

Fundamentals of Radar Signal Processing

Mark A. Richards





Fundamentals of Radar Signal Processing

Fundamentals of Radar Signal Processing

Mark A. Richards, Ph.D.

Georgia Institute of Technology

Second Edition



New York Chicago San Francisco Athens London Madrid Mexico City Milan New Delhi Singapore Sydney Toronto Copyright © 2014 by McGraw-Hill Education. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-179833-4

MHID: 0-07-179833-1

e-Book conversion by Cenveo[®] Publisher Services

Version 1.0

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-179832-7, MHID: 0-07-179832-3.

McGraw-Hill Education eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative, please visit the Contact Us page at <u>www.mhprofessional.com</u>.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

Information has been obtained by McGraw-Hill Education from sources believed to be reliable. However, because of the possibility of human or mechanical error by our sources, McGraw-Hill Education, or others, McGraw-Hill Education does not guarantee the accuracy, adequacy, or completeness of any information and is not responsible for any errors or omissions or the results obtained from the use of such information.

TERMS OF USE

This is a copyrighted work and McGraw-Hill Education and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill Education's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL EDUCATION AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill Education and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill Education nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill Education has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill Education and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

To Theresa, with love and gratitude for all our time together, past and future. And to Jessica and Benjamin and the first of our newest generation, Amelia, with pride, joy, and high hopes for the future you all will bring.

About the Author

MARK A. RICHARDS, PH.D., is Principal Research Engineer and Adjunct Professor, School of Electrical and Computer Engineering (ECE), Georgia Institute of Technology, and a Fellow of the IEEE cited for contributions to radar signal processing education. He is engaged in academic and continuing education teaching and course development in the fields of digital signal processing and radar signal processing. Previous positions include Chief of the Radar Systems Division and Head of the Signal Processing Branch in the Sensors and Electromagnetic Applications Laboratory of the Georgia Tech Research Institute, and Program Manager for Advanced Signal Processing at the U.S. Defense Advanced Research Projects Agency (DARPA).

Contents

Preface Acknowledgments Symbols Acronyms

<u>1</u> Introduction to Radar Systems and Signal Processing

- 1.1 History and Applications of Radar
- 1.2 Basic Radar Functions
- 1.3 Elements of a Pulsed Radar
 - 1.3.1 Transmitter and Waveform Generator
 - 1.3.2 Antennas
 - 1.3.3 Receivers
- 1.4 Common Threads in Radar Signal Processing
 - 1.4.1 Signal-to-Interference Ratio and Integration
 - 1.4.2 Resolution
 - 1.4.3 Data Integration and Phase History Modeling
 - 1.4.4 Bandwidth Expansion
- 1.5 A Preview of Basic Radar Signal Processing
 - 1.5.1 Radar Time Scales
 - 1.5.2 Phenomenology
 - 1.5.3 Signal Conditioning and Interference Suppression
 - 1.5.4 Imaging
 - 1.5.5 Detection
 - 1.5.6 Measurements and Track Filtering
- 1.6 Radar Literature
 - 1.6.1 Radar Systems and Components
 - 1.6.2 Basic Radar Signal Processing
 - 1.6.3 Advanced Radar Signal Processing
 - <u>1.6.4 Radar Applications</u>
 - 1.6.5 Current Radar Research

<u>References</u>

Problems

2 Signal Models

- 2.1 Components of a Radar Signal
- 2.2 Amplitude
 - 2.2.1 Simple Point Target Radar Range Equation
 - 2.2.2 Distributed Target Forms of the Range Equation
 - 2.2.3 Radar Cross Section
 - 2.2.4 Radar Cross Section for Meteorological Targets

- 2.2.5 Statistical Description of Radar Cross Section
- 2.2.6 Target Fluctuation Models
- 2.2.7 Swerling Models
- 2.2.8 Effect of Target Fluctuations on Doppler Spectrum
- 2.3 Clutter
 - <u>2.3.1 Behavior of $\sigma^{\underline{0}}$ </u>
 - 2.3.2 Signal-to-Clutter Ratio
 - 2.3.3 Temporal and Spatial Correlation of Clutter
 - 2.3.4 Compound Models of Radar Cross Section
- 2.4 Noise Model and Signal-to-Noise Ratio
- 2.5 Jamming
- 2.6 Frequency Models: The Doppler Shift
 - 2.6.1 Doppler Shift
 - 2.6.2 The Stop-and-Hop Approximation and Phase History
 - 2.6.3 Measuring Doppler Shift: Spatial Doppler
- 2.7 Spatial Models
 - 2.7.1 Coherent Scattering
 - 2.7.2 Variation with Angle
 - 2.7.3 Variation with Range
 - 2.7.4 Noncoherent Scattering
 - 2.7.5 Projections
 - 2.7.6 Multipath
- 2.8 Spectral Model
- 2.9 Summary
- <u>References</u>
- Problems

3 Pulsed Radar Data Acquisition

- 3.1 Acquiring and Organizing Pulsed Radar Data
 - 3.1.1 One Pulse: Fast Time
 - 3.1.2 Multiple Pulses: Slow Time and the CPI
 - 3.1.3 Doppler and Range Ambiguities
 - 3.1.4 Multiple Channels: The Datacube
 - <u>3.1.5 Dwells</u>
- 3.2 Sampling the Doppler Spectrum
 - 3.2.1 The Nyquist Rate in Doppler
 - 3.2.2 Straddle Loss
- 3.3 Sampling in the Spatial and Angle Dimensions
 - 3.3.1 Spatial Array Sampling
 - 3.3.2 Sampling in Angle
- 3.4 I/Q Imbalance and Digital I/Q
 - 3.4.1 I/Q Imbalance and Offset
 - 3.4.2 Correcting I/Q Errors

3.4.3 Digital I/Q

References Problems

4 Radar Waveforms

- <u>4.1 Introduction</u>
- 4.2 The Waveform Matched Filter
 - 4.2.1 The Matched Filter
 - 4.2.2 Matched Filter for the Simple Pulse
 - 4.2.3 All-Range Matched Filtering
 - 4.2.4 Straddle Loss
 - 4.2.5 Range Resolution of the Matched Filter
- 4.3 Matched Filtering of Moving Targets
- 4.4 The Ambiguity Function
 - 4.4.1 Definition and Properties of the Ambiguity Function
 - 4.4.2 Ambiguity Function of the Simple Pulse
- 4.5 The Pulse Burst Waveform
 - 4.5.1 Matched Filter for the Pulse Burst Waveform
 - 4.5.2 Pulse-by-Pulse Processing
 - 4.5.3 Range Ambiguity
 - 4.5.4 Doppler Response of the Pulse Burst Waveform
 - 4.5.5 Ambiguity Function for the Pulse Burst Waveform
 - 4.5.6 The Slow-Time Spectrum and the Periodic Ambiguity Function
- 4.6 Frequency-Modulated Pulse Compression Waveforms
 - 4.6.1 Linear Frequency Modulation
 - 4.6.2 The Principle of Stationary Phase
 - 4.6.3 Ambiguity Function of the LFM Waveform
 - 4.6.4 Range-Doppler Coupling
 - 4.6.5 Stretch Processing
- 4.7 Range Sidelobe Control for FM Waveforms
 - 4.7.1 Matched Filter Frequency Response Shaping
 - 4.7.2 Matched Filter Impulse Response Shaping
 - 4.7.3 Waveform Spectrum Shaping
- 4.8 The Stepped Frequency Waveform
- 4.9 The Stepped Chirp Waveform
- 4.10 Phase-Modulated Pulse Compression Waveforms
 - 4.10.1 Biphase Codes
 - 4.10.2 Polyphase Codes
 - 4.10.3 Mismatched Phase Code Filters
- 4.11 Costas Frequency Codes
- 4.12 Continuous Wave Radar

<u>References</u>

Problems

5 Doppler Processing

- 5.1 Moving Platform Effects on the Doppler Spectrum
- 5.2 Moving Target Indication
 - 5.2.1 Pulse Cancellers
 - 5.2.2 Vector Formulation of the Matched Filter
 - 5.2.3 Matched Filters for Clutter Suppression
 - 5.2.4 Blind Speeds and Staggered PRFs
 - 5.2.5 MTI Figures of Merit
 - 5.2.6 Limitations to MTI Performance
- 5.3 Pulse Doppler Processing
 - 5.3.1 The Discrete-Time Fourier Transform of a Moving Target
 - 5.3.2 Sampling the DTFT: The Discrete Fourier Transform
 - 5.3.3 The DFT of Noise
 - 5.3.4 Pulse Doppler Processing Gain
 - 5.3.5 Matched Filter and Filterbank Interpretations of Pulse Doppler Processing with the DF T
 - 5.3.6 Fine Doppler Estimation
 - 5.3.7 Modern Spectral Estimation in Pulse Doppler Processing
 - 5.3.8 CPI-to-CPI Stagger and Blind Zone Maps
- 5.4 Pulse Pair Processing
- 5.5 Additional Doppler Processing Issues
 - 5.5.1 Combined MTI and Pulse Doppler Processing
 - 5.5.2 Transient Effects
 - 5.5.3 PRF Regimes
 - 5.5.4 Ambiguity Resolution
- 5.6 Clutter Mapping and the Moving Target Detector
 - 5.6.1 Clutter Mapping
 - 5.6.2 The Moving Target Detector
- 5.7 MTI for Moving Platforms: Adaptive Displaced Phase Center Antenna Processing
 - 5.7.1 The DPCA Concept
 - 5.7.2 Adaptive DPCA

<u>References</u>

Problems

6 Detection Fundamentals

- 6.1 Radar Detection as Hypothesis Testing
 - 6.1.1 The Neyman-Pearson Detection Rule
 - 6.1.2 The Likelihood Ratio Test
- 6.2 Threshold Detection in Coherent Systems
 - 6.2.1 The Gaussian Case for Coherent Receivers

- 6.2.2 Unknown Parameters and Threshold Detection
- 6.2.3 Linear and Square-Law Detectors
- 6.2.4 Other Unknown Parameters
- 6.3 Threshold Detection of Radar Signals
 - 6.3.1 Coherent, Noncoherent, and Binary Integration
 - 6.3.2 Nonfluctuating Targets
 - 6.3.3 Albersheim's Equation
 - 6.3.4 Fluctuating Targets
 - 6.3.5 Shnidman's Equation
- 6.4 Binary Integration
- 6.5 Constant False Alarm Rate Detection
 - 6.5.1 The Effect of Unknown Interference Power on False Alarm Probability
 - 6.5.2 Cell-Averaging CFAR
 - 6.5.3 Analysis of Cell-Averaging CFAR
 - 6.5.4 CA CFAR Limitations
 - 6.5.5 Extensions to Cell-Averaging CFAR
 - 6.5.6 Order Statistic CFAR
 - 6.5.7 Additional CFAR Topics
- 6.6 System-Level Control of False Alarms

References

Problems

7 Measurements and Tracking

- 7.1 Estimators
 - 7.1.1 Estimator Properties
 - 7.1.2 The Cramer-Rao Lower Bound
 - 7.1.3 The CRLB and Signal-to-Noise Ratio
 - 7.1.4 Maximum Likelihood Estimators
- 7.2 Range, Doppler, and Angle Estimators
 - 7.2.1 Range Estimators
 - 7.2.2 Doppler Signal Estimators
 - 7.2.3 Angle Estimators
- 7.3 Introduction to Tracking
 - 7.3.1 Sequential Least Squares Estimation
 - <u>7.3.2 The $\alpha -\beta$ Filter</u>
 - 7.3.3 The Kalman Filter
 - 7.3.4 The Tracking Cycle

<u>References</u>

Problems

<u>8 Introduction to Synthetic Aperture Imaging</u>

8.1 Introduction to SAR Fundamentals

- 8.1.1 Cross-Range Resolution in Radar
- 8.1.2 The Synthetic Aperture Viewpoint
- 8.1.3 Doppler Viewpoint
- 8.1.4 SAR Coverage and Sampling
- 8.2 Stripmap SAR Data Characteristics
 - 8.2.1 Stripmap SAR Geometry
 - 8.2.2 Stripmap SAR Data Set
- 8.3 Stripmap SAR Image Formation Algorithms
 - 8.3.1 Doppler Beam Sharpening
 - 8.3.2 Quadratic Phase Error Effects
 - 8.3.3 Range-Doppler Algorithms
 - 8.3.4 Depth of Focus
- 8.4 Spotlight SAR Data Characteristics
- 8.5 The Polar Format Image Formation Algorithm for Spotlight SAR
- 8.6 Interferometric SAR
 - 8.6.1 The Effect of Height on a SAR Image
 - 8.6.2 IFSAR Processing Steps
- 8.7 Other Considerations
 - 8.7.1 Motion Compensation and Autofocus
 - 8.7.2 Autofocus
 - 8.7.3 Speckle Reduction

<u>References</u>

Problems

9 Introduction to Beamforming and Space-Time Adaptive Processing

- 9.1 Spatial Filtering
 - 9.1.1 Beamforming
 - 9.1.2 Adaptive Beamforming
 - 9.1.3 Adaptive Beamforming with Preprocessing
- 9.2 Space-Time Signal Environment
- 9.3 Space-Time Signal Modeling
- 9.4 Processing the Space-Time Signal
 - 9.4.1 Optimum Matched Filtering
 - 9.4.2 STAP Metrics
 - 9.4.3 Relation to Displaced Phase Center Antenna Processing
 - 9.4.4 Adaptive Matched Filtering
- 9.5 Reduced-Dimension STAP
- 9.6 Advanced STAP Algorithms and Analysis
- 9.7 Limitations to STAP

<u>References</u>

Problems

A Selected Topics in Probability and Random Processes

B Selected Topics in Digital Signal Processing

<u>Index</u>

Preface

This second edition of *Fundamentals of Radar Signal Processing* shares with the first the goal of providing in-depth coverage of fundamental topics in radar signal processing from a digital signal processing perspective. The techniques and interpretations of linear systems, filtering, sampling, and Fourier analysis are used throughout to provide a modern and unified tutorial approach. The coverage includes a full range of the basic signal processing techniques on which virtually all modern radar systems rely, including topics such as target and interference models, matched filtering, waveform design, Doppler processing, threshold detection, and measurement accuracy. Introductions are provided to track filtering and the advanced topics of synthetic aperture imaging and space-time adaptive array processing to provide a bridge to more in-depth texts on these topics.

The first edition was published in 2005 with the intention of filling what I perceived to be a void in the technical literature on radar. There existed at that time a number of excellent books on radar systems in general (e.g., Skolnik, Edde) that provided an excellent qualitative and descriptive introduction to radar systems as a whole and could be enthusiastically recommended as first texts for anyone interested in the topic. Indeed, having worked on speech enhancement in graduate school, I read the first edition of Skolnik's Introduction to Radar Systems when I accepted a job in radar, hoping to avoid appearing completely ignorant on my first day at the new job. (It didn't work, through no fault of Skolnik.) Some of these texts (e.g., Peebles, Mahafza) provided greater quantitative depth on basic radar systems and some signal processing topics. At the same time, a number of good quality texts were available on advanced topics in radar signal processing, principally synthetic aperture imaging (Jakowatz et al., Carrara et al., Soumekh) and space-time adaptive processing (Klemm, Guerci). The problem, in my view, was the existence of a substantial gap between the qualitative systems books and the quantitative advanced signal processing books. Specifically, I believed the radar community lacked a current text providing a concise, unified, and modern treatment of the basic radar signal processing techniques mentioned above on which these more advanced methods are founded. It was my hope that this book would fill that gap.

The reception accorded the first edition since its publication has been gratifying. I have received many very kind and encouraging comments and it has been adopted for use by a number of universities and companies. I believe it has largely been successful in meeting its goals. Its success, however, also quickly brought to light many ways in which the book could be improved.

New books continue to appear, particularly the excellent Principles of

Modern Radar series. I believe it remains true today, somewhat to my surprise, that most radar textbooks generally address either the entire radar system or very specialized processing topics, and that few attempt to address the full suite of basic signal processing concepts found in virtually every radar that form the basis for advanced techniques. That is, the gap still exists. The goal of the text has therefore remained the same. The specific goal of the second edition is to strengthen that coverage, broaden it slightly, correct and improve the presentation, and provide additional resources that will increase its usefulness as a textbook as well as a professional reference.

This book was originally developed and used over several years in support of two courses at Georgia Tech. It was primarily developed as a product of ECE 6272, *Fundamentals of Radar Signal Processing*, a semesterlength first-year graduate course. Elements of this book were also used in abbreviated and simplified form in the one-week professional education course of the same name taught periodically through Georgia Tech's Professional Education division. Since publication of the first edition, I have continued to use it for both courses. Through those experiences and just the passage of time I have learned more, both about the topics and how to convey them, and I have tried to incorporate that knowledge into the updated text.

