize the backscatter characteristics of a surface in radar remote sensing.

If direction x is taken to be vertical and direction y is taken to be horizontal in Eq. (4.18), we have that the backscatter cross sections σ_{VV}, σ_{HH}, σ_{HV}, and σ_{VH} are related to the elements of the polarization-scattering matrix by the relations

$$\sigma_{VV} = K\,|a_{VV}|^2\,,$$

$$\sigma_{HH} = K\,|a_{HH}|^2\,,$$

$$\sigma_{VH} = K\,|a_{VH}|^2\,,$$

$$\sigma_{HV} = K\,|a_{HV}|^2\,.$$

Here, K is a constant taking into account scaling that is due to space loss and taking into account the spatial point of reference for the electric fields in Eq. (4.18).

4.4. Statistical Models of Radar Backscatter from Land.

Radar backscatter from land can be caused by natural objects such as plants, trees, grass, rocks, hills, and bare ground, as well as by man-made objects such as roads, railroads, buildings, automobiles, fences, and power lines. When illuminating a portion of the earth's surface with an airborne or spaceborne radar, a resolution cell on the ground is generally large enough to have a large number of scatterers in it. As the radar makes successive measurements, the radar moves and so does the resolution cell along the earth's surface. As a result, from measurement to measurement, a large number of scatterers leave the resolution cell, a large number of new scatterers enter the resolution cell, and those scatterers that remain in the resolution cell undergo a significant change of phase in their radar return because of radar motion. Thus, from measurement to measurement, there can be considerable fluctuation in the backscattered signal from the surface. In addition, there may be large fixed objects that remain in the radar's field of view for several resolution cells, contributing a strong constant component to the returned signal. Thus, it appears that land scatterers viewed with airborne radars can be grouped (at least roughly) into two broad categories:

1. Small scatterers of which there are a very large number and for which no individual scatterer dominates in terms of total backscattered power, and
2. Large specular reflectors that contribute a significant portion of the total backscattered power in a resolution cell.

Terrain made up of scatterers from the first group can be modeled as a large

number of random scatterers [4.2], and it is reasonable to assume that such a surface is rough in the Rayleigh sense ($\Delta h \sin\theta \gg \lambda/8$). We will now look at the development of a theoretical model for this case.

Consider a radar illuminating a surface made up of a large number of random scatterers. Let the electric field at the receiving antenna reflected from the j-th of a large number n of scatterers be given by

$$E_j = A_j e^{i\phi_j}, \tag{4.20}$$

where

$A_j =$ the magnitude of the electric field scattered by the j-th scatterer,

$\phi_j =$ the phase of the electric field scattered by the j-th scatterer.

Assume that both transmit and receive are done with fixed, although not necessarily the same, polarizations. Since the scatterers are randomly distributed over the surface and, for surfaces of practical interest, will generally vary in size, we will assume that both A_j and ϕ_j are random variables. Assume that the positions of these scatterers are independent of each other, that the size of an individual scatterer is independent of other individual scatterers, and that scatterer sizes are independent of scatterer positions. From these three assumptions, we can reasonably conclude that:

1. A_j is statistically independent of A_k, for $j \neq k$.
2. ϕ_j is statistically independent of ϕ_k, for $j \neq k$.
3. A_j is statistically independent of ϕ_k, for all j and k.

Since the surface is made up of scatterers randomly distributed and is rough in the Rayleigh sense, we will assume that the ϕ_j are independent, identically distributed random variables uniformly distributed on $[0, 2\pi]$. We will also assume that the A_j are independent, identically distributed random variables with mean μ_A and finite variance σ_A^2. We then have that the total electric field from a large number n of scatterers is

$$E = \sum_{j=1}^{n} A_j e^{i\phi_j} = V e^{i\theta}. \tag{4.21}$$

Expressing Eq. (4.21) in rectangular form, we have

$$E = X + iY, \tag{4.22}$$

with

$$\begin{aligned} X &= V\cos\theta = \sum_{j=1}^{n} A_j \cos\phi_j = \sum_{j=1}^{n} X_j, \\ Y &= V\sin\theta = \sum_{j=1}^{n} A_j \sin\phi_j = \sum_{j=1}^{n} Y_j. \end{aligned} \tag{4.23}$$

Here,

$$\begin{aligned} X_j &= A_j \cos\phi_j, \\ Y_j &= A_j \sin\phi_j. \end{aligned} \tag{4.24}$$

Now the X_i constitute a set of independent identically distributed random variables, as do the Y_i. Their means and variances are given by

$$\mathrm{E}(X_j) = \mathrm{E}(A_j \cos\phi_j) = \mathrm{E}(A_j)\,\mathrm{E}(\cos\phi_j) = \mu_A \cdot 0 = 0, \tag{4.25}$$

$$\sigma_{X_j}^2 = \mathrm{E}(A_j^2 \cos^2\phi_j) = \mathrm{E}(A_j^2)\,\mathrm{E}(\cos^2\phi_j) = \frac{\sigma_A^2 + \mu_A^2}{2}, \tag{4.26}$$

$$\mathrm{E}(Y_j) = \mathrm{E}(A_j \sin\phi_j) = \mathrm{E}(A_j)\,\mathrm{E}(\sin\phi_j) = \mu_A \cdot 0 = 0, \tag{4.27}$$

$$\sigma_{X_j}^2 = \mathrm{E}(A_j^2 \cos^2 \phi_j) = \mathrm{E}(A_j^2)\,\mathrm{E}(\cos^2 \phi_j) = \frac{\sigma_A^2 + \mu_A^2}{2}. \tag{4.28}$$

The central limit theorem [4.13, p.193] states that if $X_1, X_2, \ldots, X_n$ denotes a random sample from a distribution that has mean μ and finite variance σ^2, then the random variable

$$S_n = \frac{\left(\sum_{j=1}^{n} X_j - n\mu\right)}{\sqrt{n}\sigma}$$

has a limiting distribution that is Gaussian with mean 0 and variance 1, as $n \to \infty$.

Applying the central limit theorem to

$$X = \sum_{j=1}^{n} X_j \tag{4.29}$$

and

$$Y = \sum_{j=1}^{n} Y_j, \tag{4.30}$$

and realizing that although n is finite, it is very large, we have that both X and Y are approximately Gaussian, both having mean 0 and variance $n(\sigma_A^2 + \mu_A^2)/2$.

We will assume the Gaussian approximation to be exact. This gives us density functions of X and Y as

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-x^2/2\sigma^2}, \tag{4.31}$$

and

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2}, \tag{4.32}$$

with

$$\sigma^2 = \frac{n(\sigma_A^2 + \mu_A^2)}{2}. \tag{4.33}$$

The important thing to note is that for all practical purposes, X and Y are zero-mean Gaussian random variables with finite variance σ^2. In addition, we have

$$\begin{aligned} \mathrm{E}(XY) &= \mathrm{E}\left(\sum_{j=1}^{n} X_j \sum_{k=1}^{n} Y_k\right) = \mathrm{E}\left(\sum_{j=1}^{n}\sum_{k=1}^{n} A_j A_k \cos\phi_j \sin\phi_k\right) \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{E}\left(A_j A_k\right) \cdot \mathrm{E}\left(\cos\phi_j \sin\phi_k\right) = 0, \end{aligned}$$

since

$$\mathrm{E}\left(\cos\phi_j \sin\phi_k\right) = \begin{cases} \int_0^{2\pi} \cos\phi_j \sin\phi_j \, d\phi_j = 0 & \text{, for } j = k; \\ \int_0^{2\pi}\int_0^{2\pi} \cos\phi_j \sin\phi_k \, d\phi_j \, d\phi_k = 0 & \text{, for } j \neq k. \end{cases}$$

But since $\mathrm{E}(X) = \mathrm{E}(Y) = 0$, we have

$$\mathrm{E}(X)\mathrm{E}(Y) = 0,$$

and thus,

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y),$$

which implies that X and Y are uncorrelated random variables. But since X and Y are Gaussian random variables, they are statistically independent [4.13, p. 120].

From Eqs. (4.21) and (4.22), we have

$$V e^{i\Theta} = X + iY, \tag{4.35}$$

and so

$$\begin{aligned} V &= \sqrt{X^2 + Y^2} \\ \Theta &= \tan^{-1}\left(Y/X\right). \end{aligned} \tag{4.36}$$

Thus, by a change of variables, we have that the joint density of V and Θ is

$$\begin{aligned} f_{V,\theta}(v,\theta) &= f_{X,Y}\left[x(v,\theta), y(v,\theta)\right] \left|\frac{\partial(x,y)}{\partial(v,\theta)}\right| \\ &= f_X\left[x(v,\theta)\right] f_Y\left[y(v,\theta)\right] \left|\frac{\partial(x,y)}{\partial(v,\theta)}\right| \\ &= \begin{cases} \frac{v}{2\pi\sigma^2}\exp\left(-v^2/2\sigma^2\right) & \text{, for } 0 \le v < \infty,\ 0 \le \theta < 2\pi; \\ 0 & \text{, elsewhere.} \end{cases} \end{aligned} \tag{4.37}$$

Integrating the joint density $f_{V,\theta}(v,\theta)$ with respect to θ on the interval $[0, 2\pi)$, we obtain

$$f_V(v) = \int_0^{2\pi} \frac{v}{2\pi\sigma^2}\exp\left(-v^2/2\sigma^2\right) d\theta = \frac{v}{\sigma^2} exp\left(\frac{-v^2}{2\sigma^2}\right).$$

Thus, we have that the magnitude V of the received electric field, that is $V = |E|$, is a random variable which is Rayleigh distributed, and has the Rayleigh PDF

$$f_V(v) = \begin{cases} \frac{v}{\sigma^2}\exp\left(\frac{-v^2}{2\sigma^2}\right) & \text{, } v \ge 0; \\ 0 & \text{, elsewhere.} \end{cases} \tag{4.38}$$

The received power fron the surface of randomly distributed scatterers is given by

$$P = V^2. \tag{4.39}$$

By a change of variable and the result of Eq. (4.38), we see that the PDF of P is given by

$$\begin{aligned} f_P(p) &= f_V\left[v(p)\right] \left|\frac{\partial v}{\partial p}\right| = \left(\frac{1}{2}p^{-1/2}\right)\left(\frac{p^{1/2}}{\sigma^2}\right)\exp\left(\frac{-p}{2\sigma^2}\right) \\ &= \frac{1}{2\sigma^2}\exp\left(\frac{-p}{2\sigma^2}\right). \end{aligned} \tag{4.40}$$

The mean of P, μ_P, is

$$\mu_P = \mathrm{E}(P) = \int_0^{\infty} p \cdot \left(1/2\sigma^2\right)\exp\left(-p/2\sigma^2\right) dp = 2\sigma^2. \tag{4.41}$$

So we can write the PDF of the received power as

$$f_P(p) = \begin{cases} (1/\mu_P)\exp(-p/\mu_P) & \text{, for } p \geq 0; \\ 0 & \text{, elsewhere.} \end{cases} \tag{4.42}$$

So the received power from a rough surface of randomly distributed scatterers is exponentially distributed. This is a well-known and often quoted result, yet a derivation is rarely given. A derivation is included here, as this result will be used in our scattering model for the information-theoretic analysis of imaging radar in Chapter 8.

As was previously noted, large specular returns can also be present in radar returns from land. In such cases, there is a relatively constant component in the return signal as well as the return from randomly distributed scatterers. We will model the amplitude W of the return signal for such surfaces as

$$W = \left| V e^{i\Theta} + a e^{i\gamma} \right|. \tag{4.43}$$

Here,

$V = \sqrt{X^2 + Y^2}$,

$X =$ a Gaussian random variable with mean 0 and variance σ^2,

$Y =$ a Gaussian random variable with mean 0 and variance σ^2,

$\Theta =$ a random variable uniformly distributed on $[0, 2\pi]$,

$\Gamma =$ a random variable uniformly distributed on $[0, 2\pi]$,

$a =$ amplitude of return from the constant scatterer.

As before, X and Y are statistically independent. In addition, we will also assume Θ and Γ to be statistically independent, which is reasonable, since we would not expect

the phase of the sum of the returns from the random scatterers to be dependent on the phase of the return from the constant scatterer.

The PDF of W is given by [4.14, pp. 106,107]:

$$f_W(w) = \frac{w}{\sigma^2} \exp\left(-\frac{w^2 + a^2}{2\sigma^2}\right) I_0\left(\frac{wa}{\sigma^2}\right), \tag{4.44}$$

where $I_0(\cdot)$ is the modified Bessel function of order zero [4.15, p. 147]. The mean power ($P = W^2$) in this case is

$$\mu_p = 2\sigma^2 + a^2. \tag{4.45}$$

If we define the parameter m^2 as the ratio of the power of the constant component to that of the random component, we have

$$m^2 = \frac{a^2}{2\sigma^2}. \tag{4.46}$$

Then we can write the PDF of the received power P as

$$f_P(p) = \left(\frac{1+m^2}{\mu_P}\right) e^{-m^2} e^{-p\left(\frac{1+m^2}{\mu_P}\right)} I_0\left(2m\sqrt{\frac{p(1+m^2)}{\mu_p}}\right). \tag{4.47}$$

When the terrain echo contains large, dominant scatters as well as a large number of randomly distributed scatterers, we would expect the PDF of the returned power to satisfy Eq. (4.47). Note that as the parameter m^2 goes to zero, Eq. (4.47) reduces to Eq. (4.42). Scattering mechanisms that have the PDF of the received power given by Eq. (4.42) are often referred to as *Rayleigh scatterers*, whereas scattering mechanisms that have the PDF of the received power given by Eq. (4.47) are often referred to as *Rician scatterers* [4.2]. Note that Rayleigh scattering is actually a special case of Rician scattering, the case where the constant scattering amplitude $a \to 0$.

4.5. A Model for Cross Polarization Measurements From Rough Surfaces.

As was mentioned in Section 4.3 and as is well established in References [4.2, 4.11, and 4.16], polarization diversity can provide a substantial increase in a radar system's ability to determine the characteristics of a target. Reference [4.16] discusses the implementation of radar systems that both transmit and receive with multiple polarizations. Reference [4.7, Ch.4] analyzes such systems in detail.

We will now consider a relatively simple form of such a multiple polarization problem for scattering from rough surfaces and will statistically model it. Consider a system that transmits with a single linear polarization, which we will designate as *polarization A*, and receives with both *polarization A* and an orthogonal linear polarization, *polarization B*. Let P_A be the received power from *polarization A* and P_B be the received power from *polarization B*. We already know from the derivation leading to Eq. (4.32) that if the surface being illuminated is made up of a large number of randomly distributed scatterers, then the PDFs of P_A and P_B are given by

$$f_{P_A}(p_A) = \frac{1}{\mu_{P_A}} \exp\left(-P_A/\mu_{P_A}\right), \tag{4.50}$$

and

$$f_{P_B}(p_B) = \frac{1}{\mu_{P_B}} \exp\left(-P_B/\mu_{P_B}\right). \tag{4.51}$$

We will assume that P_A and P_B are statistically independent random variables. If we make this assumption, the joint PDF of P_A and P_B is

$$f_{P_A P_B}(p_A, p_B) = \frac{1}{\mu_{P_A}\mu_{P_B}} \exp - \left\{ - \left(\frac{P_A}{\mu_{P_A}} + \frac{P_B}{\mu_{P_B}} \right) \right\}. \tag{4.52}$$

We now define a parameter η, which we will call the *cross polarization ratio*, by

$$\eta \stackrel{\text{def}}{=} \frac{P_B}{P_A}. \tag{4.53}$$

The parameter η has the physical interpretation of being the ratio of the depolarized power to the non-depolarized power in the backscattered radar return from the surface. We note that this is a coarse measure of surface roughness, since smooth surfaces have very little depolarization.

We are interested in finding the PDF of η. To do this, we make the following change of variable involving P_A and P_B. Let

$$\begin{aligned} \eta &= P_B/P_A, \\ \gamma &= P_A. \end{aligned} \tag{4.54}$$

Then the joint density function of η and γ is given by

$$\begin{aligned} f_{\eta,\gamma}(\eta,\gamma) &= f_{P_A,P_B}\left[p_A(\eta,\gamma), p_B(\eta,\gamma)\right] \left|\frac{\partial(p_A,p_B)}{\partial(\eta,\gamma)}\right| \\ &= \frac{\gamma}{\mu_{P_A}\mu_{P_B}} \exp\left[-\left(\frac{\gamma}{\mu_{P_A}} + \frac{\eta\gamma}{\mu_{P_B}}\right)\right] \\ &= \frac{\gamma}{\mu_{P_A}\mu_{P_B}} \exp\left[-\gamma\left(\frac{\mu_{P_B} + \eta\mu_{P_A}}{\mu_{P_A}\mu_{P_B}}\right)\right]. \end{aligned} \tag{4.55}$$

Letting $\alpha = \mu_{P_2} + \eta\mu_{P_1}$ and $\beta = \mu_{P_1}\mu_{P_2}$, and integrating with respect to γ, Eq. (4.55) becomes

$$\begin{aligned} f_\eta(\eta) &= \int_0^\infty f_{\eta,\gamma}(\eta,\gamma)\, d\gamma = \int_0^\infty \frac{\gamma}{\beta} \exp\left(-\frac{\gamma\alpha}{\beta}\right) d\gamma \\ &= \left[-\frac{\gamma}{\alpha} - \frac{\beta}{\alpha^2}\exp\left(-\frac{\gamma\alpha}{\beta}\right)\right]_0^\infty \\ &= \frac{\beta}{\alpha^2} = \frac{\mu_{P_1}\mu_{P_2}}{\left(\mu_{P_2} + \eta\mu_{P_1}\right)^2} = \frac{\mu_{P_2}/\mu_{P_1}}{\left(\eta + \mu_{P_2}/\mu_{P_1}\right)^2}. \end{aligned} \tag{4.56}$$

Thus, we have that the PDF of η is

$$f_\eta(\eta) = \begin{cases} \frac{\rho}{(\eta+\rho)^2} & \text{, for } \eta \geq 0, \\ 0 & \text{, for } \eta < 0, \end{cases} \tag{3.57}$$

where $\rho = \mu_{P_2}/\mu_{P_1}$.

4.6. Spatial Resolution Characteristics of Radar Waveforms.

A radar waveform $x(t)$ is a function of time that represents the electric field transmitted by the radar at some fixed polarization. Thus, $x(t)$ represents the magnitude of the transmitted electric field as a function of time, and the direction (polarization) of the electric field is assumed to be known. In the operation of a radar system, the transmitted field is scattered by the target and a portion of it is received by the receiver, which determines target characteristics from it. One of the target characteristics that can be determined by a radar system is the spatial extent of the target in range. The ability of a radar system to do this is a function of the transmitted radar waveform. We will now investigate how the pulse shape of a radar waveform determines its spatial resolution properties.

In discussing the spatial resolution characteristics of radar signals, a commonly quoted result is that the minimum resolvable distance between two objects is inversely proportional to the bandwidth of the transmitted radar waveform. We will now investigate this statement, using a very simple waveform of finite bandwidth, one made up of two sinusoids. Such a waveform $x(t)$ can be written as

$$x(t) = \cos 2\pi f_1 t + \cos 2\pi f_2 t. \tag{4.58}$$

The Fourier transform $X(f)$ of $x(t)$ is

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\,dt = \frac{1}{2}\left[\delta(f-f_1)+\delta(f+f_1)\right] + \frac{1}{2}\left[\delta(f-f_2)+\delta(f+f_2)\right]. \tag{4.59}$$

Here $\delta(\cdot)$ is the Dirac delta function. The spectrum $X(f)$ is shown in Fig. 4.1. As can be seen from Fig. 4.1, $x(t)$ can be viewed as a double-sideband, suppressed-carrier waveform, and can be rewritten as

$$x(t) = 2\left[\cos 2\pi\Delta f t\right]\left[\cos 2\pi f_0 t\right]. \tag{4.60}$$

Here,

$$f_0 = \frac{f_1+f_2}{2},$$

and

$$\Delta f = f_2 - f_1.$$

We now analyze this waveform's spatial resolution characteristics.

Consider a target consisting of two identical point scatterers as shown in Fig. 4.2. The scatterers are separated by a distance Δz along the direction of wave propagation of a monostatic radar illuminating the object. If the waveform $x(t)$ as given in Eq. (4.58) is illuminating the scatterers, there are interfering waves being returned from the scatterers, and as a result, an interference pattern arises. As the frequency difference Δf between the two sinusoids is varied from $\Delta f = 0$, the situation arises in which the waves from the scatterers arrive back at the radar with a phase difference of 180°, causing destructive interference at the radar antenna

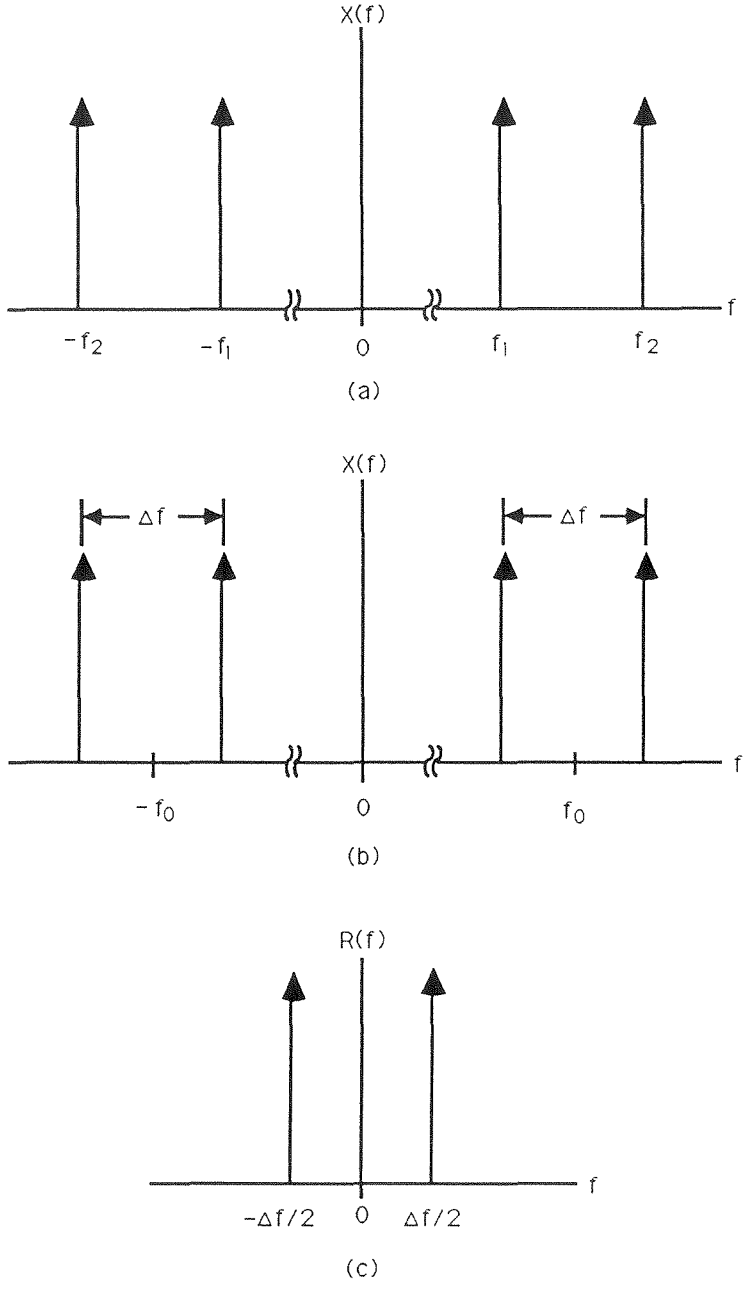


Figure 4.1. (a) Spectrum $X(f)$ of the waveform $x(t)$ of Eq. (4.58). (b) Spectrum $X(f)$ relabelled with the parameters of Eq. (4.60). (c) Spectrum of the baseband waveform $r(t)$, which modulates $2\cos 2\pi f_0 t$ to produce $x(t)$.

and thus a received field of zero magnitude. The phase difference ϕ_1 of the waves of frequency f_1 received from the scatterers is

$$\phi_1 = \frac{2\pi(2\Delta z)}{\lambda_1} = \frac{2\pi f_1(2\Delta z)}{c}, \tag{4.61}$$

and that for frequency f_2 is given by

$$\phi_2 = \frac{2\pi(2\Delta z)}{\lambda_2} = \frac{2\pi f_2(2\Delta z)}{c}. \tag{4.62}$$

Here, λ_1 is the wavelength of the sinusoid of frequency f_1, λ_2 is the wavelength of the sinusoid of frequency f_2, and c is the velocity of propagation of the electromagnetic wave.

In general, we have **constructive interference** between the waves at frequencies f_1 and f_2 when

$$\phi_2 = \phi_1 + 2n\pi, \qquad n = \text{integer}, \tag{4.63}$$

and **destructive interference** when

$$\phi_2 = \phi_1 + (2n-1)\pi, \qquad n = \text{integer}. \tag{4.64}$$

Looking at the destructive interference for the case of the smallest possible Δf (smallest positive n in Eq. (4.64)), we see that this occurs when

$$\phi_2 = \phi_1 + \pi. \tag{4.65}$$

The Δf that gives rise to this situation is

$$\begin{aligned} \Delta f = f_2 - f_1 &= \frac{\phi_2 c}{4\pi\Delta z} - \frac{\phi_1 c}{4\pi\Delta z} = \frac{(\phi_1+\pi)c}{4\pi\Delta z} - \frac{\phi_1 c}{4\pi\Delta z} \\ &= \frac{\pi c}{4\pi\Delta z} = \frac{c}{4\Delta z}. \end{aligned} \tag{4.66}$$

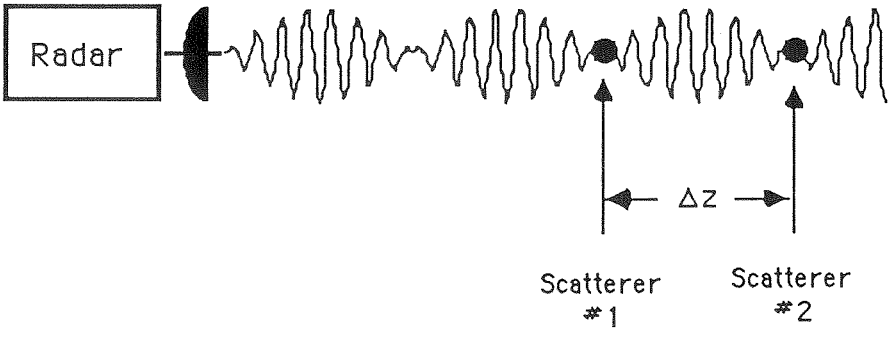


Figure 4.2. Constructive interference from two scatterers separated by a distance Δz and illuminated by two sinusoids with a frequency difference of $\Delta f = c/2\Delta z$.

Thus, we have that the frequency difference Δf from constructive interference to destructive interference is $c/4\Delta z$.

The value of the smallest non-zero frequency difference for which constructive interference occurs can be found from Eq. (4.63) taking $n = 1$. In this case, we have

$$\phi_2 = \phi_1 + 2\pi, \tag{4.67}$$

and thus,

$$\begin{aligned} \Delta f = f_2 - f_1 &= \frac{\phi_2 c}{4\pi\Delta z} - \frac{\phi_1 c}{4\pi\Delta z} = \frac{(\phi_1 + 2\pi)c}{4\pi\Delta z} - \frac{\phi_1 c}{4\pi\Delta z} \\ &= \frac{2\pi c}{4\pi\Delta z} = \frac{c}{2\Delta z}. \end{aligned} \tag{4.68}$$

From Eqs. (4.66) and (4.68), we see that for $\Delta f = 0$ we have constructive

interference in the radar return. As we increase Δf, there is partial destructive interference until we reach $\Delta f = c/4\Delta z$, at which point there is complete destructive interference. If we increase Δf further, we again have partial destructive interference until we reach $\Delta f = c/2\Delta z$, at which point we have constructive interference. Beyond this point the interference repeats periodically in Δf with period $c/2\Delta z$. We can see then that if we wish to resolve two scatterers separated by a distance Δz along the path of propagation easily, a minimum bandwidth of $\Delta f = c/2\Delta z$ is required, as this gives us the complete interference pattern in Δf generated by the two scatterers. We will write this frequncy bandwidth required to separate the two scatterers separated by distance Δz as

$$\Delta f_{\Delta z} = \frac{c}{2\Delta z}. \tag{4.69}$$

So by a simple physical argument, we have demonstrated that the bandwidth required to resolve two scatterers separated by Δz in the range is inversely proportional to Δz, and is in fact given by Eq. (4.69).

It is interesting to note the similarity between matching a pair of frequencies to a pair of scatterers to obtain constructive interference and selecting a single frequency with which to illuminate a pair of scatterers in order to obtain constructive interference. This latter situation is shown in Fig. 4.3.

Fig. 4.3 shows the case of two point scatterers separated in range by a distance Δz. There are returns at the radar receiver resulting from scattering by both of the scatterers. The signal from the second lags the signal from the first by a phase difference of $2\pi 2\Delta z/\gamma$ or equivalently $2\pi 2\Delta z f_0/c$. If $z(t)$ is the total return signal

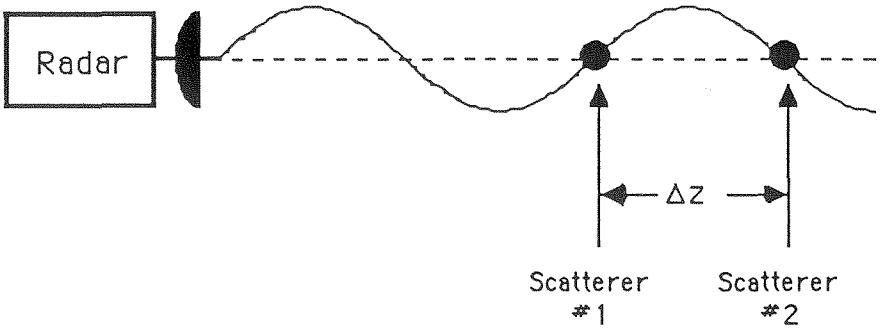


Figure 4.3. Constructive interference from two scatterers separated by a distance Δz and illuminated by a single sinusoid with a frequency of $f = c/2\Delta z$.

at the radar receiver, then ignoring the round trip delay time to and from the first scatterer,

$$\begin{aligned} z(t) &\propto x(t) + x(t - 2\Delta z/c) \\ &= \cos 2\pi f t + \cos 2\pi f(t - 2\Delta z/c) \\ &= \cos 2\pi f t + \cos(2\pi f t - 4\pi f \Delta z c). \end{aligned} \tag{4.70}$$

We can see from Eq.(4.70) that constructive interference occurs when $4\pi f_0 \Delta z/c = 2\pi n$ and destructive interference occurs when $4\pi f_0 \Delta z/c = (2n-1)\pi$, where n is any integer. Thus the frequencies at which **constructive interference** occurs are given by

$$f_{C,n} = \frac{nc}{2\Delta z}, \tag{4.71}$$

and the frequencies at which **destructive interference** occurs are given by

$$f_{D,n} = \frac{(2n-1)c}{2\Delta z}, \tag{4.72}$$

where n takes on integer values. In particular, the smallest positive frequency at which **constructive interference** occurs is

$$f_C = \frac{c}{2\Delta z}, \tag{4.73}$$

and the smallest positive frequency at which **destructive interference** occurs is

$$f_D = \frac{c}{4\Delta z}. \tag{4.74}$$

Note the correspondence between the constructive interference cases of Eqs. (4.68) and (4.73), as well as the correspondence between the destructive interference cases of Eqs. (4.66) and (4.74).

That this phenomenon of constructive and destructive interference occurs on frequency scales corresponding to both Δf and f_0 can be seen by noting the symmetry of $x(t)$ as expressed in Eq. (4.60), which is repeated here:

$$x(t) = 2\left[\cos 2\pi \Delta f t\right]\left[\cos 2\pi f_0 t\right]. \tag{4.60}$$

We see that Eq. (4.60) is symmetric in $\cos 2\pi \Delta f t$ and $\cos 2\pi f_0 t$, and that they are both product terms of Eq. (4.60). The case of matching a single sinusoid of frequency f_0 to the two scatterers is really just a special case of Eq. (4.60), in which $\Delta f = 0$ (ignoring the scale factor of 2). It is interesting that many references [4.2, 4.3, 4.4] that consider the interference from spatially distributed scatterers emphasize the interaction that is due to the carrier frequency term, but not to that of the modulation term. For most radar systems, f_0 ranges between 1 GHz and 20 GHz, and so the corresponding Δz for constructive interference would be on the

order of 0.15 m to 0.75 cm. But it is not unusual to have modulation bandwidths of 1 MHz to 10 MHz, with corresponding Δz for constructive interference of 150 m down to 15 m. So the signal given by Eq. (4.60) would experience constructive interference with, or would be "matched" to, scatterers spatially distributed on two very different scales, one scale corresponding to the carrier frequency and one scale corresponding to the modulation bandwidth.

In examining an interference pattern such as that in Fig. 4.2, components that are due to both carrier and the modulation can be seen. While the interference pattern arising from the modulation term is the physical basis for the range resolution capabilities of radar waveforms, the viewpoint of Eq. (4.60) is not often explicitly stated. Yet this viewpoint is instructive, as it shows that the interaction with scatterers that is due to waveform modulation is described by simple wave interference phenomena, and provides physical insight into the effect of the shape of the transmitted waveform in scattering from spatially distributed scatterers.

It is also significant to note that since the scales at which the carrier and modulation components of the transmitted waveform are "matched" to a surface are very different (a couple of orders of magnitude for most real radar systems), a surface may appear rough or diffuse on the frequency scale of the carrier but relatively smooth on the frequency scale of the modulation bandwidth. Thus, although a surface may appear rough compared to the wavelength corresponding to the carrier frequency, it may exhibit a larger-scale scattering variation that may be characterizable with proper waveform modulation. An example of such a surface might be a

surface made up of several patches of homogeneous rough surface, each with different scattering characteristics, much like a checkerboard, with each square consisting of a rough homogeneous scattering surface differing from neighboring squares. Another example might be crops planted in rows, where the crops themselves would appear to be rough compared to the wavelength of the carrier, but where the spacing between rows is on the order of the modulation wavelength corresponding to Δf. Yet another example might be the surface of the ocean, where wind waves and turbulence may cause the surface to be rough on the scale of the carrier frequency, yet the larger gravity waves or swells may have a periodic structure on the order of the wavelength corresponding to the waveform modulation bandwidth [4.17]. Fig. 4.4 illustrates what this surface might look like.

So one can view waveforms in terms of the interference patterns they generate on the scales of both their carrier wavelength and their modulation wavelength when considering the remote sensing of physical surfaces. The significance of this lies in the fact that many natural surfaces are rough with respect to carrier wavelength, but their material and geometric (or roughness) characteristics vary spatially. One of the main purposes of remote sensing is, in fact, to determine the boundaries between different regions and to produce an image or map that shows the distinctions between these different regions. Since the variation of importance distinguishing these regions is often occurring on a scale larger than the carrier wavelength $\gamma = c/f_0$, the interference characteristics that are due to the waveform modulation may be useful in characterizing the variation and texture of these regions.

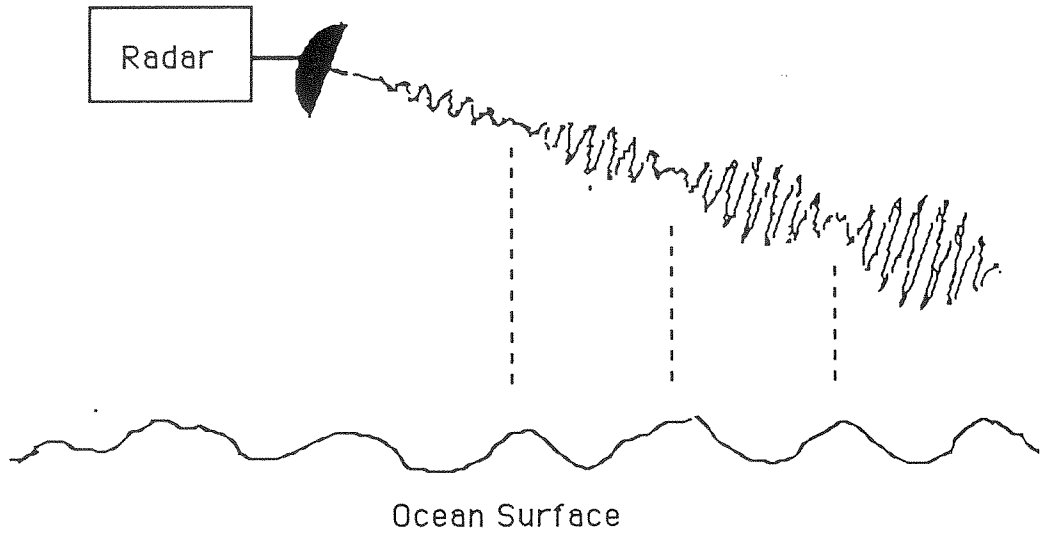


Figure 4.4. Radar making measurements of a surface that is rough on the scale corresponding to f_0 but that has variation on the scale corresponding to Δf.