There is one major change and many minor to moderate ones from the first edition. The major change is the addition of what is now Chap. 7, "Measurements and Tracking." This chapter introduces an important basic topic missing from the first edition, that of measurement accuracy. The Cramèr-Rao lower bound (CRLB) and maximum likelihood estimation are introduced and applied to measurements of time delay, frequency, phase, and angle using common techniques such as matched filtering with peak detection, leading edge pulse detection, the DFT, and monopulse angle measurement. Also included is an overview of basic track filtering covering α - β and Kalman filters. This chapter should have been in the first edition, and I am happy to remedy its absence now.

Changes to the other portions of the text are more modest. The review of basic digital signal processing concepts previously in <u>Chap. 1</u> has been relocated to <u>App. B</u> and expanded slightly. An entirely new <u>App. A</u> has been added to reference basic information from random variables and random signals needed for this text, including common probability density functions (PDFs) in radar; estimators and the CRLB; and the effect of linear shift-invariant systems on random signals.

<u>Chapter 2</u> attempts to improve the discussion of fluctuating target models. The traditional Swerling models do not apply in many situations today, both because finer-resolution radars require new PDFs and because the "scan-toscan" and "pulse-to-pulse" terminology is a poor fit to processing based on coherent processing intervals (CPIs). However, the analysis strategy remains valid. I have therefore kept the presentation of the detector design and analysis strategy based on the Neyman-Pearson criterion largely intact while reducing reliance on the "scan-to-scan" and "pulse-to-pulse" terminology for discussing noncoherent integration issues. I have not abandoned these terms completely because the student still needs to understand them to interpret the literature and apply it to modern systems. The other change to <u>Chap. 2</u> is a modest increase in the discussion of clutter reflectivity.

<u>Chapter 3</u> has been renamed "Pulsed Radar Data Acquisition" but is largely unchanged. The discussion of acquiring a datacube for one CPI has been reorganized a bit to make the sequence clearer. Similarly, <u>Chap. 4</u> on waveforms has been expanded only slightly, to add time-domain control of linear FM sidelobes and brief mentions of the stepped-chirp waveform, mismatched filters for binary phase codes, and continuous wave radar.

<u>Chapter 5</u> on Doppler processing has been significantly expanded to include more explanation of the behavior of the pulse Doppler spectrum in the presence of range and Doppler ambiguities. A short mention of the pros and cons of the low, medium, and high PRF regimes has been added. Coverage of ambiguity resolution has been increased and a discussion of blind zones added. Also, the discussion of staggered pulse repetition frequencies for moving target indication has been redone in terms of pulse repetition intervals.

<u>Chapter 6</u> on basic detection theory and <u>Chap. 7</u> on constant false alarm rate (CFAR) thresholding have been combined into the new <u>Chap. 6</u> but are otherwise little changed except for corrections and clarifications. <u>Chapter 8</u> on synthetic aperture radar has likewise been corrected and clarified, with some additional information on interferometric SAR added. Finally, <u>Chap. 9</u> on adaptive beamforming and space-time adaptive processing has also been corrected and clarified. The only significant change has been the elimination of most of the material on computational issues in STAP. (Perhaps a future third edition will have room for a new chapter that can address computational issues in all the radar signal processing techniques.) While SAR and adaptive interference suppression are extremely important in modern radar, the intent of this text is to introduce the basics and prepare the student to tackle some of the many fine books that address these topics in depth.

Throughout the text, I attempt to do a better job of identifying and bringing out common themes that arise again and again in radar signal processing, if sometimes in disguise. These include phase history, coherent integration, matched filtering, integration gain, and maximum likelihood estimation.

A one-semester course in radar signal processing can cover <u>Chaps. 1</u> through <u>7</u>, perhaps also skipping some of the later sections of <u>Chaps. 2</u> and <u>3</u> for additional time savings. Such a course provides a solid foundation for more advanced work in detection theory, adaptive array processing, synthetic aperture imaging, and more advanced radar concepts such as passive and bistatic systems. A quarter-length course could cover <u>Chaps. 1</u> through <u>5</u> and the non-CFAR portion of <u>Chap. 6</u> reasonably thoroughly. In either case, a firm

background in basic continuous and discrete signal processing and an introductory exposure to random variables and processes are advisable. In this edition, I have added homework problems to each chapter to improve the book's usefulness as an academic text. Solutions to the problems are available to instructors of courses using the text on request from the publisher.

Since publication of the first edition, I have collected and maintained a thorough list of every error reported to me by readers or that I have found myself and made it available on the textbook support website at http://www.radarsp.com. That website also provides additional support information such as occasional technical memo supplements on topics related to this book, and some simple MATLAB[®]-based software demos and projects that I have used in my classes. I have tried in this edition to eliminate all known errors without introducing new ones. Complete success is certainly impossible, but I sincerely hope I have mostly succeeded. (Being all new content, Chap. 7 and App. A remain particularly at risk for errors in this edition.) I invite readers to send any and all errata that they find to me at mrichards@gatech.edu. As always, I will make available at the website an errata document with all known errors in this edition.

Mark A. Richards, Ph.D.

Acknowledgments

I remain indebted to many colleagues and students who have helped me to learn this material, write it, and try to debug it once written. Dr. Edward Reedy hired me into my first job after graduation, at the Georgia Tech Research Institute (GTRI), converting me from a speech processor to a radar signal processor. Dr. Jim Echard had the unenviable task of actually overseeing that conversion as my first supervisor and thus my first teacher of the basics of radar. In later years, we jointly developed and taught the course that eventually became ECE 6272 and led to this book. GTRI colleagues too numerous to mention helped me along the way, but a few on whom I have relied particularly merit special mention. Dr. Byron Keel shared his expertise in waveforms and CFAR. Dr. Gregory Showman led me to a greater understanding of SAR, in addition to being the best person I know for talking about radar signal processing in general. Dr. William Melvin introduced me to STAP. I am grateful to each of them for their knowledge and friendship. Since the publication of the first edition, Drs. Keel, Melvin, and Showman have continued to be excellent sources of new information and sounding boards for my ideas on what to include and how to present it. In addition, I have come to rely on several more colleagues. Mr. James Scheer, formerly of GTRI and an outstanding all-around radar engineer, has provided comments and a sanity check on a wide range of topics. Prof. Nadav Levanon of Tel Aviv University, who wrote the lucid and concise text I used for ECE 6272 before I developed this one, has exchanged many materials and ideas with me over the last several years, for which I am grateful. Dr. W. Dale Blair of GTRI has started to teach me about tracking through the courses and books we have worked on together. My great thanks to all of them and many more who have helped me throughout my career and the writing of this book.

Symbols

The following definitions and relations between symbols are used throughout this text except as otherwise specifically noted. Some symbols, such as θ , have more than one usage; their meaning is generally clear from the context.

*	Convolution operator
\otimes	Kronecker product operator
0	Hadamard product operator
(x)	Continuous variable x
[<i>x</i>]	Discrete variable x
$((\cdot))_x$	Modulo <i>x</i>
~	"Is distributed as"
X	Vector variable
Х	Matrix variable
<i>x'</i>	Matrix or vector transpose
<i>x</i> *	Complex conjugate
X ^H	Hermitian transpose
$lpha_q$	Clutter temporal fluctuation vector
α_{GO}	Threshold multiplier, "greatest-of" CFAR
$lpha_{ m log}$	Threshold multiplier, log CFAR
α_{SO}	Threshold multiplier, "smallest-of" CFAR
α_v	Doppler time scaling factor
β	Bandwidth; standard deviation of interference process
β_{C}	Nonlinear term of nonlinear FM waveform bandwidth
β_D	Doppler bandwidth
eta_L	Linear term of nonlinear FM waveform bandwidth
β_n	Noise-equivalent receiver bandwidth
β_{nn}	Null-to-null bandwidth
β_r	Rayleigh bandwidth
$eta_{ m rms}$	Root-mean-square bandwidth
$\beta_v = v/c$	Normalized velocity
$\beta_x, \beta_y, \beta_z$	Spatial bandwidth in x , y , and z dimensions
χ	Signal-to-noise ratio
χ_1	Single sample signal-to-noise ratio
χ_N	N-sample signal-to-noise ratio

$\chi_{ m out}$ χ_{∞}	Output signal-to-noise ratio Signal-to-noise ratio with perfect noise level estimate
χ_{Σ}	Lobing antenna sum channel signal-to-noise ratio
δ	Grazing angle
$\delta[\cdot]$	Discrete impulse function
$\delta \theta$	Target angle relative to boresight
$\delta_{D}(\cdot)$	Dirac impulse ("delta") function
δR	Range error
δt	Differential delay
$\Delta heta$	Angular resolution, lobing antenna squint
$\Delta \psi$	Change in squint angle
ΔCR	Cross-range resolution
ΔF	Frequency step size
ΔF_D	Doppler frequency resolution
Δh	Height displacement
ΔR	Range resolution
ΔR_b	Range relative to central reference point
ΔR_c	Range curvature
ΔR_s	Range bin spacing
ΔR_w	Range walk
Δt	Time resolution
Δt_b	Time relative to central reference point delay
З	I/Q amplitude mismatch
${\cal E}_{\Delta\!/\!\Sigma}$	Error in lobing antenna ratio voltage $v_{\Sigma/\Delta}$
ϕ	Elevation angle; phase; baseband received signal phase
$\overline{\phi}$	Non-baseband received signal phase
ϕ_3	3-dB elevation beamwidth
$\phi_{\!f\!g}$	Interferometric phase difference
$\phi_{ m max}$	Maximum quadratic phase error
ϕ_n	Subpulse phase in phase-coded waveform
ϕ_{nn}	Null-to-null elevation beamwidth
γ	Q channel DC offset; interferogram coherence
Γ	Tracking index; gamma function
η	Volume reflectivity
κ	I channel DC offset; Doppler spectrum oversampling factor; adaptive filter scale factor
$\ell(a b)$	Likelihood function for parameter a given data b

λ	Wavelength; eigenvalue
Λ	Likelihood ratio
$\overline{\mu}$	Estimated mean
θ	Azimuth angle; phase; baseband transmitted signal phase
$\theta(t)$	Phase modulation of waveform
$\overline{\theta}$	Non-baseband transmitted signal phase
θ	Vector of random phases
$ heta_3$	3-dB azimuth beamwidth
$ heta_{az}$	Azimuth beamwidth
$ heta_{\it el}$	Elevation beamwidth
θ_{nn}	Null-to-null azimuth beamwidth
$ heta_{\scriptscriptstyle S\!A\!R}$	Effective beamwidth of synthetic aperture radar
θ_t	Target angle of arrival
Θ	Parameter to be estimated
Θ	Parameter vector to be estimated
Ô	Estimate of Θ
Ô	Estimate of Θ
$\rho = \zeta \exp[j\psi] = \rho_{\rm I} + j\rho_{\rm Q}$	Complex baseband reflectivity
$ ho_I$	Baseband reflectivity in-phase (I) component
$ ho_{ m Q}$	Baseband reflectivity quadrature-phase (Q) component
$ ho_{f}, ho_{fg}$	Normalized autocorrelation or cross-correlation function
ρ'	Effective baseband complex reflectivity
$\overline{\rho} = \overline{\varsigma} \exp\left[j\overline{\psi}\right] \\= \overline{\rho}_L + j\overline{\rho}_Q$	Complex non-baseband reflectivity
$\overline{\rho}_{I}$	Non-baseband reflectivity in-phase (I) component
$\overline{\rho}_{Q}$	Non-baseband reflectivity quadrature-phase (Q) component
ρ	Cross-range averaged effective baseband complex reflectivity
Ρ̃	Range spatial spectrum (Fourier transform of $\tilde{\rho}$)
$\sigma = \rho ^2 = \zeta^2$	Radar cross section (RCS)
σ^0	Area reflectivity
σ_h	Surface roughness
σ_x^2	Variance of random variable x
σÔ	Precision of estimate
Σ	Diagonal matrix of clutter powers
ζ	Baseband reflectivity amplitude ($\zeta \ge 0$)

ζ	Non-baseband reflectivity amplitude ($\overline{\zeta} \ge 0$)
τ	Pulse length
$ au_c$	Subpulse length in phase coded waveform
γ	Sufficient statistic
ω	Normalized frequency (radians per sample)
$\omega_{\scriptscriptstyle D}$	Normalized Doppler frequency shift (radians per sample)
W _s	Sampling interval in normalized frequency ω (samples per radian)
Ω	Frequency (radians per second)
$\Omega_ heta$	Azimuth rotation rate (radians per second)
$\Omega_{\scriptscriptstyle D}$	Doppler frequency shift (radians per second)
$\Omega_{ m diff}$	Doppler frequency mismatch (radians per second)
Ω_i	Matched Doppler frequency shift (radians per second)
Ψ	Baseband reflectivity phase; squint angle
$\overline{\psi}$	Non-baseband reflectivity phase
a	Baseband transmitted signal amplitude
$a_s(\theta)$	Spatial steering vector
$a_t(\theta)$	Temporal steering vector
ā	Non-baseband transmitted signal amplitude
A	Ambiguity function
A, \hat{A}, \overline{A}	Signal amplitude
Â	Complex ambiguity function
A_e	Effective antenna aperture size
A_q	Covariance matrix of clutter temporal fluctuations
A_n	Complex amplitude of subpulse in phase coded waveform
$AF(\theta), AF(\theta, \phi)$	Phased array antenna array factor
APB	A priori bound
b	Amplitude
b	Non-baseband received signal amplitude
В	Number of bits; interferometric baseline
B_N	Length of Barker phase code
$B_{\hat{\Theta}}(\Theta)$	Accuracy of estimator
С	Speed of light
C_q	Clutter space-time steering vector for patch q
CA	Clutter attenuation
CRLB	Cramèr-Rao lower bound

$C_x(\cdot)$	Characteristic function of random variable <i>x</i> ; centroid of
d	Phased array element spacing
$d_{\scriptscriptstyle M}$	Mahalanobis distance
d_{pc}	Phase center spacing
D	Antenna aperture size
D_{az}	Antenna size, azimuth dimension
D_{el}	Antenna size, elevation dimension
D_{SAR}	Synthetic aperture size
DOF	Degrees of freedom
DR	Dynamic range
D_x, D_y, D_z	Antenna aperture size in x , y , or z dimension
e	Eigenvector
E, E_x	Energy; energy in signal x
$E(\cdot)$	Expected value
$E(\theta, \phi)$	Electric field amplitude
$E_{el}(\theta), E_{el}(\theta, \phi)$	Phased array antenna element pattern
f	Normalized frequency (cycles per sample)
\tilde{f}	Quantized version of a function f
f_D	Normalized Doppler frequency shift (cycles per sample)
f_{Dt}	Target normalized Doppler frequency shift (cycles per sample)
$f_{ heta}$	Normalized spatial frequency (cycles per sample)
F	Fourier transform of a function f
F	Fourier transform operator; track filter state transition matrix
F	Frequency (hertz)
$F_{ heta}$	Spatial frequency (cycles per meter)
F_b	Beat frequency, blind Doppler frequency (hertz)
F_{bs}	Blind Doppler frequency using staggered PRIs
F_c	Corner frequency (hertz)
F_D	Doppler frequency shift (hertz)
F_{Da}	Apparent Doppler frequency (hertz)
F_{Dua}	Unambiguous Doppler frequency interval (hertz)
F_{g}	Greatest common divisor of a set of staggered PRFs
F_i	Instantaneous frequency (hertz)
F_n	Noise figure

F_s	Sampling frequency (samples per second)
F_t	Transmitted frequency (hertz)
F_{us}	Unstaggered blind Doppler frequency
g, G	Tracking process noise gain
G	Antenna voltage or power gain
G_{nc}	Noncoherent integration gain
G_s	Maximum receiver gain
G_{sp}	Signal processing gain
h	Height
h	Filter weight vector; beamformer weight vector
h, H	Tracking observation matrix
h(t)	Impulse response (continuous time)
h[n]	Impulse response (discrete time)
H_0	Null hypothesis (interference only)
H_1	Nonnull hypothesis (target plus interference)
$\begin{array}{c} H(f), H(F), H(\omega), \\ H(\Omega) \end{array}$	Frequency response in various units
H(z)	Discrete-time system function
$H_N(z)$	System function of N-pulse canceller
$H_{N,P}(F)$	Frequency response of <i>N</i> -pulse canceller with <i>P</i> staggered PRFs
Ι	Interference power; improvement factor
$I(\Theta)$	Fisher information matrix
I _{opt}	Improvement factor for matched filter
\mathbf{I}_N	Nth-order identity matrix
I _{sub}	Suboptimum improvement factor
Ι	In-phase channel
I_k	In-phase component, sample k
$I(\cdot, \cdot)$	Incomplete gamma function
$I(\cdot)$	Fisher information matrix
$J_n(t)$	Jammer signal
J	Jammer signal sample vector
<i>K</i> , k, K	Tracking filter gain; Kalman filter gain (symbol varies with dimensionality)
k_p	Stagger ratio
$k_{\Delta/\Sigma}$	Lobing antenna Δ/Σ error slope
	Spatial frequency (cycles per meter); DFT size; normalized

$K K_R$	quantizer step size Range spatial frequency (cycles per meter)
K_u	Cross-range spatial frequency (cycles per meter)
K_x, K_v, K_z	Spatial frequency in x, y , or z dimensions
$K_{ heta}$	Spatial frequency corresponding to AOA θ (cycles per meter)
$k_{ heta}$	Normalized spatial frequency corresponding to AOA θ (cycles per sample)
L	Number of fast time samples per pulse
L_a	Atmospheric loss factor
L_d	Target depth as viewed from the radar
L_0	Maximum acceptable signal-to-interference ratio loss
L_s	System loss factors; synthetic aperture radar swath length
L _{SIR}	Signal-to-interference ratio loss
L_w	Target width as viewed from the radar
LPG	Loss in processing gain
т	Mean of random variable
m_n	Mean of <i>n</i> th element of a random vector
$\mathbf{m} = \tilde{\mathbf{m}} \exp(j\theta)$	Mean of random vector
m	Nonrandom portion of mean of random vector random
M	Number of slow time samples per coherent processing interval
М	Tracking mean-square error estimate matrix
M_{DD}	Minimum detectable Doppler shift
M_{DD^+}, M_{DD^-}	Minimum detectable positive, negative Doppler shift
$M_{ m opt}$	Optimum value of M in " M of N " detection rule
M_s	DPCA time slip
n_P	Matched filter output noise power
Ν	Noise power; number of phase centers
N_{γ}	Number of spotlight SAR radial slices
N_R	Number of spotlight SAR range samples
$N_{ m spot}$	Number of spotlight SAR images per unit time
$p_x(\cdot)$	Probability density function for a random variable x
$P_x(\cdot)$	Cumulative density function for a random variable x
Р	Power; degrees of freedom in space-time snapshot
$P(\theta, \phi)$	Antenna one-way power pattern

$P_{\theta}(\phi)$	Azimuth one-way antenna power pattern
PA	Periodic ambiguity function
P_b	Backscattered power
P_{BD}	Binary integrated probability of detection
P_{BFA}	Binary integrated probability of false alarm
P_{CD}	Cumulative probability of detection
P_{CFA}	Cumulative probability of false alarm
P_D	Probability of detection
\overline{P}_{D}	Expected value of probability of detection
P_{FA}	Probability of false alarm
\overline{P}_{FA}	Expected value of probability of false alarm
P_M	Probability of miss
P_r	Received power; relative power of I/Q mismatch image
Pr ·	Probability of argument occurring
P_t	Transmitted power
PL	Processing loss
PRF	Pulse repetition frequency (pulses per second)
q	Quantizer step size
Q	Power density
Q	Quadrature channel
Q_b	Backscattered power density
Q_k	Quadrature component, sample k
Q_M	Marcum Q function
Q_t	Transmitted power density
R, R_0	Range
R_a	Apparent range
R_{\min}	Minimum range
R_t	True range
R_{ua}	Unambiguous range
$R_{\rm uas}$	Unambiguous range using staggered PRIs
R_w	Range window, range swath
Ŝ	Estimated standard deviation
S_A	Autocorrelation of phase code complex amplitude sequence
S_f	Autocorrelation of a function or random signal f
S _{fg}	Cross-correlation of functions or random signals f and g

$S_p(t)$	Output of filter matched to single pulse in pulse train
S	waveform Polarization scattering matrix
$S_f(\omega)$	Power spectrum of a function or random signal f
$\mathbf{S}(\mathbf{\omega})$	Cross-power spectrum of functions or random signals f and
$S_{fg}(\omega)$	g
Sx	Covariance matrix for a random vector x
$\tilde{\mathbf{S}}_x$	Transformed covariance matrix
Ŝ	Estimated covariance matrix
SIR	Signal-to-interference ratio
SQNR	Signal-to-quantization noise ratio
<i>t</i> , <i>t</i> ₀	Time
t	Target model vector
ť	Transformed target model vector
Т	Pulse repetition interval; detection threshold; track measurement update interval
Т	Transformation matrix
T'	Equivalent receiver temperature; detection threshold
\hat{T}	Estimated threshold
\hat{T}_{log}	Estimated threshold, log CFAR
$T_{ heta}$	Sampling interval in θ
T_a	Aperture time
T_{avg}	Average PRI of a set of staggered PRFs
T_M	Time of matched filter output peak
T_p	pth PRI in set of staggered PRFs
T_s	Fast time sampling interval; sampling interval in $s = \sin \theta$
$T_{ m tot}$	Sum of PRIs corresponding to set of staggered set of PRFs
T_w	Time corresponding to swath width
u	Along-track coordinate of synthetic aperture radar platform
<i>u</i> , u	Tracking process noise
UDSF	Usable Doppler space fraction
var(x)	Variance of random variable x
υ	Platform velocity
v_{Σ}, v_{Δ}	Sum and difference lobing voltages
$v_{\Sigma/\Delta}$	Lobing antenna ratio voltage
v_a	Apparent velocity
v_b	Blind speed

v_{bs}	Blind speed using staggered PRIs
v_L, v_R	Left and right lobing antenna voltages
v_{ua}	Unambiguous velocity interval
W, W	Tracking measurement noise
W _f	Temporal weight vector
$\mathbf{W}_{ heta}$	Spatial weight vector
x	Mean of random variable x
â	Estimated value of random variable x
$x = a \exp[j\theta] = x_{\rm I} + jx_{\rm O}$	Transmitted signal, baseband
$\overline{x} = \overline{a} \exp[j\tilde{\theta}] \\ = x_1 + j\overline{x}_Q$	Transmitted signal, non-baseband
x_{I}	Baseband transmitted signal in-phase (I) component
\overline{x}_{I}	Non-baseband transmitted in-phase (I) component
x_P	Along-track coordinate of synthetic aperture radar scatterer
x_Q	Baseband transmitted signal quadrature-phase (Q) component
xQ	Non-baseband transmitted quadrature-phase (Q) component
$x_p(t)$	Subpulse of phase coded waveform
$x_p(t)$	Single pulse of pulse train waveform
$y = b \exp[j\phi] = y_{\rm I} + jy_{\rm Q}$	Received signal, baseband
у	Baseband received signal sample vector
ỹ	Transformed baseband received signal sample vector
$\overline{y} = \overline{b} \exp[j\overline{\phi}] = \overline{y}_1 j \overline{y}_Q$	Received signal, non-baseband
${\cal Y}_{\rm I}$	Baseband received signal in-phase (I) component
${\cal Y}_{\rm I}$	Non-baseband received signal in-phase (I) component
\mathcal{Y}_{Q}	Baseband received signal quadrature-phase (Q) component
\overline{y}_{Q}	Non-baseband received signal quadrature-phase (Q) component
y[l, m, n]	Datacube for one coherent processing interval
y[l,m]	Fast time/slow time data matrix for one CPI
$y_s[m]$	Slow time sequence for one CPI
Ζ	Detected output
ž	Transformed detected output
Ζ	Meteorological reflectivity; altitude

Acronyms

The following acronyms are used throughout this text.

1D	One-Dimensional
2D	Two-Dimensional
AC	Alternating Current
ACF	Autocorrelation Function
A/D	Analog-to-Digital
AF	Ambiguity Function
AGC	Automatic Gain Control
AMF	Adaptive Matched Filter
AMTI	Airborne Moving Target Indication
AOA	Angle of Arrival
AR	Autoregressive
ASR	Airport Surveillance Radar
AWGN	Additive White Gaussian Noise
BPF	Bandpass Filter
BSR	Beam Sharpening Ratio
BT	Time-Bandwidth Product
CA	Clutter Attenuation
CA-CFAR	Cell-Averaging Constant False Alarm Rate
CAT	Computerized Axial Tomography
CCD	Coherent Change Detection
CDF	Cumulative Distribution Function
CF	Characteristic Function
CFAR	Constant False Alarm Rate
CMT	Covariance Matrix Taper
CNR	Clutter-to-Noise Ratio
CRLB	Cramèr-Rao Lower Bound
CRP	Central Reference Point
CRT	Chinese Remainder Theorem
CPI	Coherent Processing Interval
CUT	Cell under Test
CW	Continuous Wave
D/A	Digital-to-Analog

dB	Decibel
dBsm	Decibels relative to 1 square meter
DBS	Doppler Beam Sharpening
DC	Direct Current
DCT	Discrete Cosine Transform
DF CFAR	Distribution-Free Constant False Alarm Rate
DFT	Discrete Fourier Transform
DOF	Degrees of Freedom
DPCA	Displaced Phase Center Antenna
DSP	Digital Signal Processing
DTED	Digital Terrain Elevation Data
DTFT	Discrete Time Fourier Transform
EA	Electronic Attack
ECM	Electronic Countermeasures
EKF	Extended Kalman Filter
EM	Electromagnetic
EMI	Electromagnetic Interference
ENOB	Effective Number of Bits
ERIM	Environmental Research Institute of Michigan
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FM	Frequency Modulation
FMCW	Frequency Modulated Continuous Wave
FSK	Frequency Shift Keying
GLRT	Generalized Likelihood Ratio Test
GMTI	Ground Moving Target Indication
GOCA CFAR	Greatest-of Cell-Averaging Constant False Alarm Rate
GPS	Global Positioning System
GTRI	Georgia Tech Research Institute
HF	High Frequency
HPRF	High Pulse Repetition Frequency
Ι	In-Phase
ICM	Internal Clutter Motion; Intrinsic Clutter Motion
IDFT	Inverse Discrete Fourier Transform
IF	Intermediate Frequency

IEET I.i.d.	Inverse Fast Fourier Transform
IFSAR	Interferometric Synthetic Aperture Radar
IIR	Infinite Impulse Response
IMM	Interacting Multiple Models
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
InSAR	Interferometric Synthetic Aperture Radar
IPD	Interferometric Phase Difference
ISAR	Inverse Synthetic Aperture Radar
ISL	Integrated Sidelobe Level, Interference Subspace Leakage
JNR	Jammer-to-Noise Ratio
KA	Knowledge-Aided
KF	Kalman Filter
LEO	Low Earth Orbit
LCM	Least Common Multiple
LFM	Linear Frequency Modulation
LNA	Low Noise Amplifier
LO	Local Oscillator
LPF	Lowpass Filter
LPG	Loss in Processing Gain
LPRF	Low Pulse Repetition Frequency
LRT	Likelihood Ratio Test
LSB	Least Significant Bit
LSE	Least Squares Estimate
LSI	Linear Shift-Invariant
LTI	Linear Time-Invariant
MDD	Minimum Detectable Doppler
MDV	Minimum Detectable Velocity
MLE	Minimum Likelihood Estimate (or Estimator, or Estimation)
MMSE	Minimum Mean-Squared Error, Minimum Means-Squared Estimate
MMW	Millimeter Wave
MPRF	Medium Pulse Repetition Frequency
MSB	Most Significant Bit
MSE	Mean-Squared Error
MTD	Moving Target Detector

MTI MVU	Moving Target Indication Minimum Variance Unbiased
NASA	National Aeronautics and Space Agency
NEXRAD	Next Generation Radar
NLFM	Nonlinear Frequency Modulation
NRL	Naval Research Laboratory
OS CFAR	Order Statistic Constant False Alarm Rate
PAF	Periodic Ambiguity Function
PC	Principal Components
PDF	Probability Density Function
PFA	Polar Format Algorithm
PGA	Phase Gradient Algorithm
PL	Processing Loss
PPP	Pulse Pair Processing
PRF	Pulse Repetition Frequency
PRI	Pulse Repetition Interval
PSD	Power Spectrum (or Spectral) Density
PSL	Peak Sidelobe Level
PSM	Polarization Scattering Matrix
PSP	Principle of Stationary Phase
PSR	Point Spread Response
Q	Quadrature
RCS	Radar Cross Section
RCSR	Radar Cross Section Reduction
RD	Range-Doppler
RF	Radar Frequency
RMB	Reed-Mallet-Brennan
RMS	Root-Mean-Square
ROI	Region of Interest
RSS	Root Sum Square
RV	Random Variable
RVP	Residual Video Phase
SAR	Synthetic Aperture Radar
SB	Sampling Bound
SQNR	Signal-to-Quantization Noise Ratio
S-CFAR	Switching Constant False Alarm Rate

SCR SIR	Signal-to-Clutter Ratio Signal-to-Interference Ratio; Shuttle Imaging Radar
SMI	Sample Matrix Inverse
SMTI	Surface Moving Target Indication
SNR	Signal-to-Noise Ratio
SOCA CFAR	Smallest-of Cell-Averaging Constant False Alarm Rate
STALO	Stable Local Oscillator
STAP	Space-Time Adaptive Processing
T/R	Transmit/Receive
UDSF	Usable Doppler Space Fraction
UHF	Ultra-High Frequency
UMP	Uniformly Most Powerful
VHF	Very High Frequency
WGN	White Gaussian Noise
ZZB	Ziv-Zakai Bound

CHAPTER 1

Introduction to Radar Systems and Signal Processing

1.1 History and Applications of Radar

The word "radar" was originally an acronym, RADAR, for "*ra*dio *d*etection *a*nd *r*anging." Today, the technology is so common that the word has become a standard English noun. Many people have direct personal experience with radar in such applications as measuring fastball speeds or, often to their regret, traffic control.

The history of radar extends to the early days of modern electromagnetic theory (Swords, 1986; Skolnik, 2001). In 1886, Hertz demonstrated reflection of radio waves, and in 1900 Tesla described a concept for electromagnetic detection and velocity measurement in an interview. In 1903 and 1904, the German engineer Hülsmeyer experimented with ship detection by radio wave reflection, an idea advocated again by Marconi in 1922. In that same year, Taylor and Young of the U.S. Naval Research Laboratory (NRL) demonstrated ship detection by radar and in 1930 Hyland, also of NRL, first detected aircraft (albeit accidentally) by radar, setting off a more substantial investigation that led to a U.S. patent for what would now be called a *continuous wave* (CW) radar in 1934.

The development of radar accelerated and spread in the middle and late 1930s, with largely independent developments in the United States, Britain, France, Germany, Russia, Italy, and Japan. In the United States, R. M. Page of NRL began an effort to develop pulsed radar in 1934, with the first successful demonstrations in 1936. The year 1936 also saw the U.S. Army Signal Corps begin active radar work, leading in 1938 to its first operational system, the SCR-268 antiaircraft fire control system, and in 1939 to the SCR-270 early warning system, the detections of which were tragically ignored at Pearl Harbor. British development, spurred by the threat of war, began in earnest with work by Watson-Watt in 1935. The British demonstrated pulsed radar that year, and by 1938 established the famous Chain Home surveillance radar network that remained active until the end of World War II. They also built the first airborne interceptor radar in 1939. In 1940, the United States and Britain began to exchange information on radar development. Up to this time, most radar work was conducted at high frequency (HF) and very high frequency (VHF) wavelengths; but with the British disclosure of the critical cavity magnetron microwave power tube and the United States formation of the Radiation Laboratory at the Massachusetts Institute of Technology, the groundwork was

laid for the successful development of radar at the microwave frequencies that have predominated ever since.

Each of the other countries mentioned also carried out CW radar experiments, and each fielded operational radars at some time during the course of World War II. Efforts in France and Russia were interrupted by German occupation. On the other hand, Japanese efforts were aided by the capture of U.S. radars in the Philippines and by the disclosure of German technology. The Germans themselves deployed a variety of ground-based, shipboard, and airborne systems. By the end of the war, the value of radar and the advantages of microwave frequencies and pulsed waveforms were widely recognized.

Early radar development was driven by military necessity, and the military is still a major user and developer of radar technology. Military applications include surveillance, navigation, and weapons guidance for ground, sea, air, and space vehicles. Military radars span the range from huge ballistic missile defense systems to fist-sized tactical missile seekers.

Radar now enjoys an increasing range of applications. One of the most common is the police traffic radar used for enforcing speed limits (and measuring the speed of baseballs and tennis serves). Another is the "color weather radar" familiar to every viewer of local television news. The latter is one type of meteorological radar; more sophisticated systems are used for largescale weather monitoring and prediction and atmospheric research. Another radar application that affects many people is found in the air traffic control systems used to guide commercial aircraft both en route and in the vicinity of airports. Aviation also uses radar for determining altitude and avoiding severe weather, and may soon use it for imaging runway approaches in poor weather. Radar is commonly used for collision avoidance and buoy detection by ships, and is now beginning to serve the same role for the automobile and trucking industries. Finally, spaceborne (both satellite and space shuttle) and airborne radar is an important tool in mapping earth topology and environmental characteristics such as water and ice conditions, forestry conditions, land usage, and pollution. While this sketch of radar applications is far from exhaustive, it does indicate the breadth of applications of this remarkable technology.