Most real targets encountered are more complex than the two scatterers shown in Fig. 4.2. They are generally made up of many distributed scattering centers. In general, if a target is made up of n individual scatterers, there will be several different effective Δz_i between scatters (in general $n(n-1)/2$ different Δz_i). With a different Δf_i for constructive interference corresponding to each distinct Δz_i, it seems feasible to characterize the structure of a group of distributed scatterers by measuring the frequency response of the surface with respect to Δf. Fig. 4.5 illustrates how this might be done for a real physical target. In this example, the structure of the target on two different scales is determined with multiple Δf

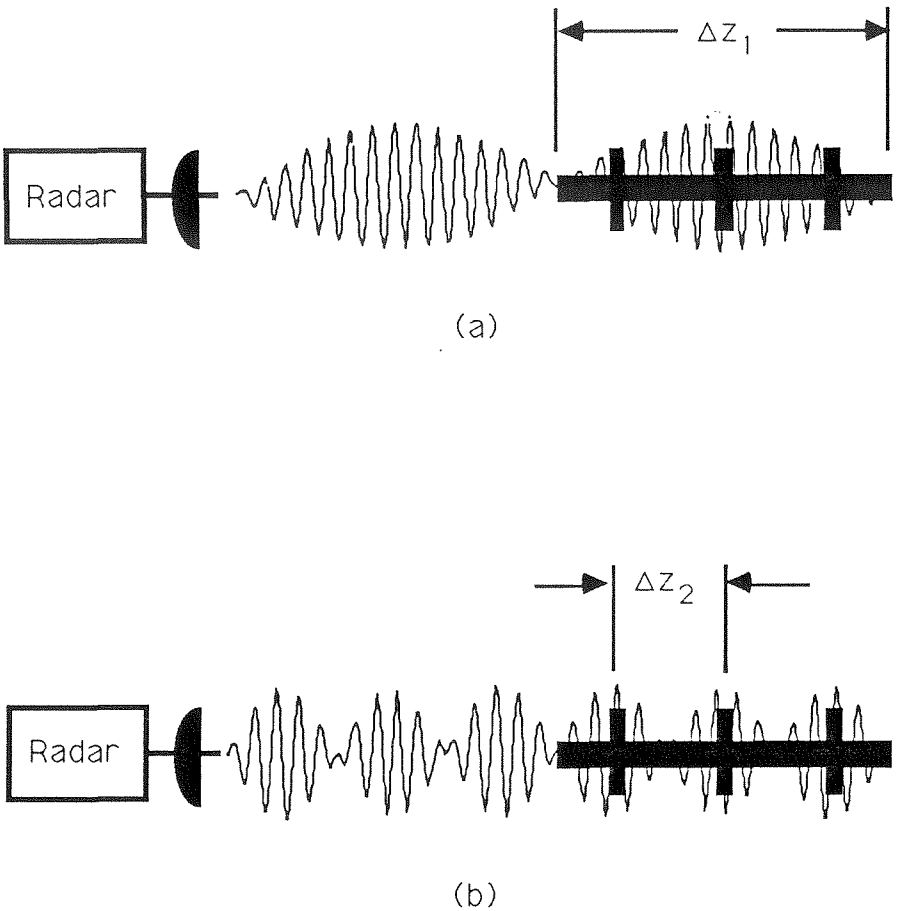


Figure 4.5. Different values of Δf cause constructive interference for scatterers of different resolution scales on a single target made up of complex scatterers. Note that $\Delta f_1 = c/2\Delta z_1$ in (a) is less than $\Delta f_2 = c/2\Delta z_2$ in (b), since $\Delta z_1 > \Delta z_2$.

measurements.

The characterization of distributed targets as a function of their frequency response measurements is similar to the problem of characterizing linear systems by frequency response measurements. The similarity arises because the scattering

of electromagnetic fields is a linear process. This can be seen by noting that if $\vec{E}^i$ is the incident electric field on a scatterer and $\vec{E}^s$ is the scattered electric field at some point in space, the fields are related through the polarization scattering matrix A, as specified in Eqs. (4.18) and (4.19), by the relation

$$\vec{E}^s = \mathbf{A}\vec{E}^i. \tag{4.75}$$

The elements of the polarization scattering matrix A are complex scalars. From Eq. (4.75), we can see that both homogeneity and superposition hold in electromagnetic scattering, since if $\vec{E}_1^s = \mathbf{A}\vec{E}_1^i$ and $\vec{E}_2^s = \mathbf{A}\vec{E}_2^i$, then for any complex scalars α and β,

$$\mathbf{A}\left(\alpha\vec{E}_1^i + \beta\vec{E}_2^i\right) = \alpha\mathbf{A}\vec{E}_1^i + \beta\mathbf{A}\vec{E}_2^i = \alpha\vec{E}_1^s + \beta\vec{E}_2^s.$$

Thus, electromagnetic scattering can be modeled as a linear process.

In engineering, one common method of studying linear processes is to view them as linear systems and to study the system input/output relationships. In addition to being linear, the system can also be time-invariant, as would be true if the radar and target were not moving rapidly with respect to each other. The system impulse response is then a convenient tool for characterizing the input/output relationships of the system. We will now examine the application of linear systems analysis to electromagnetic scattering problems.

To apply linear systems analysis to scattering problems, we will first define the input and output quantities to be the electric field magnitudes at a pair of points in space. A block diagram of this linear time-invariant system is shown in

Fig. 4.7. We will assume a fixed, although not necessarily identical, polarization at each point. The input $e(t)$ is the $\hat{\mathbf{x}}$ electric field intensity at point $\mathcal{P}_1$. We assume that the plane wave is moving along the line connecting $\mathcal{P}_1$ and the origin. Now if the plane wave is incident on the target located at the origin, as shown in Fig. 4.6, a scattered electric field will be present at an arbitrarily chosen observation point $\mathcal{P}_2$. We select an arbitrary polarization $\hat{\mathbf{x}}'$ at $\mathcal{P}_2$ and view as the output of our linear system the electric field $v(t)$ at point $\mathcal{P}_2$ with polarization $\hat{\mathbf{x}}'$. Thus, restricting the direction and polarization of the incident plane wave and selecting a point $\mathcal{P}_2$ 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 $h(t)$. In principle, the impulse response of our linear time-invariant scattering system can be obtained at the system output $v(t)$ if the incident electric field is an impulsive plane wave, that is, if $e(t) = \delta(t)$, where $\delta(t)$ is the Dirac delta function. For general $e(t)$, the output $v(t)$ of the linear system is given by

$$v(t) = \int_{-\infty}^{\infty} h(\tau)e(t-\tau)\,d\tau. \tag{4.76}$$

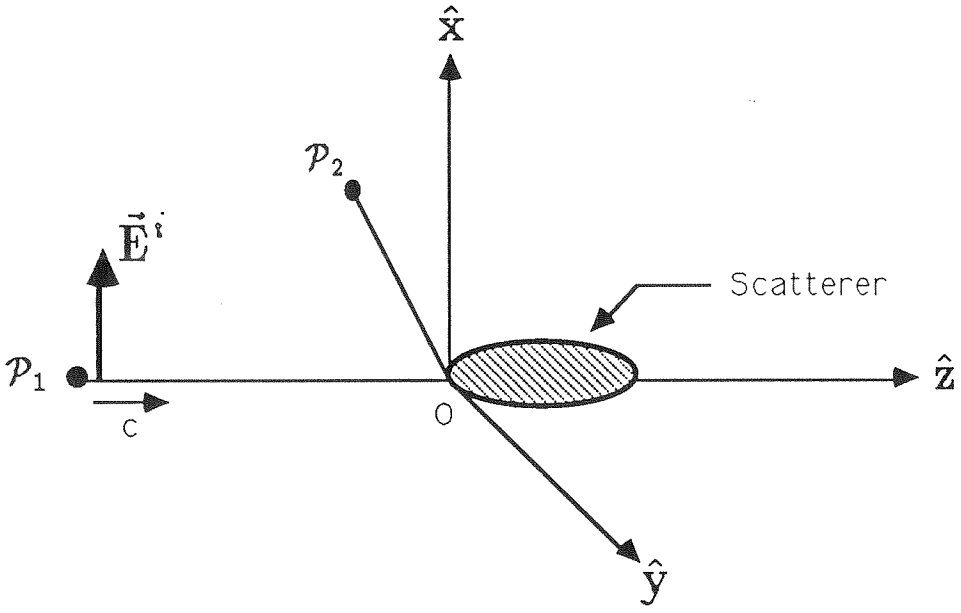


Figure 4.6. Coordinate System of Scattering Problem.

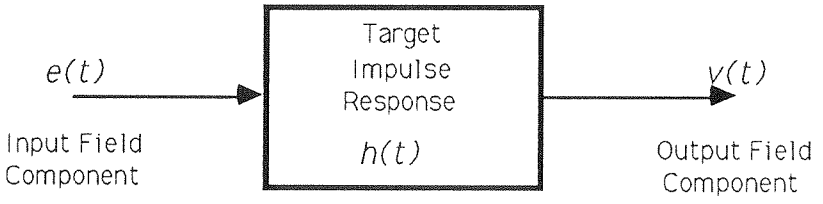


Figure 4.7. Linear Time-Invariant System Representation of the Stationary Target-Scattering Mechanism.

Let the Fourier transforms of $e(t)$, $v(t)$, and $h(t)$ be

$$\begin{aligned} E(f) &= \int_{-\infty}^{\infty} e(t)e^{-i2\pi ft}\,dt, \\ V(f) &= \int_{-\infty}^{\infty} v(t)e^{-i2\pi ft}\,dt, \\ H(f) &= \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,dt. \end{aligned} \tag{4.77}$$

Then by the convolution theorem of Fourier transforms [4.18], we have

$$V(f) = E(f)H(f). \tag{4.78}$$

The significance of Eq. (4.78) to our analysis of scattering is that $H(f)$ corresponds to a frequency response measurement of the target or scatterer, and the impulse response can be obtained from $H(f)$ by the inverse Fourier transform:

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{i2\pi ft}\,df. \tag{4.78}$$

So the impulse response characterization of a target is equivalent to measuring its frequency response for all frequencies. Practically, of course, this cannot be done. It can, however, be done approximately, using large-bandwidth radar waveforms.

The impulse response characterization of targets also allows one to characterize not only discrete targets made up of a discrete collection of scatterers but also distributed targets made up of a continuous scattering structure. The wideband waveforms used in characterizing the impulse response of scatterers will in general have continuous frequency spectra $X(f)$, as opposed to the line spectra used when

illuminating targets with a pair of sinusoids as was previously done. Such a continuous spectrum might be similar to the one shown in Fig. 4.8a. These signals, however, are still generated by modulating a baseband signal, such as that in Fig. 4.8b, up to a higher frequency.

What is significant to note is that in the case of $x(t)$ with continuous spectrum $X(f)$, there is still interference resulting from the carrier term $\cos 2\pi f_0 t$ and also from the modulating waveform $r(t)$, which provides information about the target on a frequency scale corresponding to the bandwidth of $r(t)$ or more specifically to the frequency differences present in the spectrum $R(f)$ of $r(t)$.

In the next two chapters, we will examine the design of waveforms optimized for two different tasks. We will use the notion of target impulse response in both cases. First, we will examine the design of waveforms with optimum detection properties for targets either with a given impulse response or with a given family of impulse responses. After that, we will examine the design of radar waveforms that maximize the mutual information between an ensemble of targets and the received signal at the radar receiver.

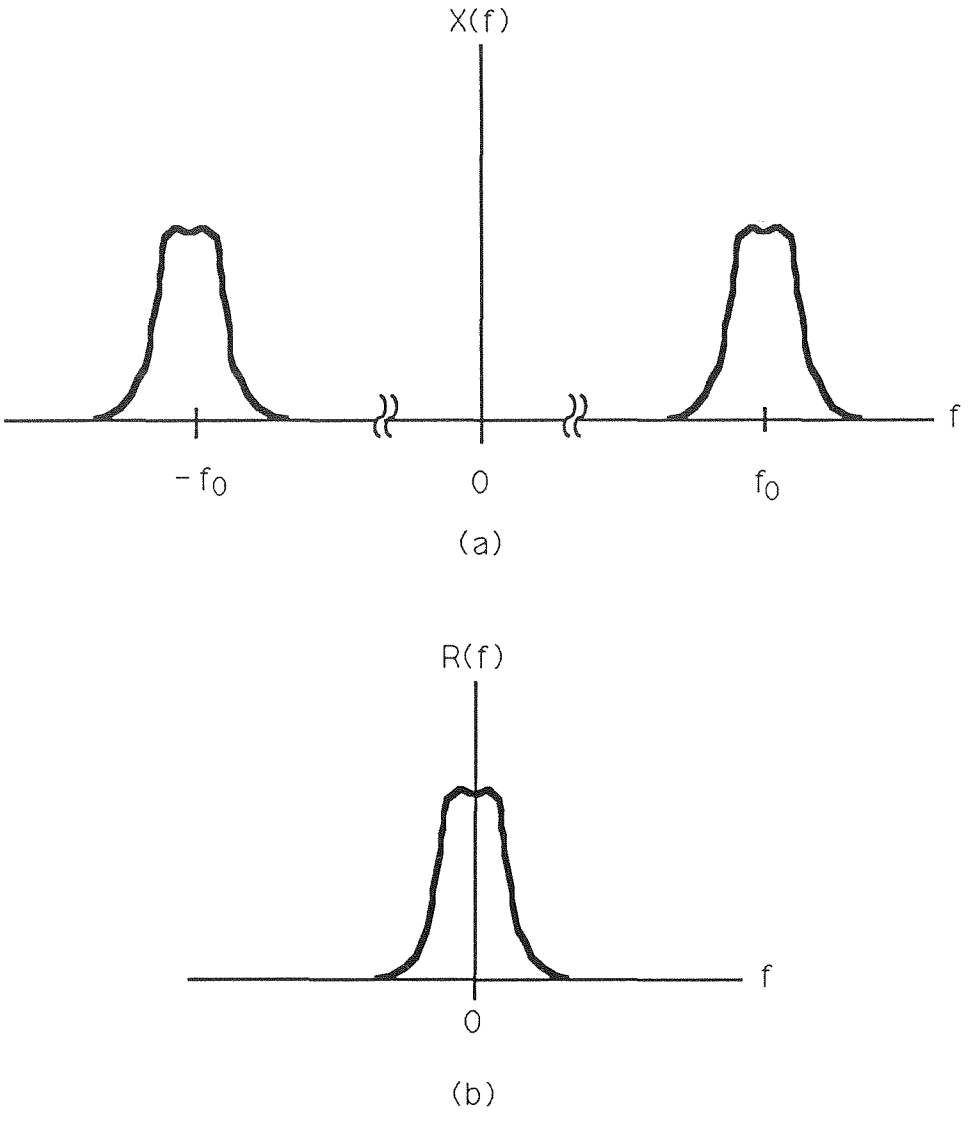


Figure 4.8. (a) Continuous Spectrum $X(f)$ of Radar Waveform. (b) Spectrum $R(f)$ of Baseband Modulating Signal Used to Generate Waveform $x(t)$ with Spectrum $X(f)$.

4.7. Chapter 4 References.

4.1 Elachi, C., *Physics and Techniques of Remote Sensing*, John Wiley and Sons, New York, NY, 1987.

4.2 Long, M. W., *Radar Reflectivity of Land and Sea*, Artech House, Dedham, MA, 1983.

4.3 Ulaby, F. T., *Microwave Remote Sensing*, vol. 2, Addison-Wesley, Reading, MA, 1982.

4.4 Ulaby, F. T., *Microwave Remote Sensing*, vol. 3, Artech House, Dedham, MA, 1986.

4.5 Tsang, L., J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing*, John Wiley and Sons, New York, NY, 1985.

4.6 Ruck, G. T., D. E. Barrick, W. D. Stuart, and C. K. Krichbaum, *Radar Cross Section Handbook*, vols. 1 and 2, Plenum Press, New York, NY, 1970.

4.7 Ostrovityanov, R. V., and F. A. Basalov, *Statistical Theory of Extended Radar Targets*, Artech House, Dedham, MA, 1985.

4.8 Ramo, S., J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communications Electronics*, 2nd. ed., John Wiley and Sons, New York, NY, 1984.

4.9 Pierce, J. R. and E. C. Posner, *Introduction to Communication Science and Systems*, Plenum, New York, NY, 1980. 1984.

4.10 Cook, C. E., and M. Bernfield, *Radar Signals*, Academic Press, New York, NY, 1967.

4.10 Rihaczek, A. W., *Principles of High Resolution Radar*, Mark Resources Inc.,

Marina del Rey, CA, 1977.

4.11 Van Zyl, J., On the Importance of Polarization in Radar Scattering Problems, Ph.D Dissertation, California Institute of Technology, Pasadena, CA, 1986.

4.12 Berkowitz, R. S., *Modern Radar*, John Wiley and Sons, New York, NY, 1965.

4.13 Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics*, Macmillan, New York, NY, 1970.

4.14 Rice, S. O., "Mathematical Analysis of Random Noise," *Bell System Tech. J.*, **23**, pp. 282-322; **24**, 46-156. Reprinted in *Noise and Stochastic Processes*, N. Wax, ed., Dover Publications, New York, NY, 1954.

4.16 Giuli, D., "Polarization Diversity in Radar Systems," *Proceedings of the IEEE*, vol. **74**, No. 2, Feb., 1986.

4.17 Gjessing, D. J., *Target Adaptive Matched Illumination Radar: Principles and Applications*, Peter Peregrinus Ltd., London, U. K., 1986

4.18 Bracewell, R. N., *The Fourier Transform and Its Applications*, McGraw-Hill, New York, NY, 1986.

CHAPTER 5

MATCHING A RADAR WAVEFORM/RECEIVER-FILTER PAIR TO A TARGET OF KNOWN IMPULSE RESPONSE

In this chapter, we will take a brief detour from our study of the information-theoretic analysis of radar systems in order to apply the target impulse response discussed in the previous chapter to the problem of optimum waveform and receiver-filter design for radar target detection. While the matched filter is typically used in radar detection problems, the matched filter derivation in radar assumes that the target under observation is a point target. As a result, the solution achieving maximum signal-to-noise ratio at the receiver output is independent of the transmitted waveform's wave-shape and is a function only of its energy.

Real radar targets, however, are not point targets and may have considerable spatial extent. As a result, the electric field scattered from such a target is scattered from a multiple number or even a continuum of points in space. When this occurs, the resultant scattered field is the superposition of the field scattered from the various points on the target. The target impulse response discussed in Chapter 4 provides a convenient representation of this scattering from extended targets. In this chapter, we will assume that the target impulse response of the target we wish to detect is known, and we will design a waveform and a receiver-filter that maximizes the signal-to-noise ratio at the receiver output. We will constrain the transmitted waveform to a finite time interval and constrain the waveform energy

to a fixed value. It will be assumed that the scattered signal is received in the presence of stationary additive noise with power spectral density $S_{nn}(f)$. We will see that the resulting waveforms are realizable.

Before we solve for the optimum detection waveforms, we will briefly review radar detection using a threshold test on the energy in the received waveform. We will see that for the two most common decision rules applied in radar detection problems, the *Neyman-Pearson* and *Bayes* decision rules, the detection performance improves as the signal-to-noise ratio increases. This is what we would intuitively expect.

5.1. Radar Target Detection by Energy Threshold Test.

In the detection of 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. This is because when a target is present, the received signal has a component that is due to reflection of the transmitted wave by the radar target and a component that is due to noise. However, when no target is present, all of the energy in the received signal is due to noise.

We would expect intuitively that in order to obtain the best target detection performance, the ratio of the energy in the received signal that is due to reflection from the target to the energy due to the noise—the *signal-to-noise ratio*—should be as large as possible. Formally, it can be shown that this is also the case for radar systems that use the Neyman-Pearson criterion and the Bayes criterion in order

to decide if a target is present [5.1, 5.2]. In order to see why this is the case, we consider the notion of an energy threshold test in radar detection problems. Let E be the energy in the received radar signal. As we have already noted, there are two possible components to the energy in the received signal. There is a component E_T that is due to the reflection of the transmitted waveform from the target if a target is present, and a noise component E_N, which is present whether or not a target is present. Generally, E_T and E_N will both be random variables, E_N being random, since it is a functional of a random process, and E_T being random because of inexact knowledge of the target's nature and distance from the radar. We will assume that E_T and E_N are statistically independent, and we note that since they represent the energy components of the received signal, they are non-negative. We will assume that the probability density function (PDF) of E_N is $f_N(e_N)$ and, given that a target is present, the PDF of E_T is $f_T(E_T)$.

In the detection of radar targets, we either have the hypothesis H_0, that no target is present, or we have the hypothesis H_1, that a target is present. We observe the energy E in the received signal. Based on observation of E we make a decision as to whether the hypothesis H_0 or the hypothesis H_1 is true. We note that in the case that H_0 is true, the conditional PDF of E is

$$f(e|H_0) = f_N(e),$$

and in the case that H_1 is true, the conditional PDF of E is

$$f(e|H_1) = f_N(e) * f_T(e).$$

Here "$*$" denotes convolution. In making an observation and decision as to the presence or absence of a target, there are four possible outcomes:

1. H_0 true, H_0 decided.
2. H_0 true, H_1 decided.
3. H_1 true, H_0 decided.
4. H_1 true, H_1 decided.

Outcomes 1 and 4 correspond to correct decisions, whereas outcomes 2 and 3 correspond to incorrect decisions or errors.

In order to make a decision as to whether H_0 or H_1 is true based on a threshold test on the observation of E, we define a decision rule $D(E)$ as

$$D(E) = \begin{cases} H_0 & \text{, for } E < \xi; \\ H_1 & \text{, for } E \geq \xi. \end{cases}$$

Here, the threshold ξ is a non-negative real number. Let p_{jk} be the probability of deciding H_j based on observation of E, given that H_k is true, for $j, k = 0, 1$. Then we have that the probabilities p_{jk} for the four possible outcomes are as follows:

$$p_{00} = \int_0^{\xi} f(e|H_0)\, de,$$

$$p_{10} = \int_{\xi}^{\infty} f(e|H_0)\, de,$$

$$p_{01} = \int_0^{\xi} f(e|H_1)\, de,$$

$$p_{11} = \int_{\xi}^{\infty} f(e|H_1)\, de.$$

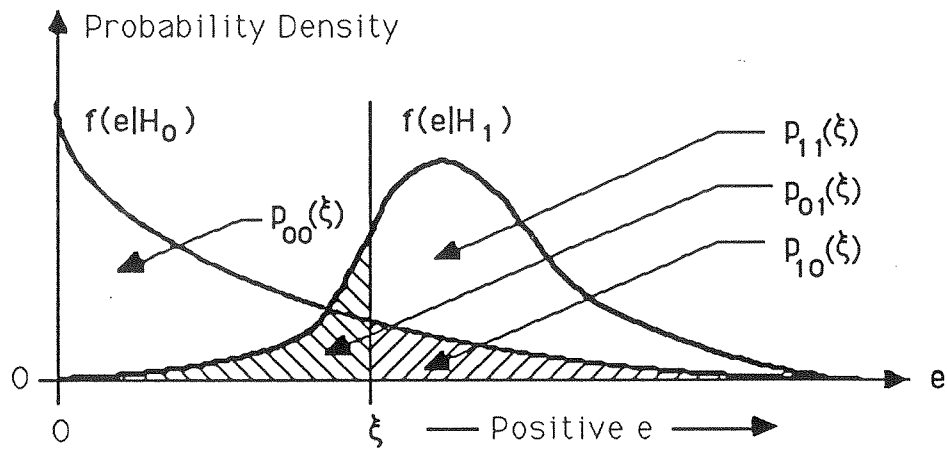


Figure 5.1. Probability Density Functions $f(e|H_0)$ and $f(e|H_1)$.

Fig. 5.1 illustrates these regions.

The question arises as to how to pick the threshold ξ in this decision rule. The *Neyman-Pearson decision criterion* and the *Bayes decision criterion* are methods of selecting this threshold. We will now examine these two methods.

The *Bayes decision criterion* determines the optimum value of the threshold ξ by selecting it to minimize a cost function $C(\xi)$. This cost function is constructed by assigning a set of costs $\{C_{jk}\}$ to the possible outcomes, where

$$C_{jk} = \text{cost of deciding } H_j \text{ given } H_k \text{ is true, for } j,k = 0,1.$$

It is then assumed that there is an *a priori* probability P_0 that H_0 actually occurred and an *a priori* probability that H_0 actually occurred. The expected value of the

cost associated with an observation and decision based on a threshold ξ is then given by

$$C(\xi) = P_0 C_{00} p_{00}(\xi) + P_0 C_{10} p_{10}(\xi)$$

$$+ P_1 C_{01} p_{01}(\xi) + P_1 C_{11} p_{11}(\xi).$$

The *Bayes decision criterion* selects the decision threshold ξ such that $C(\xi)$ is minimum. As is discussed in Reference [5.2, p. 82], the ξ that minimizes $C(\xi)$ can be found by setting

$$\frac{f(\xi|H_1)}{f(\xi|H_0)} = \frac{P_0 (C_{10} - C_{00})}{P_1 (C_{01} - C_{11})}.$$

The greater the signal energy, the farther to the right the PDF $f(e|H_1)$, while the density $f(e|H_0)$ remains unchanged, since the noise energy is assumed to be independent of the signal energy. As a result, increasing the signal energy decreases both $p_{10}(\xi)$ and $p_{01}(\xi)$, while increasing $p_{00}(\xi)$ and $p_{11}(\xi)$ for the optimum ξ. As a result, for $C_{10} > C_{00}$ and $C_{01} > C_{11}$, $C(\xi)$ decreases as the signal energy increases for the optimally chosen ξ. So the Bayes decision rule gives better performance as the signal-to-noise ratio increases if $C_{10} > C_{00}$ and $C_{01} > C_{11}$, that is, when it is more costly to make an incorrect decision about the presence or absence of a target than it is to make a correct decision. We would expect this to be true of all reasonable cost functions. The Bayes decision criterion is not often used in real radar detection problems because of the difficulty in determining the *a priori* probabilities P_0 and P_1 as well as a meaningful set of costs $\{C_{jk}\}$.

The *Neyman-Pearson* decision criterion, when applied to radar detection problems, gets around the problem of determining the *a priori* probabilities P_0 and P_1 and the set of costs $\{C_{jk}\}$. Here instead we look at the problem of radar detection

as one in which we want to maximize the probability $p_{11}(\xi)$ of deciding that a target is present when one is present, while holding the probability $p_{10}(\xi)$ of deciding that a target is present when one is not present to a reasonably small level. In radar target detection problems, $p_{10}(\xi)$ is often referred to as the "false alarm" probability. Ideally, we would, of course, like to select ξ to simultaneously make $p_{11}(\xi)$ as large as possible and $p_{10}(\xi)$ as small as possible, but these are usually conflicting objectives. As a result, we select as our maximum allowable value of $p_{10}(\xi)$ a probability P_F, and we then maximize $p_{11}(\xi)$ with respect to ξ under the constraint that $p_{10}(\xi) \leq P_F$. For all energy threshold tests of the type we are considering, we note that $p_{11}(\xi)$ is a monotonically decreasing function of ξ and $p_{10}(\xi)$ is a monotonically decreasing function of ξ. As a result, the Neyman-Pearson solution corresponds to selecting the optimal threshold ξ such that

$$p_{10}(\xi) = \int_0^{\xi} f(e|H_0)\, de = P_F.$$

Then $p_{11}(\xi)$ is as large as possible under the constraint on $p_{10}(\xi)$, and this maximum value is

$$p_{11}(\xi) = \int_{\xi}^{\infty} f(e|H_1)\, de.$$

Again, we see that as the energy in the reflected signal from the target increases, the PDF $f(e|H_1)$ shifts to the right, while $f(e|H_0)$ remains unchanged. As a result, the performance of the Neyman-Pearson decision improves (that is, $p_{11}(\xi)$ increases for a fixed $p_{10}(\xi)$) as the signal-to-noise ratio increases.

We have seen that for radar detection systems that use energy thresholding of the received signal as the method of determining the presence or absence of a target,

the performance of the detection system improves with the signal-to-noise ratio of the received signal, whether the threshold is chosen using a Bayes decision rule or a Neyman-Pearson decision rule. The method by which the maximum signal-to-noise ratio of the received signal is obtained in radar sytems is that of *matched filtering*[5.3].

In the derivation of the matched filter in radar [5.1], it is assumed that the target reflecting the radar wave behaves as a point scatterer. The result is that the backscattered waveform received by the receiver is a scaled and delayed version of the transmitted waveform. Real radar targets, however, may be of significant physical extent, and the assumption that the target behaves as a point scatterer may, in fact, result in the design of a filter that is actually mismatched to the received signal because of the interference pattern generated by the waveform and target in combination. The matched filter in radar, derived under the assumption that the target behaves as a point scatterer, leads to the conclusion that the wave shape of the transmitted waveform has no effect on the signal-to-noise ratio at the output of the receiver. As long as the receiver filter is a matched filter matched to the transmitted waveform and the power spectral density (PSD) of the additive noise, the signal-to-noise ratio is a function only of the transmitted waveform's energy and not its wave-shape. Thus, for a point target, all waveforms of equal energy provide an equal signal-to-noise ratio at the receiver output if detected with a matched filter.

But as we have noted, the point target assumption is not realistic for many real

radar targets. Because these extended targets interact with the transmitted waveform to generate interference patterns that are a function both of the transmitted waveform and of the target's spatial characteristics, as we saw in Section 4.5, the signal-to-noise ratio for a given target is a function of both the wave shape of the transmitted waveform and the impulse response of the receiver filter. Thus, in order to maximize the signal-to-noise ratio at the receiver output for a given target, we must optimize both the transmitted waveform and the receiver filter.

In this chapter, we will investigate radar waveform/receiver-filter pairs matched to targets that exhibit behavior differing from that of a point target. This will be done through the use of the *target impulse response*, developed in Section 4.6. As a result of our investigation, a procedure for the design of optimum realizable radar waveform/receiver-filter pairs for a given target in stationary additive Gaussian noise of arbitrary power spectral density will be developed.

5.2. Matching a Waveform/Receiver-Filter Pair to a Target of Known Impulse Response.

In this section, we will consider the problem of selecting a transmitted waveform and receiver filter such that the signal-to-noise ratio at the output of a radar receiver is maximized. In order to do this, we will assume that the target we are attempting to detect has a known impulse response $h(t)$. We will also assume that the signal reflected from the target is contaminated by additive Gaussian noise with power spectral density (PSD) $S_{nn}(f)$.

The problem of interest is shown in Fig. 5.2. Here we have a radar transmitter

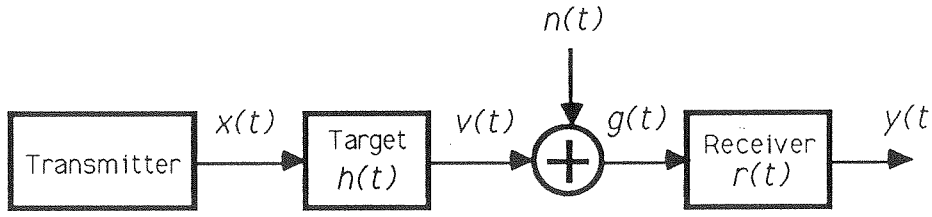


Figure 5.2. Block Diagram of Radar Waveform/Receiver-Filter Design Problem.

transmitting a waveform $x(t)$. The transmitted waveform is then scattered by the target with a target impulse response $h(t)$. The resulting scattered waveform $v(t)$ is then corrupted with additive Gaussian noise $n(t)$ having PSD $S_{nn}(f)$, producing the signal $g(t)$ at the input to the receiver filter. Finally, the receiver filter, with impulse response $r(t)$, filters $g(t)$ to produce the receiver output $y(t)$.

The problem of interest can be stated as follows: *Given a target impulse response $h(t)$ and stationary additive Gaussian noise $n(t)$ with PSD $S_{nn}(f)$, find a transmitted waveform $x(t)$ with total energy $\mathcal{E}$ 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, since real radar waveforms are of finite duration, restrict the waveform $x(t)$ such that it is zero outside the interval $[-T/2, T/2]$.*

From Fig. 5.2 we see that the receiver output is given by

$$\begin{aligned} y(t) &= r(t) * g(t) \\ &= r(t) * [x(t) * h(t) + n(t)] \\ &= r(t) * x(t) * h(t) + r(t) * n(t) \\ &= y_s(t) + y_n(t). \end{aligned} \tag{5.1}$$

Here, "$*$" denotes convolution, $y_s(t)$ is the signal component in the receiver output, and $y_n(t)$ is the noise component in the receiver output. These two components are given by

$$y_s(t) = r(t) * x(t) * h(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(\tau) h(p - \tau) r(t - p)\, d\tau\, dp, \tag{5.2}$$

and

$$y_n(t) = r(t) * n(t) = \int_{-\infty}^{\infty} n(p) r(t - p)\, dp. \tag{5.3}$$

Now as previously stated, we are interested in finding a waveform/receiver-filter pair that will maximize the output signal-to-noise ratio $\left(\frac{S}{N}\right)_{t_0}$ at time t_0. This signal-to-noise ratio is defined as

$$\left(\frac{S}{N}\right)_{t_0} \stackrel{\text{def}}{=} \frac{|y_s(t_0)|^2}{\mathrm{E}\,|y_n(t_0)|^2}. \tag{5.4}$$

The problem of finding the waveform/filter pair that maximizes the output signal-to-noise ratio is most easily solved in the frequency domain. Thus, we define

the following Fourier transforms:

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\,dt, \\ H(f) &= \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,dt, \\ V(f) &= \int_{-\infty}^{\infty} v(t)e^{-i2\pi ft}\,dt, \\ G(f) &= \int_{-\infty}^{\infty} g(t)e^{-i2\pi ft}\,dt, \\ Y(f) &= \int_{-\infty}^{\infty} y(t)e^{-i2\pi ft}\,dt. \end{aligned} \tag{5.5}$$

The corresponding inverse Fourier transforms are

$$\begin{aligned} x(t) &= \int_{-\infty}^{\infty} X(f)e^{i2\pi ft}\,dt, \\ h(t) &= \int_{-\infty}^{\infty} H(f)e^{i2\pi ft}\,dt, \\ v(t) &= \int_{-\infty}^{\infty} V(f)e^{i2\pi ft}\,dt, \\ g(t) &= \int_{-\infty}^{\infty} G(f)e^{i2\pi ft}\,dt, \\ y(t) &= \int_{-\infty}^{\infty} Y(f)e^{i2\pi ft}\,dt. \end{aligned} \tag{5.6}$$

The convolution theorem of Fourier transforms [5.4] states that if $a(t)$ has Fourier transform $A(f)$ and $b(t)$ has Fourier transform $B(f)$, then the Fourier transform of the convolution $a(t) * b(t)$ is $A(f)B(f)$. That is,

$$\int_{-\infty}^{\infty} [a(t) * b(t)]\, e^{-i2\pi ft}\,dt = A(f)B(f). \tag{5.7}$$

So from Eqs. (5.2), (5.6), and (5.7), we have

$$|y_s(t_0)|^2 = \left| \int_{-\infty}^{\infty} X(f)H(f)R(f)e^{i2\pi f t_0}\, df \right|^2 . \tag{5.8}$$

The mean-square value of $y_n(t_0)$ is given by

$$\begin{aligned} \mathrm{E}\,|y_n(t_0)|^2 &= \mathrm{E}\left\{ y_n(t_0)\overline{y_n(t_0)} \right\} = \mathcal{R}_{y_n y_n}(0) \\ &= \int_{-\infty}^{\infty} S_{y_n y_n}(f)\, df = \int_{-\infty}^{\infty} |R(f)|^2\, S_{nn}(f)\, df. \end{aligned} \tag{5.9}$$

Here, $\mathcal{R}_{y_n y_n}(\tau)$ is the autocorrelation function of $y_n(t)$.

In what follows, we will need to make use of the Schwartz Inequality [5.4, p.159], which says that if $A_1(f)$ and $A_2(f)$ are square integrable functions on the real line, then

$$\left| \int_{-\infty}^{\infty} A_1(f)A_2(f)\, df \right|^2 \le \int_{-\infty}^{\infty} |A_1(f)|^2\, df \int_{-\infty}^{\infty} |A_2(f)|^2\, df, \tag{5.10}$$

with equality if and only if $A_1(f) = K\overline{A_2(f)}$, where K is a complex constant. We will use this inequality to maximize the signal-to-noise ratio.

From Eqs. (5.4), (5.8), and (5.9), we have

$$\left(\frac{S}{N}\right)_{t_0} = \frac{\left| \int_{-\infty}^{\infty} X(f)H(f)R(f)e^{i2\pi f t}\, df \right|^2}{\int_{-\infty}^{\infty} |R(f)|^2\, S_{nn}(f)\, df}. \tag{5.11}$$

This can be rewritten as

$$\left(\frac{S}{N}\right)_{t_0} = \frac{\left| \int_{-\infty}^{\infty} \left[\frac{H(f)X(f)e^{i2\pi f t_0}}{\sqrt{S_{nn}(f)}} \right] \left[R(f)\sqrt{S_{nn}(f)} \right] df \right|^2}{\int_{-\infty}^{\infty} |R(f)|^2\, S_{nn}(f)\, df}. \tag{5.12}$$

Applying the Schwartz Inequality to the numerator, we have

$$\left| \int_{-\infty}^{\infty} \left[\frac{H(f)X(f)e^{i2\pi f t_0}}{\sqrt{S_{nn}(f)}} \right] \left[R(f)\sqrt{S_{nn}(f)} \right] df \right|^2$$

$$\leq \int_{-\infty}^{\infty} \frac{|H(f)X(f)|^2}{S_{nn}(f)} df \int_{-\infty}^{\infty} |R(f)|^2 S_{nn}(f)\, df, \quad (5.13)$$

with equality if and only if

$$R(f)\sqrt{S_{nn}(f)} = K\overline{\left[\frac{H(f)X(f)e^{i2\pi f t_0}}{\sqrt{S_{nn}(f)} } \right]}, \quad (5.14)$$

or equivalently,

$$R(f) = \frac{K\overline{X(f)H(f)}e^{-i2\pi f t_0}}{S_{nn}(f)}, \quad (5.15)$$

where K is a complex constant. Thus, we have

$$\left(\frac{S}{N}\right)_{t_0} \leq \frac{\int_{-\infty}^{\infty} |R(f)|^2 S_{nn}(f)\, df \int_{-\infty}^{\infty} \frac{|X(f)H(f)e^{i2\pi f t_0}|^2}{S_{nn}(f)} df}{\int_{-\infty}^{\infty} |R(f)|^2 S_{nn}(f)\, df} \quad (5.16)$$

$$= \int_{-\infty}^{\infty} \frac{|X(f)H(f)|^2}{S_{nn}(f)} df,$$

with equality if and only if Eq. (5.15) holds.

Thus, we have that Eq. (5.15) is a necessary condition to obtain the maximum signal-to-noise ratio at the output of the receiver. However, it actually insures only that

$$\left(\frac{S}{N}\right)_{t_0} = \int_{-\infty}^{\infty} \frac{|X(f)H(f)|^2}{S_{nn}(f)} df. \quad (5.17)$$

In order to obtain the maximum possible signal-to-noise ratio under the specified energy and time constraints on the transmitted signal $x(t)$, we must maximize Eq.

(5.17) under the constraints

$$\int_{-\infty}^{\infty} |X(f)|^2 \, df = \mathcal{E},$$

$$x(t) = 0, \quad \text{for all } t \notin [-T/2, T/2]. \tag{5.18}$$

Here, $\mathcal{E}$ is the total energy available for the transmitted waveform $x(t)$.

We can rewrite Eq. (5.17) as

$$\left(\frac{S}{N}\right)_{t_0} = \int_{-\infty}^{\infty} \left| X(f) \left[\frac{H(f)}{\sqrt{S_{nn}(f)}} \right] \right|^2 df = \int_{-\infty}^{\infty} |X(f)B(f)|^2 \, df, \tag{5.19}$$

where

$$B(f) = \frac{H(f)}{\sqrt{S_{nn}(f)}}. \tag{5.20}$$

Examining Eq. (5.20), we see that we can view its maximization as that of maximizing an integeral of the form

$$\int_{-\infty}^{\infty} |q(t)|^2 \, dt = \int_{-\infty}^{\infty} |Q(f)|^2 \, df = \int_{-\infty}^{\infty} |X(f)B(f)|^2 \, df, \tag{5.21}$$

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,

$$Q(f) = \int_{-\infty}^{\infty} q(t) e^{-i2\pi ft} \, dt,$$

and

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

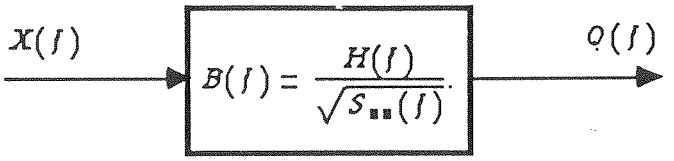


Figure 5.3. Linear System Representation of the Relation Between $Q(f)$ and $X(f)$.

Fig. 5.3 is a block diagram illustrating this situation. Since the system of Fig. 5.3 is a linear, time-invariant system, we have that

$$Q(f) = X(f)B(f). \tag{5.22}$$

We must maximize Eq. (5.21) with respect to all $X(f)$ that satisfy the constraints of Eq. (5.18). We can rewrite Eq. (5.21) for the energy in $q(t)$, noting that both $x(t)$ and $q(t)$ are real, as

$$\int_{-\infty}^{\infty} q^2(t)\,dt = \int_{-\infty}^{\infty} \left[\int_{-T/2}^{T/2} x(\tau) e^{-i2\pi f\tau}\,d\tau \right] \left[\int_{-T/2}^{T/2} x(p) e^{i2\pi fp}\,dp \right] |B(f)|^2\,df. \tag{5.23}$$

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

$$L(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |B(f)|^2\, e^{i2\pi ft}\,df. \tag{5.24}$$

Then from Eqs. (5.23) and (5.24), we have, changing the order of integration,

$$\int_{-\infty}^{\infty} q^2(t)\,dt = \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} x(\tau)x(p)L(p-\tau)\,d\tau\,dp. \tag{5.25}$$

We wish to find the function $\hat{x}(t)$ on $[-T/2, T/2]$, which maximizes Eq. (5.25). It can be shown [5.2, p.125] that the optimum function $\hat{x}(t)$ satisfies the integral equation

$$\lambda\hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau)L(t-\tau)\,d\tau, \tag{5.26}$$

where λ is the maximum eigenvalue of Eq. (5.26). So $\hat{x}(t)$ is an eigenfunction corresponding to λ, the maximum eigenvalue of Eq. (5.26), and having energy $\mathcal{E}$. From Eqs. (5.25) and (5.26), we then have

$$\begin{aligned}\int_{-\infty}^{\infty} q^2(t)\,dt &= \int_{-T/2}^{T/2} \hat{x}(\tau) \int_{-T/2}^{T/2} \hat{x}(p)L(p-\tau)\,dp\,d\tau \\ &= \int_{-T/2}^{T/2} \hat{x}(\tau)\lambda\hat{x}(\tau)\,d\tau = \lambda \int_{-T/2}^{T/2} \hat{x}(\tau)\hat{x}(\tau)\,d\tau \\ &= \lambda\mathcal{E}.\end{aligned} \tag{5.27}$$

Interpreting these results in terms of our problem, we substitute $B(f) = H(f)/\sqrt{S_{nn}(f)}$ and then define $L(t)$ as in Eq. (5.24). This yields

$$L(t) = \int_{-\infty}^{\infty} \frac{|H(f)|^2}{S_{nn}(f)} e^{i2\pi ft}\,df. \tag{5.28}$$

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

$$\lambda_{\max}\hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(t)L(t-\tau)\,d\tau, \tag{5.29}$$

where $\lambda_{\max}$ is the maximum eigenvalue of Eq. (5.29) and $\hat{x}(t)$ is a corresponding eigenfunction scaled to have energy $\mathcal{E}$.

From these results, we now have an optimum waveform/receiver-filter pair design algorithm. It is as follows:

1. Compute
$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,dt.$$
Here, $h(t)$ is the impulse response of the target.

2. Compute
$$L(t) = \int_{-\infty}^{\infty} \frac{|H(f)|^2}{S_{nn}(f)} e^{i2\pi ft}\,df.$$
Here $S_{nn}(f)$ is the two-sided power spectral density of the noise $n(t)$.

3. 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 $\mathcal{E}$.

4. Compute the spectrum $\hat{X}(f)$ corresponding to the optimum waveform $\hat{x}(t)$:
$$\hat{X}(f) = \int_{-\infty}^{\infty} \hat{x}(t)e^{-i2\pi ft}\,dt.$$

5. Implement a receiver filter of the form
$$R(f) = \frac{K\overline{\hat{X}(f)H(f)}e^{-i2\pi ft_0}}{S_{nn}(f)},$$
where K is a complex constant.

6. The resulting signal-to-noise ratio for this design, which is the maximum obtainable under the specified constraints, is
$$\left(\frac{S}{N}\right)_{t_0} = \lambda_{\max}\mathcal{E}. \tag{5.30}$$

We note that since the waveform $x(t)$ is designed to take on non-zero values only on the interval $[-T/2, T/2]$, it can take on non-zero values for $t < 0$, and so the waveform $x(t)$ is not necessarily causal (that is, it cannot in general be generated by exciting a realizable filter). However, since $x(t) = 0$ for all $t < -T/2$, we can obtain a causal waveform $\tilde{x}(t) = \hat{x}(t - T/2)$, which will also yield the optimal response at the receiver output, except with delay $T/2$, that is, at time $t_0 + T/2$. To see that this waveform also maximizes the signal to noise ratio, we note that $\tilde{X}(f) = \hat{X}(f)e^{-i\pi fT}$. But from Eq. (5.17) or intuition, we see that the phase term $e^{-i\pi fT}$, resulting from the delay of duration $T/2$, does not affect the resulting signal-to-noise ratio. We do, however, note from Eq. (5.15) that the response occurs at time $t_0 + T/2$ instead of time t_0.

The important thing to note here is that an optimal waveform $\tilde{x}(t)$ 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 we note that from the optimum waveform/receiver-filter solution, we can obtain a waveform/receiver-filter pair that is physically realizable.

We will now consider an example that both 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 noise with PSD

$$S_{nn}(f) = \frac{N_0}{2},$$

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 we will take to be equal to $\sqrt{N_0/2}$, so that

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

This being the case, we have

$$L(t) = \int_{-W}^{W} 1 \, df = 2W \frac{\sin 2\pi W t}{2\pi W t} = \frac{\alpha}{\pi} \frac{\sin \alpha t}{\alpha t}.$$

Here $\alpha = 2\pi W$. So we have that $\hat{x}(t)$ 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 to have a countable number of solutions for $n = 0, 1, 2, \ldots$ [5.1]. 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 λ_n can be written in terms of the *radial prolate spheroidal functions*, which are designated $R_{0n}^{(1)}(\alpha T/2, 1)$. The eigenvalues and their associated eigenfunctions are given by

$$\lambda_n = 2WT \, R_{0n}^{(1)}(\alpha T/2, 1), \qquad \text{for } n = 0, 1, 2, \ldots.$$

$$x_n(t) = S_{0n}(\alpha T/2, 2t/T),$$

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

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

Table 5-1

Signal-To-Noise Ratio for $x_n(t) = \sqrt{\mathcal{E}} S_{0n}(\alpha T/2, 2t/T)$.		
n	$2WT = 2.55$	$2WT = 5.10$
0	$0.966\mathcal{E}$	$1.000\mathcal{E}$
1	$0.912\mathcal{E}$	$0.999\mathcal{E}$
2	$0.519\mathcal{E}$	$0.997\mathcal{E}$
3	$0.110\mathcal{E}$	$0.961\mathcal{E}$
4	$0.009\mathcal{E}$	$0.748\mathcal{E}$
5	$0.004\mathcal{E}$	$0.321\mathcal{E}$
6	—	$0.061\mathcal{E}$
7	—	$0.006\mathcal{E}$

and its associated eigenvalue is

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

The signal-to noise ratio in this case is

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

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 $x_n(t) = \sqrt{\mathcal{E}} S_{0n}(\alpha T/2, 2t/T)$, for $n = 0, 1, 2, \ldots$. All of these waveforms have transmitted energy $\mathcal{E}$, but the resulting signal-to-noise ratios are $\lambda_n \mathcal{E}$. As we noted previously, $\{\lambda_n\}$ is a positive decreasing sequence of n, and thus $\lambda_0\mathcal{E} > \lambda_1\mathcal{E} > \lambda_2\mathcal{E} > \cdots$. So we see that the output signal-to-noise ratio is definitely a function of the transmitted waveform. In Table 5-1, we show the resulting

signal-to-noise ratio for the cases of $2WT = 2.55$ and $2WT = 5.10$. (The eigenvalues were obtained from Reference [5.1, 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 mode of 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 should, however, note that other modes of the target may contain significant information about the target, so putting as much energy as possible into the mode with the largest eigenvalue will not in general be the best method of obtaining the maximum amount of information about 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. We will examine this problem in the next chapter.

5.3. Optimum Waveform/Receiver-Filter Pairs for Sphere Detection.

In this section, we will apply the results of Section 5.2 to the problem of designing a waveform/receiver-filter pair that is optimum for detecting a perfectly conducting metal sphere of radius a in the presence of stationary white noise. We will examine the shape of the resulting waveform and compare its performance 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 (transmit and receive antennas collocated) and thus we are interested in the backscatter impulse response. We will also assume that the transmit and receive antennas employ identical linear polarization. The backscatter impulse response of a perfectly conducting sphere, when both transmit and receive antennas have identical linear polarization, has been calculated by Kennaugh and Moffatt, using the physical optics approximation [5.6]. This impulse response can be written as

$$h(t) = -\frac{a}{c}\delta(t) + \frac{1}{2}\left[u(t) - u(t - 2a/c)\right]. \tag{5.31}$$

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

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

A plot of $h(t)$ appears in Fig. 5.4.

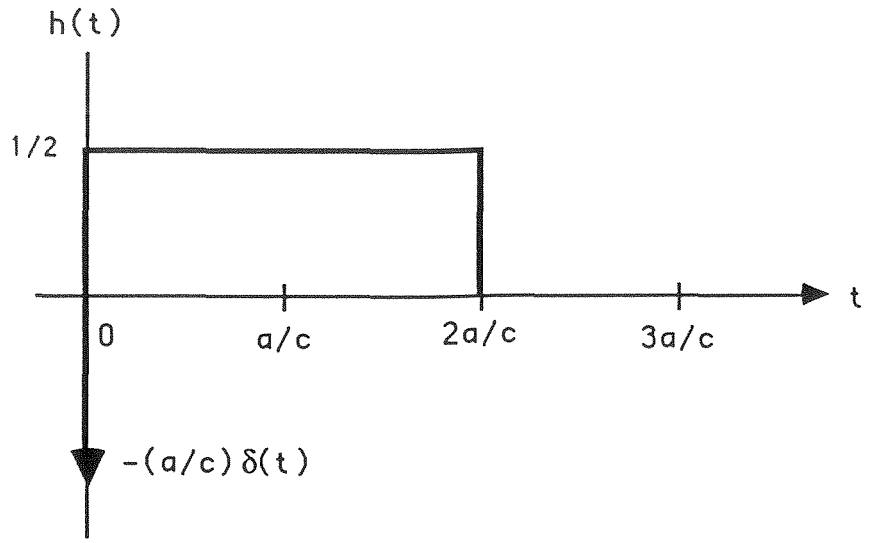


Figure 5.4. The Backscatter Impulse Response $h(t)$ of a Sphere of Radius a.

Applying our waveform/receiver-filter design procedure (pp.155–6), we must first calculate the Fourier Transform of $h(t)$:

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}\,dt \\ &= \int_{-\infty}^{\infty} \left[-\frac{a}{c}\delta(t) + \frac{1}{2}\left[u(t) - u(t-2a/c)\right]\right] e^{-i2\pi ft}\,dt \\ &= -\frac{a}{c}\int_{-\infty}^{\infty} \delta(t)e^{-i2\pi ft}\,dt + \frac{1}{2}\int_{0}^{2a/c} e^{-i2\pi ft}\,dt \\ &= -\frac{a}{c} + \frac{1}{2}\left[\frac{e^{-i4\pi fa/c} - 1}{-i2\pi f}\right] \\ &= -\frac{a}{c} + \frac{1}{2}e^{-i2\pi fa/c}\left[\frac{e^{i2\pi fa/c} - e^{-i2\pi fa/c}}{i2\pi f}\right] \\ &= -\frac{a}{c} + \frac{a}{c}e^{-i2\pi fa/c}\left[\frac{\sin 2\pi fa/c}{2\pi fa/c}\right]. \end{aligned} \tag{5.32}$$

If we define $\hat{a}$ as

$$\hat{a} \stackrel{\text{def}}{=} \frac{a}{c}, \tag{5.33}$$

then we can rewrite Eq. (5.32) as

$$H(f) = -\hat{a} + \hat{a}e^{-i2\pi f\hat{a}}\left[\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right]. \tag{5.34}$$

The magnitude-squared spectrum $|H(f)|^2$ is thus

$$|H(f)|^2 = \hat{a}^2\left[1 + \left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)^2 - 2\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)\cos 2\pi f\hat{a}\right]. \tag{5.35}$$

A plot of $|H(f)|^2$ appears in Fig. 5.5.

The next step in our procedure is to 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

$$\begin{aligned} \frac{|H(f)|^2}{N_0/2} &= \frac{2\hat{a}^2}{N_0}\left[1 + \left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)^2 - 2\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)\cos 2\pi f\hat{a}\right] \\ &= \frac{2\hat{a}^2}{N_0} + \frac{2\hat{a}^2}{N_0}\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)^2 - \frac{4\hat{a}^2}{N_0}\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)\cos 2\pi f\hat{a}. \end{aligned} \tag{5.36}$$

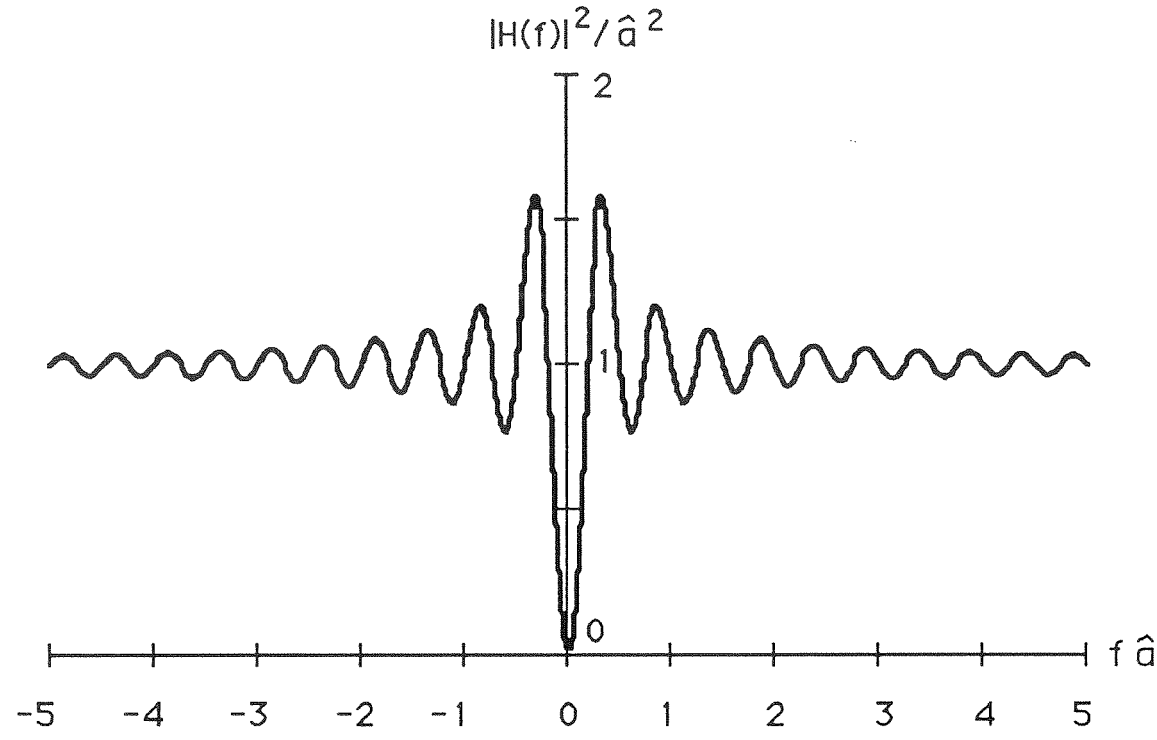


Figure 5.5. Magnitude-Squared Spectrum $|H(f)|^2$ versus f.

Calculating the inverse Fourier transform of Eq. (5.36), we now obtain the $L(t)$ as specified in step 2 of our design procedure. To simplify notation, define the two functions

$$V_\alpha(t) = \begin{cases} 1 - |t|/\alpha, & \text{for } |t| \leq \alpha; \\ 0, & \text{elsewhere.} \end{cases} \tag{5.37}$$

and

$$B_\alpha(t) = \begin{cases} 1, & \text{for } |t| \leq \alpha; \\ 0, & \text{elsewhere.} \end{cases} \tag{5.38}$$

Then, since

$$\int_{-\infty}^{\infty} B_{\hat{a}}(t)e^{-i2\pi ft}\,dt = 2\hat{a}\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}} \tag{5.39}$$

and

$$\int_{-\infty}^{\infty} V_{2\hat{a}}(t)e^{-i2\pi ft}\,dt = 2\hat{a}\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)^2, \tag{5.40}$$

we can write

$$\begin{aligned} L(t) &= \int_{\infty}^{\infty} \frac{|H(f)|^2}{N_0/2} e^{i2\pi ft}\,dt \\ &= \int_{\infty}^{\infty} \left[\frac{2\hat{a}^2}{N_0} + \frac{2\hat{a}^2}{N_0}\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)^2 - \frac{4\hat{a}^2}{N_0}\left(\frac{\sin 2\pi f\hat{a}}{2\pi f\hat{a}}\right)\cos 2\pi f\hat{a}\right] e^{i2\pi ft}\,dt \\ &= \frac{2\hat{a}^2}{N_0}\delta(t) + \frac{\hat{a}}{N_0}V_{2\hat{a}}(t) - \frac{\hat{a}}{N_0}B_{\hat{a}}(t) * [\delta(t-\hat{a}) + \delta(t+\hat{a})] \\ &= \frac{2\hat{a}^2}{N_0}\delta(t) + \frac{\hat{a}}{N_0}V_{2\hat{a}}(t) - \frac{\hat{a}}{N_0}B_{2\hat{a}}(t). \end{aligned} \tag{5.41}$$

We will now use this $L(t)$ in our design procedure.

The next step in our design procedure is to solve the integral equation

$$\lambda_{\max}\hat{x}(t) = \int_{-T/2}^{T/2} \hat{x}(\tau)L(t-\tau)\,d\tau \tag{5.42}$$

for an eigenfunction $\hat{x}(t)$ corresponding to the maximum eigenvalue $\lambda_{\max}$. For our particular $L(t)$ as given in Eq. (5.41), this becomes

$$\begin{aligned} \lambda_{\max}\hat{x}(t) &= \int_{-T/2}^{T/2} \left[\frac{2\hat{a}^2}{N_0}\delta(t) + \frac{\hat{a}}{N_0}V_{2\hat{a}}(t) - \frac{\hat{a}}{N_0}B_{2\hat{a}}(t)\right]\hat{x}(t)\,d\tau \\ &= \frac{2\hat{a}^2}{N_0}\int_{-T/2}^{T/2} \delta(t-\tau)\hat{x}(\tau)\,d\tau + \frac{\hat{a}}{N_0}\int_{-T/2}^{T/2} [V_{2\hat{a}}(t-\tau) - B_{2\hat{a}}(t-\tau)]\,\hat{x}(\tau)\,d\tau. \end{aligned} \tag{5.43}$$

From Eq. (5.43), we see that if $\hat{x}(t)$ is an eigenfunction corresponding to the maximum eigenvalue λ_{max}, it must also be an eigenvector of the integral equation

$$\mu_{max}\hat{x}(t) = \frac{\hat{a}}{N_0}\int_{-T/2}^{T/2} [V_{2\hat{a}}(t-\tau) - B_{2\hat{a}}(t-\tau)]\,\hat{x}(\tau)\,d\tau \tag{5.44}$$

corresponding to the maximum eigenvalue μ_{max}. Note that

$$\lambda_{max} = \frac{2\hat{a}^2}{N_0} + \mu_{max}. \tag{5.45}$$

For convenience in our analysis, we will assume that the waveform $\hat{x}(t)$ that we are designing has unit energy. This corresponds to the case where $\mathcal{E} = 1$ in our design procedure. In addition, for computational convenience, we will assume that $\hat{a} = 1$. Although such a value of $\hat{a}$ does not correspond to values typically encountered in practice (such a sphere would have a radius of 3×10^8 m), these results can be applied to more typical values by normalizing both the *amplitude* of the received signal and the time axis linearly in the length unit.

Solving the integral equation of Eq. (5.44) numerically for $T = 1$, 25, 50, and 100, we obtain the eigenvalues μ_{max} of Eq.(5.44) and thus the eigenvalues λ_{max} of Eq. (5.43) as given in Table 5-2. The eigenfunctions corresponding to these maximum eigenvalues have wave shapes as given in Figs. 5.6–5.9.

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 and having 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.

Table 5-2

Eigenvalues μ_{max} and λ_{max} for Various T.		
T	$\mu_{max} \cdot N_0$	$\lambda_{max} \cdot N_0$
1	0.1737	2.1737
25	1.9108	3.9108
50	1.8682	3.8682
100	1.7477	3.7477

Such a waveform can be expressed as

$$x(t) = \beta B_{T/2}(t) \cos 2\pi f_0 t. \tag{5.46}$$

Here, for fixed T, β is a normalizing constant such that the waveform has unit energy. For $T = n/f_0$, $\beta = \sqrt{2/T}$, and for all $T \gg 1/f_0$, $\beta \approx \sqrt{2/T}$. For our analysis, however, the value of β need not be explicitly known, as it can be conveniently calculated numerically in the process of evaluating the resulting signal-to-noise ratio.

In order to obtain the most favorable result when transmitting a waveform as specified in Eq. (5.46), we must select the carrier frequency in order to take advantage of the resonance of the spherical scatterer as expressed by $|H(f)|^2$. Looking at the plot of $|H(f)|^2/\hat{a}$ as shown in Fig. 5.5, we see that $|H(f)|^2$ has its peak value at a frequency between $0.25/\hat{a}$ and $0.5/\hat{a}$. Solving Eq. (5.35) numerically in order to obtain the frequency at which this maximum occurs, we find that $|H(f)|^2$ takes on a maximum value of 1.5862 at a frequency of $f = 0.3251/\hat{a}$.

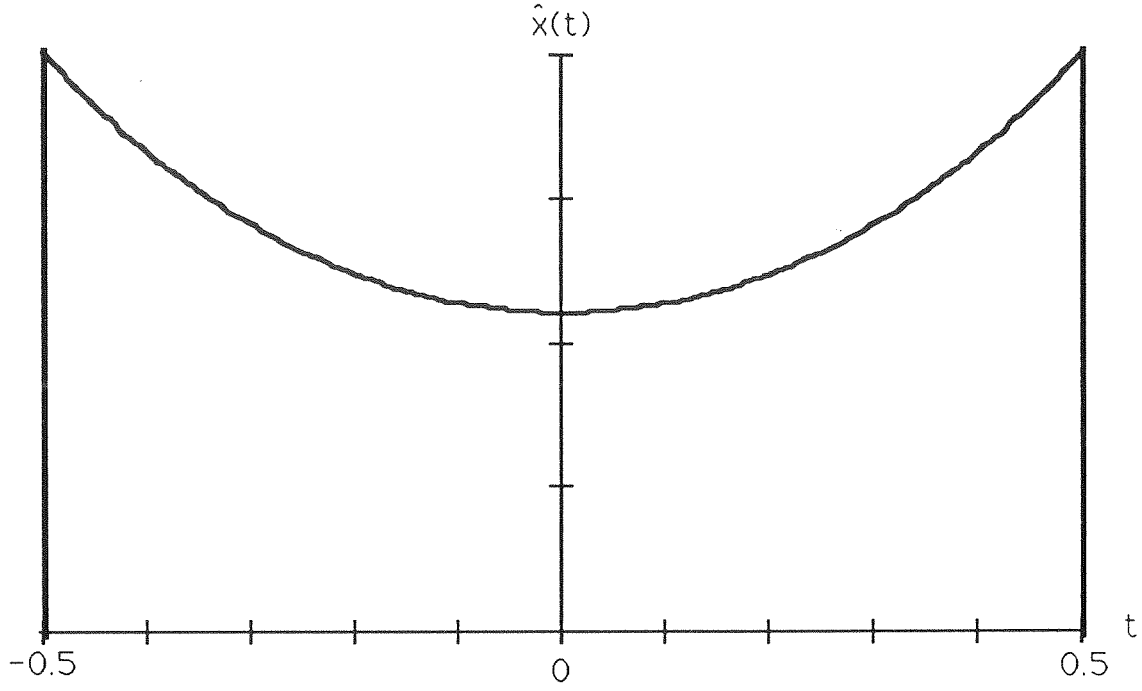


Figure 5.6. Wave shape of $\hat{x}(t)$ for $T = 1$.

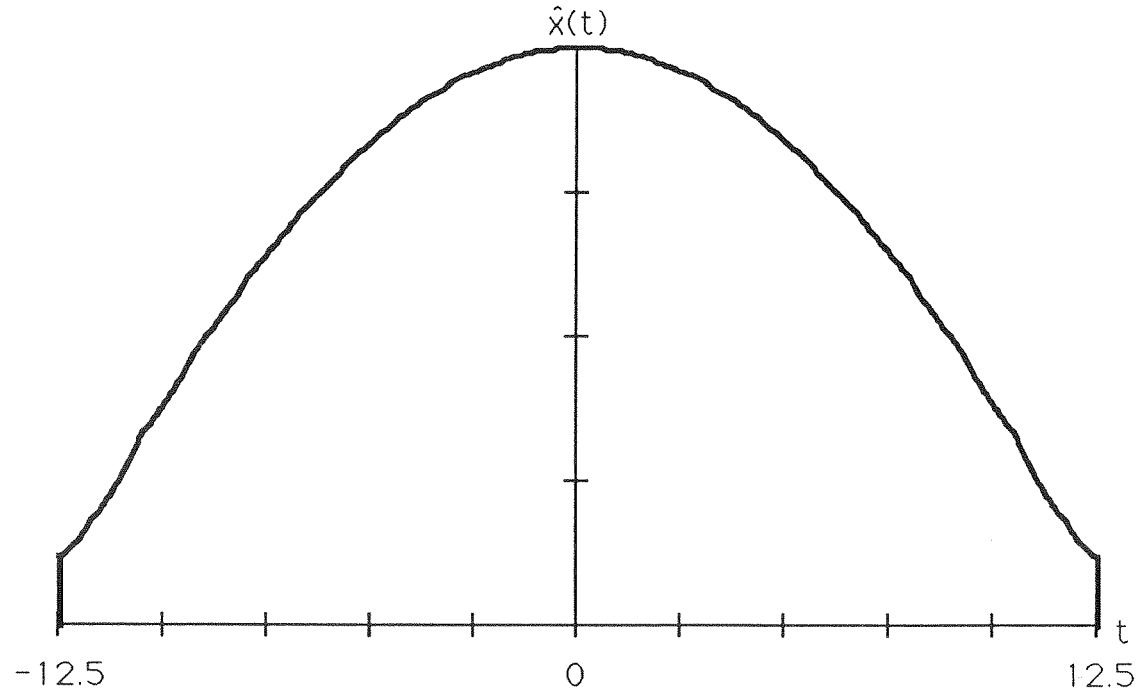


Figure 5.7. Wave shape of $\hat{x}(t)$ for $T = 25$.

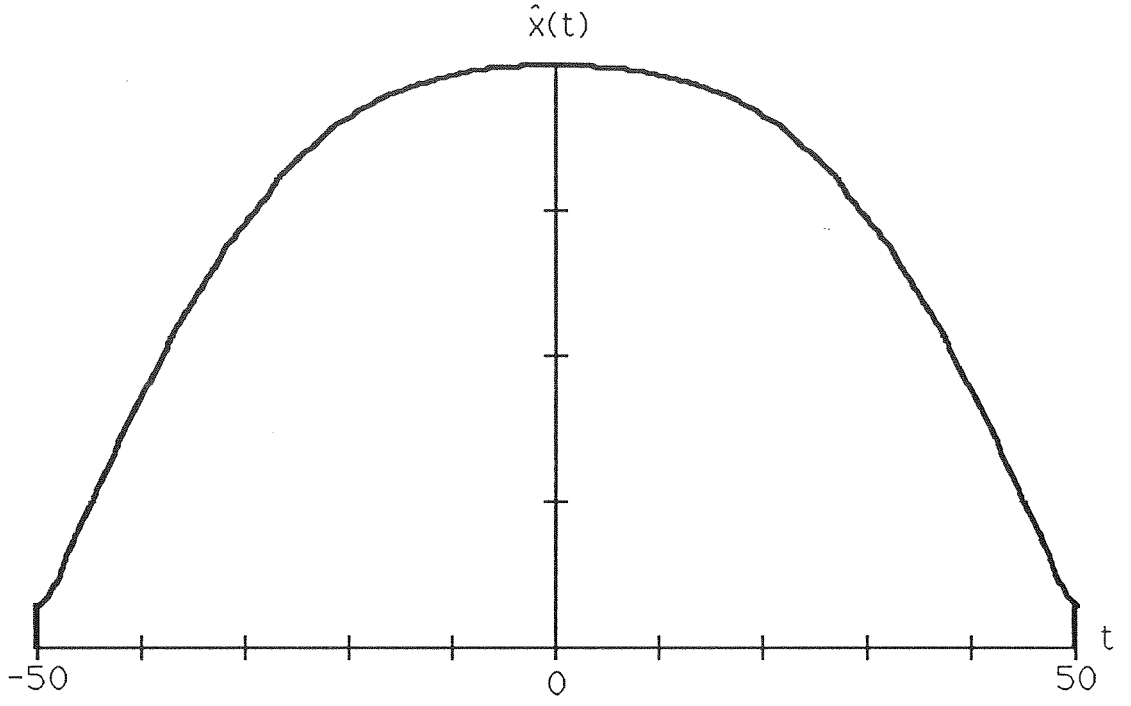


Figure 5.8. Wave shape of $\hat{x}(t)$ for $T = 50$.

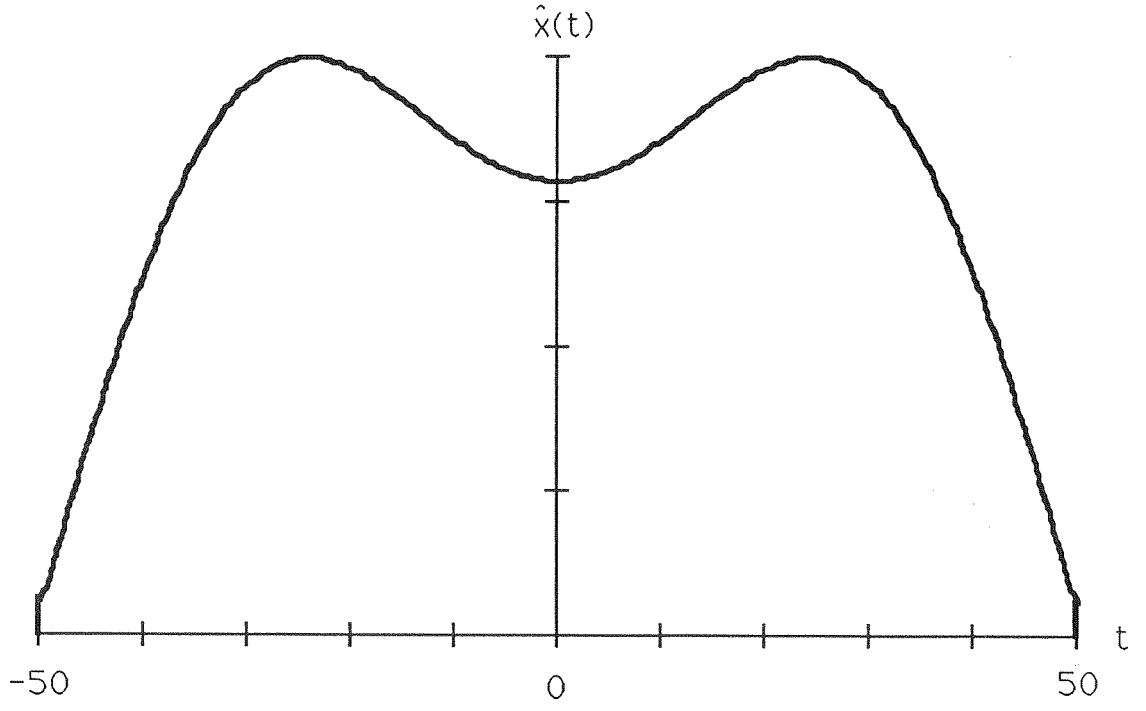


Figure 5.9. Wave shape of $\hat{x}(t)$ for $T = 100$.

Let $X(f)$ be the Fourier transform of the transmitted waveform $x(t)$ as given in Eq. (5.46). 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

$$R(f) = \frac{k\overline{X(f)}e^{-i2\pi f t_0}}{S_{nn}(f)}, \tag{5.47}$$

where k is any non-zero constant. From Fig. 5.2, we see that the signal-to-noise ratio at time t_0 is given by

$$\begin{aligned}\left(\frac{S}{N}\right)_{t_0} &= \frac{|y_s(t_0)|^2}{\mathrm{E}|y_n(t_0)|^2} \\ &= \frac{\left|\int_{-\infty}^{\infty} X(f)H(f)R(f)e^{i2\pi f t_0}\,df\right|^2}{\int_{-\infty}^{\infty} |R(f)|^2 S_{nn}(f)\,df}.\end{aligned} \tag{5.48}$$

For white noise with power spectral density $S_{nn}(f) = N_0/2$ and $R(f)$ as given in Eq. (5.47), noting that $H(f)$ is a conjugate-symmetric function of f, this simplifies to

$$\left(\frac{S}{N}\right)_{t_0} = \left(\frac{4}{N_0}\right) \frac{\left|\int_0^{\infty} |X(f)|^2 \mathrm{Re}\{H(f)\}\,df\right|^2}{\int_0^{\infty} |X(f)|^2\,df}. \tag{5.49}$$

From Eq. (5.34), we have that when $\hat{a} = 1$ as in our example,

$$\mathrm{Re}\{H(f)\} = \left(\frac{\sin 2\pi f}{2\pi f}\right)\cos 2\pi f - 1. \tag{5.50}$$

For $x(t)$ as given in Eq. (5.46), $X(f)$ is given by

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

Table 5-3

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

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

$$|X(f)|^2 = \frac{\beta^2 T^2}{4}\left[\left(\frac{\sin\pi(f-f_0)T}{\pi(f-f_0)T}\right)^2 + \left(\frac{\sin\pi(f+f_0)T}{\pi(f+f_0)T}\right)^2 + 2\left(\frac{\sin\pi(f-f_0)T}{\pi(f-f_0)T}\right)\left(\frac{\sin\pi(f+f_0)T}{\pi(f+f_0)T}\right)\right]. \tag{5.51}$$

Using Eqs. (5.49), (5.50), and (5.51), we can solve for the signal-to-noise ratio that results when $x(t)$ is the unit energy pulsed sinusoid given in Eq. (5.46). This is done for $T = 1$, 25, 50, and 100. The values of β that provides a unit-energy waveform for each of these T are calculated numerically. Table 5-3 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 optimum 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 optimum waveform/receiver-filter pair over the pulsed sinusoid and its associated matched filter.

As can be seen from Table 5-3, there is a significant improvement in the resulting signal-to-noise ratio when the optimum waveform/receiver-filter pair is used over that which occurs when a more typical *ad hoc* procedure is used. For the range of T examined, the optimum 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 to 17 percent. This is a significant increase in detection performance. 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 outlined in this chapter. Such an improvement could be very helpful in marginal radar detection scenarios.