This text tries to present a thorough, straightforward, and consistent description of the signal processing aspects of radar technology, focusing primarily on the more fundamental functions common to most radar systems. Pulsed radars are emphasized over CW radars, though many of the ideas are applicable to both. Similarly, *monostatic* radars, where the transmitter and receiver antennas are collocated (and in fact are usually the same antenna), are emphasized over *bistatic* radars, where they are significantly separated, though again many of the results apply to both. The reason for this focus is that the majority of radar systems are monostatic, pulsed designs. Finally, the subject is approached from a digital signal processing (DSP) viewpoint as much as practicable, both because most new radar designs rely heavily on digital
processing and because this approach can unify concepts and results often treated separately.

1.2 Basic Radar Functions

Most uses of radar can be classified as *detection, tracking*, or *imaging*. This text addresses all three, as well as the techniques of signal acquisition and interference reduction necessary to perform these tasks.

The most fundamental problem in radar is detection of an object or physical phenomenon. This requires determining whether the receiver output at a given time represents the echo from a reflecting object or only noise. Detection decisions are usually made by comparing the amplitude A(t) of the receiver output (where *t* represents time) to a threshold T(t), which may be set a priori in the radar design or may be computed adaptively from the radar data; in Chap. 6 it will be seen why this detection technique is appropriate. The time required for a pulse to propagate a distance *R* and return, thus traveling a total distance 2*R*, is just 2R/c; thus, if A(t) > T(t) at some time delay t_0 after a pulse is transmitted, it is assumed that a target is present at range

$$R = \frac{ct_0}{2} \tag{1.1}$$

where *c* is the speed of light.¹

Once an object has been detected, it may be desirable to track its location or velocity. A monostatic radar naturally measures position in a spherical coordinate system with its origin at the radar antenna's phase center, as shown in Fig. 1.1. In this coordinate system, the antenna look direction, sometimes called the *boresight* direction, is along the +x axis. The angle θ is called *azimuth* angle, while ϕ is called *elevation* angle. Range R to the object follows directly from the elapsed time from transmission to detection as just described. Elevation and azimuth angle ϕ and θ are determined from the antenna orientation, since the target must normally be in the antenna main beam to be detected. Velocity is estimated by measuring the Doppler shift of the target echoes. Doppler shift provides only the radial velocity component, but a series of measurements of position and radial velocity can be used to infer target dynamics in all three dimensions.



FIGURE 1.1 Spherical coordinate system for radar measurements.

Because most people are familiar with the idea of following the movement of a "blip" on the radar screen, detection and tracking are the functions most commonly associated with radar. Increasingly, however, radars are being used to generate two-dimensional images of an area. Such images can be analyzed for intelligence and surveillance purposes, for topology mapping, or for analysis of earth resources issues such as mapping, land use, ice cover analysis, deforestation monitoring, and so forth. They can also be used for "terrain following" navigation by correlating measured imagery with stored maps. While radar images have not achieved the resolution of optical images, the very low attenuation of electromagnetic waves at microwave frequencies gives radar the important advantage of "seeing" through clouds, fog, and precipitation very well. Consequently, imaging radars generate useful imagery when optical instruments cannot be used at all.

The quality of a radar system is quantified with a variety of figures of merit, depending on the function being considered. In analyzing detection performance, the fundamental parameters are the *probability of detection* P_D and the *probability of false alarm* P_{FA} . If other system parameters are fixed, increasing P_D always requires accepting a higher P_{FA} as well. The achievable combinations are determined by the signal and interference statistics, especially the *signal-to-interference ratio* (SIR). When multiple targets are present in the radar field of view, additional considerations of resolution and sidelobes arise in evaluating detection performance. For example, if two targets cannot be resolved by a radar, they will be registered as a single object. If sidelobes are high, the echo from one strongly reflecting target may mask the echo from a nearby but weaker target, so that again only one target is registered when two

are present. Resolution and sidelobes in range are determined by the radar waveform, while those in angle are determined by the antenna pattern.

In radar tracking, the basic figure of merit is *accuracy* of range, angle, and velocity estimation. While resolution presents a crude limit on accuracy, with appropriate signal processing the achievable accuracy is ultimately limited in each case by the SIR.

In imaging, the principal figures of merit are spatial resolution and dynamic range. Spatial resolution determines what size objects can be identified in the final image, and therefore to what uses the image can be put. For example, a radar map with 1 km by 1 km resolution would be useful for land use studies, but useless for military surveillance of airfields or missile sites. Dynamic range determines image contrast, which also contributes to the amount of information that can be extracted from an image.

The purpose of signal processing in radar is to improve these figures of merit. SIR can be improved by pulse integration. Resolution and SIR can be jointly improved by pulse compression and other waveform design techniques, such as frequency agility. Accuracy benefits from increased SIR and interpolation methods. Sidelobe behavior can be improved with the same windowing techniques used in virtually every application of signal processing. Each of these topics are discussed in the chapters that follow.

Radar signal processing draws on many of the same techniques and concepts used in other signal processing areas, from such closely related fields as communications and sonar to very different applications such as speech and image processing. Linear filtering and statistical detection theory are central to radar's most fundamental task of target detection. Fourier transforms, implemented using *fast Fourier transform* (FFT) techniques, are ubiquitous, being used for everything from fast convolution implementations of matched filters, to Doppler spectrum estimation, to radar imaging. Modern model-based spectral estimation and adaptive filtering techniques are used for beamforming and jammer cancellation. Pattern recognition techniques are used for target/clutter discrimination and target identification.

At the same time, radar signal processing has several unique qualities that differentiate it from most other signal processing fields. Most modern radars are coherent, meaning that the received signal, once demodulated to baseband, is complex-valued rather than real-valued. Radar signals have very high dynamic ranges of several tens of decibels, in some extreme cases approaching 100 dB. Thus, gain control schemes are common, and sidelobe control is often critical to avoid having weak signals masked by stronger ones. SIR ratios are often relatively low. For example, the SIR at the point of detection may be only 10 to 20 dB, while the SIR for a single received pulse prior to signal processing is frequently less than 0 dB.

Especially important is the fact that, compared to most other DSP applications, radar signal bandwidths are large. Instantaneous bandwidths for an

individual pulse are frequently on the order of a few megahertz, and in some fine-resolution² radars may reach several hundred megahertz and even as high as 1 GHz. This fact has several implications for digital signal processing. For example, very fast analog-to-digital (A/D) converters are required. The difficulty of designing good converters at multi-megahertz sample rates has historically slowed the introduction of digital techniques into radar signal processing. Even now, when digital techniques are common in new designs, radar word lengths in high-bandwidth systems are usually a relatively short 8 to 12 bits, rather than the 16 bits common in many other areas. The high data rates have also historically meant that it has often been necessary to design custom hardware for the digital processor in order to obtain adequate throughput, that is, to "keep up with" the onslaught of data. This same problem of providing adequate throughput has resulted in radar signal processing algorithms being relatively simple compared to, say, sonar processing techniques. Only in the late 1990s has Moore's Law³ provided enough computing power to host radar algorithms for a wide range of systems on commercial hardware. Equally important, this same technological progress has allowed the application of new, more complex algorithms to radar signals, enabling major improvements in detection, tracking, and imaging capability.

1.3 Elements of a Pulsed Radar

Figure 1.2 is one possible block diagram of a simple pulsed monostatic radar. The waveform generator output is the desired pulse waveform. The transmitter modulates this waveform to the desired radio frequency (RF) and amplifies it to a useful power level. The transmitter output is routed to the antenna through a duplexer, also called a *circulator* or T/R switch (for transmit/receive). The returning echoes are routed, again by the duplexer, into the radar receiver. The receiver is usually a superheterodyne design, and often the first stage is a lownoise RF amplifier. This is followed by one or more stages of modulation of the received signal to successively lower intermediate frequencies (IFs) and ultimately to baseband, where the signal is not modulated onto any carrier frequency. Each modulation is carried out with a *mixer* and a *local oscillator* (LO). The baseband signal is next sent to the signal processor, which performs some or all of a variety of functions such as pulse compression, matched filtering, Doppler filtering, integration, and motion compensation. The output of the signal processor takes various forms, depending on the radar purpose. A tracking radar would output a stream of detections with measured range and angle coordinates, while an imaging radar would output a two- or threedimensional image. The processor output is sent to the system display, the data processor, or both as appropriate.



FIGURE 1.2 Block diagram of a pulsed monostatic radar.

The configuration of Fig. 1.2 is not unique. For example, many systems perform some of the signal processing functions at IF rather than baseband; matched filtering, pulse compression, and some forms of Doppler filtering are very common examples. The list of signal processing functions is redundant as well. For example, pulse compression and Doppler filtering can both be considered part of the matched filtering process. Another characteristic which differs among radars is at what point in the system the analog signal is digitized. Older systems are, of course, all analog, and many currently operational systems do not digitize the signal until it is converted to baseband. Thus, any signal processing performed at IF must be done with analog techniques. Increasingly, new designs digitize the signal at an IF stage, thus moving the A/D converter closer to the radar front end and enabling digital processing at IF. Finally, the distinction between signal processing and data processing is sometimes unclear or artificial.

In the next few subsections, the major characteristics of these principal radar subsystems are briefly discussed.

1.3.1 Transmitter and Waveform Generator

The transmitter and waveform generator play a major role in determining the sensitivity and range resolution of radar. Radar systems have been operated at frequencies as low as 2 MHz and as high as 220 GHz (Skolnik, 2001); laser radars operate at frequencies on the order of 10^{12} to 10^{15} Hz, corresponding to wavelengths on the order of 0.3 to 30 µm (Jelalian, 1992). However, most

radars operate in the microwave frequency region of about 200 MHz to about 95 GHz, with corresponding wavelengths of 0.67 m to 3.16 mm. <u>Table 1.1</u> summarizes the letter nomenclature used for the common nominal radar bands (IEEE, 1976). The millimeter wave band is sometimes further decomposed into approximate subbands of 36 to 46 GHz (Q band), 46 to 56 GHz (V band), and 56 to 100 GHz (W band) (Richards et al., 2010).

Band	Frequencies	Wavelengths				
HF	3–30 MHz	100–10 m				
VHF	30–300 MHz	10–1 m				
UHF	300 MHz-1 GHz	1–30 cm				
L	1–2 GHz	30–15 cm				
S	2–4 GHz	15–7.5 cm				
С	4–8 GHz	7.5–3.75 cm				
Х	8–12 GHz	3.75–2.5 cm				
Ku	12–18 GHz	2.5–1.67 cm				
К	18–27 GHz	1.67–1.11 cm				
Ka	27–40 GHz	1.11 cm-7.5 mm				
mm	40–300 GHz	7.5–1 mm				

TABLE 1.1 LE	etter Nomenclature	for Nominal	Radar Free	quency Bands
--------------	--------------------	-------------	------------	--------------

Within the HF to K_a bands, specific frequencies are allocated by international agreement to radar operation. In addition, at frequencies above X band, atmospheric attenuation of electromagnetic waves becomes significant. Consequently, radar in these bands usually operates at an "atmospheric window" frequency where attenuation is relatively low. Figure 1.3 illustrates the atmospheric attenuation for one-way propagation over the most common radar frequency ranges under one set of atmospheric conditions. Most K_a band radars operate near 35 GHz and most W band systems operate near 95 GHz because of the relatively low atmospheric attenuation at these wavelengths.



FIGURE 1.3 One-way atmospheric attenuation of electromagnetic waves. (*Source:* EW and Radar Systems Engineering Handbook, Naval Air Warfare Center, Weapons Division, <u>http://ewhdbks.mugu.navy.mil/</u>)

Lower radar frequencies tend to be preferred for longer range surveillance applications because of the low atmospheric attenuation and high available powers. Higher frequencies tend to be preferred for finer resolution, shorter range applications due to the smaller achievable antenna beamwidths for a given antenna size, higher attenuation, and lower available powers.

Weather conditions can also have a significant effect on radar signal propagation. Figure 1.4 illustrates the additional one-way loss as a function of RF frequency for rain rates ranging from a drizzle to a tropical downpour. X-band frequencies (typically 10 GHz) and below are affected significantly only by very severe rainfall, while millimeter wave frequencies suffer severe losses for even light-to-medium rain rates.



FIGURE 1.4 Effect of different rates of precipitation on one-way atmospheric attenuation of electromagnetic waves. (*Source:* EW and Radar Systems Engineering Handbook, Naval Air Warfare Center, Weapons Division, <u>http://ewhdbks.mugu.navy.mil/</u>)

Radar transmitters operate at peak powers ranging from milliwatts to in excess of 10 MW. One of the more powerful existing transmitters is found in the AN/FPS-108 COBRA DANE radar, which has a peak power of 15.4 MW (Brookner, 1988). The interval between pulses is called the *pulse repetition interval* (PRI), and its inverse is the *pulse repetition frequency* (PRF). PRF varies widely but is typically between several hundred pulses per second (pps) and several tens of thousands of pulses per second. The duty cycle of pulsed systems is usually relatively low and often well below 1 percent, so that average powers rarely exceed 10 to 20 kW. COBRA DANE again offers an extreme example with its average power of 0.92 MW. Pulse lengths are most often between about 100 ns and 100 μ s, though some systems use pulses as short as a few nanoseconds while others have extremely long pulses, on the order of 1 ms.

It will be seen (Chap. 6) that the detection performance achievable by a radar improves with the amount of energy in the transmitted waveform. To maximize detection range, most radar systems try to maximize the transmitted power. One way to do this is to always operate the transmitter at full power during a pulse. Thus, radars generally do not use amplitude modulation of the transmitted pulse. On the other hand, the nominal range resolution ΔR is determined by the waveform bandwidth β according to Chap. 4.

$$\Delta R = \frac{c}{2\beta} \tag{1.2}$$

For an unmodulated pulse, the bandwidth is inversely proportional to its duration. To increase waveform bandwidth for a given pulse length without sacrificing energy, many radars routinely use phase or frequency modulation of the pulse.

Desirable values of range resolution vary from a few kilometers in longrange surveillance systems, which tend to operate at lower RFs, to a meter or less in very fine-resolution imaging systems, which tend to operate at high RFs. Corresponding waveform bandwidths are on the order of 100 kHz to 1 GHz, and are typically 1 percent or less of the RF. Few radars achieve 10 percent bandwidth. Thus, most radar waveforms can be considered narrowband, bandpass functions.

1.3.2 Antennas

The antenna plays a major role in determining the sensitivity and angular resolution of the radar. A wide variety of antenna types are used in radar systems. Some of the more common types are parabolic reflector antennas, scanning feed antennas, lens antennas, and phased array antennas.

From a signal processing perspective, the most important properties of an antenna are its gain, beamwidth, and sidelobe levels. Each of these follows from consideration of the antenna *power pattern*. The power pattern $P(\theta, \phi)$ describes the radiation intensity during transmission in the direction (θ, ϕ) relative to the antenna boresight. Aside from scale factors, which are unimportant for normalized patterns, it is related to the radiated electric field intensity $E(\theta, \phi)$, known as the antenna *voltage pattern*, according to

$$P(\theta, \phi) = |E(\theta, \phi)|^2$$
(1.3)

For a rectangular aperture with an illumination function that is separable in the two aperture dimensions, $P(\theta, \phi)$ can be factored as the product of separate onedimensional patterns (Stutzman and Thiele, 1998):

$$P(\theta, \phi) = P_{\theta}(\theta) P_{\phi}(\phi)$$
(1.4)

For most radar scenarios, only the *far-field* (also called *Fraunhofer*) power pattern is of interest. The far-field is conventionally defined to begin at a range of D^2/λ or $2D^2/\lambda$ for an antenna of aperture size *D*. Consider the azimuth (θ) pattern of the one-dimensional linear aperture geometry shown in Fig. 1.5. From a signal processing viewpoint, an important property of aperture antennas (such as flat plate arrays and parabolic reflectors) is that the electric field intensity as a function of azimuth $E(\theta)$ in the far field is just the inverse Fourier transform of the distribution A(y) of current across the aperture in the azimuth



FIGURE 1.5 Geometry for one-dimensional electric field calculation on a rectangular aperture.

$$E(\theta) = \int_{-D_y/2}^{D_y/2} A(y) e^{j(2\pi y/\lambda)\sin\theta} dy$$
(1.5)

where the "frequency" variable is $(2\pi/\lambda) \sin\theta$ and is in radians per meter. The idea of spatial frequency is introduced in <u>App. B</u>.