5.4. Chapter 5 References.

5.1 Van Trees, H. L., *Detection, Estimation, and Modulation Theory*, Part 1, John Wiley and Sons, New York, NY, 1968.

5.2 Helstrom, C. W., *Statistical Theory of Signal Detection*, Pergamon Press, New York, NY, 1968.

5.3 North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Report PTR-6C; reprinted in *Proc. IEEE*, **51** (July 1963), pp. 1016-27.

5.4 Bracewell, R. N., *The Fourier Transform and Its Applications*, McGraw-Hill, New York, NY, 1986.

5.5 Papoulis, A., *The Fourier Integral and its Applications*, McGraw-Hill, New York, NY, 1962.

5.6 Kennaugh, E. M., and D. L. Moffatt, "Transient and Impulse Response Approximations," *Proc. IEEE*, **51**, No. 8 (August 1965), pp. 893–901.

CHAPTER 6

INFORMATION THEORY AND RADAR WAVEFORM DESIGN

In this chapter, we will examine the design of radar waveforms that maximize the mutual information between the target and the received radar signal in the radar measurement process. This is done using the target impulse response defined in Section 4.6 to model the interaction of the transmitted radar waveform with the target. However, now we will consider a target impulse response that is a random process, in order to model our uncertainty of the target's scattering characteristics.

Using the model of a random target impulse response that is a finite energy Gaussian random process, we will solve for a family of radar waveforms that maximize the mutual information between the random, target impulse response and the received radar signal. This will be done assuming that the signal is received in additive Gaussian noise of arbitrary power spectral density $P_{nn}(f)$. Next, we will examine the characteristics of these waveforms through the use of several examples. We will then address some of the considerations involved in implementing these waveforms in real radar systems. Finally, we will compare these waveforms to those derived in Chapter 5 for the optimum detection of radar targets with known impulse response.

6.1. Radar Waveform Design for Maximum Information Transfer Between the Target and the Received Radar Signal.

Consider the radar target channel model shown in Fig. 6.1. Here, $x(t)$, a finite-energy deterministic waveform with energy $\mathcal{E}$ 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,

$$\mathcal{E} = \int_{-T/2}^{T/2} |x(t)|^2 \, dt. \tag{6.1}$$

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 have $\mathcal{E} = P_x T$, where P_x is the average power of $x(t)$. Since $\mathcal{E} = P_x T$, we note that for a given average power P_x, the longer the duration T of $x(t)$, the greater the total energy in the transmitted waveform. We will also assume that the waveform $x(t)$ is confined to a frequency interval $\mathcal{W} = [f_0, f_0 + W]$. While strictly speaking, we cannot have a waveform $x(t)$ confined to both a finite time interval and its Fourier transform confined to a finite frequency interval, we will assume that $\mathcal{W}$ is selected such that only a negligible fraction of the total energy resides outside the frequency interval $\mathcal{W}$. In real radar systems, the frequency band of operation corresponding to $\mathcal{W}$ is ultimately determined by the bandwidth of the components making up the radar system.

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

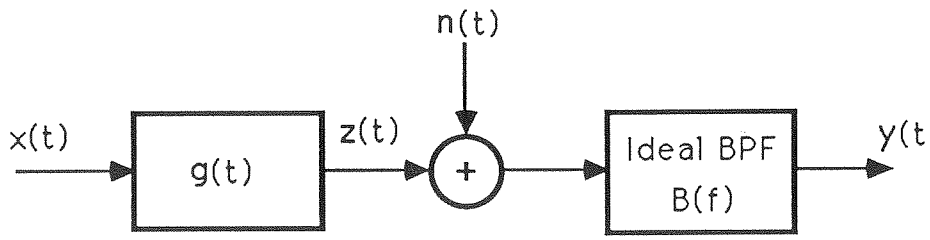


Figure 6.1. Block Diagram of The Radar Target Channel.

process, and is given by the convolution integral

$$\mathbf{z}(t) = \int_{-\infty}^{\infty} \mathbf{g}(\tau) x(t-\tau)\, d\tau. \tag{6.2}$$

The signal $\mathbf{z}(t)$ is received at the receiver in the presence of the zero-mean additive Gaussian noise process $\mathbf{n}(t)$. This noise process is assumed to be stationary and ergodic, and to have one-sided power spectral density (PSD) $P_{nn}(f)$. In addition, $\mathbf{n}(t)$ is assumed to be statistically independent of both the transmitted waveform $x(t)$ and the target impulse response $\mathbf{g}(t)$. We note that this assumption may not be valid in all radar remote sensing problems, as thermal noise emission from a surface in the microwave region is a function of both the surface geometry and the surface's material composition. Since the random impulse response $\mathbf{g}(t)$ is also a function of surface geometry and material composition, it does not follow that $\mathbf{g}(t)$ and $\mathbf{n}(t)$ are necessarily statistically independent. In many cases, however, the target impulse response $\mathbf{g}(t)$ and the additive noise $\mathbf{n}(t)$ are statistically independent. Examples of such cases include targets such as aircraft or ships providing

the scattering mechanism and additive noise coming from thermal noise radiated by the environment and by the radar receiver itself. In any case, when sensing surfaces with radar, independence of the additive noise from the backscattered signal is commonly assumed [6.1, pp. 492-5]. In our analysis, we will therefore assume that $\mathbf{g}(t)$ and $\mathbf{n}(t)$ are statistically independent.

The waveform received at the receiver is shown in Fig. 6.1 to be $\mathbf{z}(t) + \mathbf{n}(t)$ filtered by the ideal bandpass filter $B(f)$, passing only frequencies in the band $\mathcal{W}$. This 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 $\mathbf{z}(t)$, since it is the response of a linear time-invariant system to the transmitted signal. Hence we will not consider frequencies outside this interval.

The random process $\mathbf{g}(t)$ can be thought of as an ensemble $\{g(t)\}$ of functions with a probability measure assigned to each sample function $g(t)$. (Actually, the sample functions $g(t)$ are more properly written as $g(t, \omega)$, where $\omega \in \Omega$ and Ω is the underlying probability space.) We will now examine some properties of the random impulse response $\mathbf{g}(t)$.

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

$$\int_{-\infty}^{\infty} |g(t, \omega)|^2 \, dt \leq 1. \tag{6.3}$$

This follows from conservation of energy and the fact that electromagnetic scattering is a passive process. The next property of $\mathbf{g}(t)$ we will assume is that all of its sample

functions are causal impulse responses; that is,

$$g(t,\omega) = 0,\ \forall t < 0,\ \forall \omega \in \Omega. \tag{6.4}$$

This is a property of all *physical* linear time-invariant systems.

We define the Fourier transform $\mathrm{G}(f)$ of $\mathbf{g}(t)$ as

$$\mathrm{G}(f) = \int_{-\infty}^{\infty} \mathbf{g}(t) e^{-i2\pi ft}\, dt. \tag{6.5}$$

Thus, for any sample point $\omega_0 \in \Omega$, there is a sample function $g(t,\omega_0)$ of $\mathbf{g}(t)$ and a sample function $G(f,\omega_0)$ of $\mathrm{G}(f)$, and these two sample functions are related by

$$G(f,\omega_0) = \int_{-\infty}^{\infty} g(t,\omega_0) e^{-i2\pi ft}\, dt. \tag{6.6}$$

We will also assume that $\mathbf{g}(t)$ is a Gaussian random process. As discussed in Section 4.4, this is a reasonable assumption for targets made up 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 Gaussian random processes. This model is particularly good for natural surfaces as well as for many complex, manmade targets. As will be seen in the next chapter, such a scattering process is also responsible for the "speckle" seen in synthetic aperture radar images.

Returning to Fig. 6.1, we see that for a given sample function $g(t,\omega_0)$ with Fourier transform $G(f,\omega_0)$, the resulting spectrum of the scattered signal $\mathbf{z}(t)$ is given by

$$Z(f,\omega_0) = X(f)G(f,\omega_0). \tag{6.7}$$

The magnitude squared of this spectrum is

$$|Z(f,\omega_0)|^2 = |X(f)|^2 \, |G(f,\omega_0)|^2 \, . \tag{6.8}$$

Taking the expectation with respect to $\mathrm{G}(f)$, the mean-square spectrum of $\mathbf{z}(t)$ is

$$\mathrm{E}\,|\mathrm{Z}(f)|^2 = |X(f)|^2 \, |\mathrm{G}(f)|^2 \, . \tag{6.9}$$

Now,

$$\mathrm{E}\left\{|\mathrm{G}(f)|^2\right\} = |\mu_G(f)|^2 + \sigma_G^2(f), \tag{6.10}$$

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

$$\mu_G(f) = \mathrm{E}\left\{\mathrm{G}(f)\right\}, \tag{6.11}$$

and

$$\sigma_G^2(f) = \mathrm{E}\left\{|\mathrm{G}(f) - \mu_G(f)|^2\right\}. \tag{6.12}$$

We are interested primarily in $\sigma_G^2(f)$ for the Gaussian target model, as the signal component of $\mathbf{z}(t)$ corresponding to the mean $\mu_G(f)$ is known since $\mathbf{x}(t)$ is known. It thus provides no information about the target (assuming $\mu_G(f)$ is known). In most cases, $\mu_G(f) = 0$, since there is a random delay $\mathbf{d}$ in $\mathbf{g}(t)$ because of the target's random position in space. This corresponds to a random phase factor of $\exp\{-i2\pi f\mathbf{d}\}$, which has expectation zero for a wide class of distributions on $\mathbf{d}$. We will thus assume that $\mu_G(f) = 0$ throughout the rest of this analysis. If this were not the case, however, the analysis could be easily carried out in an identical manner by replacing $\mathbf{z}(t)$ with $\tilde{\mathbf{z}}(t) = \mathbf{z}(t) - \mathrm{E}\{\mathbf{z}(t)\}$. This is the component of

$\mathbf{z}(t)$ that remains when the mean component $\mathrm{E}\{\mathbf{z}(t)\}$ corresponding to $\mu_G(f)$ is removed.

Similarly, if we define

$$\mu_Z(f) = \mathrm{E}\{\mathrm{Z}(f)\} \tag{6.13}$$

and

$$\sigma_Z^2(f) = \mathrm{E}\left\{|\mathrm{Z}(f) - \mu_Z(f)|^2\right\}, \tag{6.14}$$

then

$$\mathrm{E}\,|\mathrm{Z}(f)|^2 = |\mu_Z(f)|^2 + \sigma_Z^2(f). \tag{6.15}$$

Referring again to Fig. 6.1, we will assume that the radar receiver observes $\mathbf{y}(t)$ for a period $\widetilde{T}$ in order to obtain information about the target. The duration of observation $\widetilde{T}$ must be long enough to allow the receiver to capture all but a negligible portion of the energy in the scattered signal $\mathbf{z}(t)$. We know that the duration of the transmitted waveform is T, and we know that $\mathbf{z}(t)$ 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 $\mathbf{g}(t)$, then the duration of $\mathbf{z}(t)$ is $T + T_g$.

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

$$h_{\mathcal{W}}(t) = W\,\frac{\sin \pi W t}{\pi W t}\cos\,(f_0 + W/2)t. \tag{6.16}$$

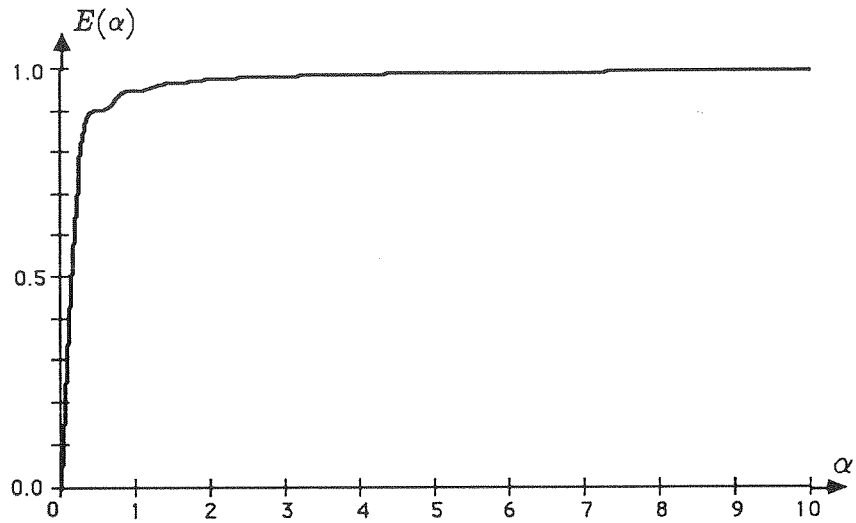


Figure 6.2. $E(\alpha)$ as a function of $\alpha = T/W$.

The duration of this pulse is infinite, but most of the energy is concentrated in an interval of duration $1/W$. Taking the duration to be $T_W(\alpha) = \alpha/W$, the fraction of the energy in the interval of duration $T_W(\alpha)$ centered about $t = 0$ is plotted as a function of α in Fig. 6.2. These values were calculated numerically. As this figure illustrates, it is reasonable to assume the impulse response duration T_W of the ideal bandpass filter at the receiver to be $T_W \approx 1/W$.

It is reasonable to assume that the bandwidth over which most real radar targets exhibit significant scattering of electromagnetic waves is much larger than the bandwidth of a radar system we are able to build in order to make the scattering

measurements [6.3, Sec. 27.6]. Hence it is reasonable to assume

$$T_g \ll T_W. \tag{6.17}$$

Also, for most radar signals, the duration of the transmitted signal is much larger than $1/W$. This is often necessary in radar detection problems in order to get enough energy into the signal 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; spectral analysis of the Doppler spectrum is one of the more common techniques in radar target recognition. For such signals, the actual duration T of transmission is much larger than the $1/W$. For such systems,

$$T \gg T_W \approx \frac{1}{W}. \tag{6.18}$$

So in summary, the observation interval $\widetilde{T}$ is

$$\begin{aligned} \widetilde{T} &= T + T_g + T_W \\ &\approx T + T_W \\ &\approx T + \frac{1}{W}, \end{aligned} \tag{6.19}$$

and for systems that satisfy the condition $T \gg 1/W$,

$$\widetilde{T} \approx T. \tag{6.20}$$

We are interested in finding the mutual information $I(\mathbf{g}(t); \mathbf{y}(t)|x(t))$, that is, the mutual information between the random target impulse response and the

received radar waveform. We note that $x(t)$ is a deterministic waveform. It is explicitly denoted in $I(\mathrm{g}(t);\mathrm{y}(t)|x(t))$ because the mutual information is a function of $x(t)$, and we are interested in finding those functions $x(t)$ that maximize $I(\mathrm{g}(t);\mathrm{y}(t)|x(t))$ under constraints on their energy and bandwidth. In order to find the functions $x(t)$ that maximize $I(\mathrm{g}(t);\mathrm{y}(t)|x(t))$, we will first find $I(\mathrm{z}(t);\mathrm{y}(t)|x(t))$ and those functions $x(t)$ that maximize it. We will then show that the functions $x(t)$ that maximize $I(\mathrm{z}(t);\mathrm{y}(t)|x(t))$ also maximize $I(\mathrm{g}(t);\mathrm{y}(t)|x(t))$, and that for these $x(t)$, $I(\mathrm{g}(t);\mathrm{y}(t)|x(t))$=$I(\mathrm{z}(t);\mathrm{y}(t)|x(t))$.

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

$$I(\hat{\mathrm{y}}_k(t);\hat{\mathrm{z}}_k(t)|x(t)) = \widetilde{T}\Delta f \ln\left[1 + \frac{2|X(f_k)|^2\sigma_G^2(f_k)}{P_{nn}(f_k)\widetilde{T}}\right] \tag{6.21}$$

To see why this is true, consider the continuous communication channel of Fig. 6.3.

Here we have an additive Gaussian noise channel with input Z, a zero-mean Gaussian random variable with variance σ_Z^2, and additive zero-mean Gaussian noise N with variance σ_N^2. Y and N are statistically independent. The mutual information

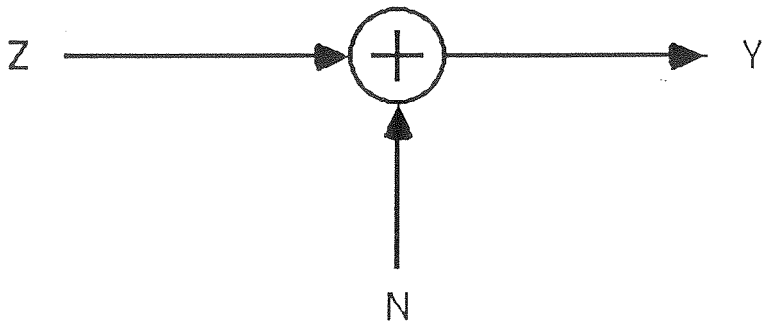


Figure 6.3. Additive Gaussian Noise Channel.

$I(Y;Z)$ between Y and Z is

$$I(Y;Z) = h(Y) - h(Y|Z). \tag{6.22}$$

Now Y, being the sum of two zero-mean Gaussian random variables, is itself a zero-mean Gaussian random variable. Since Z and N are statistically independent, the variance of Y, σ_Y^2, is

$$\sigma_Y^2 = \sigma_Z^2 + \sigma_N^2. \tag{6.23}$$

This being the case, the differential entropy $h(Y)$ of Y is

$$h(Y) = \frac{1}{2} \ln 2\pi\sigma_Y^2 = \frac{1}{2} \ln 2\pi \left(\sigma_Z^2 + \sigma_N^2\right) \quad \text{(nats)}. \tag{6.24}$$

The differential entropy $h(Y|Z)$ is equal to the differential entropy $h(N)$, since the conditional density

$$f(y|z) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\left[-\frac{(y-z)^2}{2\sigma_N^2}\right],$$

is a Gaussian density for a random variable with mean z and variance σ_N^2. But the differential entropy of a Gaussian random variable is a function only of its variance and not its mean; hence,

$$h(Y|Z) = h(N) = \frac{1}{2} \ln 2\pi\sigma_N^2 \quad \text{(nats)}. \tag{6.25}$$

So the mutual information $I(Y;Z)$ is given by the expression

$$\begin{aligned} I(Y;Z) &= h(Y) - h(Y|Z) \\ &= \frac{1}{2} \ln 2\pi \left(\sigma_Z^2 + \sigma_N^2\right) - \frac{1}{2} \ln 2\pi\sigma_N^2 \\ &= \frac{1}{2} \ln \left(\frac{\sigma_N^2 + \sigma_Z^2}{\sigma_N^2}\right) \\ &= \frac{1}{2} \ln \left(1 + \frac{\sigma_Z^2}{\sigma_N^2}\right) \quad \text{(nats)}. \end{aligned} \tag{6.26}$$

Referring again to the signals $\hat{\mathbf{z}}_k(t)$, $\hat{\mathbf{y}}_k(t)$, and $\hat{\mathbf{n}}_k(t)$ with frequency components confined to the interval $\mathcal{F}_k = [f_k, f_k + \Delta f]$, we have from the Sampling Theorem [6.2, p.194] that each of the signals can be represented by a sequence of samples taken at a uniform sampling rate of $2\Delta f$. Since we assume that the spectra $X(f)$, $\mathrm{Z}(f)$, and $\mathrm{Y}(f)$ are smooth and have a constant value (at least approximately) for all $f \in \mathcal{F}_k$, the samples of the Gaussian process sampled at a uniform rate $2\Delta f$ are statistically independent.

The samples $\hat{\mathbf{z}}_k(t)$ are independent, identically distributed random variables with zero mean and variance σ_Z^2. In order to calculate this variance σ_Z^2, we note that the total energy $\mathcal{E}_{\hat{z}}(\mathcal{F}_k)$ in $\hat{\mathbf{z}}_k(t)$ is

$$\mathcal{E}_z(\mathcal{F}_k) = 2\Delta f \, |X(f_k)|^2 \, \sigma_G^2(f_k). \tag{6.27}$$

Here, the factor of 2 is due to the fact that $X(f_k)$ is the two-sided spectrum of $x(t)$ and that we are carrying out our calculations using only positive frequencies. This energy is evenly spread among $2\Delta f\widetilde{T}$ statistically independent samples. Hence, the variance of each sample, σ_Z^2, is

$$\sigma_Z^2 = \frac{\mathcal{E}_z(\mathcal{F}_k)}{2\Delta f\widetilde{T}} = \frac{2\Delta f\,|X(f_k)|^2\,\sigma_G^2(f_k)}{2\Delta f\widetilde{T}} = \frac{|X(f_k)|^2\,\sigma_G^2(f_k)}{\widetilde{T}}. \tag{6.28}$$

Similarly, the noise process $\hat{\mathbf{n}}_k(t)$ has total energy $\mathcal{E}_{\hat{n}}(\mathcal{F}_k)$ on the interval $\mathcal{T}$ given by

$$\mathcal{E}_{\hat{n}}(\mathcal{F}_k) = \Delta f P_{nn}(f_k)\widetilde{T}, \tag{6.29}$$

and this energy is evenly distributed among the $2\Delta f\widetilde{T}$ statistically independent, zero-mean samples representing $\hat{\mathbf{n}}_k(t)$. Hence, the variance σ_N^2 of each sample is

$$\sigma_N^2 = \frac{\Delta f P_{nn}(f_k)\widetilde{T}}{2\Delta f\widetilde{T}} = \frac{P_{nn}(f_k)}{2}. \tag{6.30}$$

Substituting the results of Eqs. (6.28) and (6.30) into Eq. (6.26), we have that for each sample Z_m of $\hat{\mathbf{z}}_k(t)$ and the corresponding sample Y_m of $\hat{\mathbf{y}}_k(t)$, the mutual information between Z_m and Y_m is

$$\begin{aligned} I(Y_m;Z_m) &= \frac{1}{2}\ln\left[1+\frac{|X(f_k)|^2\sigma_G^2(f_k)/\widetilde{T}}{P_{nn}(f_k)/2}\right] \quad \text{(nats)} \\ &= \frac{1}{2}\ln\left[1+\frac{2|X(f_k)|^2\sigma_G^2(f_k)}{P_{nn}(f_k)\widetilde{T}}\right] \quad \text{(nats)}. \end{aligned} \tag{6.31}$$

Now there are $2\Delta f\widetilde{T}$ statistically independent sample values for both $\hat{\mathbf{z}}_k(t)$ and $\hat{\mathbf{y}}_k(t)$ in the observation interval $\mathcal{T}$. Thus,

$$\begin{aligned} I(\hat{\mathbf{y}}_k(t);\hat{\mathbf{z}}_k(t)|x(t)) &= 2\Delta f\widetilde{T}I(Y_m;Z_m) \\ &= \widetilde{T}\Delta f\ln\left[1+\frac{2|X(f_k)|^2\sigma_G^2(f_k)}{P_{nn}(f_k)\widetilde{T}}\right] \quad \text{(nats)}. \end{aligned} \tag{6.32}$$

This is the relation stated in Eq. (6.21) that we wished to show.

Now, 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 PSD's, and such processes are known to be statistically independent [6.4, 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:

$$I\left([\hat{\mathbf{y}}_j(t), \hat{\mathbf{y}}_k(t)]; [\hat{\mathbf{z}}_j(t), \hat{\mathbf{z}}_k(t)] | x(t)\right) = I(\hat{\mathbf{y}}_j(t); \hat{\mathbf{z}}_j(t) | x(t)) + I(\hat{\mathbf{y}}_k(t); \hat{\mathbf{z}}_k(t) | x(t)). \tag{6.33}$$

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 Δf, and then let the number of intervals increase as $\Delta f \to 0$, 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

$$I(\mathbf{y}(t); \mathbf{z}(t) | x(t)) = \widetilde{T} \int_{\mathcal{W}} \ln \left[1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f) \widetilde{T}} \right] df. \tag{6.34}$$

We wish to maximize $I(\mathbf{y}(t); \mathbf{z}(t) | x(t))$ with respect to $|X(f)|^2$ under the aver-

age power constraint,

$$\int_{\mathcal{W}} |X(f)|^2 \, df = \mathcal{E}. \tag{6.35}$$

Using the Lagrange multiplier technique [3.4, p.357-9], we form the objective function

$$\Phi\left(|X(\cdot)|^2\right) = \widetilde{T} \int_{\mathcal{W}} \ln\left[1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f)\widetilde{T}}\right] df - \lambda\left[\int_{\mathcal{W}} |X(f)|^2 \, df - \mathcal{E}\right]. \tag{6.36}$$

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

$$\phi\left(|X(f)|^2\right) = \widetilde{T} \ln\left[1 + \frac{2|X(f)|^2 \sigma_G^2(f)}{P_{nn}(f)\widetilde{T}}\right] - \lambda|X(f)|^2, \tag{6.37}$$

and λ is the Lagrange multiplier, to be determined from the constraint of Eq. (6.35). This equivalence can be seen by noting that

$$\Phi\left(|X(\cdot)|^2\right) = \int_{\mathcal{W}} \phi\left(|X(f)|^2\right) \, df.$$

Thus, when

$$\frac{\partial \phi\left(|X(f)|^2\right)}{\partial |X(f)|^2} = 0, \quad \forall f \in \mathcal{W},$$

it is also true that

$$\frac{\partial \Phi\left(|X(f)|^2\right)}{\partial |X(f)|^2} = 0, \quad \forall f \in \mathcal{W}.$$

So we obtain an $|X(f)|^2$ that maximizes Eq. (6.36) when we solve for an $|X(f)|^2$ that maximizes Eq. (6.37). Thus, the $|X(f)|^2$ that maximizes $\Phi\left(|X(f)|^2\right)$ satisfies the relation

$$\frac{\partial \phi\left(|X(f)|^2\right)}{\partial |X(f)|^2} = 0, \quad \forall f \in \mathcal{W}. \tag{6.38}$$

Taking the partial derivatives, we have

$$\begin{aligned}\frac{\partial\phi\left(|X(f)|^2\right)}{\partial|X(f)|^2} &= \frac{2\sigma_G^2(f)}{P_{nn}(f)}\left[1+\frac{2|X(f)|^2\sigma_G^2(f)}{P_{nn}(f)\widetilde{T}}\right]^{-1}-\lambda \\ &= \left[\frac{P_{nn}(f)}{2\sigma_G^2(f)}+\frac{|X(f)|^2}{\widetilde{T}}\right]^{-1}-\lambda.\end{aligned} \tag{6.39}$$

This must equal zero. Solving for $|X(f)|^2$ then yields

$$|X(f)|^2 = A - \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}. \tag{6.40}$$

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

Substituting the expression for $|X(f)|^2$ of Eq. (6.40) into the constraint of Eq. (6.35), we obtain

$$\begin{aligned}\int_W |X(f)|^2\,df &= \int_W \left[A-\frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right]df \\ &= WA - \int_W \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\,df \\ &= \mathcal{E}.\end{aligned} \tag{6.41}$$

Solving for the constant A, we have

$$A = \frac{1}{W}\left[\mathcal{E}+\int_W \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\,df\right]. \tag{6.42}$$

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

$$|X(f)|^2 = \frac{1}{W}\left[\mathcal{E}+\int_W \frac{P_{nn}(\tilde{f})\widetilde{T}}{2\sigma_G^2(\tilde{f})}\,d\tilde{f}\right] - \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}. \tag{6.43}$$

If we define $r(f)$ as

$$r(f) \stackrel{\text{def}}{=} \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}, \tag{6.44}$$

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

$$|X(f)|^2 = A - r(f). \tag{6.45}$$

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

$$\begin{aligned} I_{\max}(\mathbf{y}(t);\mathbf{z}(t)|x(t)) &= \tilde{T}\int_{\mathcal{W}} \ln\left[1+\frac{2|X(f)|^2\sigma_G^2(f)}{P_{nn}(f)\tilde{T}}\right] df \\ &= \tilde{T}\int_{\mathcal{W}} \ln\left[1+\frac{A-r(f)}{r(f)}\right] df \\ &= \tilde{T}\int_{\mathcal{W}} \ln\left[\frac{A}{r(f)}\right] df \\ &= \tilde{T}\int_{\mathcal{W}} [\ln A - \ln r(f)]\, df \\ &= \tilde{T}W\ln A - \tilde{T}\int_{\mathcal{W}} \ln r(f)\, df \quad \text{(nats)}. \end{aligned} \tag{6.46}$$

Note that since $|X(f)|^2$ is the magnitude squared of the transmitted signal spectrum, it must be real and non-negative for all $f \in \mathcal{W}$ (we assume it to be zero for all $f \notin \mathcal{W}$). Yet from Eq. (6.40),

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

It may be possible that for a given $\mathcal{E}$, $\sigma_G^2(f)$, 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

$$\mathcal{E} = \int_{\mathcal{W}} \max\left[0, A - \frac{P_{nn}(f)\tilde{T}}{2\sigma_G^2(f)}\right] df. \tag{6.47}$$

In most cases, the solution of this equation will have to be done numerically. However, for any (positive) value of $\mathcal{E}$, A can be bounded as follows:

$$\max_{f\in\mathcal{W}}\left\{\frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right\} \le A \le \frac{\mathcal{E}}{W} - \min_{f\in\mathcal{W}}\left\{\frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right\}. \tag{6.48}$$

Once A has been solved for using Eq. (6.47), we have

$$\begin{aligned} I_{\max}(\mathbf{y}(t);\mathbf{z}(t)|x(t)) &= \widetilde{T}\int_{\mathcal{W}} \max\left[0, \ln A - \ln\left(\frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right)\right] df \\ &= \widetilde{T}\int_{\mathcal{W}} \max\left[0, \ln A - \ln r(f)\right] df. \end{aligned} \tag{6.49}$$

The associated magnitude-squared spectrum $|X(f)|^2$ that achieves it is

$$\begin{aligned} |X(f)|^2 &= \max\left[0, A - \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right] \\ &= \max\left[0, A - r(f)\right]. \end{aligned} \tag{6.50}$$

We note as a check that the solution for $|X(f)|^2$ obtained by solving Eq. (6.47) for A and substituting this A into Eq. (6.50) to obtain the $|X(f)|^2$ that achieves the maximum value of $I_{\max}(\mathbf{y}(t);\mathbf{z}(t)|x(t))$ as given in Eq. (6.49), is equivalent to that of Eqs. (6.41) and (6.46), when $A \ge P_{nn}(f)\widetilde{T}/2\sigma_G^2(f)$ for all $f \in \mathcal{W}$.

We have shown that the mutual information $I(\mathbf{y}(t);\mathbf{z}(t)|x(t))$ is maximized by an $x(t)$ with $|X(f)|^2$ as given in Eq. (6.49). In reality, however, we are interested in the mutual information $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$, the mutual information between the target and the received signal $\mathbf{y}(t)$. We will now show that when $|X(f)|^2$ satisfies Eq. (6.49), $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ is maximized, and the resulting maximum value $I_{\max}(\mathbf{y}(t);\mathbf{z}(t)|x(t))$ is the same as $I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))$. In order to show this, we will make use of the following two theorems:

Theorem 1: *Let* $\mathrm{a}(t)$ *and* $\mathrm{b}(t)$ *be finite-energy random processes and let* D *be a reversible transformation of* $\mathrm{a}(t)$ *to a finite-energy random process* $\mathrm{c}(r)$ *(where* r *is a new independent variable, but* r *could equal* t*). Then*

$$I(\mathrm{a}(t);\mathrm{b}(t)) = I(\mathrm{c}(r);\mathrm{b}(t)). \tag{6.54}$$

Theorem 2: *Let* $\mathrm{a}(t)$ *and* $\mathrm{b}(t)$ *be finite-energy random processes with Fourier transforms* $\mathrm{A}(f)$ *and* $\mathrm{B}(f)$*, respectively. Then if* $I(\mathrm{a}(t);\mathrm{b}(t))$ *is the mutual information between* $\mathrm{a}(t)$ *and* $\mathrm{b}(t)$ *and* $I(\mathrm{A}(f);\mathrm{B}(f))$ *is the mutual information between* $\mathrm{A}(f)$ *and* $\mathrm{B}(f)$*, we have*

$$I(\mathrm{a}(t);\mathrm{b}(t)) = I(\mathrm{A}(f);\mathrm{B}(f)). \tag{6.55}$$

To see that **Theorem 1** is true, we note that if D is a reversible transformation between $\mathrm{a}(t)$ and $\mathrm{c}(r)$, that is $\mathrm{c}(r) = D\{\mathrm{a}(t)\}$, then [6.5, p. 90-1]

$$h(\mathrm{c}(r)) = h(\mathrm{a}(t)) + \mathcal{K}(D), \tag{6.56}$$

and

$$h(\mathrm{c}(r)|\mathrm{b}(t)) = h(\mathrm{a}(t)|\mathrm{b}(t)) + \mathcal{K}(D). \tag{6.57}$$

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

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

To see that **Theorem 2** is true, we choose as our reversible transform D the Fourier transform, and apply it to both $\mathbf{a}(t)$ and $\mathbf{b}(t)$, yielding

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

and

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

Applying Theorem 1 first to the transformation $D\{\mathbf{a}(t)\} = \mathrm{A}(f)$, we get

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

Applying Theorem 1 again, this time to the transformation $D\{\mathbf{b}(t)\} = \mathrm{B}(f)$, we have

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

So it follows that

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

In order to show that $I_{\max}(\mathbf{y}(t);\mathbf{z}(t)|x(t)) = I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))$, we define the two-sided set of frequencies

$$\hat{\mathcal{W}}_2 \stackrel{\text{def}}{=} \left\{f : |f| \in \mathcal{W},\, |X(f)|^2 \neq 0\right\}. \tag{6.58}$$

Note that $\hat{\mathcal{W}}_2$ is a *two-sided* set of frequencies, including both positive and negative frequencies, whereas $\mathcal{W}$ is a *one-sided* set of frequencies, containing only positive frequencies. We also note that if $X(f)$ has frequencies limited to $\hat{\mathcal{W}}_2$, then so does $\mathrm{Z}(f)$, since

$$\mathrm{Z}(f) = X(f)\mathrm{G}(f). \tag{6.59}$$

So for $f \in \hat{\mathcal{W}}_2$, we can determine $\mathrm{G}(f)$ from $\mathrm{Z}(f)$, since

$$\mathrm{G}(f) = \frac{\mathrm{Z}(f)}{X(f)}, \quad \forall f \in \hat{\mathcal{W}}_2. \tag{6.60}$$

For $f \notin \hat{\mathcal{W}}_2$, we cannot determine $\mathrm{G}(f)$ from $\mathrm{Z}(f)$, since $X(f) \neq 0$ and thus $1/X(f)$ is indeterminate. However, physically we would not expect to determine anything about $\mathrm{G}(f)$ for $f \notin \hat{\mathcal{W}}_2$. This is because the sample functions of $\mathrm{G}(f)$ describe linear, time-invariant systems. From such systems we would not expect an output response at frequencies that have no input excitation [6.2, p. 185-7].

We now define

$$\hat{\mathrm{G}}(f) = \begin{cases} \mathrm{G}(f) & \text{, for } f \in \hat{\mathcal{W}}_2; \\ 0 & \text{, elsewhere;} \end{cases} \tag{6.61}$$

$$\hat{\mathrm{Z}}(f) = \begin{cases} \mathrm{Z}(f) & \text{, for } f \in \hat{\mathcal{W}}_2; \\ 0 & \text{, elsewhere;} \end{cases} \tag{6.62}$$

$$\hat{\mathrm{g}}(t) = \int_{-\infty}^{\infty} \hat{\mathrm{G}}(f) e^{i2\pi f t}\, dt; \tag{6.63}$$

$$\hat{\mathrm{z}}(t) = \int_{-\infty}^{\infty} \hat{\mathrm{Z}}(f) e^{i2\pi f t}\, dt. \tag{6.64}$$

Then, from **Theorem 2** we have

$$I_{\max}(\mathrm{y}(t); \mathrm{z}(t) | x(t)) = I_{\max}(\mathrm{Y}(f); \mathrm{Z}(f) | x(t)). \tag{6.65}$$

Note that for $f \notin \hat{\mathcal{W}}_2$, $\mathrm{Z}(f) = 0$, since $\mathrm{Z}(f) = X(f)\mathrm{G}(f)$ and $X(f) = 0$ for $f \notin \hat{\mathcal{W}}_2$. Thus, $\hat{\mathrm{Z}}(f) = \mathrm{Z}(f)$ from the definition of $\hat{\mathrm{Z}}(f)$ in Eq. (6.55). So from **Theorem 1**,

$$I_{\max}(\mathrm{y}(t); \mathrm{z}(t) | x(t)) = I_{\max}(\mathrm{Y}(f); \hat{\mathrm{Z}}(f) | x(t)). \tag{6.66}$$

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

$$\hat{\mathrm{G}}(f) = \frac{\mathrm{Z}(f)}{X(f)}, \quad \forall f \in \hat{\mathcal{W}}_2. \tag{6.67}$$

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

$$D\left[\hat{\mathrm{Z}}(f)\right] = \begin{cases} \dfrac{\hat{\mathrm{Z}}(f)}{X(f)} & \text{, for } f \in \hat{\mathcal{W}}_2; \\ 0 & \text{, elsewhere.} \end{cases} \tag{6.68}$$

But as we know from **Theorem 1**, mutual information is invariant under reversible transformations, so

$$I(\mathrm{Y}(f); \hat{\mathrm{Z}}(f)|x(t)) = I(\mathrm{Y}(f); \hat{\mathrm{G}}(f)|x(t)). \tag{6.69}$$

Eqs. (6.66) and (6.69) yield

$$I(\mathbf{y}(t); \mathbf{z}(t)|x(t)) = I(\mathrm{Y}(f); \hat{\mathrm{G}}(f)|x(t)). \tag{6.70}$$

Note that $\hat{\mathrm{G}}(f)$, defined by Eq. (6.61), is equal to $\mathrm{G}(f)$ for $f \in \hat{\mathcal{W}}_2$ and is zero elsewhere. Thus, $\mathcal{I}_0$ nats of information about $\hat{\mathrm{G}}(f)$ is also $\mathcal{I}_0$ nats of information about $\mathrm{G}(f)$. So it follows that any information obtained about $\hat{\mathrm{G}}(f)$ by observation of $\mathrm{Y}(f)$ is an equivalent amount of information about $\hat{\mathrm{G}}(f)$, and so immediately we see that

$$I(\mathrm{Y}(f); \mathrm{G}(f)|x(t)) \geq I(\mathrm{Y}(f); \hat{\mathrm{G}}(f)|x(t)). \tag{6.71}$$

But for $f \notin \hat{\mathcal{W}}_2$, $\mathrm{Y}(f)$ provides no additional information about $\mathrm{G}(f)$. To see that this is true, we note that for $f \notin \hat{\mathcal{W}}_2$, $\mathrm{Z}(f) = 0$. Thus,

$$\mathrm{Y}(f) = N_{\widetilde{T}}(f) = \int_{t_0}^{t_0+\widetilde{T}} \mathbf{n}(t) e^{-2\pi f t}\, dt, \quad \text{for } f \in \hat{\mathcal{W}}_2. \tag{6.72}$$

Thus, for $f \notin \hat{\mathcal{W}}_2$, $\mathrm{Y}(f)$ 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 Eq. (6.71) is actually an equality:

$$I(\mathrm{Y}(f); \mathrm{G}(f)|x(t)) = I(\mathrm{Y}(f); \hat{\mathrm{G}}(f)|x(t)). \tag{6.73}$$

From Theorem 2 we have

$$I(\mathrm{Y}(f); \mathrm{G}(f)|x(t)) = I(\mathrm{y}(t); \mathrm{g}(t)|x(t)). \tag{6.74}$$

Thus, from Eqs. (6.70), (6.73), and (6.74), we have

$$I(\mathrm{y}(t); \mathrm{g}(t)|x(t)) = I_{\max}(\mathrm{y}(t); \mathrm{z}(t)|x(t)). \tag{6.75}$$

We have now shown that for the class of functions $x(t)$ that maximize $I(\mathrm{y}(t); \mathrm{z}(t)|x(t))$ (that is, those $x(t)$ with $t \in [-T/2, T/2]$ and with $|X(f)|^2$ given by Eq. (6.40)), we have $I(\mathrm{y}(t); \mathrm{g}(t)|x(t)) = I(\mathrm{y}(t); \mathrm{z}(t)|x(t))$. But we have not yet shown that some other transmitted waveform $\tilde{x}(t)$ confined to the time interval $[-T/2, T/2]$ satisfying the energy constraint of Eq. (6.1) with magnitude-squared spectrum $\left|\tilde{X}(f)\right|^2$ does not result in a larger mutual information between $\mathrm{g}(t)$ and $\mathrm{y}(t)$. In other words, we must show that there is not another waveform $\tilde{x}(t)$ satisfying the constraints on the transmitted waveform for which $I(\mathrm{y}(t); \mathrm{g}(t)|\tilde{x}(t)) > I(\mathrm{y}(t); \mathrm{g}(t)|x(t))$.

In order to show that no such $\tilde{x}(t)$ exists, we redraw the target channel model of Fig. 6.1 as shown in Fig. 6.4. Here we view both $\mathrm{g}(t)$ and $x(t)$ as inputs. The target impulse response $\mathrm{g}(t)$ is observed by illuminating the target, resulting in the

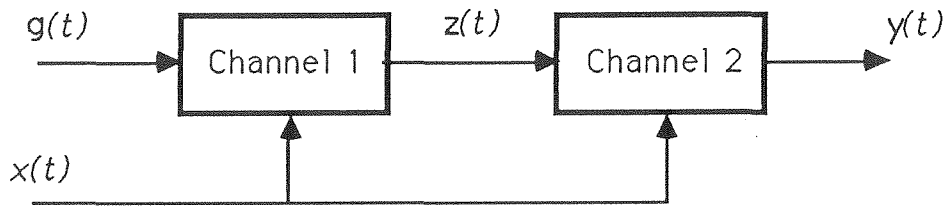


Figure 6.4. Another Interpretation the of Radar Target Channel.

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 information theory [6.6, p. 31], we have that, for any $x(t)$ transmitted,

$$I(\mathbf{y}(t);\mathbf{g}(t)|x(t)) \leq I(\mathbf{y}(t);\mathbf{z}(t)|x(t)). \tag{6.76}$$

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 Eq. (6.76), 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 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 Eq. (6.40),

$I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ is maximized, and the maximum value $I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ is

$$I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = I_{\max}(\mathbf{y}(t);\mathbf{z}(t)|x(t)). \tag{6.77}$$

In summary then, if $x(t)$ is a finite-energy waveform with energy $\mathcal{E}$ confined to the symmetric time interval $[-T/2, T/2]$, and with all but a negligible fraction of its energy confined to the frequency interval $\mathcal{W} = [f_0, f_0 + W]$, the mutual information $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ between $\mathbf{y}(t)$ and $\mathbf{g}(t)$ is maximized by an $x(t)$ with a magnitude-squared spectrum

$$\begin{aligned} |X(f)|^2 &= \max\left[0, A - \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right] \\ &= \max\left[0, A - r(f)\right], \end{aligned} \tag{6.78}$$

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

$$\mathcal{E} = \int_{\mathcal{W}} \max\left[0, A - \frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right] df. \tag{6.79}$$

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

$$\begin{aligned} I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)) &= \widetilde{T}\int_{\mathcal{W}} \max\left[0, \ln A - \ln\left(\frac{P_{nn}(f)\widetilde{T}}{2\sigma_G^2(f)}\right)\right] df \\ &= \widetilde{T}\int_{\mathcal{W}} \max\left[0, \ln A - \ln r(f)\right] df. \end{aligned} \tag{6.80}$$

Note the behavior of the magnitude-square spectrum

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

which maximizes $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$. If the variance $\sigma_G^2(f)$ of the frequency spectrum $G(f)$ of the scattering function $\mathbf{g}(t)$ 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 $P_{nn}(f) \geq 2A\sigma_G^2(f)/\widetilde{T}$. 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 $\sigma_G^2(f) \gg P_{nn}(f)\widetilde{T}/2A$ and $|X(f)|^2 = 0$ for $\sigma_G^2(f) \leq P_{nn}(f)\widetilde{T}/2A$. In order to interpret this behavior physically, recall that $\sigma_G^2(f)$ is the variance of the frequency spectrum $\mathrm{G}(f)$. 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 $A = A(\mathcal{E}, \sigma_G^2(f), P_{nn}(f))$ is a function of the transmitted energy $\mathcal{E}$, of the size and shape of the spectral variance $\sigma_G^2(f)$ of the target, and of the noise power spectral density $P_{nn}(f)$. The fact that $|X(f)|^2 = 0$ for all f such that $\sigma_G^2(f) \leq P_{nn}(f)\widetilde{T}/2A$ 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.

An interesting interpretation of the relationship between $|X(f)|^2$, A, $P_{nn}(f)$, and $\sigma_G^2(f)$ is shown in Fig. 6.4. Comparing Eq. (6.79) and Fig. 6.4, we see that the total energy $\mathcal{E}$ corresponds to the shaded area in Fig. 6.5a. 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. 6.5b.

The interpretation of Fig. 6.5 is called the "water-filling" interpretation and arises in many problems dealing with the spectral distribution of power and energy in information theory [6.14]. Think of $r(f) = P_{nn}(f)\widetilde{T}/2\sigma_G^2(f)$ as being the shape of the bottom of a container and think of $\mathcal{E}$ as the volume of "water" poured into the container (actually, we are looking at a two-dimensional profile of the container). We assume that the water can flow such that the surface height is the same everywhere in the container (all regions of the container are connected). Then the "water" will distribute itself so as to give a $|X(f)|^2$ that maximizes the mutual information $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$. We note also that the set of values on the frequency axis that have values corresponding to a shaded region make up the set $\hat{\mathcal{W}}_2$, the set of frequencies that have non-zero $|X(f)|^2$.

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 hypothetical example of Fig. 6.6. In Fig. 6.6a we have the spectral variance $\sigma_G^2(f)$. In Fig. 6.6b we have the power spectral density $P_{nn}(f)$. In Fig. 6.6c we have $r(f) = P_{nn}(f)\widetilde{T}/2\sigma_G^2(f)$, a function of both the power spectral density $P_{nn}(f)$ of the noise and of the spectral variance $\sigma_G^2(f)$. Finally, in Fig. 6.6d 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

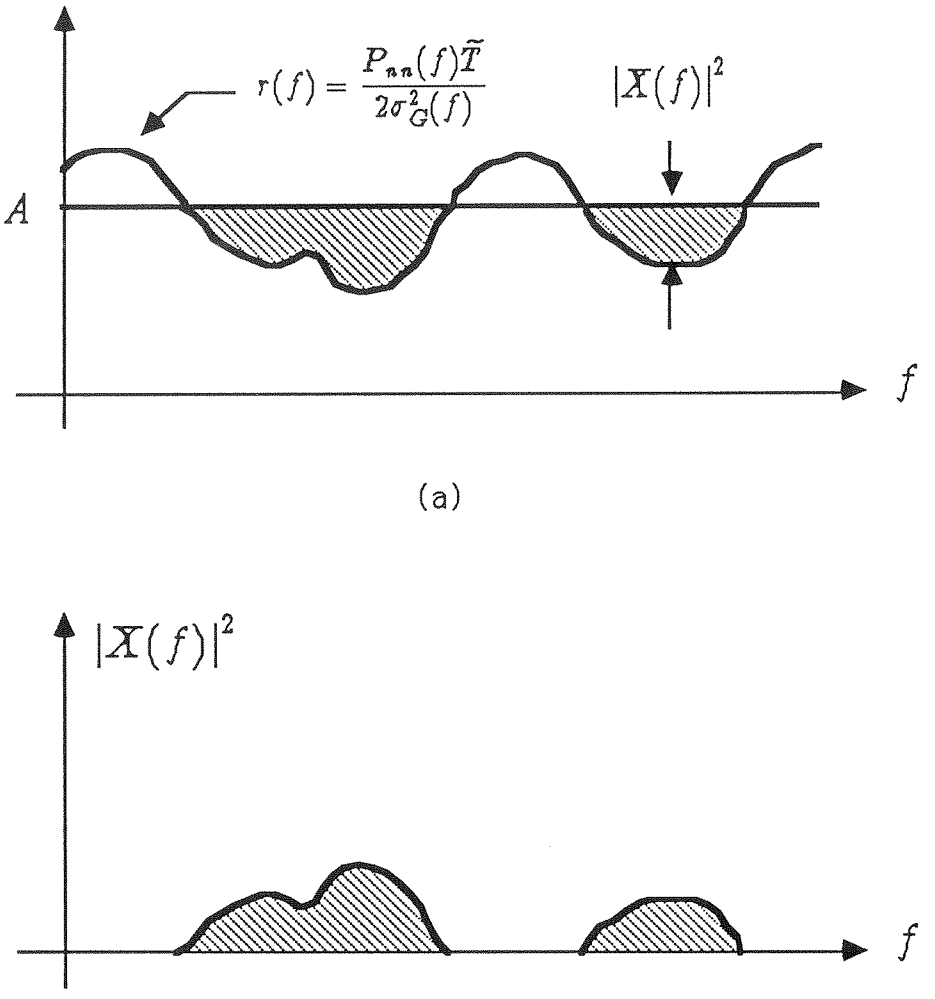


Figure 6.5. (a) "Water-Filling" Interpretation of the Magnitude-Squared Spectrum $|X(f)|^2$ that maximizes the Mutual Information $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$. (b) Magnitude-Squared Spectrum $|X(f)|^2$ that maximizes $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$. Note the Relationship to the Shaded Area in (a).

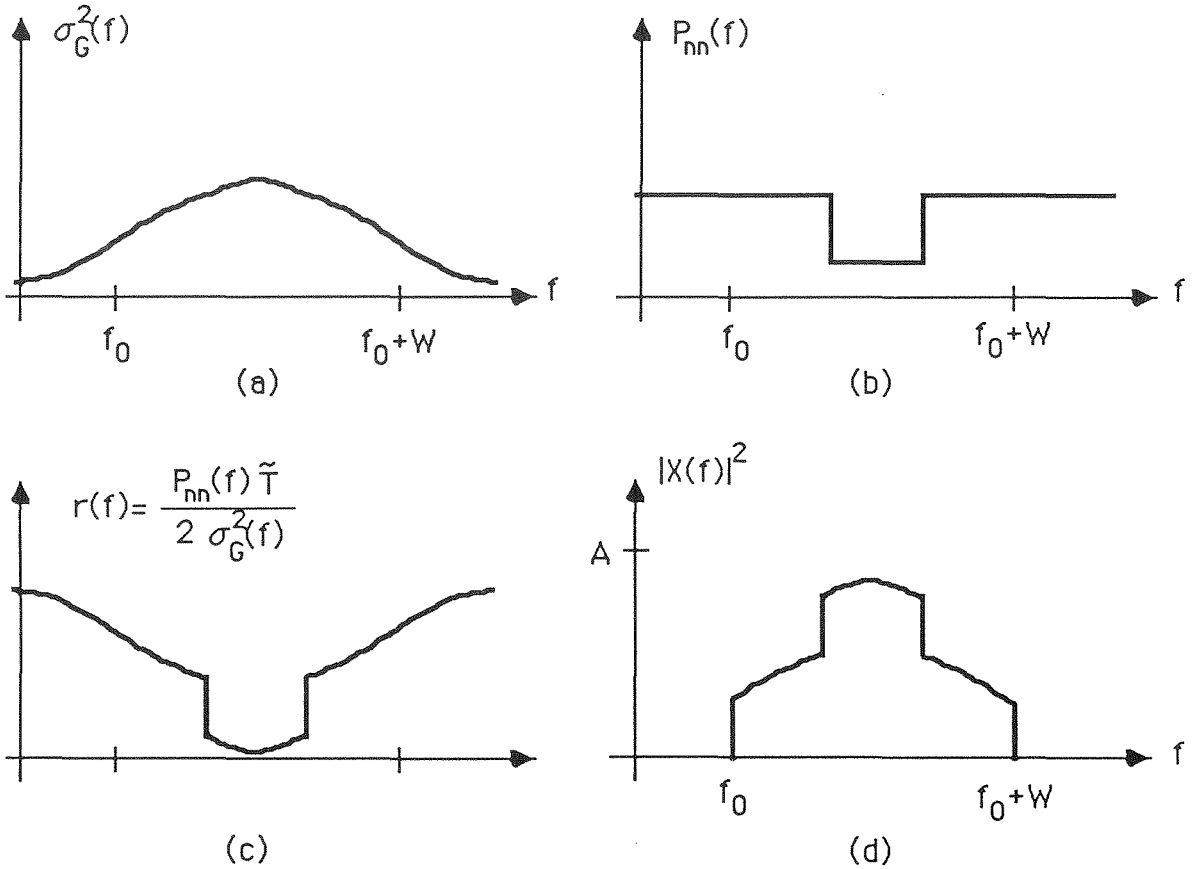


Figure 6.6. Example illustrating the resulting $|X(f)|^2$ for a given $\sigma_G^2(f)$ and $P_{nn}(f)$. (a) Example $\sigma_G^2(f)$. (b) Example $P_{nn}(f)$. (c) Resulting $r(f) = P_{nn}(f)\tilde{T}/2\sigma_G^2(f)$. (d) Resulting $|X(f)|^2$.

more clearly.

A point to remember in this analysis is that we have assumed that the random impulse response $\mathbf{g}(t)$ has been assumed to be 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 we noted in Chapter 2 in the case of the additive Gaussian noise channel, for a channel input with a given variance σ^2, the mutual information between the channel input and the channel output is maximized when the input is Gaussian. Then in 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. Then in solving for the maximum mutual information $I_{\max}(\mathbf{y}(t); \mathbf{g}(t)|x(t))$, we have derived an upper bound in 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.

6.2. A Numerical Example.

We will now examine how the results of the previous section can be applied to a realistic radar problem. In the process, we will examine the characteristics of the optimal transmitted signal's spectrum and the amount of information obtained.

For our example, 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 $A_e = 3\,\mathrm{m}^2$, 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

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

In the following analysis, we will analyze this radar system for a range of average power constraints ranging from 1 W to 1000 W and a range of observation times from 10 μs to 100 ms. However, for the purpose of the narrative, we will look at the case where the observation time $\tilde{T} \approx T = 10\,\mathrm{ms}$, a fairly typical value for radar target recognition systems, and for a few specific values of average power P_x. Numerically calculated results for the full ranges noted above will be presented after going through these sample calculations.

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

$$\sigma_G^2(f) = B \exp\left\{-\alpha(f - f_p)^2\right\}. \tag{6.81}$$

Here, B and α 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

$$\alpha = 10^{-13}\,\mathrm{s}^2, \tag{6.82}$$

a value illustrating well the effect of the transmitted waveform's spectral characteristics for the 10 MHz system bandwidth being considered. In order to determine a reasonable value of B for our example, we will assume that the spectral variance $\sigma_G^2(f)$ at frequency f_p corresponds to that of a $1\,\mathrm{m}^2$ variance in the radar cross section σ of the target at this frequency. From the *radar equation*, Eq. (4.1), and the fact that the gain G of an antenna of effective area A_e is [6.6, p. 11]

$$G = \frac{4\pi A_e}{\lambda^2},$$

it follows that the ratio of the received power to the transmitted power is

$$\frac{P_R}{P_T} = \frac{A_e^2 \sigma}{4\pi\lambda^2 R^4}. \tag{6.83}$$

Note that, in our example, we have

$$\lambda = \frac{c}{f_p} = \frac{2.998 \times 10^8\,\mathrm{m/s}}{1.000 \times 10^9\,\mathrm{Hz}} = 0.30\,\mathrm{m}.$$

This being the case, the ratio of the change in received power that is due to a $1\,\mathrm{m}^2$ change in radar cross section, ΔP_R, to the transmitted power, P_T, is

$$B = \frac{\Delta P_R}{P_T} = \frac{(3\,\mathrm{m}^2)^2(1\,\mathrm{m}^2)}{4\pi(0.3\,\mathrm{m})^2(10\,\mathrm{km})^4} = 7.9577 \times 10^{-16}. \tag{6.84}$$

We will assume that the additive Gaussian noise present at the radar receiver is thermal noise that is white over the frequency interval $\mathcal{W}$. We will assume a system noise tempertature of $T_s = 300\,K$. Hence, the resulting one-sided noise PSD is [6.6, p. 29]

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

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

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

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

and so in our case, with $\widetilde{T} = T = 10\,\mathrm{ms}$,

$$\begin{aligned} r(f) &= \frac{(4.1430 \times 10^{-21}\,\mathrm{J})(0.01\,\mathrm{s})}{2(7.9577 \times 10^{-16}) \exp\left[-\alpha(f - f_p)^2\right]} \\ &= (2.7835 \times 10^{-8}\,\mathrm{J\text{-}s}) \exp\left[\alpha(f - f_p)^2\right], \end{aligned} \tag{6.85}$$

where again, $\alpha = 10^{-13}\,\mathrm{s}^2$ and $f_p = 1\,\mathrm{GHz}$. A plot of this $r(f)$ is shown in Fig. 6.7.

From Eq. (6.42) we know that

$$\mathcal{E} = \int_{\mathcal{W}} \max\left[0, A - r(f)\right] df,$$

where

$$\mathcal{E} = P_x T = (1000\,\mathrm{W})(10\,\mathrm{ms}) = 10\,\mathrm{J}. \tag{6.86}$$

Solving for A numerically, we obtain

$$A = 1.0870 \times 10^{-6}\,\mathrm{J\text{-}s}. \tag{6.87}$$

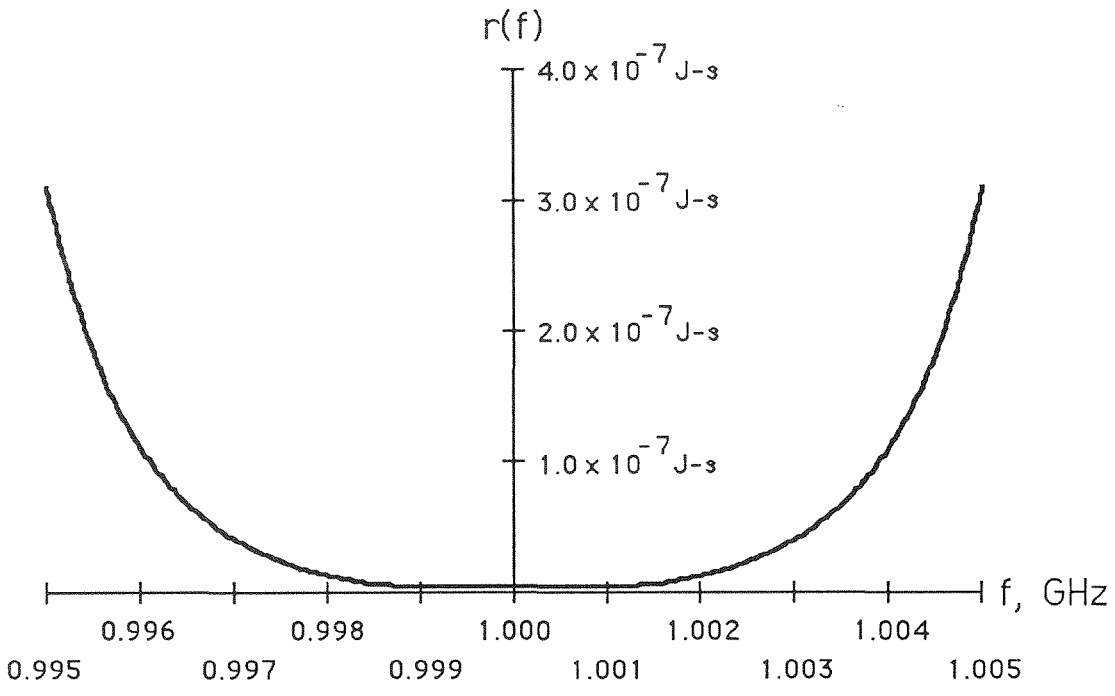


Figure 6.7. $r(f) = P_{nn}(f)\widetilde{T}/2\sigma_g^2(f)$ as a function of f.

Thus, we have that $|X(f)|^2 = \max[0, A - r(f)]$ is

$$|X(f)|^2 = \begin{cases} 1.0870 \times 10^{-6}\,\text{J-s} - (2.7835 \times 10^{-8}\,\text{J-s}) \exp[\alpha(f - f_p)^2] & \text{, for } f \in \mathcal{W}; \\ 0 & \text{, elsewhere.} \end{cases} \tag{6.88}$$

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 $f < 0$. A plot of $|X(f)|^2$ for positive f is shown in Fig. 6.8.

From Eq. (6.49), the mutual information $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$ is given by the inte-

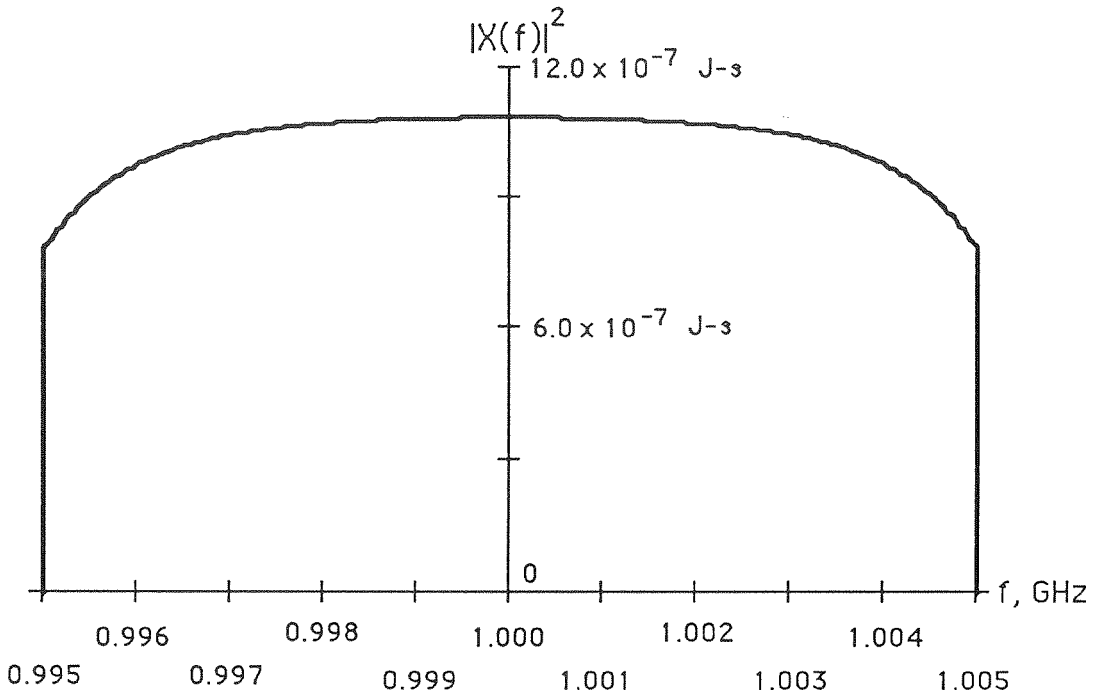


Figure 6.8. $|X(f)|^2$ for $T = 10\,\text{ms}$ and $P_x = 1000\,\text{W}$.

gral

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

Substituting A as given in Eq. 6.87 and $r(f)$ as given in Eq. 6.88 into this equation and integrating numerically over $\mathcal{W}$, we obtain the result

$$I(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = 2.3815 \times 10^5\,\text{nats}. \tag{6.89}$$

If we repeat the calculation, reducing the available average power P_x to 500 W,

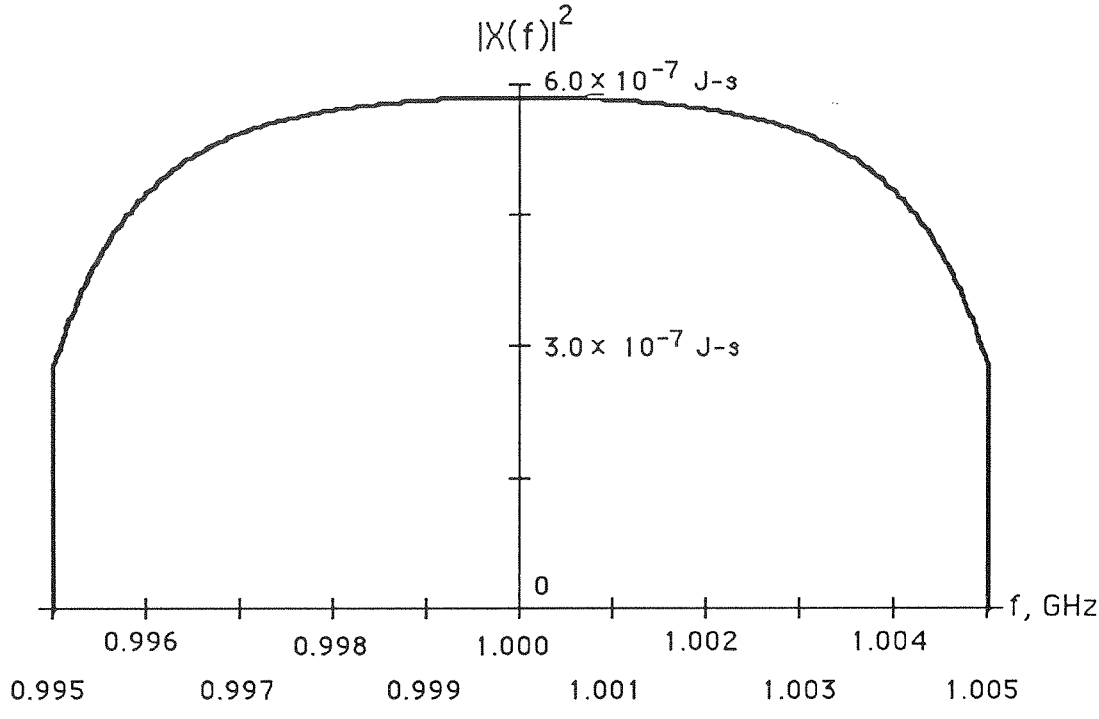


Figure 6.9. $|X(f)|^2$ for $T = 10\,\text{ms}$ and $P_x = 500\,\text{W}$.

maintaining $T = \widetilde{T} = 10\,\text{ms}$, we find

$$A = 5.8691 \times 10^{-7}\,\text{J-s}, \tag{6.90}$$

and the resulting $|X(f)|^2$ for positive frequencies is

$$|X(f)|^2 = \begin{cases} 5.8691 \times 10^{-7}\,\text{J-s} - (2.7835 \times 10^{-8}\,\text{J-s}) \exp\left[\alpha(f - f_p)^2\right] & \text{, for } f \in \mathcal{W}; \\ 0 & \text{, elsewhere.} \end{cases} \tag{6.91}$$

A plot of this $|X(f)|^2$ for positive frequencies is given in Fig. 6.9. The mutual

information $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ for this case, obtained by numerical integration, is

$$I(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = 2.2152 \times 10^5 \text{ nats}. \tag{6.92}$$

We now investigate a case in which for some $f \in \mathcal{W}$, $r(f) > A$. We assume that the average available average power P_x is 100 W, still maintaining $T = \widetilde{T} = 10\,\text{ms}$. Solving Eq. 6.47 numerically for A, where now

$$\mathcal{E} = P_x T = (100\,\text{W})(10\,\text{ms}) = 1\,\text{J}, \tag{6.93}$$

we have

$$A = 1.7709 \times 10^{-7}\,\text{J-s}, \tag{6.94}$$

and the resulting $|X(f)|^2$ for positive frequencies is

$$|X(f)|^2 = \begin{cases} 1.7709 \times 10^{-7}\,\text{J-s} - (2.7835 \times 10^{-8}\,\text{J-s}) \exp\left[\alpha(f - f_p)^2\right] & \text{, for } f \in \hat{\mathcal{W}}; \\ 0 & \text{, elsewhere.} \end{cases} \tag{6.95}$$

Here, $\hat{\mathcal{W}} \subset \mathcal{W}$ is given by

$$\hat{\mathcal{W}} = \{f \in \mathcal{W} : A \geq r(f)\} = [0.995698\,\text{GHz}, 1.004302\,\text{GHz}]. \tag{6.96}$$

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 Eq. 6.96 optimizes this tradeoff between the measured number of

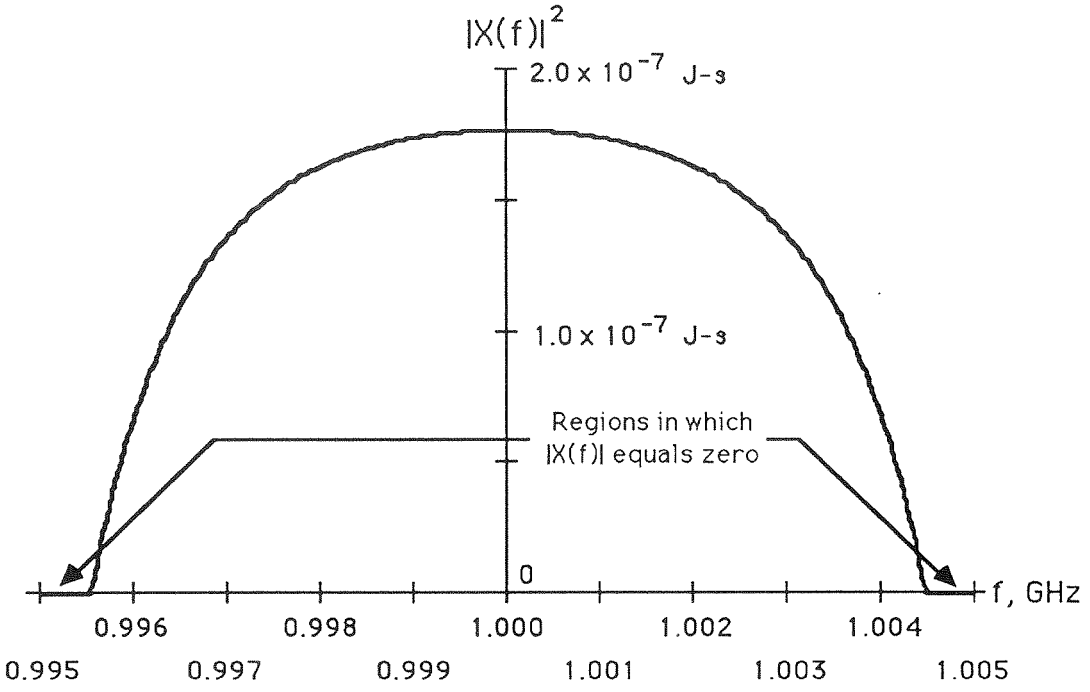


Figure 6.10. $|X(f)|^2$ for $T = 10\,\text{ms}$ and $P_x = 100\,\text{W}$.

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 $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$. A plot of $|X(f)|^2$ in this case is shown in Fig. 6.10. The resulting value of $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$ for this $|X(f)|^2$ is

$$I(\mathbf{y}(t); \mathbf{g}(t)|x(t)) = 2.2152 \times 10^5 \,\text{nats}. \tag{6.97}$$

We will now display the results of the numerical solution of Eqs. (6.47) and (6.49) for the mutual information $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$ as a function of both T and the

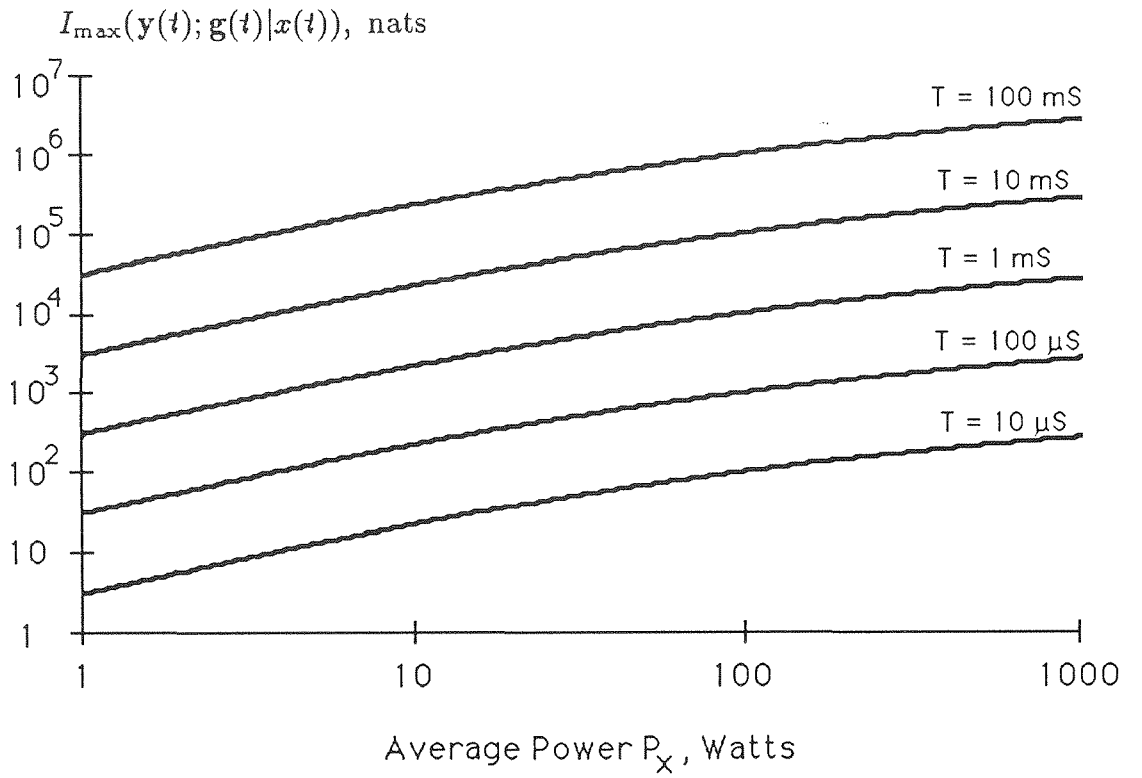


Figure 6.11. $I_{max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ as a function of T and P_x.

average available power P_x. This numerical solution was carried out for values of T equal to 10 μs, 100 μs, 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* [6.7, pp. 322-6]. The resulting maximum values of $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ are plotted in Fig. 6.11. Note that both scales in Fig. 6.11 are logarithmic.

Several points are worth noting in Fig. 6.11. 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 $\widetilde{T} = T$. This is true for all T considered in Fig. 6.11, since $1/W = 0.1\,\mu$s. This proportionality is reflected in Fig. 6.11 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 Eq. (6.80), we have

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

But for $T \gg 1/W$, which is true for all values of T considered in our example, it is reasonable to assume that $\widetilde{T} = T$. This gives

$$I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = T\int_{W} \max\left[0, \ln\left(\frac{2A\sigma_G^2(f)}{P_{nn}(f)T}\right)\right] df. \tag{6.98}$$

But as can be seen from Eq. (6.42), A is proportional to $\widetilde{T}$, and so assuming $\widetilde{T} = T$, A is proportional to T. If we define

$$a \stackrel{\text{def}}{=} \frac{A}{T}, \tag{6.99}$$

then we can write Eq. (6.98) as

$$I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = T\int_{W} \max\left[0, \ln\left(\frac{2a\sigma_G^2(f)}{P_{nn}(f)}\right)\right] df. \tag{6.100}$$

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

$$R_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = \int_{W} \max\left[0, \ln\left(\frac{2a\sigma_G^2(f)}{P_{nn}(f)}\right)\right] df, \tag{6.101}$$

which is not a function of T. We then have

$$I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)) = TR_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t)). \tag{6.102}$$

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. In Section 2.4, where we examined the relationship between mutual information and radar-measurement performance, we noted in Eq. (2.36) that, if $I(\mathbf{X};\mathbf{Y})$ is the mutual information between a set of parameters $\mathbf{X}$ to be measured and their measurement $\mathbf{Y}$, the maximum number of equiprobable classes N into which X can be assigned with statistical reliability by observation of $\mathbf{Y}$ is

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

($\lfloor\alpha\rfloor$ denotes the largest integer less than or equal to α). 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

$$\begin{aligned} N &= \left\lfloor e^{I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))} \right\rfloor \\ &= \left\lfloor e^{TR_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))} \right\rfloor . \end{aligned} \tag{6.103}$$

Note that 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 Eq. (6.103), 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. Assume that the frequency interval over which this is done has bandwidth W. Then, using classical methods of spectral analysis [6.8], the frequency resolution Δf of the measured spectrum is inversely proportional to T. Thus, the number of frequency bins of bandwidth Δf that span the interval of bandwidth W is proportional to T. Call this number of frequency bins M, as is illustrated in Fig. 6.12.