To be more explicit about this point, define $s = \sin\theta$ and $\zeta = y/\lambda$. Substituting these definitions in Eq. (1.5) gives

$$\frac{1}{\lambda} \int_{-D_y/2\lambda}^{D_y/2\lambda} A(\lambda\zeta) e^{j2\pi\zeta s} d\zeta = \hat{E}(s)$$
(1.6)

which is clearly of the form of an inverse Fourier transform. (The finite integral limits are due to the finite support of the aperture.) Because of the definitions of ζ and s, this transform relates the current distribution as a function of aperture position normalized by the wavelength to a spatial frequency variable that is related to the azimuth angle through a nonlinear mapping. It of course follows that

$$A(\lambda\zeta) = \int_{-\infty}^{+\infty} \hat{E}(s)e^{-j2\pi\zeta s}ds$$
(1.7)

The infinite limits in Eq. (1.7) are misleading, since the variable of integration s

= $\sin\theta$ can only range from -1 to +1. Because of this, $\hat{E}(s)$ is zero outside of this range on *s*.

Equation (1.5) is a somewhat simplified expression that neglects a rangedependent overall phase factor and a slight amplitude dependence on range (Balanis, 2005). This Fourier transform property of antenna patterns will, in <u>Chap. 2</u>, allow the use of linear system concepts to understand the effects of the antenna on cross-range resolution and the pulse repetition frequencies needed to avoid spatial aliasing.

An important special case of Eq. (1.5) occurs when the aperture current illumination is a constant, $A(y) = A_0$. The normalized far-field voltage pattern is then the familiar sinc function,

$$E(\theta) = \frac{\sin[\pi (D_y/\lambda)\sin\theta]}{\pi (D_y/\lambda)\sin\theta}$$
(1.8)

If the aperture current illumination is separable, then the far-field is the product of two Fourier transforms, one in azimuth (θ) and one in elevation (ϕ).

The magnitude of $E(\theta)$ is illustrated in Fig. 1.6, along with the definitions for two important figures of merit of an antenna pattern. The angular resolution of the antenna is determined by the width of its mainlobe, and is conventionally expressed in terms of the 3-dB beamwidth. This can be found by setting $E(\theta) = 1/\sqrt{2}$ and solving for the argument $\alpha = \pi(D_y/\lambda) \sin\theta$. The answer can be found numerically to be $\alpha = 1.4$, which gives the value of θ at the -3-dB point as $\theta_0 = 0.445\lambda/D_y$. The 3-dB beamwidth extends from $-\theta_0$ to $+\theta_0$ and is therefore



FIGURE 1.6 One-way radiation pattern of a uniformly illuminated aperture. The 3-dB beamwidth and peak sidelobe definitions are illustrated.

3-dB beamwidth =
$$\theta_3 = 2 \sin^{-1} \left(\frac{1.4\lambda}{\pi D_y} \right) \approx 0.89 \frac{\lambda}{D_y}$$
 radians (1.9)

Thus, the 3-dB beamwidth is 0.89 divided by the aperture size in wavelengths. Note that a smaller beamwidth requires a larger aperture or a shorter wavelength. Typical beamwidths range from as little as a few tenths of a degree to several degrees for a *pencil beam antenna* where the beam is made as narrow as possible in both azimuth and elevation. Some antennas are deliberately designed to have broad vertical beamwidths of several tens of degrees for convenience in wide area search; these designs are called *fan beam antennas*.

The *peak sidelobe* of the pattern affects how echoes from one scatterer affect the detection of neighboring scatterers. For the uniform illumination pattern, the peak sidelobe is 13.2 dB below the mainlobe peak. This is often considered too high in radar systems. Antenna sidelobes can be reduced by use of a nonuniform aperture distribution (Skolnik, 2001), sometimes referred to as *tapering* or *shading* the antenna. In fact, this is no different from the window or weighting functions used for sidelobe control in other areas of signal processing such as digital filter design, and peak sidelobes can easily be reduced to around 25 to 40 dB at the expense of an increase in mainlobe width. Lower sidelobes are possible, but are difficult to achieve due to manufacturing imperfections and inherent design limitations.

The factor of 0.89 in Eq. (1.9) is often dropped, thus roughly estimating the 3-dB beamwidth of the uniformly illuminated aperture as just λ/D_y radians. In fact, this is the 4-dB beamwidth, but since aperture weighting spreads the mainlobe it is a good rule of thumb.

The antenna *power gain* G is the ratio of peak radiation intensity from the antenna to the radiation that would be observed from a lossless, isotropic (omnidirectional) antenna if both have the same input power. Power gain is determined by both the antenna pattern and by losses in the antenna. A useful rule of thumb for a typical antenna is (Stutzman, 1998)

$$G \approx \frac{26,000}{\theta_3 \phi_3} \qquad (\theta_3, \phi_3 \text{ in degrees})$$
$$= \frac{7.9}{\theta_3, \phi_3} \qquad (\theta_3, \phi_3 \text{ in radians})$$

Though both higher and lower values are possible, typical radar antennas have gains from about 10 dB for a broad fan-beam search antenna to approximately 40 dB for a pencil beam that might be used for both search and track.

Effective aperture A_e is an important characteristic in describing the behavior of an antenna being used for reception. If a wave with power density $W \text{ W/m}^2$ is incident on the antenna, and the power delivered to the antenna load is *P*, the effective aperture is defined as the ratio (Balanis, 2005)

$$A_e = \frac{P}{W} \qquad m^2 \tag{1.11}$$

Thus, the effective aperture is the area A_e such that, if all of the power incident on the area was collected and delivered to the load with no loss, it would account for all of the observed power output of the actual antenna. (Note, however, that A_e is *not* the actual physical area of the antenna. It is a fictional area that accounts for the amount of incident power density captured by the receiving antenna.) Effective aperture is directly related to antenna directivity, which in turn is related to antenna gain and efficiency. For most antennas, the efficiency is near unity and the effective aperture and gain are related by (Balanis, 2005)

$$G = \frac{4\pi}{\lambda^2} A_e \tag{1.12}$$

Two more useful antenna concepts are the antenna *phase front* (or *wave front*) and *phase center* (Balanis, 2005; Sherman, 1984). A phase front of a radiating antenna is any surface on which the phase of the field is a constant. In the far-field, the phase fronts are usually approximately spherical, at least over localized regions. The phase center of the antenna is the center of curvature of the phase fronts. Put another way, the phase center is the point at which an isotropic radiator should be located so that the resulting phase fronts best match those of the actual antenna. The phase center concept is useful because it defines an effective location of the antenna, which can in turn be used for analyzing effective path lengths, Doppler shifts, and so forth. For symmetrically illuminated aperture antennas, the phase center will be centered in the aperture plane, but may be displaced forward or backward from the actual aperture. Referring to Fig. 1.5, the phase center would occur at y = 0, but possibly $x \neq 0$, depending on the detailed antenna shape.

Another important type of antenna is the *array antenna*. An array antenna is one composed of a collection of individual antennas called array *elements*.

The elements are typically identical dipoles or other simple antennas with very broad patterns. Usually, the elements are evenly spaced to form a *uniform linear array* as shown in Fig. 1.7. Figure 1.8 illustrates examples of real array and aperture antennas.



FIGURE 1.7 Geometry of the uniform linear array antenna.



FIGURE 1.8 Examples of typical array and aperture antennas. (*a*) Slotted phased array in the nose of an F/A-18 aircraft. This antenna is part of the AN/APG-73 radar system. (*b*) A Cassegrain reflector antenna. (*Image (a) courtesy of Raytheon Corp. Image (b) courtesy of Quinstar Corp. Used with permission.*)

The voltage pattern for the linear array is most easily arrived at by

considering the antenna in its receive, rather than transmit mode. Suppose the leftmost element is taken as a reference point, there are *N* elements in the array, and the elements are isotropic (unity gain for all θ). The signal in branch *n* is weighted with the complex weight a_n . For an incoming electric field $E_0 \exp(j\Omega t)$ at the reference element, the total output voltage *E* can easily be shown to be (Stutzman and Thiele, 1998; Skolnik, 2001)

$$E(\theta) = E_0 \sum_{n=0}^{N-1} a_n e^{j(2\pi/\lambda)nd \sin\theta}$$
(1.13)

This is similar in form to the *discrete Fourier transform* (DFT) of the weight sequence $\{a_n\}$. Like the aperture antenna, the antenna pattern of the linear array thus involves a Fourier transform, this time of the weight sequence (which determines the current distribution in the antenna). For the case where all the $a_n = 1$, the pattern is the familiar "aliased sinc" function, whose magnitude is

$$|E(\theta)| = E_0 \left| \frac{\sin[N(\pi d/\lambda)\sin\theta]}{\sin[(\pi d/\lambda)\sin\theta]} \right|$$
(1.14)

This function is very similar to that of Eq. (1.8) and Fig. 1.6. If the number of elements N is reasonably large (nine or more) and the product Nd is considered to be the total aperture size D, the 3-dB beamwidth is $0.89\lambda/D$, and the first sidelobe is 13.2 dB below the mainlobe peak; both numbers are the same as those of the uniformly illuminated aperture antenna. Of course, by varying the amplitudes of the weights a_n , it is possible to reduce the sidelobes at the expense of a broader mainlobe. The phase center is at the center of the array.

Actual array elements are not isotropic radiators. A simple model often used as a first-order approximation to a typical element pattern $E_{el}(\theta)$ is

$$E_{el}(\theta) \approx \cos\theta \tag{1.15}$$

The right-hand side of Eq. (1.13) is then called the *array factor* $AF(\theta)$, and the composite radiation pattern becomes

$$E(\theta) = AF(\theta)E_{el}(\theta)$$
(1.16)

Because the cosine function is slowly varying in θ , the beamwidth and first sidelobe level are not greatly changed by including the element pattern for signals arriving at angles near broadside (near $\theta = 0^{\circ}$). The element pattern does reduce distant sidelobes, thereby reducing sensitivity to waves impinging on the

array from off broadside.

The discussion so far has been phrased in terms of the transmit antenna pattern (for aperture antennas) or the receive pattern (for arrays), but not both. The patterns described have been *one-way antenna patterns*. The reciprocity theorem guarantees that the receive antenna pattern is identical to the transmit antenna pattern (Balanis, 2005). Consequently, for a monostatic radar, the *two-way antenna pattern* (power or voltage) is just the square of the corresponding one-way pattern. It also follows that the antenna phase center is the same in both transmit and receive modes.

1.3.3 Receivers

It was shown in Sec. 1.3.1 that radar signals are usually narrowband, bandpass, phase- or frequency-modulated functions. This means that the echo waveform r(t) received from a single scatterer is of the form

$$r(t) = A(t) \sin[\Omega t + \theta(t)]$$
(1.17)

where the amplitude modulation A(t) represents only the pulse envelope. The major function of the receiver processing is demodulation of the information bearing part of the radar signal to baseband, with the goal of measuring $\theta(t)$. Figure 1.9 illustrates the conventional approach to receiver design used in most classical radars.



FIGURE 1.9 Conventional quadrature channel receiver model. In this illustration, the lower channel is the in-phase ("I") channel, and the upper is the quadrature phase ("Q") channel.

The received signal is split into two channels. One channel, called the *in-phase* or "I" channel of the receiver (the lower branch in Fig. 1.9) mixes the received signal with an oscillator, called the *local oscillator* (LO), at the radar frequency. This generates both sum and difference frequency components:

$$2\sin(\Omega t) A(t) \sin[\Omega t + \theta(t)] = A(t) \cos[\theta(t)] - A(t) \cos[2\Omega t + \theta(t)]$$

(1.18)

The sum term is then removed by the lowpass filter, leaving only the modulation term $A(t)\cos[\theta(t)]$. The other channel, called the *quadrature* phase or "Q" channel, mixes the signal with an oscillator having the same frequency but a 90° phase shift from the I channel oscillator. The Q channel mixer output is

$$2\cos(\Omega t) A(t)\sin[\Omega t + \theta(t)] = A(t)\sin[\theta(t)] + A(t)\sin[2\Omega t + \theta(t)]$$
(1.19)

which, after filtering, leaves the modulation term $A(t)\sin[\theta(t)]$. If the input r(t) is written as $A(t)\cos[\Omega t + \theta(t)]$ instead, the upper channel of Fig. 1.9 becomes the I channel and the lower the Q channel, with outputs $A(t)\cos[\theta(t)]$ and $-A(t)\sin[\theta(t)]$, respectively. In general, the I channel is the one where the oscillator function (sine or cosine) is the same as that used in modeling the signal.

The reason that both the I and Q channels are needed is that either one alone does not provide sufficient information to determine the phase modulation $\theta(t)$ unambiguously. Figure 1.10 illustrates the problem. Consider the case shown in Fig. 1.10*a*. The signal phase $\theta(t)$ is represented as a solid black phasor in the complex plane. If only the I channel is implemented in the receiver, only the cosine of $\theta(t)$ will be measured. In this case, the true phasor will be indistinguishable from the gray phasor $-\theta(t)$. Similarly, if only the Q channel is implemented so that only the sine of $\theta(t)$ is measured, then the true phasor will be indistinguishable from the gray phasor of Fig. 1.10*b*, which corresponds to $\pi - \theta(t)$. When both the I and Q channels are implemented, the phasor quadrant is determined unambiguously.⁴ In fact, the signal processor will normally assign the I signal to be the real part of a complex signal and the Q signal to be the imaginary part, forming a single complex signal



FIGURE 1.10 (*a*) The I channel of the receiver in Fig. 1.9 measures only the cosine of the phasor $\theta(t)$. (*b*) The Q channel measures only the sine of the phasor.

$$x(t) = I(t) + jQ(t) = e^{j\theta(t)}$$
(1.20)

Equation (1.20) implies a more convenient way of representing the effect of an ideal coherent receiver on a transmitted signal. Instead of representing the transmitted signal by a sine function, an equivalent complex exponential function is used instead.⁵ The echo signal of (1.17) is thus replaced by

$$r(t) = A(t)e^{j[\Omega t + \theta(t)]}$$
(1.21)

The receiver structure of Fig. 1.9 is then replaced with the simplified model of Fig. 1.11, where the echo is demodulated by multiplication with a complex reference oscillator $exp(-j\Omega t)$. This technique of assuming a complex transmitted signal and corresponding complex demodulator produces exactly the same result obtained in Eq. (1.20) by explicitly modeling the real-valued signals and the I and Q channels, but is much simpler and more compact. This complex exponential analysis approach is used throughout the remainder of the book. It is important to remember that this is an analysis technique; actual analog hardware must still operate with real-valued signals only. However, once signals are digitized, they may be treated explicitly as complex signals in the digital processor.



FIGURE 1.11 Simplified transmission and receiver model using complex exponential signals.

Figure 1.9 implies several requirements on a high-quality receiver design. For example, the local oscillator and the transmitter frequencies must be identical. This is usually ensured by having a single *stable local oscillator* (STALO) in the radar system that provides a frequency reference for both the transmitter and the receiver. Furthermore, many types of radar processing require *coherent* operation. The IEEE *Standard Radar Definitions* defines

"coherent signal processing" as "echo integration, filtering, or detection using amplitude and phase of the signal referred to a coherent oscillator" (emphasis added) (IEEE, 1982). Coherency is a stronger requirement than frequency stability. In practice, it means that the transmitted carrier signal must have a fixed phase reference for several, perhaps many, consecutive pulses. Consider a pulse transmitted at time t_1 of the form $a(t - t_1) \sin[\Omega(t - t_1) + \phi]$, where a(t) is the pulse shape. In a coherent system, a pulse transmitted at time t_2 will be of the form $a(t - t_2) \sin[\Omega(t - t_1) + \phi]$. Note that both pulses have the same argument (t $(-t_1) + \phi$ for their sine term; only the envelope term changes location on the time axis. Thus, both sinusoids are referenced to the same absolute starting time and phase. This is as opposed to the second pulse being of the form $a(t - t_2) \sin[\Omega(t - t_2)]$ $(t_1 - t_2) + \phi$, which is nonzero over the same time interval as the coherent pulse $a(t_1 - t_2) + \phi$ $(t - t_2) \sin[\Omega(t - t_1) + \phi]$ and has the same frequency, but has a different phase at any instant in time. Figure 1.12 illustrates the difference visually. In the coherent case, the two pulses appear as if they were excised from the same continuous, stable sinusoid; in the noncoherent case, the second pulse is not in phase with the extension of the first pulse. Because of the phase ambiguity discussed earlier, coherency also implies a system having both I and Q channels.



FIGURE 1.12 Illustration of the concept of a fixed phase reference in coherent signals. (*a*) Coherent pulse pair generated from the reference sinusoid. (*b*) Reference sinusoid. (*c*) Noncoherent pulse pair.

Another requirement is that the I and Q channels have perfectly matched transfer functions over the signal bandwidth. Thus, the gain through each of the two signal paths must be identical, as must be the phase delay (electrical length) of the two channels. Of course, real receivers do not have perfectly matched channels. The effect of gain and phase imbalances will be considered in <u>Chap.</u> <u>3</u>. Finally, a related requirement is that the oscillators used to demodulate the I and Q channels must be exactly in quadrature, that is, 90° out of phase with one another.