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

$$\tilde{N} = Q^M . \tag{6.104}$$

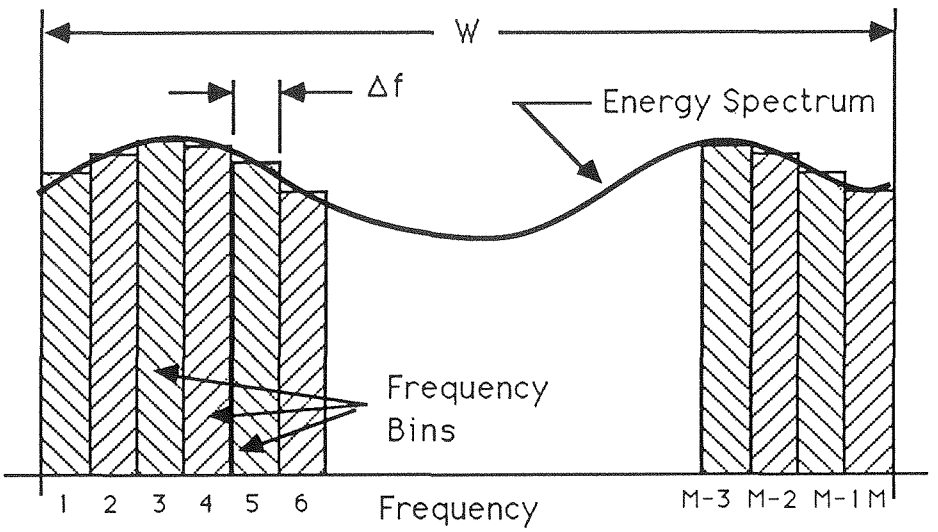


Figure 6.12. Spectral Analysis of Bandwidth W Using M Bins of Bandwidth Δf.

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

$$M(T) = mT, \tag{6.105}$$

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 Eqs. (6.104) and (6.105), we have

$$\tilde{N} = Q^{M(T)} = Q^{mT}. \tag{6.106}$$

Hence, the number of discernible frequency spectra also increases exponentially with T. This is not to say that there is any equivalence between the discernible frequency spectra and the equiprobable classes of $\mathbf{g}(t)$ to which Eq. (6.103) 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 Eq. (6.103) does make intuitive sense.

It is important to note that when a waveform $x(t)$ that achieves $I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ is transmitted, the N equiprobable classes referred to in Eq. (6.103) are not under the control of the radar but are a function of the target. In actuality, what Eq. (6.103) states is that the probability space Ω can be partitioned into N subsets $\Omega_1, \Omega_2, \ldots, \Omega_N$, where

$$\Pr\{\omega \in \Omega_k\} = \frac{1}{N}, \quad \text{for } k = 1, \ldots, N, \tag{6.107}$$

where N is given by Eq. (6.103). These N subsets $\Omega_1, \Omega_2, \ldots, \Omega_N$, which form a partition of Ω, correspond to a set of N classes into which $\mathbf{g}(t)$ can be reliably classified by observation of $\mathbf{g}(t)$. 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 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_V$ based on observation of $\mathbf{y}(t)$.

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 $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_V\}$ based on observation of $\mathbf{y}(t)$ can be viewed indirectly as finding a mapping from $\mathcal{C} : \Omega \rightarrow \mathcal{A}$, such that for $k = 1, \ldots, N$, each Ω_k is mapped to one of the $\mathcal{A}_j$ with a reasonable probability of error in determining the proper target class. In general, as N becomes larger, we would expect the performance of the best mapping $\mathcal{C}$ to improve for fixed V. Since N is an exponentially increasing function of $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$, we would expect the probability of correct classification into one of the classes in $\mathcal{A}$ to improve as $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$ becomes larger. In actually designing radar target classifiers, features that characterize $\mathbf{y}(t)$ are usually determined empirically and decision rules based on these features are constructed, allowing the classification of targets based on observation of $\mathbf{y}(t)$. Reference [6.9] describes the approaches typically taken in designing such classifiers, and Reference [6.10] gives a detailed account of such classification techniques.

Returning again to the results in Fig. 6.11, we see that a very large amount of information is obtained in the radar measurements in our example. For example, for $T = 100\,\mu$s and $P_x = 10\,W$, $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$ is approximately 100 nats, which equals approximately 144 bits. The corresponding N calculated from Eq. (6.103) is 2.69×10^{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 application in which

the radar measurements will be used.

When processing the received signal into a form that is useful, a significant amount of information can also be destroyed. We will see an example of this in the next chapter, in which synthetic aperture radar measurements are used to determine the radar reflectivity of a natural surface.

6.3. Radar Waveform Design and Implementation.

We now turn to the problem of designing specific radar waveforms $x(t)$ that have a magnitude-squared spectrum $|X(f)|^2$, to be used in order to probe the target so that the greatest amount of target information may be extracted from the received signal scattered by the target. In the first section of this chapter, we determined that any $x(t)$ with $|X(f)|^2$ as given by Eq. (6.78) would maximize $I(\mathbf{y}(t); \mathbf{g}(t)|x(t))$. Now we must address the design of specific transmitter waveforms that will do this. Some considerations in this waveform design will now be addressed.

The most theoretically obvious, although not the most practical, method of designing a waveform $x(t)$ with magnitude-squared spectrum $|X(f)|^2$ is to excite a filter with transfer function $H(f)$ such that $|H(f)|^2 = |X(f)|^2$ with an impulse or delta function. Given that $|H(f)|^2 = |X(f)|^2$, we see that such an $H(f)$ must satisfy

$$H(f) = \sqrt{|X(f)|^2} e^{i\phi(f)}, \tag{6.108}$$

where $\phi(f)$ can be any arbitrary phase function of f. A major shortcoming of this approach is that it requires the design of high-frequency (microwave) filters with very specific frequency characteristics. This can be very difficult to do and is also

an inefficient use of microwave energy, as it requires the broadband generation of microwaves to be fed into the filter. Here the frequencies to be attenuated are filtered out to the degree required to obtain the desired spectral characteristic $|X(f)|^2$, and this energy is lost. In addition, radar systems often employ signals of very high power, and the direct generation of complex, high-power, broadband microwave signals is not easily accomplished.

Typically in the design of radar systems, waveforms are generated at baseband frequencies much lower that that of the microwave transmitter frequency and then are used to modulate a microwave carrier in order to obtain the desired waveform at the radar operating frequency. Typical modulation techniques used in radar signal generation are amplitude, frequency, and phase modulation, with a large number of systems using phase (phase-coded waveforms) and frequency (both linear and non-linear FM chirp) modulation. This is because some constant-amplitude microwave amplifiers have greater efficiency than their variable-amplitude counterparts.

Generally, radar waveforms are bandpass waveforms because of practical bandwidth constraints on the components that generate and process them. As a result, $H(f)$ will typically have a bandpass filter response, and thus $x(t)$ can be generated by amplitude modulation of the carrier with a baseband signal. Such a bandpass signal can also be obtained by phase or frequency modulation of the carrier with a baseband waveform. However, synthesizing a desired spectrum using phase or frequency modulation is mathematically a much more difficult problem, since both phase modulation and frequency modulation are non-linear modulation processes.

As a result, it is more difficult to find a baseband modulating waveform to obtain the desired RF spectral characteristics. Reference [6.6] gives descriptions of these modulation processes, and Reference [6.11] gives an analysis of the spectral characteristics of phase modulation and frequency modulation for deterministic signals.

A problem of a more fundamental nature that arises in the design of $x(t)$ is due to the form of $|X(f)|^2$ in Eq. (6.78). The problem lies with the realizability of a waveform $x(t)$ with magnitude-squared spectrum $|X(f)|^2$. The realizability problem is due to the fact that there may not be an $x(t)$ with a magnitude-squared spectrum $|X(f)|^2$ that is causal. In practice, we get around this problem by generating a waveform $\tilde{x}(t)$ having a magnitude-squared spectrum $\left|\tilde{X}(f)\right|^2 \approx |X(f)|^2$. This approximation is normally carried out by by delaying a non-causal $x(t)$ corresponding to $|X(f)|^2$ and truncating the result prior to time zero in the delayed version. Since $x(t)$ is considered to be confined to the interval $\mathcal{T} = [-T/2, T/2]$, this should, in principle, be possible and a realizable waveform with magnitude-squared spectrum $|X(f)|^2$ (or at least approximately, ignoring some energy outside the frequency interval $\mathcal{W}$) will exist. For completeness however, we will investigate those $|X(f)|^2$ that have no corresponding causal $x(t)$.

In order for a linear system with transfer function $H(f)$ and impulse response $h(t)$ to be physically realizable, it must be a causal system. A causal system is one in which the output at time t does not depend on future input values. Mathematically, if the system with impulse response $h(t)$ is causal, then, if the input is $w(t)$ and the

output is $x(t)$, we have

$$\begin{aligned} x(t_0) &= \int_{-\infty}^{t_0} w(t)h(t-t_0)\,dt, \quad \forall t_0 \\ &= \int_{-\infty}^{t_0} w(\tau)h(t_0-\tau)\,dt, \quad \forall t_0. \end{aligned} \tag{6.109}$$

Thus, we have that an equivalent definition of a causal system is one in which the the impulse response $h(t) = 0$, for all $t < 0$.

The magnitude-squared transfer function, $|H(f)|^2$, can also be tested directly for causality using the *Paley-Wiener Criterion* [6.12]. The Paley-Wiener Criterion states that for a magnitude-squared function $|H(f)|^2$ to have an associated impulse response $h(t)$ that is causal and thus realizable, a necessary and sufficient condition on $|H(f)|^2$ is

$$\int_{-\infty}^{\infty} \frac{\left|\ln |H(f)|^2\right|}{1+f^2}\,df < \infty. \tag{6.110}$$

This says that for there to be *any* physically realizable $h(t)$ with a magnitude-squared spectrum $|H(f)|^2$, the area under the curve $\left|\ln |H(f)|^2\right|/(1+f^2)$ must be finite. This places two major constraints on the magnitude-squared function $|H(f)|^2$. The first, which will not concern us, requires that $|H(f)|^2 < K\exp(f^2)$ as $|f| \to \infty$, for some positive constant K. This is obviously true in the case of $|X(f)|^2$, since $|X(f)|^2$ satisfies the finite-energy constraint of Eq. (6.2). The second major restriction imposed by the Paley-Wiener Criterion is that the magnitude-squared function $|H(f)|^2$ *cannot be zero over any band or interval of frequencies.* $|H(f)|^2$ may, however, have a countable number of zeroes in f without violating the Paley-Wiener Criterion.

The significance of the Paley-Wiener Criterion to us is that it states that if $|X(f)|^2$ has any more than a countable number of zeroes in f, there is no realizable $x(t)$ that has a magnitude-squared spectrum $|X(f)|^2$ (n.b., the Paley-Wiener Criterion does not say that all $x(t)$ with a magnitude-squared spectrum $|X(f)|^2$ are realizable, only that there exist realizable $x(t)$ with a magnitude-squared spectrum $|X(f)|^2$). Since we specify that $|X(f)|^2 = 0$ for all $f \notin \mathcal{W}$, it would appear that none of the $|X(f)|^2$ obtained for a finite frequency interval $\mathcal{W}$ would have realizable $x(t)$. Recall that, from our discussion in the first section of this chapter, we noted that if $x(t)$ is zero outside the time interval $\mathcal{T} = [-T/2, T/2]$, then $X(f)$ cannot be limited strictly to a finite frequency interval $\mathcal{W}$. In reality, when we design our waveforms to be limited to the time interval $\mathcal{T}$, we attempt to concentrate as much of the energy as possible into the frequency interval $\mathcal{W}$, but some small portion of the energy falls outside of $\mathcal{W}$, and according to the Paley-Wiener Criterion, this is unavoidable. In addition, within the interval $\mathcal{W}$, for solutions yielding an $|X(f)|^2$ that is non-zero only on some $\hat{\mathcal{W}} \subset \mathcal{W}$, $|X(f)|^2$ will not have a corresponding realizable $x(t)$ if the set $\{\mathcal{W} - \hat{\mathcal{W}}\}$ has more than a countable number of points. Again, the solution to this apparent problem lies in the fact that since we cannot actually build these filters, they must be approximated. These realizable approximations do not violate the Paley-Wiener Criterion.

These approximations to the ideal, unrealizable filter can be done in several different ways. Many of the methods are covered in References [6.12] and [6.13]. Perhaps the most commonly used technique to approximate a non-causal impulse

response in radar and communication systems is to delay the response until it is causal or, in the case of impulse responses that are not of finite duration, delay them as much as is feasible, and then truncate the part of the delayed impulse response for time T less than zero. For any finite-energy impulse response, this technique can be used in principle to obtain arbitrarily small error in the magnitude-squared spectrum $|H(f)|^2$. In practice, however, this may not be practical because of the cost of implementation. Note also that these delays in generating the signal at the transmitter must be taken into account at the receiver.

6.4. Comparison of Waveforms for Optimum Detection and Maximum Information Extraction.

In Chapter 5, we examined the design of radar waveforms and receiver filters for optimum detection of targets with a known impulse response. In this chapter, we have examined the design of radar waveforms that maximize the mutual information between a target with random impulse response and the received signal at the radar receiver. The question arises: How do these two types of waveform compare?

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 in Chapter 5 that the solution corresponded to the eigenfunction with energy $\mathcal{E}$ corresponding to the largest eigenvalue of the integral equation of Eq. (5.29). 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 there, 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. Of course, when the impulse response of the target is known *a priori*, as was assumed in Chapter 5, there is nothing to be gained by this approach.

In this chapter, we have assumed that the target has a random impulse response $\mathbf{g}(t)$ and we distribute the energy in the transmitted signal as specified in Eq. (6.78). If we interpret $|H(f)|^2$ as "target response" in Chapter 5 and $\sigma_G^2(f)$ as "target response" in this chapter, 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. Using the two-sided power spectral density $S_{nn}(f)$ of the noise, we had

$$|X(f)|^2 \approx \alpha \frac{|H(f)|^2}{S_{nn}(f)}, \qquad \text{(Chapter 5)}$$

where α was a constant, and here we have

$$|X(f)|^2 = \max\left[0, A - \frac{S_{nn}(f)\widetilde{T}}{\sigma_G^2(f)}\right]. \qquad \text{(Chapter 6)}$$

The power spectral density of the noise enters into the solution in two quite different ways in the two solutions, and hence the form of the magnitude-squared spectrum of the two waveforms is quite different. While the waveform design of Chapter 5

attempted to put as much of the energy as possible into the mode of the target that gave the largest response when weighted with respect to the noise, the waveform design of this chapter distributes the available energy in order to maximize the information obtained about the target.

6.5. Chapter 6 References.

6.1 Ulaby, F. T., R. K. Moore, A. K. Fung, *Microwave Remote Sensing*, vol. 2, Addison-Wesley, Reading, MA, 1982.

6.2 Bracewell, R. N., *The Fourier Transform and its Applications*, McGraw-Hill, New York, NY, 1986.

6.3 Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, NY, 1965.

6.4 Skolnik, M. I., *Radar Handbook*, McGraw-Hill, New York, NY, 1970.

6.5 Shannon, C. E., "A Mathematical Theory of Communication," *Bell Sys. Tech. J.* 27, 1948. 379-423, 623-656. Reprinted in C. E. Shannon and W. W. Weaver, *The Mathematical Theory of Communication*, Univ. Ill. Press, Urbana, IL, 1949.

6.6 Pierce, J. R. and E. C. Posner, *Introduction to Communication Science and Systems*, Plenum, New York, NY, 1980.

6.7 Hamming, R. W., *Numerical Methods for Scientists and Engineers*, 2d. ed., Dover Publications, New York, NY, 1973.

6.8 Blackman, R. B. and J. W. Tukey, *The Measurement of Power Spectra*, Dover Publications, New York, NY,1958.

6.9 Duda, R. O. and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, NY, 1973.

6.10 Simon, J. C., *Patterns and Operators*, McGraw-Hill, New York, NY, 1986.

6.11 Stark, H. and F. B. Tuteur, *Modern Electrical Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

6.12 Lindquist, C. S., *Active Network Design with Signal Filtering Applications*, Steward and Sons, Long Beach, CA, 1977.

6.13 Lam, H. Y-F., *Analog and Digital Filters*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

6.14 Gallager, R. G., *Information Theory and Reliable Communication*, John Wiley and Sons, New York, NY, 1968.

CHAPTER 7

THE INFORMATION CONTENT OF RADAR IMAGES

In this chapter, we will investigate the information content of radar images. We will do this by defining the *spatial information channel* and examining its characteristics. We will then use these results and the results of Chapters 2, 3, and 4 to examine quantitatively how much information is conveyed by a radar image. This analysis will substantiate quantitatively certain qualitative results in the processing of radar images, particularly in the area of machine analysis of radar images.

In Section 7.1, we will examine the general and information-theoretic characteristics of radar images. We will examine the general characteristics of images and the way that discrete images arise as the result of the spatial sampling of two-dimensional, continuous images. We will then examine how such sampling arises naturally an imaging system as a result of the spatial resolution of the imaging system. Next, we will study the information-theoretic characteristics of discrete images in order to determine their information content. This will provide us with an understanding of the effects of spatial dependencies between pixels in a radar image on the image's information content. In Section 7.1, we perform this analysis, assuming a general discrete distribution on the resolution cell or pixel intensity values in the image.

In Section 7.2, we examine the information-theoretic characteristics of the intensity parameter of the individual pixels of a radar image when the displayed

quantity is received power from a diffuse surface. In particular, we examine the mutual information between the average surface reflectivity and displayed power in the image. We note the effects of "speckle noise" in limiting this mutual information. We also note the effect of averaging independent observations of a homogeneous region of the surface to combat speckle noise and the effect of this procedure on the mutual information between the average reflectivity of the surface and the displayed image intensity. From this mutual information, we are able to determine the maximum number of equiprobable classes into which we can classify the surface reflectivity, based on our observation of received power. Coupling these results with those of Section 7.1, we conclude that in order to obtain significant ability to classify regions of a surface into categories that are a function of surface reflectivity, a region made up of a large number of pixels or resolution cells is required. This is an information-theoretic explanation of a well-known heuristic result.

7.1 Radar Images.

In this section, we will examine some general characteristics of radar images, that is, images generated by either *real aperture* or *synthetic aperture* radars. We will not present an in-depth analysis of imaging radar systems, as there are several good books that give excellent in-depth coverage of the underlying physical principles of radar imaging [7.1–7.6]. We will instead investigate the information-theoretic characteristics of radar images.

An *image*, according to the IEEE definition [7.7, p. 361], is "a spatial distribution of a physical property such as radiation, electric charge, conductivity or

reflectivity, mapped from another distribution of either the same or another physical property." A *radar image*, then, can be defined as an image obtained with the use of a sensor employing active electromagnetic scattering. A radar image is a spatial mapping of the scattering characteristics of the area being illuminated by the radar transmitter. Actually, the radar image consists of some intensity parameter $a(x, y)$ displayed as a function of two orthogonal coordinates x and y, and this parameter is a function of the scattering characteristics of the target at the spatial point corresponding to these image coordinates. The displayed intensity parameter $a(x, y)$ could represent the reflected signal power, the ratio of received signal powers from two orthogonal polarizations, or some other functions of the scattered radar signal. Examples of these would include soil moisture, wind speed over water, crop characteristics, or surface texture metrics [7.8]. If the intensity pattern $a(x, y)$ is viewed as a continuous random field with finite energy, the two-dimensional spectrum $A(\mu, \nu)$ of $a(x, y)$ is given by the two-dimensional Fourier Transform

$$A(\mu, \nu) = \int_{-\infty}^{\infty} a(x, y) e^{-i2\pi(\mu x + \nu y)}\, dx\, dy. \tag{7.1}$$

Here μ is the spatial frequency in the direction of the x-coordinate and ν is the spatial frequency in the direction of the y-coordinate.

In practice, the radar images obtained using imaging radars are discrete random fields. The samples that make up these discrete images correspond to the discrete sampling of a spatially bandlimited continuous image in (x, y) with spatial bandwidths $\mu_{\max}$ and $\nu_{\max}$ corresponding to the x-coordinate and y-coordinate, respectively. For sampling intervals of Δx in the x-coordinate and Δy in the y-

coordinate, $\mu_{\max}$ and $\nu_{\max}$ are

$$\begin{aligned} \mu_{\max} &= \frac{1}{2\Delta x}, \\ \nu_{\max} &= \frac{1}{2\Delta y}. \end{aligned} \tag{7.2}$$

This follows from the Sampling Theorem [7.8, pp. 127-8]. So we assume that $A(\mu,\nu) = 0$ for $|\mu| > \mu_{\max}$ and $|\nu| > \nu_{\max}$. Note that the spatial bandlimiting is not a function of the surface alone, but is also a function of the imaging system —in particular, the imaging system's spatial resolution. In an imaging system, Δx and Δy correspond to the spatial resolution in the x and y directions, respectively. Thus, although there may be spatial fluctuations in the physical scattering characteristics of a surface with spatial frequencies greater than $\mu_{\max}$ and $\nu_{\max}$, the spatial resolution of the imaging system effectively low-pass-filters these fluctuations out of the image produced by the imaging system. We can thus represent the spatial sampling of the image as shown in Fig. 7.1. Here, $H(\mu,\nu)$ represents the two-dimensional, low-pass filtering effect that is due to the limited spatial resolution of the imaging system, and the two-dimensional sampler samples the resulting spatially bandlimited signal at the *Nyquist sampling rates*

$$\begin{aligned} \mu_N &= 2\mu_{\max} = \frac{1}{\Delta x}, \\ \nu_N &= 2\nu_{\max} = \frac{1}{\Delta y}. \end{aligned} \tag{7.3}$$

We can view the resulting image as an $m \times n$ array of resolution cells as shown in Fig. 7.2, with an intensity value $a(i,j)$ for the discrete pair of coordinates (i,j), where $i \in \{1,\ldots,m\}$ and $j \in \{1,\ldots,n\}$. The intensity level corresponding to each of these resolution cells is most easily modeled as a continuous parameter, and in our

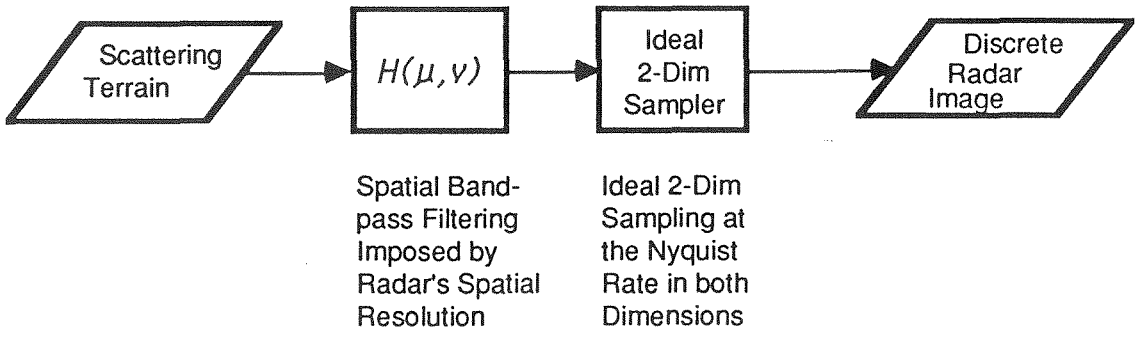


Figure 7.1. Spatial Sampling in an Imaging Radar System.

analysis we will often do so. However, in most real imaging systems and especially those that use digital processing, these levels will be quantized into discrete levels. For our present analysis, we will assume that there are K such levels. We will designate the set of K possible values that the intensity parameter can take on in each resolution cell as $\mathbf{E} = \{E_1, \ldots, E_K\}$. The image under consideration can be considered to be an $m \times n$ array, with each element of the array taking on one of the K elements of $\mathbf{E}$. The elements of the array correspond to the resolution cells or "pixels" of the image.

If we view the image under consideration as a noiseless communication channel, we note that there are N unique messages that can be conveyed by such a channel, with

$$N = K^{mn}. \tag{7.4}$$

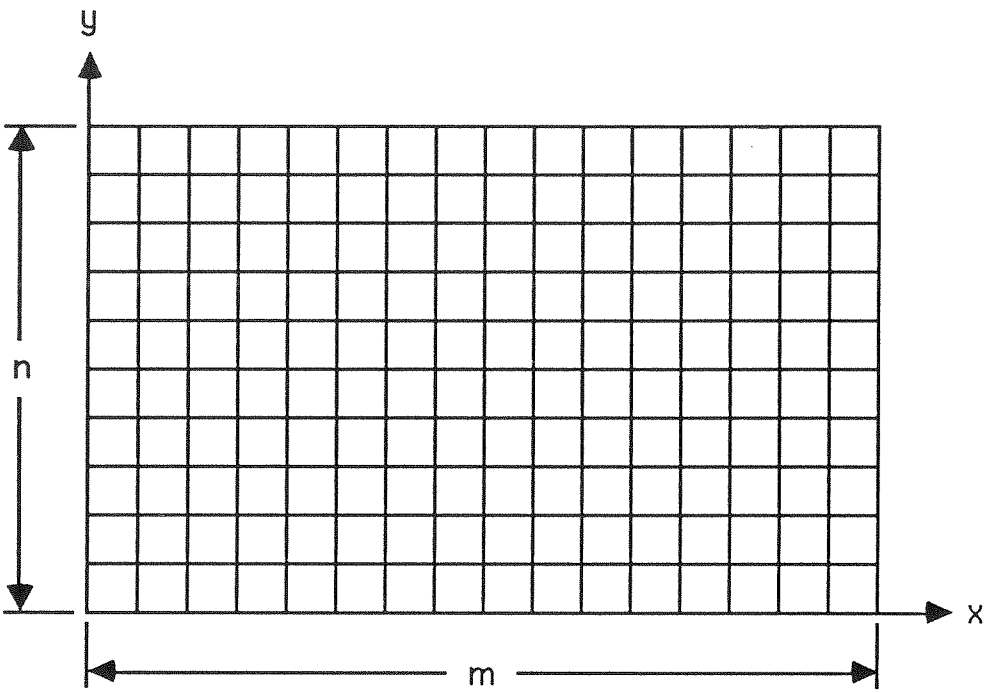


Figure 7.2. Discrete Radar Image as an $m \times n$ Array of Resolution Cells.

Equivalently, we see that there are N distinguishable images that can be constructed. By viewing the image as a spatial information channel, we are led to consider the quantity of information that can be conveyed by the image. The amount of information I that can be conveyed by this image viewed as a noiseless channel is

$$\begin{aligned} I &= \log N \\ &= \log K^{mn} \\ &= mn \log K. \end{aligned} \tag{7.5}$$

If we take the logarithm to have base-e, we have

$$I = mn \ln K \quad \text{(nats/image)}. \tag{7.6}$$

Assume that we generate the image $\mathbf{F}$ by assigning a value from $\mathbf{E}$ to each of the $m \times n$ resolution cells independently, with each cell having probabilities $P_1, \ldots, P_K$ of taking on values $E_1, \ldots, E_K$ from $\mathbf{E}$, respectively. Then when the number of resolution cells mn grows large, the entropy $H(\mathbf{F})$ of the image, and thus the amount of information it conveys when generated by such a source, is

$$H(\mathbf{F}) = mn\, H(E), \tag{7.7}$$

where $H(E)$, the entropy of an individual resolution cell, is

$$H(E) = -\sum_{k=1}^{K} P_k \log P_k. \tag{7.8}$$

To see that such an image does, in fact, convey this much information, assume that there are a total of $Q = mn$ resolution cells in the image under consideration. Let $n_1, \ldots, n_K$ be the number of resolution cells with intensities $E_1, \ldots, E_K$, respectively. Then we have $Q = n_1 + \cdots + n_K$, and the number N of distinguishable images with a given set of numbers $\{n_1, \ldots, n_K\}$ is

$$N = \frac{M!}{n_1! \cdots n_K!}. \tag{7.9}$$

Note that for large $Q = mn$, we have from the law of large numbers that

$$P_j \approx \frac{n_j}{Q}. \tag{7.10}$$

We saw previously that the information conveyed by an image is given by $\log N$. Taking the logarithm of Eq. (7.9), we obtain

$$\begin{aligned} \log N &= \log\left(\frac{Q!}{n_1! \cdots n_K!}\right) \\ &= \log Q! - [\log n_1! + \cdots + \log n_K!]\,. \end{aligned} \tag{7.11}$$

Using Sterling's approximation [7.9, p.], which states that

$$k! \approx \sqrt{2\pi}\, k^{k+1/2} e^{-k},$$

to approximate the factorials in Eq. (7.11), we have for large Q that

$$\begin{aligned} \log N \approx\ & -(K-1)\log\sqrt{2\pi} - Q + (n_1 + \cdots + n_K) \\ & + (Q+1/2)\log Q - [(n_1+1/2)\log n_1 + \cdots + (n_K+1/2)\log n_K] \\ =\ & (Q+1/2)\log Q - [(n_1+1/2)\log n_1 + \cdots + (n_K+1/2)\log n_K] \\ & - (K-1)\log\sqrt{2\pi}. \end{aligned} \tag{7.12}$$

Since Q is very large, we will assume that $(Q+1/2)\log Q \gg K\log\sqrt{2\pi}$, and thus we have

$$\log N \approx (Q+1/2)\log Q - [(n_1+1/2)\log n_1 + \cdots + (n_K+1/2)\log n_K]\,. \tag{7.13}$$

From Eq. (7.10),

$$n_j \approx QP_j, \quad \text{for } j = 1, \ldots, K. \tag{7.14}$$

This being the case, we can rewrite Eq. (7.13) as

$$\log N \approx (Q+1/2)\log Q - [(QP_1+1/2)\log QP_1 + \cdots + (QP_K+1/2)\log QP_K]\,. \tag{7.15}$$

We will assume that for the probabilities P_j chosen, Q is sufficiently large such that $QP_j \gg 1/2$. Then we can write

$$\begin{aligned}
\log N &\approx Q \log Q - [QP_1 \log QP_1 + \cdots + QP_K \log QP_K] \\
&= Q \log Q - \sum_{k=1}^{K} QP_k \, [\log Q + \log P_k] \\
&= Q \log Q - Q \log Q - Q \sum_{k=1}^{K} P_k \log P_k \\
&= -Q \sum_{k=1}^{K} P_k \log P_k \\
&= QH(E) \\
&= mnH(E).
\end{aligned} \tag{7.16}$$

So as Q grows large, we have

$$\log N \approx QH(E), \tag{7.17}$$

and thus in the limit, as $Q \to \infty$, we have

$$\log N = mnH(E). \tag{7.18}$$

So we see that the amount of information conveyed by an image in which all mn resolution cells are statistically independent and in which each resolution cell has entropy $H(E)$ is given by Eq. (7.18). Thus, in order for this amount of information to be maximum, it follows from Eq. 7.18 that the probability distribution $(P_1, \ldots, P_K)$ must be selected to maximize $H(E)$. It is easily seen that this occurs when $P_j = 1/K$, for all $j = 1, \ldots, K$. When the intensity levels are selected according to this probability distribution, we have

$$\begin{aligned}
H(E) &= -\sum_{k=1}^{K} \frac{1}{K} \log \frac{1}{K} \\
&= \log K,
\end{aligned} \tag{6.19}$$

and thus the maximum amount of information that can be conveyed by an image is

$$\log N = mn \log K, \tag{7.20}$$

which is in agreement with the maximum amount of information that could be conveyed by an image as given in Eq. (7.5).

In examining the information content of images, we have seen that the maximum amount of information a K-level $m \times n$ image can convey is $mn \log K$. Our primary interest here, however, is to consider the information content of radar images. In finding the maximum amount of information that a K-level $m \times n$ image can convey, we generated the image by selecting the intensity values of the resolution cells independently of that of all other resolution cells and such that all of the K intensity levels were equiprobable. As a result, we have that all of the K^{mn} possible images are equiprobable. Examination of real radar images indicates that that there is significant correlation between adjacent pixels in the image. This is due to the fact that there are usually significant similarities in the scattering characteristics of the physical terrain being imaged for adjacent resolution cells.

As an illustration of this, consider the radar image in Fig. 7.3. This image was obtained using a side-looking synthetic-aperture radar mounted on a NASA CV990 aircraft and processed at the Jet Propulsion Laboratory. The operating characteristics of the radar are given in Table 7.1. The image was generated using linear cross-polarization between transmit and receive and a grazing angle (the angle between the direction of wave propagation and the surface) of 45°. In this image, four independent looks were averaged in order to reduce speckle. (Speckle

(*Courtesy of Jet Propulsion Laboratory*)

Figure 7.3. Radar Image of San Francisco Area.

Table 7.1: Radar Operating Characteristics.

Parameter	Value
Wavelength	24.5 cm
Resolution	approx. 10 m × 10 m with 4 looks
Waveform Bandwidth	20 MHz nominal
Peak Power	5 kW
Pulse Repetition Rate	1 per 34 cm along track
Ground Track Velocity	200–300 m/s

is a multiplicative noise that will be discussed in this next section.) Note that the image obviously has correlation between pixels. Thus, a model that treats the returns from each resolution cell as statistically independent is not realistic. This means that in considering all of the K^{mn} possible K-level $m \times n$ images, some of them will be much more probable than others, and some will have a truly negligible probability of occurring (of being observed by an imaging radar). We will now investigate these characteristics quantitatively by examining the entropy of random images that represent the radar image of a random surface. This analysis follows that in Reference [7.12, Sect. 7.6].

Let an $m \times n$ image matrix $\mathbf{F}$ of resolution cells quantized to one of K intensity levels be represented by the $Q \times 1$ column vector $\mathbf{f}$, where $Q = mn$ and $\mathbf{f}$ is obtained from $\mathbf{F}$ by column scanning $\mathbf{F}$. We can think of the vector $\mathbf{f}$ as the output of an information source capable of producing any of the K^{mn} possible vectors $\mathbf{f}$. The majority of the K^{mn} possible vectors have no discernible structure and appear to be

similar to a random noise field (much like the output of a television receiver tuned to a channel in which no signal is present). For typical values of m,n, and K that one would find in an imaging system, only a small fraction of the K^{mn} possible images correspond to images that could be generated by a real radar system imaging terrain. It is reasonable then, at least in principle, to consider the *a priori* probability of occurrence for each of the $T = K^{mn}$ possible images $\mathbf{f}_t$. This probability will be designated $P(\mathbf{f}_t)$, where $t = 1, \ldots, T$. The entropy $H(\mathbf{f})$ of the image source is thus

$$H(\mathbf{f}) = -\sum_{t=1}^{T} P(\mathbf{f}_t) \log P(\mathbf{f}_t). \tag{7.21}$$

The significance of $H(\mathbf{f})$ is derived from Shannon's *noiseless channel coding theorem* [7.10], which states that it is possible to encode a source of entropy $H(\mathbf{f})$ bits without error using $H(\mathbf{f}) + \epsilon$ code bits, where ϵ is a small positive quantity. Conversely, it states that it is not possible to encode a source of entropy $H(\mathbf{f})$ bits with $H(\mathbf{f}) - \epsilon$ bits without error. In fact, if the source produces successive images independently, then the information $I(\mathbf{f}_{t1}, \ldots, \mathbf{f}_{tL})$ required to specify L source images (that is, a sequence of L source images) exhibits the following asymptotic behavior:

$$\frac{I(\mathbf{f}_{t1}, \ldots, \mathbf{f}_{tL})}{L} \xrightarrow{L \to \infty} H(\mathbf{f}). \tag{7.22}$$

We note that $H(\mathbf{f}) \leq mn \log K$, with equality if and only if $P(\mathbf{f}_t) = K^{-mn} = 1/T$, for $t = 1, \ldots, T$. In the case of radar images as well as most other physically generated images, $H(\mathbf{f}) \ll mn \log K$.

The probability of occurrence of $\mathbf{f}_t$, $P(\mathbf{f}_t)$, can be expressed as the joint prob-

ability distribution of the individual resolution cell intensities:

$$P(\mathbf{f}_i) = \Pr\{f(1) = r_{j1}(1), \ldots, f(Q) = r_{jQ}(Q)\}. \tag{7.23}$$

Here, $r_{jp}(p)$ represents the j-th intensity value at resolution cell p. An alternative expression for $P(\mathbf{f}_i)$ can be obtained by expanding Eq. (7.23) in terms of conditional probabilities:

$$\begin{aligned} P(\mathbf{f}_i) =& \Pr\{f(1) = r_{j1}(1)\} \Pr\{f(2) = r_{j2}(2) | f(1) = r_{j1}(1)\} \\ & \cdots \Pr\{f(q) = r_{jq}(q) | f(q-1), f(q-2), \ldots, f(1)\} \\ & \cdots \Pr\{f(Q) = r_{jQ}(Q) | f(Q-1), f(Q-2), \ldots, f(1)\}. \end{aligned} \tag{7.24}$$

Using this expression and the definition of $H(\mathbf{f})$ given in Eq. (7.21), we have

$$\begin{aligned} H(\mathbf{f}) = & -\sum_{i=1}^{T} P(\mathbf{f}_i) \log\left[\Pr\{f(1) = r_{j1}(1)\}\right] \\ & -\sum_{i=1}^{T} P(\mathbf{f}_i) \log\left[\Pr\{f(2) = r_{j2}(2) | f(1)\}\right] \\ & - \cdots \\ & -\sum_{i=1}^{T} P(\mathbf{f}_i) \log\left[\Pr\{f(q) = r_{jq}(q) | f(q-1), \ldots, f(1)\}\right] \\ & - \cdots \\ & -\sum_{i=1}^{T} P(\mathbf{f}_i) \log\left[\Pr\{f(Q) = r_{jQ}(Q) | f(Q-1), \ldots, f(1)\}\right]. \end{aligned} \tag{7.25}$$

The general term in Eq. (7.25), which will be given the notation

$$H\left[f(q) | f(q-1), \ldots, f(1)\right],$$

can be interpreted as the average information provided by the resolution cell at image vector coordinate q when the values of the preceding $q-1$ resolution cells

are known. We can rewrite Eq. (7.25) as

$$H(\mathbf{f}) = \sum_{q=1}^{Q} H\left[f(q)|f(q-1),\ldots,f(1)\right]. \tag{7.26}$$

Eq. (7.26) gives the image source entropy of the radar image and thus gives a measure of the amount of information or "information content" of the image.

We will now make a simplification in our analysis. Recall that we are considering a column scanned image vector $\mathbf{f}$. For such a random vector, it is easily shown [7.11] that for $j > k$,

$$H\left[f(q)|f(q-1),\ldots,f(q-j)\right] \leq H\left[f(q)|f(q-1),\ldots,f(q-k)\right]. \tag{7.27}$$

We can interpret this inequality as saying that the more knowledge we have of pixels preceding pixel q, the less uncertainty we have about the value of pixel q. To be more precise, the inequality states that if we increase our knowledge of the pixel values preceding q, our uncertainty of $f(q)$ cannot increase, and it will either decrease or remain constant. The uncertainty in $f(q)$ remains constant as knowledge of the history of preceding pixel values is increased if and only if the pixels in the increased history are statistically independent of $f(q)$.

For most physically generated images, an individual term of Eq. (7.26) of the form $H\left[f(q)|f(q-1),\ldots,f(1)\right]$ approaches a non-zero limiting value as j becomes large (i.e., as the history of previous pixel values becomes greater) [7.12, p. 187]. We will denote this limiting pixel value as $H\left[f(q)|\infty\right]$. If we now assume that end-of-sequence effects are negligible, then from Eq. (7.26), the image entropy $H(\mathbf{f})$ can

be approximated by

$$H(\mathbf{f}) \approx \sum_{q=1}^{Q} H\,[f(q)|\infty]. \tag{7.28}$$

Furthermore, if $\mathbf{F}$ behaves as an $m \times n$ element area of a stationary, two-dimensional process, then $H\,[f(q)|\infty]$ is constant for all q, and Eq. (7.28) becomes

$$H(\mathbf{f}) \approx QH\,[f(q)|\infty]. \tag{7.29}$$

Eqs. (7.28) amd (7.29) provide a means for estimating the image entropy $H(\mathbf{f})$, and thus, of estimating the image's information content.

In order to use Eqs. (7.28) and (7.29) to estimate the entropy of a column-scanned image, we must determine the limiting conditional entropy $H\,[f(q)|\infty]$ of $f(q)$ from a finite sequence of pixels. So we will use an approximation of the form

$$\begin{aligned} H\,[f(q)|\infty] &\approx H\,[f(q)|f(q-1),\ldots,f(q-p)] \\ &= \sum_{j_0=1}^{K} \cdots \sum_{j_p=1}^{K} P\,[f(q),\ldots,f(q-p)] \log \frac{P\,[f(q),\ldots,f(q-p)]}{P\,[f(q-1),\ldots,f(q-p)]}, \end{aligned} \tag{7.30}$$

where

$$P\,[f(q),\ldots,f(q-p)] = \Pr\{f(q) = r_{j0}(q),\ldots,f(q-p) = r_{jp}(q)\}. \tag{7.31}$$

We note from Eq. (7.27) that the approximation given by Eq. (7.30) is an upper bound.

In order to make use of the (upper-bound) estimate of $H\,[f(q)|\infty]$ given by Eq. (7.30), we must estimate the distribution given in Eq. (7.31). This can be done either by mathematically modeling the image source and deriving the source

statistics from this model or by histogram estimation of the distribution from real radar image data. Mathematical modeling of the image source may be inaccurate if detailed knowledge of the surface is not available, as there are many different surface models having widely varying scattering behavior [7.1, 7.3, 7.11, 7.13]. On the other hand, the histogram estimation of the distribution has extreme computational and image data requirements. The joint distribution $P[f(q),\ldots,f(q-p)]$ for a k-level image has k^{p+1} distinct sets of arguments at which it must be estimated. Thus, for typical values of K, such as $k = 256$ corresponding to 8-bit image level quantization, the computational requirements become excessive for even relatively small p. In addition, the number of sample values required for reliable estimation of the distribution from the histograms can be quite large [7.14, Ch. 4].

An appreciation for the difficulties of the estimation of image distributions using histograms can be gained from the work of Schrieber [7.15], who performed such estimations on television images with $K = 64$ and $p = 0$, 1, and 2. The results obtained in that study showed that the uncertainty in $f(q)$ was significantly reduced in going from $p = 0$ to $p = 1$ (from 4.39 bits to 1.91 bits), while there was only a slight additional reduction in going from $p = 1$ to $p = 2$ (from 1.91 bits to 1.49 bits). The joint entropy of $f(q)$, $f(q-1)$, and $f(q-2)$ was 7.80 bits. These results seem to indicate that most of the information that could be obtained from the preceding pixels could be obtained from only one preceding pixel, although the rate at which $H[f(q)|f(q-1),\ldots,f(q-p)]$ converged to $H[f(q)|\infty]$ could not be determined. This was due to the computational difficulties of histogram estimation for larger p:

$H\,[f(q)|f(q-1),\ldots,f(q-p)]$ could not be easily determined for p greater than 2.

So far, we have been considering the conditional entropy of the pixel value $f(q)$ conditioned on the values of previous pixels in the column-scanned image vector $\mathbf{f}$ of $\mathbf{F}$. This situation is shown in Fig. 7.4a. It is also possible, and perhaps physically more meaningful, to condition the entropy of $f(q)$ on the values of other sets of pixels, for example, those pixels directly neighboring the pixel q or some larger neighborhood of the pixel q. Typical examples are shown in Fig. 7.4b, which shows the neighborhood

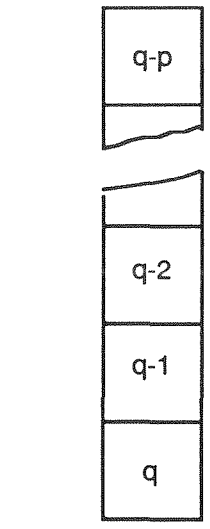
$$\{f(q); f(q-1), f(q+1), f(q-m), f(q+m)\},$$

and in Fig. 7.4c, which shows the neighborhood

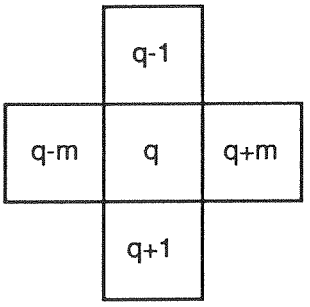
$$\left\{\begin{array}{l} f(q); f(q-1), f(q+1), f(q-m-1), f(q-m), \\ f(q-n+1, f(q+m-1), f(q+m), f(q+m+1) \end{array}\right\}.$$

Recall that m is the number of resolution cells per row in the image matrix $\mathbf{F}$. Larger neighborhoods, such as the one shown in Fig. 7.4d, could also be constructed in principle, although the computational burden of computing the histogram estimates of their associated conditional distributions would be prohibitive in practice for typical numbers of quantization levels K.

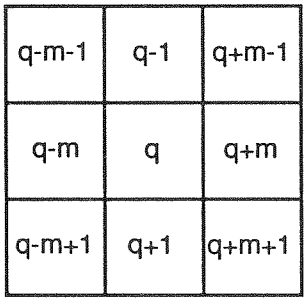
In general, in order to estimate the image entropy $H(\mathbf{f})$, we will use one of the previously mentioned estimates of $H\,[f(q)|\infty]$ and then use either Eq. (7.28) or Eq. (7.29) in order to determine $H(\mathbf{f})$. In practice, in order to simplify the computation so that it is reasonable, we will most likely have to assume that the $m \times n$ image $\mathbf{F}$ under consideration is a section of a stationary, two-dimensional,



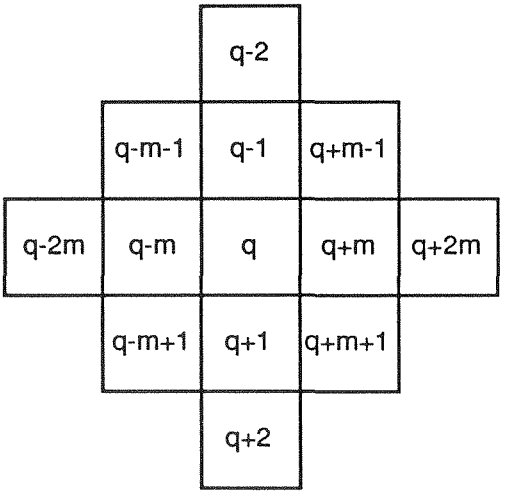
(a)



(b)



(c)



(d)

Figure 7.4. Resolution Cell Neighborhoods on Which $f(q)$ Is Conditioned.

random process and will use the approximation

$$H(\mathbf{f}) \approx QH\,[f(q)|\infty]$$

given in Eq. (7.29).

As a simple example to clarify these ideas, consider a 10×10 binary image F with a $Q = 100$ element column-scanned image vector $\mathbf{f} = [f(1), \ldots, f(100)]$. The pixel intensities $f(q)$ take on values of either "0" or "1." We will assume that the image is generated by a binary Markov process, such that the probability distribution of $f(q)$ for $2 \leq q \leq 100$ is dependent only on the previous pixel value $f(q-1)$. We will assume that the conditional distribution governing these probabilities is

$$\Pr\{f(q) = 0 | f(q-1) = 0\} = 0.75,$$

$$\Pr\{f(q) = 1 | f(q-1) = 0\} = 0.25,$$

$$\Pr\{f(q) = 0 | f(q-1) = 1\} = 0.25,$$

$$\Pr\{f(q) = 1 | f(q-1) = 1\} = 0.75.$$

We will assume that $\Pr\{f(1) = 0\} = \Pr\{f(1) = 1\} = 1/2$. A typical image constructed in this manner is shown in Fig. 7.5, where "1"" is represented by black in the image and "0" is represented by white. Since the value of $f(q)$ is dependent only on the previous pixel value $f(q-1)$, we have that

$$\begin{aligned} H\,[f(q)|\infty] &= H\,[f(q)|f(q-1)] \\ &= -0.25 \ln 0.25 - 0.75 \ln 0.75 \\ &= 0.5623 \text{ (nats)}. \end{aligned}$$

Thus, since the probability distribution is stationary in q, we have that the entropy

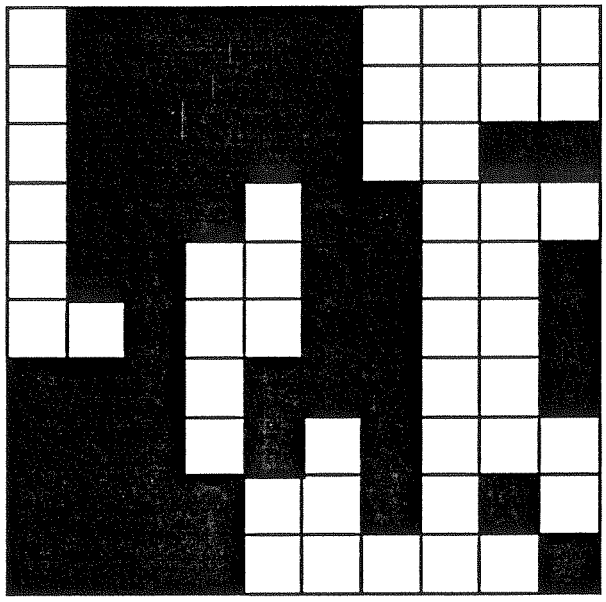


Figure 7.5. 10×10 Binary Markov Image.

of the image $H(\mathbf{f})$, as given by Eq. (7.29), is

$$H(\mathbf{f}) \approx QH\,[f(q)|\infty]$$

$$= 100 \cdot 0.5623\,(\text{nats})$$

$$= 56.23\,(\text{nats}).$$

To calculate the exact entropy of the image, we note that the entropy of $f(1)$ is

$$H\,[f(1)] = -0.5 \ln 0.5 - 0.5 \ln 0.5$$

$$= \ln 2$$

$$= 0.6931\,(\text{nats}).$$

Each of the remaining 99 pixels has entropy $H\,[f(q)|f(q-1)] = 0.5623$ nats condi-

tioned on the previous pixel, and thus we have

$$H(\mathbf{f}) = H\,[f(1)] + 99 \cdot H\,[f(q)|\infty]$$

$$= 0.6931 + 99 \cdot 0.5623$$

$$= 56.36\,(\text{nats}).$$

So we see that the approximation of Eq. (7.29) is very good in this example.

We note that the marginal probability that any pixel value $f(q)$ equals "0" or "1" is 1/2. If each of the pixels were statistically independent of all other pixels, such a 100 pixel image would have an entropy of 69.31 nats. So we see that the dependence between successive pixels in the binary Markov process that generates this image accounts for a significant reduction in the entropy of the image. We can see this visually in Fig. 7.5 by noting the presence of a significant number of "runs" of both "0" and "1" in the columns of the image.

Having considered the basic information-theoretic concepts of radar images, we now look at some specific details determining the information content of radar images.

7.2 Information Theory and Speckle Noise in Radar Images.

In the previous section, we examined the general information-theoretic characteristics of radar images. However, we did not address the actual information content of each individual resolution cell or pixel. We assumed that each resolution cell displayed an intensity that was quantized to one of K levels, and we determined that the maximum amount of information that could be obtained per resolution cell was $\log K$, but we did not actually investigate the information that could be expected from a radar measurement of a resolution cell. In this section, we will examine the information per resolution cell when making measurements of the radar reflectivity of a surface. We will do this not only for the case of an image generated by a single look at the resolution cell, but also for an image in which each resolution cell is made up of the arithmetic mean of multiple measurements of the reflected power from a resolution cell, a technique used to reduce "speckle noise" in radar images. Much of what is presented in this section is the result of an investigation by Frost and Shanmugan [7.16].

Consider an imaging radar, with either real or synthetic aperture, making measurements of a homogeneous diffuse surface. By homogeneous, we mean that the radar reflectivity Z of the surface is a wide-sense-stationary, 2-dimensional random process. By diffuse, we mean that the scattering characteristics of the surface at the wavelength of the radar radiation satisfiy the characteristics of diffuse scatterers as outlined in Section 4.4. That is, the surface area within a radar resolution cell is assumed to be made up of a large number of independent scatterers, none of

which are dominant in terms of the total scattered power. In addition, they are assumed to be spatially distributed such that the surface is rough in the Rayleigh sense, as described in Section 4.1. Such a surface would be a *Category 1* surface in the nomenclature of Section 4.4.

The radar reflectivity of the surface is a random variable Z, which we will assume to be uniformly distributed over the interval $[\beta_1, \beta_2]$. This assumption is more than a mathematical convenience. In Reference [7.15], this assumption is shown to be in good agreement with synthetic-aperture radar-measurements of terrain.

We will now assume that a radar reflectivity measurement is made by measuring the received power from the surface by using one of m "waveforms" from the set $\mathbf{A} = \{\alpha_1, \ldots, \alpha_m\}$. Here, the term "waveform" has a special meaning. We will assume that the actual waveform transmitted by the target is of any of a number of typical forms used in imaging radar, such as a chirp or pulsed sinusoid waveforms. However when we say that a "waveform" α_k is used to make a measurement of the radar reflectivity of the surface, we mean that k independent measurements of the reflected power from a homogeneous region of the surface are made, and that their arithmetic mean is taken as the output of the radar receiver. In terms of the model presented in Chapter 3 for the *Radar/Information Theory Problem* for continuous target channels, we are using the target channel *only one time* ($N = 1$), but the proper definition of the "waveform" α_k allows us to average the results of k independent observations of the received power reflected from the surface in the

use of a single "waveform." So the homogeneous region of the surface is assumed to have a reflectivity Z, and when a waveform $X = \alpha_k$ is used to measure this reflectivity, the resulting receiver output Y is the arithmetic mean of the received power from each of the k independent measurements of received power scattered from the surface.

We are interested in determining the mutual information $I(Y;Z|X = \alpha_k)$ in order to determine how much information is obtained about the surface reflectivity Z by observation of Y. Y and Z are parameters of primary interest in radar imaging. Y, as previously defined, represents the radar measurement of the surface reflectivity, whereas Z represents the mean reflectivity of the region (which is unknown and modeled as a random variable). As was discussed in Chapter 4, the radar reflectivity of an object is both a function of its material composition and geometric or spatial structure. As we will see, Z corresponds to the surface material characteristics, but Y includes the effects of a multiplicative noise called "speckle," resulting from the constructive and destructive interference occurring at the radar receiver that is due to the roughness of the surface being imaged.

The mutual information $I(Y;Z|X = \alpha_k)$ can be written as

$$I(Y;Z|X = \alpha_k) = h(Y|X = \alpha_k) - h(Y|Z, X = \alpha_k). \tag{7.32}$$

Here, the differential entropies $h(Y|X = \alpha_k)$ and $h(Y|Z, X = \alpha_k)$ (in nats) are given by

$$h(Y|X = \alpha_k) = -\int_{-\infty}^{\infty} f(y|X = \alpha_k) \ln f(y|X = \alpha_k)\, dy, \tag{7.33}$$

and

$$h(Y|Z, X = \alpha_k) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y, z|X = \alpha_k) \ln f(y|z, X = \alpha_k)\, dy\, dz. \qquad (7.34)$$

Thus, in order to evaluate $I(Y; Z|X = \alpha_k)$, we must determine the density functions $f(y|X = \alpha_k)$, $f(y, z|X = \alpha_k)$, and $f(y|z, X = \alpha_k)$. We will now determine these density functions.

Assume that the radar is illuminating the homogeneous diffuse surface as previously described and that the transmitter and receiver have fixed (although possibly different) antenna polarizations. We will assume that the received signal has sufficient energy such that it is reasonable to ignore the effects of additive noise. This assumption is reasonable, since multiplicative "speckle" noise, not additive noise, is most often the limiting factor in imaging radars [7.3, §8-7, 9-8]. Then, as we determined in Section 4.4, the received power P from a single observation is exponentially distributed, and its probability density function (PDF), given by Eq. (4.42), is

$$f_P(p) = \begin{cases} (1/\mu)\exp(-p/\mu) & \text{, for } p \geq 0; \\ 0 & \text{, elsewhere.} \end{cases} \qquad (7.35))$$

Here μ is the mean value of P. The characteristic function $\Psi_P(\omega)$ corresponding to this PDF is

$$\begin{aligned} \Psi_P(\omega) &= \int_{-\infty}^{\infty} f_P(p) e^{i\omega p}\, dp \\ &= \frac{1}{1 - i\omega\mu}. \end{aligned} \qquad (7.36)$$

When $X = \alpha_k$, Y is defined as the arithmetic means of k independent, identically distributed random variables P_j, each with a density function given by

Eq. (7.35). So we have that for $X = \alpha_k$,

$$Y = \frac{1}{k} \sum_{j=1}^{k} P_j. \tag{7.37}$$

In order to find the distribution of Y, we define the random variable W as

$$W \stackrel{\text{def}}{=} kY = \sum_{j=1}^{k} P_j. \tag{7.38}$$

Then, since the P_j are independent and identically distributed, we have that $\Psi_W(\omega)$, the characteristic function of W, is

$$\begin{aligned} \Psi_W(\omega) &= [\Psi_P(\omega)]^k \\ &= \frac{1}{(1 - i\omega\mu)^k}. \end{aligned} \tag{7.39}$$

But since

$$Y = \frac{1}{k} \cdot W,$$

we have that

$$f_Y(y) = k f_W(ky),$$

and thus we have

$$\Psi_Y(\omega) = \Psi_W\left(\frac{\omega}{k}\right). \tag{7.40}$$

Hence, we have that the characteristic function of Y is

$$\Psi_Y(\omega) = \frac{1}{(1 - i\omega\mu/k)^k}. \tag{7.41}$$

This is the characteristic function of the gamma distribution with mean μ/k and k degrees of freedom [7.17, p.104-6]. Thus, we have that the density function $f(y|X = \alpha_k)$ is

$$f(y|X = \alpha_k) = \frac{y^{k-1}\mu^{-k}\exp(-yk/\mu)}{\Gamma(k)k^{-k}}, \quad \text{for } y \geq 0. \tag{7.42}$$

Here, $\Gamma(k)$ is the gamma function. For integer k, $\Gamma(k) = (k-1)!$.

In radar imaging of terrain, we are primarily interested in finding the mean value μ. We note that what is typically done in radar images to determine μ at a given point on the surface is to average several independent observations of the received power from a resolution cell. This is done in order to reduce the effects of "speckle" or "fading" in the radar image [7.3, pp. 586–90]. But we note that in averaging k observations to obtain Y as described in Eq. (7.37), we obtain a random variable with mean

$$\mathrm{E}\{Y\} = \mu \tag{7.43}$$

and variance

$$\mathrm{E}\{(Y-\mu)^2\} = \frac{\mu^2}{k}. \tag{7.44}$$

So as many independent looks are averaged to reduce speckle, Y converges to μ with high probability. Thus, we are interested in determining μ. As we have previously noted, however, this mean value μ is itself a random variable, the random variable Z, which we have assumed to be uniformly distributed on the non-negative interval $[\beta_1, \beta_2]$. Thus, we have that the density of Y conditioned on Z is given by

$$f(y|z, X = \alpha_k) = \frac{y^{k-1} z^{-k} \exp(-yk/z)}{\Gamma(k) k^{-k}}, \quad \text{for } y \geq 0, \tag{7.45}$$

and the joint density of Y and Z is

$$f(y, z|X = \alpha_k) = \frac{y^{k-1} z^{-k} \exp(-yk/z)}{\Gamma(k) k^{-k} (\beta_2 - \beta_1)}, \quad \text{for } y, z \geq 0. \tag{7.46}$$

The density function $f(y|X = \alpha_k)$ is found by integrating Eq. (7.46) with respect

to z over the interval $[\beta_1, \beta_2]$:

$$f(y|X=\alpha_k) = \int_{\beta_1}^{\beta_2} \frac{y^{k-1} z^{-k} \exp(-yk/z)}{\Gamma(k) k^{-k} (\beta_2 - \beta_1)}\, dz. \tag{7.47}$$

For $k = 1$, this integral is given by (see Chapter 7, Appendix)

$$f(y|X=\alpha_k) = \left[\frac{\mathrm{Ei}(-y/\beta_1) - \mathrm{Ei}(-y/\beta_2)}{\beta_2 - \beta_1}\right], \tag{7.48}$$

where $\mathrm{Ei}(x)$ is the *exponential integral*, given by

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{1}{a} e^a\, da.$$

For $k \geq 2$, this integral is given by (see Chapter 7, Appendix)

$$f(y|X=\alpha_k) = \frac{k}{(k-1)(\beta_2 - \beta_1)} \left\{ \exp\left(-\frac{ky}{\beta_2}\right) \sum_{r=0}^{k-2} \left[\frac{(ky/\beta_2)^{k-r-2}}{(k-r-2)!}\right] - \exp\left(-\frac{ky}{\beta_1}\right) \sum_{r=0}^{k-2} \left[\frac{(ky/\beta_1)^{k-r-2}}{(k-r-2)!}\right] \right\}. \tag{7.49}$$

The mean and variance of Y are given by [7.16]

$$\mu_y = \frac{\beta_1 + \beta_2}{2}, \tag{7.50}$$

and

$$\sigma_y^2 = \frac{(\beta_2 - \beta_1)^2}{12} + \frac{\beta_1^2 + \beta_1\beta_2 + \beta_2^2}{3k}. \tag{7.51}$$

We now have the necessary density function to evaluate the mutual information $I(Y; Z|X = \alpha_k)$. First, however, we will examine the relationship between the density function $f(y|z, X = \alpha_k)$ and speckle noise. From Eq. (7.45), we have that the density function $f(y|z, X = \alpha_k)$ of Y conditioned on $Z = z$ and $X = \alpha_k$ is

$$f(y|z, X=\alpha_k) = \frac{y^{k-1} z^{-k} \exp(-yk/z)}{\Gamma(k) k^{-k}}, \quad \text{for } y \geq 0.$$

Now, as pointed out in [7.16], if we define the random variable V for a fixed z by

$$V \stackrel{\text{def}}{=} \frac{2kY}{z}, \tag{7.52}$$

then V is a standard chi-square random variable with $2K$ degrees of freedom [7.18, §4.32]. Hence, it follows that for a fixed $Z = z$, the random variable Y can be written as

$$Y = \left(\frac{V}{2k}\right) \cdot z. \tag{7.53}$$

Thus, we have that the observed Y is the product of the mean reflectivity $Z = z$ of interest and the multiplicative noise term $V/2k$. This multiplicative noise term is a function of the surface roughness on a length scale the order of the wavelength of the transmitted radiation. It has mean value 1 and variance $1/2K$. Because it is a function of small-scale variations within a resolution cell, its value varies from resolution cell to resolution cell and gives the radar image a speckled appearance. Hence, this multiplicative noise is given the name "speckle." The speckle phenomenon is well known and has been studied extensively in the field of coherent optical systems [7.19]. The effects of speckle on the interpretability of radar images has also been investigated [7.20, 7.21].

Having determined the necessary density functions, we now calculate the differential entropies of Eqs. (7.33) and (7.34) needed to determine $I(Y; Z|X = \alpha_k)$. Substituting the density functions of Eqs. (7.45) and (7.46) into Eq. (7.34), we

obtain (see Chapter 7 Appendix):

$$h(Y|Z,X=\alpha_k) = -\int_{\beta_1}^{\beta_2}\int_{0}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} \ln\left[\frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}}\right] dy\,dz$$

$$= \ln\Gamma(k) - \ln k + k - (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r} - \gamma\right]$$

$$+ \frac{1}{\beta_2-\beta_1}\left[\beta_2\ln\beta_2 - \beta_1\ln\beta_1 - \beta_2 + \beta_1\right]. \tag{7.54}$$

Here, γ is Euler's constant ($\gamma = 0.577\ldots$).

Next, we evaluate the differential entropy $h(Y|X=\alpha_k)$. Substitution of the density of Eq. (7.45) into Eq. (7.33) we have

$$h(Y|X=\alpha_k) = \int_{\beta_1}^{\beta_2} f(y|X=\alpha_k)\log f(y|X=\alpha_k)\,dz$$

$$= \int_{\beta_1}^{\beta_2}\left[\frac{k}{(k-1)(\beta_2-\beta_1)}\left\{\exp\left(-\frac{ky}{\beta_2}\right)\sum_{r=0}^{k-2}\left[\frac{(ky/\beta_2)^{k-r-2}}{(k-r-2)!}\right] - \exp\left(-\frac{ky}{\beta_1}\right)\sum_{r=0}^{k-2}\left[\frac{(ky/\beta_1)^{k-r-2}}{(k-r-2)!}\right]\right\}\right]$$

$$\cdot \ln\left[\frac{k}{(k-1)(\beta_2-\beta_1)}\left\{\exp\left(-\frac{ky}{\beta_2}\right)\sum_{r=0}^{k-2}\left[\frac{(ky/\beta_2)^{k-r-2}}{(k-r-2)!}\right] - \exp\left(-\frac{ky}{\beta_1}\right)\sum_{r=0}^{k-2}\left[\frac{(ky/\beta_1)^{k-r-2}}{(k-r-2)!}\right]\right\}\right] dz. \tag{7.55}$$

A closed-form solution for Eq. (7.55) has not been found. Thus, we will use two approximations in order to characterize its behavior. The first approximation we will consider is an upper bound on $h(Y|X=\alpha_k)$. It makes use of the fact that of all continuous random variables with finite variance σ^2, a Gaussian random variable has the largest differential entropy, and this differential entropy, for any Gaussian

random variable with fixed variance σ^2, is $1/2 \log 2\pi e\sigma^2$ [7.11, p. 39]. Using this fact and the result of Eq. (7.51) giving the variance of Y when $X = \alpha_k$, we have the follwing upper bound $h_U^{(k)}(Y)$ on $h(Y|X = \alpha_k)$:

$$h_U^{(k)}(Y) = \frac{1}{2} \ln \left[2\pi e \left(\frac{(\beta_2 - \beta_1)^2}{12} + \frac{\beta_1^2 + \beta_1\beta_2 + \beta_2^2}{3k} \right) \right]. \tag{7.56}$$

Asymptotically, as k becomes large, we have

$$h_U^{(k)}(Y) \overset{k \to \infty}{\longrightarrow} h_L^{(\infty)}(Y),$$

where

$$\begin{aligned} h_U^{(\infty)}(Y) &= \lim_{k \to \infty} h_L^{(k)}(Y) \\ &= \frac{1}{2} \ln 2\pi e \frac{(\beta_2 - \beta_1)^2}{12} \\ &= \ln (\beta_2 - \beta_1) + 0.1765 \quad \text{(nats)}. \end{aligned} \tag{7.57}$$

An approximation, derived by Frost and Shanmugan [7.16], will now be presented. The density function $f(y|X = \alpha_k)$ can be written in terms of the density functions $f(y|z, X = \alpha_k)$ and $f_Z(z)$ as

$$f(y|X = \alpha_k) = \int_{\beta_1}^{\beta_2} f(y|z, X = \alpha_k) f_Z(z)\, dz. \tag{7.58}$$

We note that as k increases, the conditional distribution $f(y|z, X = \alpha_k)$ becomes narrower, centered around z. This can be seen by recalling that

$$\mu_Y = z$$

and

$$\sigma_Y^2 = \frac{z^2}{k} = O\left(\frac{1}{k}\right).$$

In the limit, as k becomes large, we have

$$f(y|z, X = \alpha_k) \stackrel{k\to\infty}{\longrightarrow} \delta(y - z),$$

where $\delta(\cdot)$ is the Dirac delta function.

For even moderate values of k, the density becomes sufficiently narrow such that it is reasonable to assume that in the integral of Eq. (7.58),

$$f(y|z, X = \alpha_k) \approx \delta(y - z). \tag{7.59}$$

Using this approximation in Eq. (7.58), we obtain

$$\begin{aligned} f(y|X = \alpha_k) &\approx \int_{-\infty}^{\infty} \delta(y - z) f_Z(z)\, dz \\ &= \int_{\beta_1}^{\beta_2} \frac{\delta(y - z)}{\beta_2 - \beta_1}\, dz \\ &= \begin{cases} (\beta_2 - \beta_1)^{-1} & \text{, for } y \in [\beta_1, \beta_2]; \\ 0 & \text{, elsewhere.} \end{cases} \end{aligned} \tag{7.60}$$

This being the case, we have that as k becomes large, we have

$$\begin{aligned} h(y|X = \alpha_k) &\approx -\int_{\beta_1}^{\beta_2} (\beta_2 - \beta_1)^{-1} \ln (\beta_2 - \beta_1)^{-1}\, dz \\ &= \ln (\beta_2 - \beta_1). \end{aligned} \tag{7.61}$$

Note that the approximation of Eq. (7.61) is actually a *lower* bound on $h(y|X = \alpha_k)$, since it is equal to $h(Z)$, but Y always has greater uncertainty than Z, since Y is the product of Z and a random variable V. This V is a standard chi-square random variable with $2k$ degrees of freedom, and is statistically independent of Z, as was noted in Eq. (7.52). Thus, we have the lower bound $h_L(y)$ on $h(Y|X = \alpha_k)$ given by

$$h_L(y) = \ln (\beta_2 - \beta_1). \tag{7.62}$$

Frost and Shanmugan [7.16] note that Eq. (7.62) is a good approximation of $h(Y|X=\alpha_k)$, with an error of less than 5 percent for k greater than 4.

We can now develop an upper bound $I_U(Y;Z|X=\alpha_k)$ on $I(Y;Z|X=\alpha_k)$ based on the results of Eqs. (7.54) and (7.56). We can also develop a lower bound $I_L(Y;Z|X=\alpha_k)$ on $I(Y;Z|X=\alpha_k)$ based on Eqs.(7.54) and (7.62). The lower bound $I_L(Y;Z|X=\alpha_k)$ is given by

$$\begin{aligned} I_L(Y;Z|X=\alpha_k) &= h_L(Y) - h(Y|Z,X=\alpha_k) \\ &= \ln(\beta_2-\beta_1) - \ln\Gamma(k) + \ln k - k + (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r} - \gamma\right] \\ &\quad - \frac{1}{\beta_2-\beta_1}\left[\beta_2\ln\beta_2 - \beta_1\ln\beta_1 - \beta_2 + \beta_1\right]. \end{aligned} \tag{7.63}$$

The upper bound $I_U(Y;Z|X=\alpha_k)$ is given by

$$\begin{aligned} I_U(Y;Z|X=\alpha_k) &= h_U^{(k)}(Y) - h(Y|Z,X=\alpha_k) \\ &= \frac{1}{2}\ln\left[2\pi e\left(\frac{(\beta_2-\beta_1)^2}{12} + \frac{\beta_1^2+\beta_1\beta_2+\beta_2^2}{3k}\right)\right] - \ln\Gamma(k) + \ln k - k \\ &\quad + (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r} - \gamma\right] - \frac{1}{\beta_2-\beta_1}\left[\beta_2\ln\beta_2 - \beta_1\ln\beta_1 - \beta_2 + \beta_1\right]. \end{aligned} \tag{7.64}$$

We now plot these upper and lower bounds on the mutual information as a function of k for various intervals $[\beta_1,\beta_2]$. In order to do this, we define the dynamic range

$$D = \frac{\beta_2}{\beta_1}.$$

For the cases we will consider, we will set $\beta_1 = 1$ and vary β_2 to obtain various values of D. The bounds $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ were evaluated numerically for k ranging from 1 to 50 and for D equal to 4, 8, 10, 20, 50, and 100.

The results are plotted in Figs. 7.6 through 7.11.

Examining the plots of $I_L(Y|Z, X = \alpha_k)$ and $I_U(Y|Z, X = \alpha_k)$, we see that the mutual information per pixel of the radar image is very small. In order to see this, consider Table 7.2, which shows the minimum required mutual information $I_{\min}(Y; Z)$ needed to classify Z into one of J equiprobable classes by observation of Y with statistical reliability. Recalling that the lower bound $I_L(Y|Z, X = \alpha_k)$ is a good approximation for $k > 4$ (an error of less than 5 percent), we see that for all of the dynamic ranges D analyzed, there is not sufficient information provided by a single pixel in order to place it reliably in one of two classes until k is approximately 10. As a result, it would appear that any algorithm or procedure that would be used to characterize a region of a radar image would require the use of multiple pixels. In order to illustrate this, consider the radar image shown in Fig. 7.3. This image is made up of approximately 267,000 pixels, each of which is hard-limited such that it is either black or white. Each pixel thus has the potential of conveying one bit or 0.6931 nats of information, although as our results indicate, each probably conveys significantly less (recall that here $k = 4$). Yet several features of the San Francisco Bay area can be distinguished from this image. In particular, note the Golden Gate Bridge appearing as a vertical white line in the upper right-hand corner.

When examining the image as a whole, or even when examining smaller multi-pixel regions of the image, it becomes possible to pick out features and characterize regions in the image. Examining only a single pixel or a small number of pixels, it is difficult to characterize the pixel or a small neighborhood of pixels under consid-

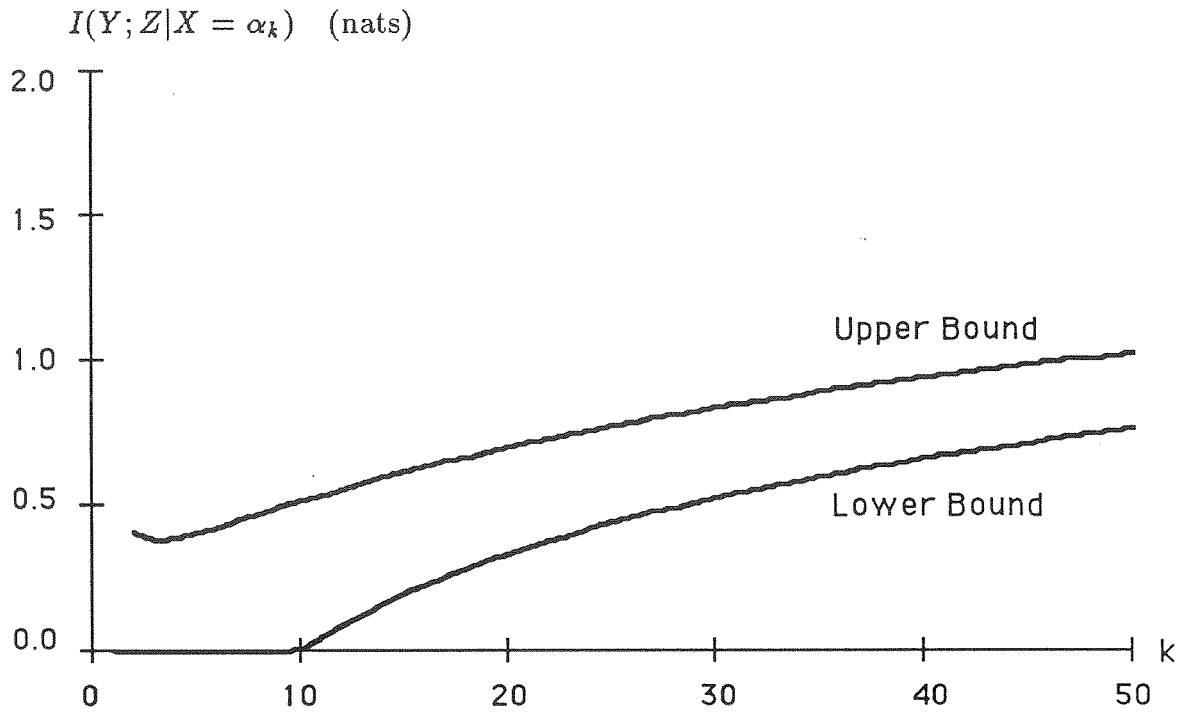


Figure 7.6. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=4$.