In the receiver structure shown in Fig. 1.9, the information-bearing portion of the signal is demodulated from the carrier frequency to baseband in a single mixing operation. While convenient for analysis, pulsed radar receivers are virtually never implemented this way in practice. One reason is that active electronic devices introduce various types of noise into their output signal, such as *shot noise* and *thermal noise*. One noise component, known as *flicker noise* or 1/F noise, has a power spectrum that behaves approximately as F^{-1} and is therefore strongest near zero frequency. Since received radar signals are very weak, they can be corrupted by 1/F noise if they are translated to baseband before being amplified.

Figure 1.13 shows a more representative superheterodyne receiver structure. The received signal, which is very weak, is amplified immediately upon reception using a low-noise amplifier (LNA). The LNA, more than any other component, determines the noise figure of the overall receiver. It will be seen in Sec. 2.3 that this is an important factor in determining the radar's signalto-noise ratio (SNR), so good design of the LNA is important. The key feature of the superheterodyne receiver is that the demodulation to baseband occurs in two or more stages. First, the signal is modulated to an IF, where it receives additional amplification. Amplification at IF is easier because of the greater percentage bandwidth of the signal and the lower cost of IF components compared to microwave components. In addition, modulation to IF rather than to baseband incurs a lower conversion loss, improving the receiver sensitivity, and the extra IF amplification also reduces the effect of flicker noise. Finally, the amplified signal is demodulated to baseband. Some receivers may use more than two demodulation stages (so that there are two or more IF frequencies), but two stages is the most common choice. One final advantage of the superheterodyne configuration is its adaptability. The same IF stages can be used with variable RFs simply by tuning the LO so as to track changes in the transmitted frequency.



FIGURE 1.13 Structure of a superheterodyne radar receiver.

1.4 Common Threads in Radar Signal Processing

A radar system's success or failure in detecting, tracking, and imaging objects or features of interest in the environment is affected by various characteristics of those objects, the environment, and the radar itself, and how they are reflected in the received signals available for processing. Two of the most basic and important signal quality metrics are the signal-to-interference ratio and the resolution. Because of their importance, improving SIR and resolution is the major goal of most of the basic radar signal processing discussed in this text.

While subsequent chapters discuss a wide variety of signal processing techniques, there are a few basic ideas that underlie most of them. These include *coherent* and *noncoherent integration*, target *phase history modeling*, *bandwidth expansion*, and *maximum likelihood estimation*. The remainder of this section gives a heuristic definition of SIR and resolution, and then illustrates the simplest forms of integration, phase history modeling, and bandwidth expansion and how they affect SIR and resolution. Maximum likelihood estimation is deferred to <u>Chap. 9</u> and <u>App. A</u>.

1.4.1 Signal-to-Interference Ratio and Integration

Consider a discrete-time signal x[n] consisting of the sum of a "desired signal" s[n] and an interfering signal w[n]:

$$x[n] = s[n] + w[n]$$
(1.22)

The discussion is identical for continuous time signals. The SIR χ of this signal is the ratio of the power of the desired signal to that of the interference. If s[n] is deterministic, the signal power is usually taken as the peak signal value, and may therefore occur at a specific time t_0 . In some deterministic cases, the average signal power may be used instead. The interference is almost invariably modeled as a random process, so that its power is the mean-square $\mathbf{E}\{|w[n]|^2\}$. If the interference is zero mean, as is very often the case, then the power also equals the variance of the interference, σ_w^2 . If the desired signal is also modeled as a random process, then its power is also taken to be its mean-square or variance.

As an example, let s[n] be a complex sinusoid $A \exp[j\omega n]$ and let w[n] be complex zero mean white Gaussian noise of variance σ_w^2 . The SIR of their sum x[n] is

$$\chi_x = \frac{A^2}{\sigma_w^2} \tag{1.23}$$

In this case, the peak and average signal power are the same. If s[n] is a real-

valued sinusoid $A\cos[\omega n]$ and w[n] is real-valued zero mean white Gaussian noise of variance σ_w^2 , the peak SIR would be the same but the average SIR would be $A^2/2\sigma_w^2$ because the average power of a real cosine or sine function of amplitude A is $A^2/2$.

A variation is the "energy SIR," defined as the ratio of the total energy $E_s = \sum |s[n]|^2$ in the signal s[n] divided by the average noise power:

$$\chi_x = \frac{E_s}{\sigma_w^2} \tag{1.24}$$

The proportionality between E_s and A depends on the signal shape. For a rectangular pulse or a complex exponential of amplitude A and duration N samples, it is just $E_s = N \cdot A^2$. It can be seen in <u>Chap. 6</u> that when matched filters are used, the peak SIR at the filter output is the energy SIR of the original signal.

SIR affects detection, tracking, and imaging performance in different ways. In general, detection performance improves with SIR in the sense that P_D increases for a given P_{FA} as SIR increases. For instance, it will be seen in <u>Chap.</u> <u>6</u> that for one particular model of the target behavior and detection algorithm, P_D is related to P_{FA} according to

$$P_D = (P_{FA})^{-1/(1+\chi)}$$
(1.25)

which shows that $P_D \rightarrow 1$ as $\chi \rightarrow \infty$ for fixed P_{FA} . As another example, the limit on precision (standard deviation of repeated measurements) due to additive noise of typical estimators of range, angle, frequency, or phase tends to decrease as $1/\sqrt{\chi}$; this behavior will be demonstrated in <u>Chap. 9</u>. In radar imaging (<u>Chap.</u> <u>8</u>), SIR directly affects the contrast or dynamic range (ratio of reflectivity of brightest to dimmest visible features) of the image. These considerations make it essential to maximize the SIR of radar data, and many radar signal processing operations discussed in this text have as their primary goal increasing the SIR. The ways in which this is done will be discussed along with each technique.

1.4.2 Resolution

The closely related concepts of *resolution* and a *resolution cell* will arise frequently. Two equal-strength scatterers are considered to be *resolved* if they produce two separately identifiable signals at the system output, as opposed to combining into a single undifferentiated output.⁶ The idea of resolution is applied in range, cross-range, Doppler shift or velocity, and angle of arrival. Two scatterers can simultaneously be resolved in one dimension, say range, and be unresolved in another, perhaps velocity.

Figure 1.14 illustrates the concept of resolution, in this case in frequency.

Part (a) of the figure shows a portion of the positive frequency spectrum of the sum of two unit amplitude cosine functions with zero initial phase, one at 1000 Hz and one at 1500 Hz. This signal could represent the Doppler spectrum of two moving targets with the same echo strength but different radial velocities. The observation time is such that the mainlobe of the sinc function contributed by each has a Rayleigh width (peak to first null width) of 100 Hz. The two vertical dotted lines mark the two cosine frequencies. There are two distinct, wellseparated peaks in the spectrum. The actual frequency of each peak is perturbed very slightly from the expected value by the sidelobes of the other sinusoid. Nonetheless, these two signal components are considered well resolved. Parts (b) through (d) of the figure repeat this measurement with the frequency spacing reduced to 100, 75, and 50 Hz. At 100 Hz spacing the two spectral peaks are still well resolved, though with more perturbation of the apparent frequencies, but as the separation drops below the Rayleigh width to 75 and then to 50 Hz, they blur into a single spectral peak. At 50 Hz, they are no longer resolved; the spectrum measurement does not show two separate signals. At 75 Hz they are marginally resolved, although a little noise added to the data would make that a precarious claim. It appears that a separation of about the Rayleigh width or greater is needed for clear resolution of the two frequencies. This demonstration also suggests that the width of the signature of a single isolated target is the major determinant of the system's resolution.



	600	800	1000	1200	1400		000	000	1000	1200	1400
(C)	Frequency (Hz)				(d)	Frequency (Hz)					

FIGURE 1.14 Resolution of two sinusoids in frequency, each having a Rayleigh width of 100 Hz. (*a*) Well resolved at 500 Hz spacing. (*b*) Well resolved at 100 Hz spacing. (*c*) Marginally resolved at 75 Hz spacing. (*d*) Unresolved at 50 Hz spacing.

The resolution of a radar in turn determines the size of a resolution cell. A resolution cell in range, velocity, or angle is the interval in that dimension that contributes to the echo received by the radar at any one instant. Figure 1.15 illustrates resolution and the resolution interval in the range dimension for a simple constant-frequency pulse. If a pulse whose leading edge is transmitted at time t = 0 has duration τ seconds, then at time t_0 the echo of the leading edge of the pulse will be received from a scatterer at range $ct_0/2$. At the same time, echoes of the trailing edge of the pulse from a scatterer at range $c(t_0 - \tau)/2$ are also received. Any scatterers at intermediate ranges would also contribute to the voltage at time t_0 . Thus, scatterers distributed over $c\tau/2$ in range contribute simultaneously to the received voltage. In order to resolve the contributions from two scatterers into different time samples, they must be spaced by more than $c\tau/2$ meters so that their individual echoes do not overlap in time. The quantity $c\tau/2$ is called the *range resolution* ΔR . Similarly, two- and threedimensional resolution cells can be defined by considering the simultaneous resolution in, say, range, azimuth angle, and elevation angle.



FIGURE 1.15 Geometry for describing conventional pulse range resolution. See text for explanation.

This description of range resolution applies only to unmodulated, constant frequency pulses. As will be seen in <u>Chap. 4</u>, pulse modulation combined with matched filtering can be used to obtain range resolution finer than $c\tau/2$.

Angular resolution in the azimuth and elevation dimensions is determined by the antenna beamwidths in the same planes. Two scatterers at the same range but different azimuth (or elevation) angles will contribute simultaneously to the received signal if they are within the antenna mainlobe and thus are both illuminated at the same time. For the purpose of estimating angular resolution, the mainlobe width is typically taken to be the 3-dB beamwidth θ_3 of the antenna. Thus, the two point scatterers in Fig. 1.16 located at the 3-dB edges of the beam define the angular resolution of the radar. The figure illustrates the relation between the angular resolution in radians and the equivalent resolution in units of distance, which will be called the *cross-range resolution* to denote resolution in a dimension orthogonal to range. The arc length at a radius *R* for an angle subtending θ_3 radians is exactly $R\theta_3$. The cross-range resolution ΔCR is the distance between two scatterers located at the 3-dB edges of the beam, corresponding to the dashed line in Fig. 1.16, and is given by



FIGURE 1.16 The angular resolution is determined by the 3-dB antenna beamwidth θ_3 .

$$\Delta CR = 2R \sin\left(\frac{\theta_3}{2}\right) \approx R\theta_3 \tag{1.26}$$

where the approximation holds when the 3-dB beamwidth is small, which is usually the case for pencil beam antennas. This result is applicable in either the azimuth or elevation dimension.

Three details bear mentioning. First, the literature frequently fails to specify whether one- or two-way 3-dB beamwidth is required or given. The two-way beamwidth should be used for monostatic radar. Second, note that cross-range resolution increases linearly with range, whereas range resolution was a constant. Finally, as with range resolution, it will be seen later (<u>Chap. 8</u>)

that signal processing techniques can be used to improve resolution far beyond the conventional $R\theta$ limit and to make it independent of range as well.

The radar resolution cell volume V is approximately the product of the total solid angle subtended by the 3-dB antenna mainlobe, converted to units of area, and the range resolution. For an antenna having an elliptical beam with azimuth and elevation beamwidths θ_3 and ϕ_3 , this is

$$\Delta V = \pi \left(\frac{R\theta_3}{2}\right) \left(\frac{R\phi_3}{2}\right) \Delta R = \frac{\pi}{4} R^2 \theta_3 \phi_3 \Delta R$$

$$\approx R^2 \theta_3 \phi_3 \Delta R \qquad (1.27)$$

The approximation in the second line of Eq. (1.27) is 27 percent larger than the expression in the first line, but is widely used. Note that resolution cell volume increases with the square of range because of the two-dimensional spreading of the beam at longer ranges.

1.4.3 Data Integration and Phase History Modeling

A fundamental operation in radar signal processing is *integration* of samples to improve the SIR. Both *coherent integration* and *noncoherent integration* are of interest. The former refers to integration of complex (i.e., magnitude and phase) data, while the latter refers to integration based only on the magnitude (or possibly the squared or log magnitude) of the data.

Suppose a pulse is transmitted, reflects off a target, and at the appropriate time the receiver output signal is measured, consisting of a complex echo amplitude $Ae^{i\phi}$ corrupted by additive noise w. The noise is assumed to be a sample of a random process with power σ_w^2 . The single-pulse SNR is

$$\chi_1 = \frac{\text{signal power}}{\text{noise power}} = \frac{A^2}{\sigma_w^2}$$
(1.28)

Now suppose the measurement is repeated N - 1 more times. One expects to measure the same deterministic echo response, but with an independent noise sample each time. Form a single measurement *z* by integrating (summing) the individual measurements; this sum of complex samples, retaining the phase information, is a coherent integration:

$$z = \sum_{n=0}^{N-1} \{Ae^{j\phi} + w[n]\}$$
$$= NAe^{j\phi} + \sum_{n=0}^{N-1} w[n]$$

The power in the integrated signal component is N^2A^2 . Provided the noise samples w[n] are independent of one another and zero mean, the power in the noise component is the sum of the power in the individual noise samples. Further assuming each has the same power σ_w^2 , the total noise power is now $N\sigma_w^2$. The integrated SNR becomes

$$\chi_N = \frac{N^2 A^2}{N \sigma_w^2} = N \left(\frac{A^2}{\sigma_w^2}\right) = N \chi_1$$
(1.30)

Coherently integrating N measurements has improved the SNR by a factor of N; this increase is called the *integration gain*. Later chapters show that, as one would expect, increasing the SNR improves detection and parameter estimation performance. The cost is the extra time, energy, and computation required to collect and combine the N pulses of data.

In coherent integration, the signal components added in phase, i.e., coherently. This is often described as adding on a *voltage* basis, since the amplitude of the integrated signal component increased by a factor of N, with the result that signal power increased by N^2 . The noise samples, whose phases varied randomly, added on a *power* basis. It is the alignment of the signal component phases that allowed the signal power to grow faster than the noise power.

Sometimes the data must be preprocessed to ensure that the signal component phases align so that a coherent integration gain can be achieved. If the target had been moving in the previous example, the signal component of the measurements would have exhibited a Doppler shift, and Eq. (1.29) would instead become

$$z = \sum_{n=0}^{N-1} \{Ae^{j(2\pi f_D n + \phi)} + w[n]\}$$
(1.31)

for some value of normalized Doppler frequency f_D . The signal power in this case will depend on the particular Doppler shift, but except in very fortunate cases will be less than $A^2 N^2$. However, if the Doppler shift is known in advance, the phase progression of the signal component can be compensated before summing:

(1.29)

$$z' = \sum_{n=0}^{N-1} e^{-j2\pi f_D n} \{ A e^{j(2\pi f_D n + \phi)} + w[n] \}$$
$$= N A e^{j\phi} + \sum_{n=0}^{N-1} e^{-j2\pi f_D n} w[n]$$
(1.32)

The phase correction aligns the signal component phases so that they add coherently. The noise phases are still random with respect to one another. Thus, the integrated signal power is again N^2A^2 , while the integrated noise power is again $N\sigma_{w}^2$, and therefore an integration gain of N is again achieved. Compensation for the phase progression so that the compensated samples add in phase is an example of phase *history modeling*: if the sample-to-sample pattern of target echo phases can be predicted or estimated (at least to within a constant overall phase), the data can be modified with a countervailing phase so that the full coherent integration gain is achieved. Phase history modeling is central to many radar signal processing functions and is essential for achieving adequate gains in SNR.

In noncoherent integration, the phases are discarded and some function of the magnitudes of the measured data samples are added, such as the magnitude, magnitude-squared, or log-magnitude. If the magnitude-squared is chosen, then z is formed as

$$z = \sum_{n=0}^{N-1} \left| Ae^{j\phi} + w[n] \right|^{2}$$

= $\sum_{n=0}^{N-1} \left| Ae^{j\phi} \right|^{2} + \sum_{n=0}^{N-1} \left| w[n] \right|^{2} + \sum_{n=0}^{N-1} 2 \operatorname{Re} \{ Ae^{j\phi} \cdot w^{*}[n] \}$
= $NA^{2} + \sum_{n=0}^{N-1} \left| w[n] \right|^{2} + \sum_{n=0}^{N-1} 2 \operatorname{Re} \{ Ae^{j\phi} \cdot w^{*}[n] \}$
(1.33)

The important fact is that phase information in the received signal samples is discarded.

The first line of Eq. (1.33) defines noncoherent square-law integration. The next two lines show that, because of the nonlinear magnitude-squared operation, z cannot be expressed as the sum of a signal-only part and a noise-only part due to the presence of the third term involving cross-products between signal and noise components. A similar situation exists if the magnitude or log-magnitude is chosen for the noncoherent integration. Consequently, a noncoherent integration gain cannot be simply defined as it was for the coherent case.

It is possible to define a noncoherent gain implicitly. For example, in <u>Chap. 6</u> it will be seen that detection of a constant-amplitude target signal in

complex Gaussian noise with a probability of detection of 0.9 and a probability of false alarm of 10^{-8} requires a single-sample SNR of 14.2 dB (about 26.3 on a linear scale). The same probabilities can be obtained by integrating the magnitude of 10 samples each having an individual SNR of only 5.8 dB (3.8 on a linear scale). The reduction of 8.4 dB (a factor of 26.3/3.8 = 6.9) in the required single-sample SNR when 10 samples are noncoherently integrated is the implied noncoherent integration gain.

Noncoherent integration is much more difficult to analyze than coherent integration, typically requiring derivation of the probability density functions of the noise-only and signal-plus-noise cases in order to determine the effect on detection and parameter estimation.

<u>Chapter 6</u> will show that in many useful cases, the noncoherent integration gain is approximately N^{α} , where α ranges from about 0.7 or 0.8 for small N to about 0.5 (\sqrt{N}) for large N, rather than in direct proportion to N. Thus, noncoherent integration is less efficient than coherent integration. This should not be surprising, since not all of the signal information is used.

1.4.4 Bandwidth Expansion

The scaling property of Fourier transforms states that if x(t) has Fourier transform $X(\Omega) = \mathbf{F}\{x(t)\}$, then

$$\mathbf{F}\{x(\alpha t)\} = \frac{1}{|\alpha|} X\left(\frac{\Omega}{\alpha}\right)$$
(1.34)

Equation (1.34) states that if the signal x is compressed in the time domain by the factor $\alpha > 1$, its Fourier transform is stretched (and scaled) in the frequency domain by the same factor (Papoulis, 1987). When $\alpha < 1$, Eq. (1.34) shows that stretching in the time domain results in compression in the frequency domain. This *reciprocal spreading* behavior is illustrated in Fig. 1.17. Part (*a*) shows a sinusoidal pulse with a frequency of 10 MHz and a duration of 1 μ s and its Fourier transform, which is a sinc function centered on 10 MHz and with a Rayleigh mainlobe width of 1 MHz, the reciprocal of the 1 μ s pulse duration. In part (*b*) the pulse has the same frequency but only one-quarter the duration. Its spectrum is still a sinc centered at 10 MHz, but the Rayleigh width is now four times larger at 4 MHz. The spectrum amplitude is also reduced by a factor of four. This effect can also be viewed in the opposite direction: if the signal gets wider in the frequency domain, it must get narrower in the time domain.



FIGURE 1.17 Illustration of reciprocal spreading property of Fourier transforms. (*a*) A sinusoidal pulse and the main portion of its Fourier transform. (*b*) A narrower pulse has a wider transform. See text for details.

Combining the reciprocal spreading property of Fourier transforms with the observation that resolution depends on signal width shows that improving resolution requires increasing "bandwidth" in the opposite Fourier domain. For example, improving range resolution for simple pulses requires using shorter pulses, as was seen in Sec. 1.4.2; but Fig. 1.17 shows that a shorter pulse implies a wider spectrum, i.e., more bandwidth. Conversely, it was also shown in Sec. 1.4.2 that improving resolution in the frequency domain requires a narrower spectrum mainlobe and thus according to Fig. 1.17, a longer observation (more "bandwidth") in the time domain. This behavior holds for any two functions related by a Fourier transform: finer resolution in one domain requires wider support in the opposite domain.

Radar designers have developed techniques for increasing the appropriate bandwidth to obtain improved resolution in various dimensions. For example, improving resolution in range requires increasing waveform bandwidth, which has led to the use of wideband phase- and frequency-modulated waveforms in place of the simple pulse (<u>Chap. 4</u>). Improving cross-range resolution requires viewing a scene over a wide angular interval to increase cross-range spatial

frequency bandwidth, and leads to the synthetic aperture techniques of <u>Chap. 8</u>. Improving velocity (equivalently, Doppler) resolution requires a long time observation and is accomplished with multipulse waveforms. Because the antenna far-field pattern is the Fourier transform of the aperture current distribution, improved angular resolution can be obtained with larger apertures, i.e., bigger antennas.

1.5 A Preview of Basic Radar Signal Processing

There are a number of instances where the design of a component early in the radar signal processing chain is driven by properties of some later component. For example, in <u>Chap. 4</u> it will be seen that the matched filter maximizes SNR; but it is not until the performance curves for the detectors that follow the matched filter are derived that it will be seen that maximizing SNR also optimizes detection performance. Until the detector is considered, it is hard to see exactly how performance depends on SNR. Having seen the major components of a typical pulsed coherent radar system, the most common signal processing operations in the radar signal processing chain are now described heuristically. By sketching out this preview of the "big picture" from beginning to end, it may be easier to understand the motivation for and interrelation of many of the processing operations to be described in later chapters.

Figure 1.18 illustrates one possible sequence of operations in a generic radar signal processor. The sequence shown is not unique, nor is the set of operations exhaustive. In addition, the point in the chain at which the signal is digitized varies in different systems; it might occur as late as the output of the clutter filtering step. The operations can be generally grouped into *signal conditioning and interference suppression; imaging; detection*; and *postprocessing*. Radar signal *phenomenology* must also be considered. In the next few subsections the basic purpose and operation of each block in this signal processing chain is described.



FIGURE 1.18 One example of a generic radar signal processor flow of operations.

1.5.1 Radar Time Scales

Radar signal processing operations take place on time scales ranging from less than a nanosecond to tens of seconds or longer, a range of 10 to 12 orders of magnitude. Different classes or levels of operations tend to operate on significantly different time scales. Figure 1.19 illustrates one possible association of operations and time scale.



FIGURE 1.19 Illustration of the range of time scales over which radar signal processing is performed.

Operations that are applied to data from a single pulse occur on the shortest time scale, often referred to as *fast time* because the sample rate, determined by the instantaneous pulse bandwidth (see <u>Chap. 2</u>), is on the order of hundreds of kilohertz (kHz) to as much as a few gigahertz in some cases. Corresponding sampling intervals range from a few microseconds down to a fraction of a nanosecond, and signal processing operations on these samples therefore tend to act over similar time intervals. Typical fast time operations are digital I/Q signal formation, beamforming, pulse compression or matched filtering, and sensitivity time control.

The next level up in signal processing operations operates on data from multiple pulses. The sampling interval between pulses (the PRI) is typically on the order of tens of microseconds to hundreds of milliseconds, so again operations that involve multiple pulses occupy similar time scales. Due to the much slower sampling rate compared to single-pulse operations, such operations are said to act in *slow time*. Typical operations include coherent and noncoherent integration, Doppler processing of all types, synthetic aperture imaging, and space-time adaptive processing. The idea of slow and fast time will be revisited in the discussion of the data organizational concept of the datacube in <u>Chap. 3</u>.

A group of pulses that are to be somehow combined coherently, for

example via Doppler processing or *synthetic aperture radar* (SAR) imaging, are said to form a *coherent processing interval* (CPI). A still higher level of radar processing acts on data from multiple CPIs and therefore operates on an even longer time scale often called a *dwell* and typically lasting milliseconds to ones or tens of seconds. Operations on this scale include multiple-CPI detection and ambiguity resolution techniques, multilook SAR imaging, and track filtering. Some radars may track detected targets for many seconds or minutes using data from multiple dwells. Track filtering operates in this regime. Finally, some imaging radars may monitor an area over days, months, or even years.

1.5.2 Phenomenology

To design a successful signal processor, the nature of the signals to be processed must be understood. *Phenomenology* refers to the characteristics of the signals received by the radar. Relevant characteristics include signal power, frequency, phase, polarization, or angle of arrival; variation in time and spatial location; and randomness. The received signal phenomenology is determined by both intrinsic features of the physical object(s) giving rise to the radar echo, such as their physical size or their orientation and velocity relative to the radar; and the characteristics of the radar itself such as its transmitted waveform, polarization, or antenna gain. For example, if more power is transmitted a more powerful received echo is expected, all other things being equal.

In <u>Chap. 2</u>, models of the behavior of typical measured signals that are relevant to the design of signal processors are developed. The radar range equation will give a means of predicting signal power. The Doppler phenomenon will predict received frequency. It will be seen that the complexity of the real world gives rise to very complex variations in radar signals; this will lead to the use of random processes to model the signals, and to particular probability density functions that match measured behavior well. A (very) brief overview of the behavior of the variation of ground and sea echo with sensing geometry and radar characteristics will be given. It will also be shown that measured signals can be represented as the convolution of the "true" signal representing the ideal measurement with the radar waveform (in the range dimension) or its antenna pattern (in the azimuth or elevation dimension, both also called cross-range dimension). Thus, a combination of random process and linear systems theory will be used to describe radar signals and to design and analyze radar signal processors.

1.5.3 Signal Conditioning and Interference Suppression

The first several blocks after the antenna in Fig. 1.18 can be considered as signal conditioning operations whose purpose is to improve the SIR of the data prior to detection, parameter measurement, or imaging operations. That is, the intent of these blocks is to "clean up" the radar data as much as possible. This is done in general with a combination of fixed and adaptive *beamforming, pulse*

compression, clutter filtering, and Doppler processing.

Beamforming is applicable when the radar antenna is an array, i.e., when there are multiple phase center signals, or *channels*, available to the signal processor. Fixed beam-forming is the process of combining the outputs of the various available phase centers to form a directive gain pattern, similar to that shown in Fig. 1.6. The high-gain mainlobe and low sidelobes selectively enhance the echo strength from scatterers in the antenna look direction while suppressing the echoes from scatterers in other directions, typically clutter. The sidelobes also provide a measure of suppression of jamming signals so long as the angle of arrival of the jammer is not within the mainlobe of the antenna. By proper choice of the weights used to combine the channels, the mainlobe of the beam can be steered to various look directions, and the tradeoff between the sidelobe level and the mainlobe width (angular resolution) can be varied.

Adaptive beamforming takes this idea a step further. By examining the correlation properties of the received data across channels, it is possible to recognize the presence of jamming and clutter entering the antenna pattern sidelobes and design a set of weights for combining the channels such that the antenna not only has a high-gain mainlobe and generally low sidelobes, but also has a null in the antenna pattern at the angle of arrival of the jammer. Much greater jammer suppression can be obtained in this way. Similarly, it is also possible to increase clutter suppression by this technique. *Space-time adaptive filtering* (STAP) combines adaptive beamforming in both angle and Doppler for simultaneous suppression of clutter and jammer interference. Figure 1.20 illustrates interference suppression using STAP, allowing a previously invisible target signal to be seen and perhaps detected. The two vertical bands in Fig. 1.20*a* represent jammer energy, which comes from a fixed angle of arrival but is usually in the form of relatively wideband noise; thus it is present at all Doppler frequencies observed by the radar. The diagonal band in Fig. 1.20a is due to ground clutter, for which the Doppler shift depends on the angle from the radar to the ground patch contributing energy. Figure 1.20b shows that the adaptive filtering has created nulls along the loci of the jammer and clutter energy, making the target at 0° angle of arrival and 400 Hz Doppler shift apparent. Adaptive interference suppression will be introduced in Chap. 9.


FIGURE 1.20 Example of effect of adaptive beamforming. (*a*) Map of received signal power as a function of angle of arrival and signal Doppler shift. (*b*) Angle-Doppler map after adaptive processing. (*Images courtesy of Dr. W. L. Melvin. Used with permission.*)

Pulse compression is a special case of *matched filtering*. Many radar system designs strive for both high sensitivity in detecting targets and fine range resolution (the ability to distinguish closely spaced targets). Upcoming chapters show that target detectability improves as the transmitted energy increases, and that range resolution improves as the transmitted waveform's instantaneous bandwidth increases. If the radar employs a simple, constant-frequency rectangular envelope pulse as its transmitted waveform the pulse must be lengthened to increase the transmitted energy for a given power level. However, lengthening the pulse also decreases its instantaneous bandwidth, degrading the range resolution. Thus sensitivity and range resolution appear to be in conflict with one another.

Pulse compression provides a way out of this dilemma by decoupling the waveform bandwidth from its duration, thus allowing both to be independently specified. This is done by abandoning the constant-frequency pulse and instead designing a modulated waveform. A very common choice is the linear frequency modulated (linear FM, LFM, or "chirp") waveform, shown in Fig. 1.21*a*. The instantaneous frequency of an LFM pulse is swept over the desired bandwidth during the pulse duration; the frequency may be swept either up or down, but the rate of frequency change is constant.



FIGURE 1.21 (*a*) Linear FM waveform modulation function, showing an increasing instantaneous frequency. (*b*) Output of the matched filter for the LFM waveform of (a).

The matched filter is by definition a filter in the radar receiver designed to maximize the SNR at its output. The impulse response of the filter having this property turns out to be a replica of the transmitted waveform's modulation function that has been reversed in time and conjugated; thus the impulse response is "matched" to the particular transmitted waveform modulation. Pulse compression is the process of designing a waveform and its corresponding matched filter so that the matched filter output in response to the echo from a single point scatterer concentrates most of its energy in a very short duration, thus providing good range resolution while still allowing the high transmitted energy of a long pulse. Figure 1.21b shows the output of the matched filter corresponding to the LFM pulse of Fig. 1.21a; note that the mainlobe of the response is much narrower than the duration of the original pulse. The concepts of matched filtering, pulse compression, and waveform design, as well as the properties of linear FM and other common waveforms, are described in Chap. 4. There it is seen that the 3-dB width of the mainlobe in time is approximately $1/\beta$ seconds, where β is the instantaneous bandwidth of the waveform used. This width determines the ability of the waveform to resolve targets in range. Converted to equivalent range units, the range resolution is given by

$$\Delta R = \frac{c}{2\beta} \tag{1.35}$$

[This is the same as Eq. (1.2) presented earlier.]

Clutter filtering and Doppler processing are closely related. Both are techniques for improving the detectability of moving targets by suppressing interference from clutter echoes, usually from the terrain in the antenna field of view, based on differences in the Doppler shift of the echoes from the clutter and from the targets of interest. The techniques differ primarily in whether they are implemented in the time or frequency domain and in historical usage of the terminology.

Clutter filtering usually takes the form of *moving target indication*, or MTI, which is simply pulse-to-pulse highpass filtering of the radar echoes at a given range to suppress constant components, which are assumed to be due to nonmoving clutter. Extremely simple, very low-order (most commonly first- or second-order) digital filters are applied in the time domain to samples taken at a fixed range but on successive transmitted pulses.

The term "Doppler processing" generally implies the use of the fast Fourier transform algorithm, or occasionally some other spectral estimation technique, to explicitly compute the spectrum of the echo data for a fixed range across multiple pulses. Due to their different Doppler shifts, energy from moving targets is concentrated in different parts of the spectrum from the clutter energy, allowing detection and separation of the targets. Doppler processing obtains more information from the radar signals, such as number and approximate velocity of moving targets, than does MTI filtering. The cost is more required radar pulses, thus consuming energy and timeline, and greater processing complexity. Many systems use both techniques in series. Clutter filtering and Doppler processing are the subjects of <u>Chap. 5</u>.

1.5.4 Imaging

Most people are familiar with the idea of a radar producing "blips" on a screen to represent targets, and in fact systems designed to detect and track moving targets may do exactly that. However, radars can also be designed to compute fine-resolution images of a scene. Figure 1.22 compares the quality routinely obtainable in SAR imagery in the mid-1990s to that of an aerial photograph of the same scene; close examination reveals many similarities and many significant differences in the appearance of the scene at radar and visible wavelengths. Not surprisingly, the photograph is easier for a human to interpret and analyze, since the imaging wavelengths (visible light) and phenomenology are the same as the human visual system. In contrast, the radar image, while remarkable, is monochromatic, offers less detail, and exhibits a "speckled" texture, some seemingly unnatural contrast reversals, and some missing features such as the runway stripes. Given these drawbacks, why is radar imaging of interest?



FIGURE 1.22 Comparison of optical and SAR images of the Albuquerque airport. (*a*) K_u band (15 GHz) SAR image, 3-m resolution. (*b*) Aerial photograph. (Images courtesy of Sandia National Laboratories. Used with permission.)

While radars do not obtain the resolution or image quality of photographic

systems, they have two powerful advantages. First, they can image a scene through clouds and inclement weather due to the superior propagation of RF wavelengths. Second, they can image equally well 24 hours a day since they do not rely on the sun for illumination; they provide their own "light" via the transmitted pulse. If the example of Fig. 1.21 were repeated in the middle of a rainy night, the SAR image on the left would not be affected in any noticeable way, but the optical image on the right would disappear entirely.