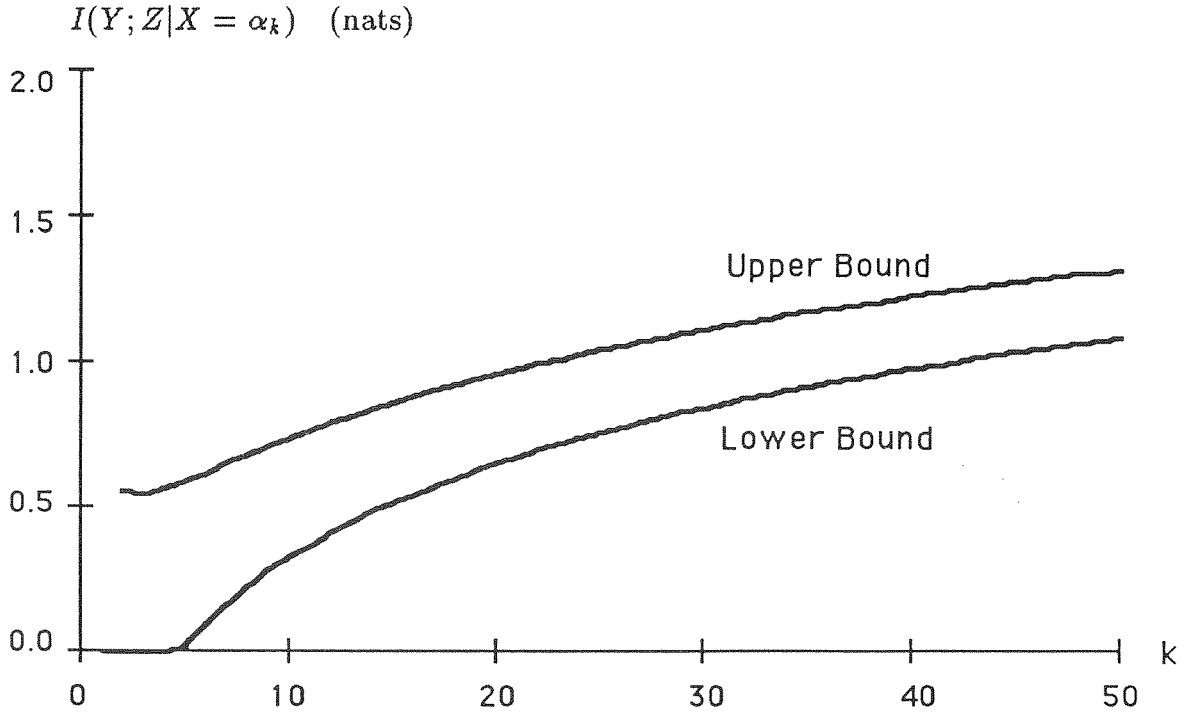


Figure 7.7. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=8$.

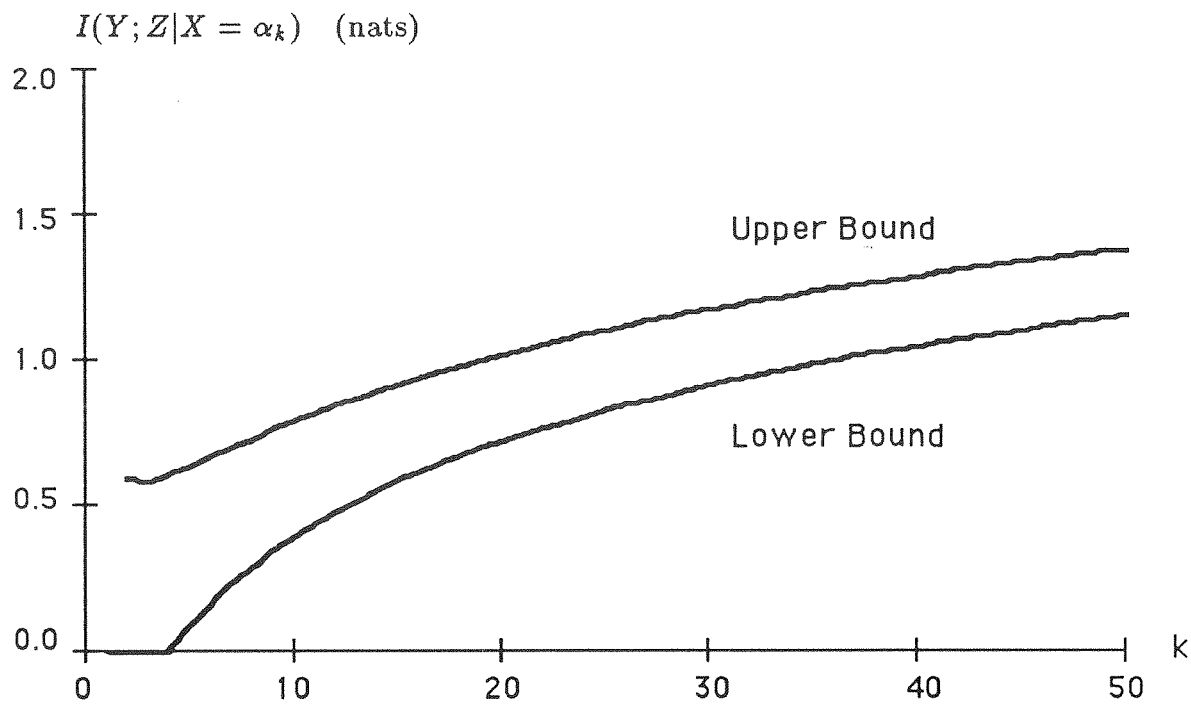


Figure 7.8. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=10$.

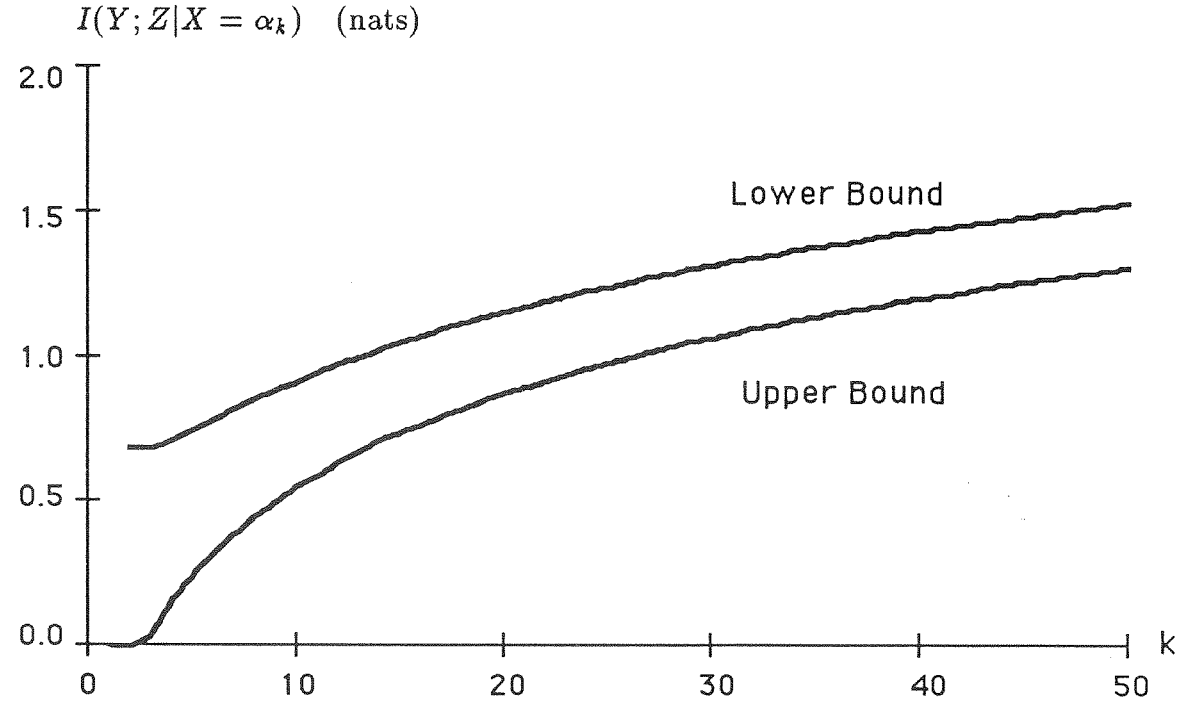


Figure 7.9. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=20$.

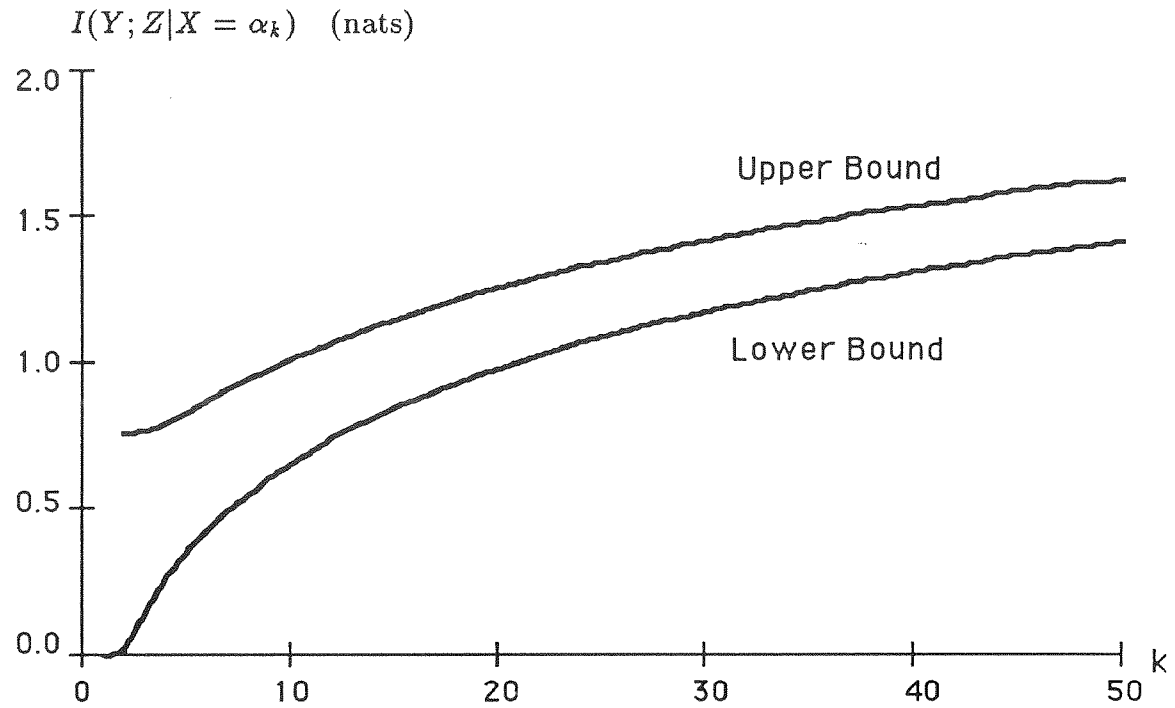


Figure 7.10. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=50$.

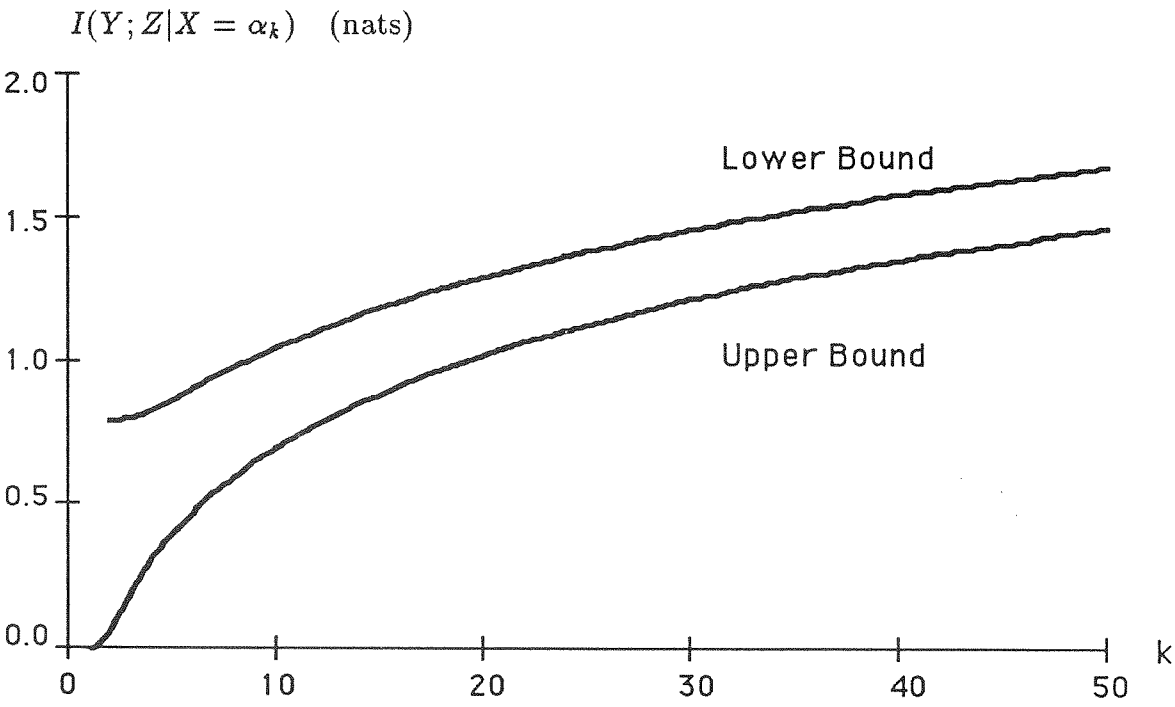


Figure 7.11. $I_L(Y;Z|X=\alpha_k)$ and $I_U(Y;Z|X=\alpha_k)$ versus k for $D=100$.

Table 7-2: Minimum Required Mutual Information for Classification into One of J Equiprobable Classes.

J	Required $I_{\min}(Y;Z)$ (nats)
2	0.6931
3	1.0986
4	1.3863
5	1.6094

eration. This corresponds to the results of our analysis of the information content of radar images, which shows that the information content of a single pixel is small. In addition, as a result of the analysis in Section 7.1, we know that the information contained in a group of pixels is less than or equal to (and most often much less than) the sum of the individual, mutual informations conveyed individually by the pixels. As a result, the classification and characterization of image regions by either human or machine will generally have to be done on an extensive region of pixels, with the extent of the region growing as the number of classes involved increases.

7.3. Chapter 7 Appendix.

7.3.1. Verification of Eq. (7.48).

Eq. (7.48) states that for $k = 1$,

$$f(y|X = \alpha_k) = \left[\frac{\mathrm{Ei}\,(-y/\beta_1) - \mathrm{Ei}\,(-y/\beta_2)}{\beta_2 - \beta_1}\right], \tag{7.48}$$

where $\mathrm{Ei}(x)$ is the *exponential integral function*[7.22, p.93]

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^a}{a}\, da. \tag{A-7.1}$$

In order to see that this is true, we note from Eq. (7.47) that for $k = 1$,

$$f(y|X = \alpha_1) = \int_{\beta_1}^{\beta_2} \frac{z^{-1}\exp\,(-y/z)}{(\beta_2 - \beta_1)}\, dz. \tag{A-7.2}$$

Making the change of variable $z = 1/w$, Eq. (A-7.2) becomes

$$\begin{aligned} f(y|X = \alpha_1) &= -\int_{1/\beta_1}^{1/\beta_2} \frac{w\exp\,(-yw)}{(\beta_2 - \beta_1)}\,\frac{dw}{w^2} \\ &= \int_{1/\beta_2}^{1/\beta_1} \frac{\exp\,(-yw)}{w(\beta_2 - \beta_1)}\, dw. \end{aligned} \tag{A-7.3}$$

Now by a change of variable in Eq. (A-7.1), we have

$$\int_{-\infty}^{x} \frac{e^{ba}}{a}\, da = \mathrm{Ei}(bx). \tag{A-7.4}$$

Applying the relation given by Eq. (A-7.4) to Eq. (A-7.3), we get

$$\begin{aligned} f(y|X = \alpha_1) &= \frac{1}{(\beta_2 - \beta_1)} \int_{1/\beta_2}^{1/\beta_1} \frac{\exp\,(-yw)}{w}\, dw \\ &= \frac{1}{(\beta_2 - \beta_1)} \mathrm{Ei}(-yw)\Big|_{1/\beta_2}^{1/\beta_1} \\ &= \left[\frac{\mathrm{Ei}\,(-y/\beta_1) - \mathrm{Ei}\,(-y/\beta_2)}{\beta_2 - \beta_1}\right]. \end{aligned} \tag{A-7.5}$$

This is the result stated in Eq. (7.48).

7.3.2. Verification of Eq. (7.49).

Eq. (7.49) states that for $k \geq 2$,

$$f(y|X=\alpha_k) = \frac{k}{(k-1)(\beta_2-\beta_1)}\left\{\exp\left(-\frac{ky}{\beta_2}\right)\sum_{r=0}^{k-2}\left[\frac{(ky/\beta_2)^{k-r-2}}{(k-r-2)!}\right] - \exp\left(-\frac{ky}{\beta_1}\right)\sum_{r=0}^{k-2}\left[\frac{(ky/\beta_1)^{k-r-2}}{(k-r-2)!}\right]\right\}. \quad (7.49)$$

In order to see that this is true, we note from Eq. (7.47) that

$$f(y|X=\alpha_k) = \int_{\beta_1}^{\beta_2} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)}\,dz. \quad (A\text{-}7.6)$$

Making the change of variable $z = 1/w$, we have

$$\begin{aligned} f(y|X=\alpha_k) &= -\int_{1/\beta_1}^{1/\beta_2} \frac{y^{k-1}w^k k^k \exp(-kyw)}{\Gamma(k)(\beta_2-\beta_1)}\,\frac{dw}{w^2} \\ &= \frac{-y^{k-1}k^k}{(k-1)!\,(\beta_2-\beta_1)}\int_{1/\beta_1}^{1/\beta_2} w^{k-2}\exp(-kyw)\,dw. \end{aligned} \quad (A\text{-}7.7)$$

From Reference [7.22, p. 92], we have that for any integer $n \geq 0$,

$$\int x^n e^{ax}\,dx = e^{ax}\left(\frac{x^n}{a} + \sum_{r=1}^{n}(-1)^r \frac{n(n-1)(n-2)\cdots(n-r+1)}{a^{r+1}}x^{n-r}\right). \quad (A\text{-}7.8)$$

Hence, for $k \geq 2$, Eq. (A-7.7) can be written as

$$
\begin{aligned}
f(y|X=\alpha_k) &= \frac{-y^{k-1}k^k}{(k-1)!\,(\beta_2-\beta_1)} \exp(-kyw) \\
&\qquad \times \left[\frac{w^{k-2}}{-ky} + \sum_{r=1}^{k-2}(-1)^r \frac{(k-2)\cdots(k-1-1)}{(-ky)^{r+1}} w^{k-r-2}\right]\Bigg|_{1/\beta_1}^{1/\beta_2} \\
&= \frac{y^{k-1}k^k}{(k-1)!\,(\beta_2-\beta_1)} \exp(-kyw) \\
&\qquad \times \left[\frac{w^{k-2}}{ky} + \sum_{r=1}^{k-2} \frac{(k-2)\cdots(k-r-1)}{(ky)^{r+1}} w^{k-r-2}\right]\Bigg|_{1/\beta_1}^{1/\beta_2} \\
&= \frac{y^{k-1}k^k}{(k-1)(\beta_2-\beta_1)} \exp(-kyw) \left[\frac{w^{k-2}}{ky(k-2)!} + \sum_{r=1}^{k-2} \frac{w^{k-r-2}}{(ky)^{r+1}(k-r-2)!}\right]\Bigg|_{1/\beta_1}^{1/\beta_2} \\
&= \frac{y^{k-1}k^k}{(k-1)(\beta_2-\beta_1)} \exp(-kyw) \left[\sum_{r=0}^{k-2} \frac{w^{k-r-2}}{(ky)^{r+1}(k-r-2)!}\right]\Bigg|_{1/\beta_1}^{1/\beta_2} \\
&= \frac{k}{(k-1)(\beta_2-\beta_1)} \exp(-kyw) \left[\sum_{r=0}^{k-2} \frac{(kyw)^{k-r-2}}{(k-r-2)!}\right]\Bigg|_{1/\beta_1}^{1/\beta_2} \\
&= \frac{k}{(k-1)(\beta_2-\beta_1)} \left\{ \exp\left(-\frac{ky}{\beta_2}\right) \sum_{r=0}^{k-2} \left[\frac{(ky/\beta_2)^{k-r-2}}{(k-r-2)!}\right] \right. \\
&\qquad \left. - \exp\left(-\frac{ky}{\beta_1}\right) \sum_{r=0}^{k-2} \left[\frac{(ky/\beta_1)^{k-r-2}}{(k-r-2)!}\right] \right\}.
\end{aligned}
\tag{A-7.9}
$$

This is the result stated in Eq. (7.49).

7.3.3. Verification of Eq. (7.54).

Eq. (7.54) states that

$$h(Y|Z,X=\alpha_k) = -\int_{\beta_1}^{\beta_2}\int_{-\infty}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} \ln\left[\frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}}\right] dy\,dz$$

$$= \ln\Gamma(k) - \ln k + k - (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r} - \gamma\right]$$

$$+ \frac{1}{\beta_2-\beta_1}\left[\beta_2\ln\beta_2 - \beta_1\ln\beta_1 - \beta_2 + \beta_1\right]. \tag{7.54}$$

Here γ is Euler's constant ($\gamma = 0.577\ldots$). In order to verify that Eq. (7.54) is true, we note that

$$h(Y|Z,X=\alpha_k) = -\int_{\beta_1}^{\beta_2}\int_{0}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} \ln\left[\frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}}\right] dy\,dz$$

$$= -\int_{\beta_1}^{\beta_2}\int_{0}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} \left\{\begin{array}{c}(k-1)\ln y - k\ln z \\ -yk/z - \ln\Gamma(k) + k\ln k\end{array}\right\} dy\,dz$$

$$= \ln\Gamma(k) - k\ln k - \int_{\beta_1}^{\beta_2}\int_{0}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)}(k-1)\ln y\,dy\,dz$$

$$+ \int_{\beta_1}^{\beta_2}\int_{0}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} k\ln z\,dy\,dz$$

$$+ \int_{\beta_1}^{\beta_2}\int_{0}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} \frac{yk}{z}\,dy\,dz. \tag{A-7.10}$$

As we can see from Eq. (A-7.10), we must evaluate three integrals in order to

obtain Eq. (7.54).The first of these is evaluated as follows:

$$\int_0^\infty \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)}(k-1)\ln y\,dy\,dz$$

$$= \frac{(k-1)}{(\beta_2-\beta_1)}\int_{\beta_1}^{\beta_2}\int_0^\infty \frac{y^{k-1}z^{-k}\exp(-yk/z)\ln y}{\Gamma(k)k^{-k}}\,dy\,dz. \tag{A-7.11}$$

From [7.22, §4.352.2], we have that for $\mu > 0$

$$\int_0^\infty x^n e^{-\mu x}\ln x\,dx = \frac{n!}{\mu^{n+1}}\left[1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}-\gamma-\ln\mu\right]. \tag{A-7.12}$$

Here γ is Euler's constant. Applying this result to Eq. (A-7.11), we have

$$\int_{\beta_1}^{\beta_2}\int_0^\infty \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)}(k-1)\ln y\,dy\,dz$$

$$= \frac{(k-1)}{(\beta_2-\beta_1)\Gamma(k)}\int_{\beta_1}^{\beta_2}\frac{z^{-k}}{k^{-k}}\int_0^\infty y^{k-1}\exp\left[-\left(\frac{k}{z}\right)y\right]\ln y\,dy\,dz$$

$$= \frac{(k-1)}{(\beta_2-\beta_1)}\int_{\beta_1}^{\beta_2}\left\{\sum_{r=1}^{k-1}\frac{1}{r}-\gamma-\ln k+\ln z\right\}dz$$

$$= (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r}-\gamma-\ln k\right]+\frac{(k-1)}{(\beta_2-\beta_1)}\int_{\beta_1}^{\beta_2}\ln z\,dz$$

$$= (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r}-\gamma-\ln k\right]+\frac{(k-1)}{(\beta_2-\beta_1)}\left[z\ln z\Big|_{\beta_1}^{\beta_2}-\int_{\beta_1}^{\beta_2}dz\right]$$

$$= (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r}-\gamma-\ln k\right]+\frac{(k-1)}{(\beta_2-\beta_1)}\left[\beta_2\ln\beta_2-\beta_1\ln\beta_1-\beta_2+\beta_1\right]. \tag{A-7.13}$$

The second integral of Eq. (A-7.10) can be evaluated as follows:

$$\int_{\beta_1}^{\beta_2}\int_{0}^{\infty}\frac{y^{k-1}z^{-k}\exp\left(-yk/z\right)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)}k\ln z\,dy\,dz$$

$$\begin{aligned}
&= \int_{\beta_1}^{\beta_2}\frac{k\ln z}{(\beta_2-\beta_1)}\int_{0}^{\infty}\frac{y^{k-1}z^{-k}\exp\left(-yk/z\right)}{\Gamma(k)k^{-k}}\,dy\,dz \\
&= \int_{\beta_1}^{\beta_2}\frac{k\ln z}{(\beta_2-\beta_1)}\,dz \\
&= \frac{k}{(\beta_2-\beta_1)}\int_{\beta_1}^{\beta_2}\ln z\,dz \\
&= \frac{k}{(\beta_2-\beta_1)}\left[z\ln z\Big|_{\beta_1}^{\beta_2}-\int_{\beta_1}^{\beta_2}dz\right] \\
&= \frac{k}{(\beta_2-\beta_1)}\left[\beta_2\ln\beta_2-\beta_1\ln\beta_1-\beta_2+\beta_1\right].
\end{aligned} \tag{A-7.14}$$

The third integral of Eq. (A-7.10) can be evaluated as follows:

$$\begin{aligned}
&\int_{\beta_1}^{\beta_2} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)}\left(\frac{yk}{z}\right)dy\,dz \\
&\qquad = \int_{\beta_1}^{\beta_2} \frac{k}{z(\beta_2-\beta_1)} \int_0^{\infty} \frac{y^k z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}}\,dy\,dz \\
&\qquad = \int_{\beta_1}^{\beta_2} \frac{k}{z(\beta_2-\beta_1)} \left[\int_0^{\infty} \frac{w^k\exp(-w)}{\Gamma(k)}\cdot\frac{z}{k}\,dw\right] dz \\
&\qquad = \int_{\beta_1}^{\beta_2} \frac{k}{(\beta_2-\beta_1)k\Gamma(k)} \int_0^{\infty} w^k\exp(-w)\,dw\,dz \\
&\qquad = \int_{\beta_1}^{\beta_2} \frac{k}{(\beta_2-\beta_1)\Gamma(k+1)}\cdot\Gamma(k+1)\,dz \\
&\qquad = k\int_{\beta_1}^{\beta_2} \frac{dz}{\beta_2-\beta_1} = k.
\end{aligned} \tag{A-7.15}$$

Substituting the results of Eqs. (A-7.13), (A-7.14), and (A-7.15) into Eq.(A-7.10), we have

$$\begin{aligned}
h(Y|Z,X=\alpha_k) &= -\int_{\beta_1}^{\beta_2}\int_{-\infty}^{\infty} \frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}(\beta_2-\beta_1)} \ln\left[\frac{y^{k-1}z^{-k}\exp(-yk/z)}{\Gamma(k)k^{-k}}\right] dy\,dz \\
&= \ln\Gamma(k) - \ln k + k - (k-1)\left[\sum_{r=1}^{k-1}\frac{1}{r} - \gamma\right] \\
&\quad + \frac{1}{\beta_2-\beta_1}\left[\beta_2\ln\beta_2 - \beta_1\ln\beta_1 - \beta_2 + \beta_1\right].
\end{aligned}$$

This is the result stated in Eq. (7.54).

7.4. Chapter 7 References.

7.1 Elachi, C., *Physics and Techniques of Remote Sensing*, John Wiley and Sons, New York, NY, 1987.

7.2 Mensa, D. L., *High Resolution Radar Imaging*, Artech House, Dedham, MA, 1981.

7.3 Ullaby, F. T., R. K. Moore, and A. K. Fung, *Microwave Remote Sensing*, vol. 2, Addison-Wesley, Reading, MA, 1982.

7.4 Stimson, G. W., *Introduction to Airborne Radar*, Hughes Aircraft Company, El Segundo, CA, 1983.

7.5 Harger, R. O., *Synthetic Aperture Radar Systems, Theory and Design*, Academic Press, New York, NY, 1970.

7.6 Hovanessian, S. A., *Introduction to Synthetic Array and Imaging Radars*, Artech House, Dedham, MA, 1980.

7.7 Ullaby, F. T., R. K. Moore, and A. K. Fung, *Microwave Remote Sensing*, vol. 3, Artech House, Dedham, MA, 1986.

7.8 Pierce, J. R. and E. C. Posner, *Introduction to Communication Science and Systems*, Plenum, New York, NY, 1980.

7.9 Feller, W., *An Introduction to Probability Theory and Its Applications*, 3d. ed., John Wiley and Sons, New York, NY, 1968.

7.10 Shannon, C. E., "A Mathematical Theory of Communication," *Bell Sys. Tech. J.* 27, 1948. 379-423, 623-656. Reprinted in C. E. Shannon and W. W. Weaver, *The Mathematical Theory of Communication*, Univ. Ill. Press, Urbana, IL, 1949.

7.11 McEliece, R. J., *The Theory of Information and Coding*, Addison-Wesley, Reading, MA, 1977.

7.12 Pratt, W. K., *Digital Image Processing*, John Wiley and Sons, New York, NY, 1978.

7.13 Long, M. W., *Radar Reflectivity of Land and Sea*, Artech House, Dedham, MA, 1983.

7.14 Duda, R. O. and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, NY, 1973.

7.15 Schreiber, W. F., "The Measurement of Third Order Probability Distributions of Television Signals," *IRE Trans. Inf. Theory*, **IT-2**, 3, September 1956, 94–105.

7.16 Frost, V. S. and K. S. Shanmugan, "The Information Content of Synthetic Aperture Radar Images of Terrain," *IEEE Trans. on Aero. and Elec. Systems*, **AES-19**, 5, September 1983, 768-74.

7.17 Hogg, R. V. and A. T. Craig, *Introduction to Mathematical Statistics*, Macmillan, New York, NY, 1970.

7.18 Manoukian, E. B., *Modern Concepts and Theorems of Mathematical Statistics*, Springer-Verlag, New York, NY, 1986.

7.19 *Laser Speckle and Related Phenomenon*, 2d. ed., ed. J. C. Dainty, Springer-Verlag, New York, NY, 1984.

7.20 Korwar, V. N. and J. R. Pierce, "Detection of gratings and small features in images corrupted by speckle," *Applied Optics*, Vol. 20, No. 2, January, 1981, 312–19.

7.21 Korwar, V. N. and J. R. Pierce, "Discrimination of form in images corrupted by speckle," *Applied Optics*, Vol. 20, No. 2, January, 1981, 320–25.

7.22 Gradshteyn, I. S., and I. M. Ryzhik, *Table of Integrals, Series, and Products*, corrected and enlarged ed., Academic Press, New York, NY, 1980.

CHAPTER 8

SUMMARY AND CONCLUSIONS

In this thesis, we have investigated the use of information theory in the analysis and design of radar systems. Our motivation for such an investigation was taken from the similarities between radar systems and communications systems and the great success of information theory in the design and analysis of communications systems.

We examined the use of the mutual information between the radar target and the received radar signal as a measure of radar performance. In particular, we considered that for problems in which the radar system was being used for target identification or target parameter measurement, mutual information was an appropriate measure of radar system performance, in addition to more common radar performance measures such as signal-to-noise ratio and probability of detection. This is due to the fact that the mutual information between the target and the received signal determines the maximum number of equiprobable classes N into which a radar target can be classified based on observation of the received radar signal.We showed that if this mutual information, in nats, is I_0, then

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

In addition, we noted that the rate distortion function $R(\delta)$, which is defined in terms of mutual information and an average distortion measure, could be used to

determine the minimum required mutual information between the target and the received radar waveform in order to obtain an average error or distortion less than some specified value δ in the radar measurement.

We then defined and solved the *Radar/Information Theory Problem*, which gave us a mathematical framework for the problem of waveform design for maximizing the mutual information between the target and the received radar waveform. Specifically, we looked at the problem of finding a distribution on an ensemble of transmitted waveforms that maximized the mutual information between the target and the received waveform. We solved the Radar/Information Theory Problem for a number of general target types, both discrete and continuous. We also showed that a deterministic solution also exists; that is, that from the family of distributions of transmitted waveforms which maximize the mutual information between target and received signal, there is a distribution that corresponds to sending a waveform or sequence of waveforms with certainty.

Next, we examined statistical electromagnetic scattering models that would allow us to apply the results of the Radar/Information Theory Problem to practical radar problems. We also introduced the notion of target impulse response as a description of linear time-invariant electromagnetic scattering.

In Chapter 5, we digressed from our information-theoretic analysis of radar systems to apply the target impulse response to the problem of designing realizable *waveform/receiver filter pairs* that maximize the signal-to-noise ratio at the receiver output when a target is present under contstraints on bandwidth and waveform

energy. This new result is an extension of North's matched filter, which provides the maximum signal-to-noise ratio for point targets but not for extended targets. The resulting waveform/ receiver-filter pair allows for the optimal detection of time-invariant, extended targets in additive noise of arbitrary power spectral density. We developed the following procedure for designing a waveform/receiver-filter pair for a target of known impulse response $h(t)$ in the presence of additive noise with noise power spectral density $S_{nn}(f)$, given the energy constraint that the total energy in the transmitted signal is $\mathcal{E}$ and the constraint that the waveform is confined to the time interval $[-T/2, T/2]$:

1. Compute

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi ft}\, dt.$$

   Here, $h(t)$ is the impulse response of the target and $H(f)$ its Fourier transform.

2. Compute

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

   Here, $S_{nn}(f)$ is the two-sided power spectral density of the noise $n(t)$, and $L(t)$ is the inverse Fourier transform of $|H(f)|^2/S_{nn}(f)$.

3. 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 $\mathcal{E}$. This is the optimum radar detection waveform.

4. Compute the spectrum $\hat{X}(f)$ corresponding to the optimum waveform $\hat{x}(t)$:

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

5. Implement a receiver filter of the form

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

where K is any convenient complex constant and t_0 is the time at which the receiver output is observed.

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

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

We noted that a waveform designed using this procedure in the presence of additive white noise puts as much of the energy as is possible into the mode of the target having the greatest response (largest eigenvalue). In the case of non-white noise, this solution takes into account the relative strength of the noise components in each mode and weights the distribution of energy among the target modes in order to obtain the largest signal-to-noise ratio.

Next, we addressed the problem of designing radar waveforms that provide maximum mutual information between a random target and the received radar waveform in the presence of additive Gaussian noise of arbitrary power spectral

density. This was done by extending the concept of target impulse response to include random scattering. Random scattering was incorporated by considering finite-energy, random processes as random impulse responses and using them to describe the response from random scatterers. We then considered the case of targets described by the finite-energy, Gaussian, random process $\mathbf{g}(t)$ with Fourier transform $\mathrm{G}(f)$ having spectral variance $\sigma_G^2(f)$. We showed that the waveforms $x(t)$ that have energy $\mathcal{E}$ and that maximize the mutual information $I(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ with bandwidth concentrated in a frequency interval $\mathcal{W}$ and a time interval $[-T/2,T/2]$ have a magnitude-squared spectrum

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

Here, $\widetilde{T}$ is the receiver observation time, which for most waveforms of interest satisfies $\widetilde{T} \approx T$; P_{nn} is the one-sided power spectral density of the additive Gaussian noise; and A is a constant found by solving the following relation for A:

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

The resulting maximum mutual information $I_{\max}(\mathbf{y}(t);\mathbf{g}(t)|x(t))$ was found to be

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

We noted that in the case of additive white Gaussian noise, these waveforms, in contrast to the waveforms of Chapter 5 for optimal detection, which attempt to put as much energy as possible into the largest target mode, distribute energy among the target modes in order to maximize the mutual information between the target

and the received signal. It is not surprising, then, that the form of the solution is different from that for optimal detection, as modes besides the largest may carry significant information about the target. As a result, we would expect energy to be allotted to them for their measurement.

Finally, we examined some information-theoretic characteristics of radar images. We did this by first considering the general information-theoretic properties of images in general and radar images in particular, and then examining the information per pixel of a radar image resulting from homogeneous diffuse terrain. We saw that on a per-pixel basis, a very small amount of information is provided in radar-imaging measurements for such a surface. As a result, we noted that multiple pixel regions have to be examined in order to classify or characterize such a surface based on its radar measurement. This was noted to be in agreement with heuristic results on surface characterization based on radar images.