To obtain fine-resolution imagery, radars use a combination of highbandwidth waveforms to obtain good resolution in the range dimension and the synthetic aperture radar technique to obtain good resolution in the cross-range dimension. The desired range resolution is obtained while maintaining adequate signal energy by using pulse compression waveforms, usually linear FM. A long pulse that is swept over a large enough bandwidth β and processed using a matched filter can provide very good range resolution according to Eq. (1.35). For example, range resolution of 1 m can be obtained with a waveform swept over 150 MHz. Depending on their applications, modern imaging radars usually have range resolution of 30 m or better; many systems have 10 m or better resolution, and some advanced systems have resolution under 1 m.

For a conventional nonimaging radar, referred to as a *real aperture* radar, the resolution in cross-range is determined by the width of the antenna beam at the range of interest and is given by $R\theta_3$ as shown in Eq. (1.26). Realistic antenna beamwidths for narrow-beam antennas are typically 1° to 3°, or about 17 to 52 mrad. Even at a relatively short imaging range of 10 km, the cross-range resolution that results would be 170 to 520 m, much worse than typical range resolutions and too coarse to produce useful imagery. This poor cross-range resolution is overcome by using SAR techniques.

The synthetic aperture technique refers to the concept of synthesizing the effect of a very large antenna by having the actual physical radar antenna move in relation to the area being imaged. Thus, SAR is most commonly associated with moving airborne or space-based radars, rather than with fixed ground-based radars. Figure 1.23 illustrates the concept. By transmitting pulses at each indicated location, collecting the range data for each pulse, and properly processing it together, a SAR system creates the effect of a much larger phased array antenna being flown along the aircraft flight path. As suggested by Eq. (1.9) (though some details differ in the SAR case), a very large aperture size produces a very narrowly focused effective antenna beam, thus making possible very fine cross-range resolution. The SAR concept is explained more fully in Chap. 8. A more modern and robust viewpoint based on integrating over a range of angles is also given there.



FIGURE 1.23 The concept of synthetic aperture radar.

1.5.5 Detection

The most basic function of a radar signal processor is detection of the presence of one or more targets of interest. Information about the presence of targets is contained in the echoes of the radar pulses. These echoes compete with receiver noise, undesired echoes from clutter signals, and possibly intentional or unintentional jamming. The signal processor must somehow analyze the total received signal and determine whether it contains a desirable target echo and, if so, at what range, angle, and velocity.

Because the complexity of radar signals leads to the use of statistical models, detection of target echoes in the presence of competing interference signals is a problem in statistical decision theory. The theory as applied to radar detection will be developed in <u>Chap. 6</u>. There it will be seen that in most cases optimal performance can be obtained using *threshold detection*. In this method, the magnitude of each complex sample of the radar echo signal, possibly after signal conditioning and interference suppression, is compared to a precomputed threshold. If the signal amplitude is below the threshold, it is assumed to be due to interference signals only. If it is above the threshold, it is assumed that the stronger signal is due to the presence of a target echo in addition to the interference, and a detection or "hit" is declared. In essence, the detector makes a decision as to whether the energy in each received signal sample is too large to likely have resulted from interference alone; if so, it is assumed a target echo contributed to that sample. Figure 1.24 illustrates the concept. The "clutter + target" signal might represent the variation in received signal strength versus

range (fast time) for a single transmitted pulse. It crosses the threshold at three different times, suggesting the presence of three targets at different ranges.



FIGURE 1.24 Illustration of threshold detection.

Because they are the result of a statistical process, threshold detection decisions have a finite probability of being wrong. For example, a noise spike could cross the threshold, leading to a false target declaration, commonly called a *false alarm*. These errors are minimized if the target spikes stand out strongly from the background interference, i.e., if the SIR is as large as possible. If this is the case the threshold can be set relatively high, resulting in few false alarms while still detecting most targets. This fact also accounts for the importance of matched filtering in radar systems. The matched filter maximizes the SIR, thus providing the best threshold detection performance. Furthermore, the achievable SIR increases monotonically with the transmitted pulse energy *E*, thus encouraging the use of longer pulses to get more energy on the target. Since longer simple pulses reduce range resolution, the technique of pulse compression is also important so that fine resolution can be obtained while maintaining good detection performance.

The concept of threshold detection can be applied to many different radar signal processing systems. Figure 1.24 illustrates its application to a fast-time (range) signal trace, but it can be equally well applied to a signal composed of measurements at different Doppler frequencies for a fixed range, or in a two-dimensional form to combined range-Doppler data or to SAR imagery.

There are numerous significant details in implementing threshold detection. Various detector designs work on the magnitude, squared-magnitude, or even log-magnitude of the complex signal samples. The threshold is computed from knowledge of the interference statistics so as to limit false alarms to an acceptable rate. However, in real systems the interference statistics are rarely known accurately enough to allow for precomputing a fixed threshold. Instead, the required threshold is set using interference statistics estimated from the data itself, a process called *constant-false-alarm rate* (CFAR) detection. Detection processing is described in detail in <u>Chap. 6</u>.

1.5.6 Measurements and Track Filtering

Radar systems employ a wide variety of processing operations after the point of target detection. One of the most common postdetection processing steps, and one of the three major functions of interest in this text, is *tracking* of targets, an essential component of many radar systems. Tracking is comprised of (usually multiple) *measurements* of the position of detected targets followed by *track filtering*.

The radar signal processor detects the presence of targets using threshold detection methods. The range, angle, and Doppler resolution cell in which a target is detected provide a coarse estimate of its location in those coordinates. Once detected, the radar will seek to refine the estimated range by using signal processing methods to more precisely estimate the time delay after pulse transmission at which the threshold crossing occurred, the angle of the target relative to the antenna mainbeam direction, or its radial velocity. Individual measurements will have some error due to interference, and so provide a noisy snapshot of the target location and motion at one instant in time.

The term track filtering describes a higher-level, longer time scale process of integrating a series of such measurements to estimate a complete trajectory of the target over time. It is often described as *data processing* rather than signal processing. Because there may be multiple targets with crossing or closely spaced trajectories, track filtering must deal with the problems of determining which measurements to associate with which targets being tracked, and with correctly resolving nearby and crossing trajectories. A variety of optimal estimation techniques have been developed to perform track filtering. An excellent reference in this area is Bar-Shalom (1988).

Figure 1.25 illustrates a series of noisy measurements in one dimension of the position of two targets and the filtering of that noisy trajectory using an extremely simple alpha-beta filter, to be discussed in Chap. 9. The position in the x dimension versus time for each target is shown by the gray lines, so the two targets are moving at different velocities along the x axis and one passes the other at around time step 32. The circle and diamond markers indicate the noisy radar measurements of position for each. The solid black lines are the smoothed estimates of position produced by the alpha-beta filter. In part (a) of the figure, the filter correctly associates the measurements with each target over the observation time. In Fig. 1.25b, the noise variance is higher, causing the filter to incorrectly swap the tracks around time step 40. This represents an error in measurement-to-track *data association*. A variety of techniques are available to attempt to address this problem; a few are discussed in Chap. 9.



FIGURE 1.25 Track filtering of noisy measurements for two targets in one dimension using an alpha-beta filter. Markers show individual measurements. Gray lines are actual position, black lines are filtered position estimates. (*a*) Low measurement noise. (*b*) Tracks incorrectly switch targets in higher measurement noise.

1.6 Radar Literature

This text covers a middle ground in radar technology. It focuses on basic radar signal processing from a digital signal processing point of view. It does not address radar systems, components, or phenomenology in any great depth except where needed to explain the signal processing aspects; nor does it provide indepth coverage of advanced radar signal processing specialties. Fortunately, there are many excellent radar reference books that address both needs. Good books appear every year; those listed in the paragraphs that follow are current as of the year 2013.

1.6.1 Radar Systems and Components

Probably the most classic introductory text to radar systems, now in its third edition, is by Skolnik (2001). The newest and one of the best "radar 101" introductions is the new text by Richards et al. (2010), the first of a three-volume series. The 1990s saw the introduction of several general radar system textbooks. The text by Edde (1995) also has an associated self-study course. Peebles (1998) provides a recent, comprehensive introduction, while Mahafza (2000) provides a number of useful MATLAB[®] files to aid in simulation and experimentation. Morris and Harkness (1996) provides a good introduction to airborne pulsed Doppler systems specifically. A newer discussion of pulsed Doppler systems is given by Alabaster (2012). An up-to-date survey of a broad range of traditional and modern radar applications is given in Scheer and Melvin (2014), showing how many of the techniques discussed in both these introductory texts and the more specialized ones discussed below are brought together into complete systems.

1.6.2 Basic Radar Signal Processing

It is this author's opinion that there are a number of excellent books about radar systems in general, including coverage of components and system designs, and several on advanced radar signal processing topics, especially in the area of synthetic aperture imaging. There have been few books that address the middle ground of basic radar signal processing, such as pulse compression, Doppler filtering, and CFAR detection. Such books are needed to provide greater quantitative depth than is available in the radar systems books without restricting themselves to in-depth coverage of a single advanced application area, and this text aims to fill that gap. Nonetheless, there are a few texts that fit somewhat into this middle area. Nathanson (1991) wrote a classic book, now in its second edition, that covers radar systems in general but in fact concentrates on signal processing issues, especially RCS and clutter modeling, waveforms, MTI, and detection. Probably the closest text in intent to this one is by Levanon (1988), which provides excellent analyses of many basic signal processing functions. The new text by Levanon and Mozeson (2004) addresses the widening variety of radar waveforms in detail. A recent text by Sullivan (2000) is interesting especially for its introductory coverage of both SAR and spacetime adaptive processing (STAP), thus providing a bridge between basic signal processing and more advanced texts specializing in SAR and STAP.

1.6.3 Advanced Radar Signal Processing

Two very active areas of advanced radar signal processing research are SAR imaging and STAP. SAR research extends back to 1951, but only in the 1990s did open literature textbooks begin to appear in the market. There are now many good textbooks on SAR. The first comprehensive text was by Curlander and McDonough (1991). Based on experience gained at the NASA Jet Propulsion Laboratory, it emphasizes space-based SAR and includes a strong component of scattering theory as well. Cumming and Wong (2005) is a newer text that also emphasizes spaced-based SAR. The spotlight SAR mode received considerable development in the 1990s, and two major groups published competing texts in the mid-1990s. Carrara, Goodman, and Majewski (1995) represented the work of the group at the Environmental Research Institute of Michigan (ERIM, now a part of General Dynamics, Inc.); Jakowatz, Jr., et al. (1996) represented the work of a group at Sandia National Laboratories, a unit of the U.S. Department of Energy. Franceschetti and Lanari (1999) provide a compact, unified treatment of both major modes of SAR imaging, namely stripmap and spotlight. The book by Soumekh (1999) is the most complete academic reference on synthetic aperture imaging and includes a number of MATLAB® simulation resources.

STAP, one of the most active radar signal processing research areas, began in earnest in 1973 and is correspondingly less mature than SAR processing. Klemm (1998) wrote the first significant open literature text on the subject. Just as with the Curlander and McDonough book in the SAR community, this book was the first sign that a series of STAP texts can be expected as that research topic matures and reaches mainstream use. The book by Guerci (2003) is the newest primer on this subject at this writing, while Van Trees (2002) prepared a detailed text that continues his classic series on detection and estimation. Additionally, there are other texts on more limited forms of adaptive interference rejection. A good example is the one by Nitzberg (1999), which discusses several forms of sidelobe cancellers. An excellent new book covering a wide range of advanced radar signal processing techniques, including such new topics as multi-input, multi-output radar, and compressive sensing, is the companion volume to (Richards et al., 2010) by Melvin and Scheer (2013).

1.6.4 Radar Applications

The preceding sections have cited a number of books addressing general radar applications, such as imaging or pulse Doppler. There are a number of books in the literature devoted to more specific application areas. The forthcoming text by Melvin and Scheer (2014) will provide an excellent survey of and introduction to a wide range of applications in a single text, and will complete the *Principles of Modern Radar* series.

1.6.5 Current Radar Research

Current radar research appears in a number of scientific and technical journals. The most important in the United States are the Institute of Electrical and Electronics Engineers (IEEE) *Transactions on Aerospace and Electronic Systems, Transactions on Geoscience and Remote Sensing, Transactions on Signal Processing*, and *Transactions on Image Processing*. Radar-related material in the latter is generally limited to papers related to SAR processing, especially interferometric three-dimensional SAR. In the United Kingdom, radar technology papers are often published in the Institution of Engineering and Technology (IET) [formerly the Institution of Electrical Engineers (IEE)] journal *IET Radar, Sonar, and Navigation*.

References

- Alabaster, C. M., *Pulse Doppler Radar: Principles, Technology, Applications.* SciTech Publishing, Raleigh, NC, 2012.
- Balanis, C. A., Antenna Theory, 3d ed., Harper & Row, New York, 2005.

Bar-Shalom, Y., and T. E. Fortmann, *Tracking and Data Association*.

Academic Press, Boston, MA, 1988.

- Bracewell, R. N., *The Fourier Transform and Its Applications*, 3d ed., McGraw-Hill, New York, 1999.
- Brookner, E. (ed.), Aspects of Modern Radar. Artech House, Boston, MA, 1988.
- Carrara, W. G., R. S. Goodman, and R. M. Majewski, Spotlight Synthetic

Aperture Radar. Artech House, Norwood, MA, 1995.

- Cumming, I. G., and F. N. Wong, *Digital Processing of Synthetic Aperture Radar Data*. Artech House, Norwood, MA, 2005.
- Curlander, J. C., and R. N. McDonough, *Synthetic Aperture Radar. J. Wiley*, New York, 1991.
- Edde, B., *Radar: Principles, Technology, Applications*. Prentice Hall PTR, Upper Saddle River, NJ, 1995.
- *EW and Radar Systems Engineering Handbook*, Naval Air Warfare Center, Weapons Division. Available at <u>http://ewhdbks.mugu.navy.mil/</u>.
- Franceschetti, G., and R. Lanari, *Synthetic Aperture Radar Processing*. CRC Press, New York, 1999.
- Guerci, J. R., *Space-Time Adaptive Processing for Radar*. Artech House, Norwood, MA, 2003.
- Institute of Electrical and Electronics Engineers, "IEEE Standard Letter Designations for Radar-Frequency Bands," Standard 521-1976, Nov. 30, 1976.
- Institute of Electrical and Electronics Engineers, "IEEE Standard Radar Definitions," Standard 686-1982, Nov. 30, 1982.
- Jakowatz, C. V., Jr., et al., Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach. Kluwer, Boston, MA, 1996.
- Jelalian, A. V., Laser Radar Systems. Artech House, Boston, MA, 1992.
- Johnson, D. H., and D. E. Dudgeon, Array Signal Processing: Concepts and Techniques. Prentice Hall, Englewood Cliffs, NJ, 1993.
- Klemm, R., *Space-Time Adaptive Processing: Principles and Applications*. INSPEC/IEEE, London, 1998.
- Levanon, N., Radar Principles. J. Wiley, New York, 1988.
- Levanon, N., and E. Mozeson, Radar Signals. J. Wiley, New York, 2004.
- Mahafza, B. R., *Radar Systems Analysis and Design Using MATLAB*[®]. Chapman & Hall/CRC, New York, 2000.
- Melvin, W. L., and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Raleigh, NC, 2013.
- Mir, H. S., and J. D. Wilkinson, "Radar Target Resolution Probability in a Noise-Limited Environment," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44(3), pp. 1234–1239, July 2008.
- Morris, G. V., and L. Harkness (eds.), *Airborne Pulsed Doppler Radar*, 2d ed. Artech House, Boston, MA, 1996.
- Nathanson, F. E., (with J. P. Reilly and M. N. Cohen), *Radar Design Principles*, 2d edition. McGraw-Hill, New York, 1991.
- Nitzberg, R., *Radar Signal Processing and Adaptive Systems*, 2d ed. Artech House, Boston, MA, 1999.
- Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing*, 2d ed. Prentice Hall, Englewood Cliffs, NJ, 1999.
- Papoulis, A., The Fourier Integral and Its Applications, 2d ed. McGraw-Hill,

New York, 1987.

Peebles, Jr., P. Z., Radar Principles. Wiley, New York, 1998.

- Richards, M. A., J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.
- Scheer, J. A. and W. L. Melvin (eds.), *Principles of Modern Radar: Radar Applications*. SciTech Publishing, Edison, NJ, to appear 2014.
- Sherman, S. M., *Monopulse Principles and Techniques*. Artech House, Boston, MA, 1984.
- Skolnik, M. I., *Introduction to Radar Systems*, 3d ed. McGraw-Hill, New York, 2001.
- Soumekh, M., Synthetic Aperture Radar Signal Processing with MATLAB Algorithms. J. Wiley, New York, 1999.
- Stutzman, W. L., "Estimating Gain and Directivity of Antennas," IEEE *Transactions on Antennas and Propagation*, vol. 40(4), pp. 7–11, Aug. 1998.
- Stutzman, W. L., and G. A. Thiele, *Antenna Theory and Design*. J. Wiley, New York, 1998.
- Sullivan, R. J., *Microwave Radar: Imaging and Advanced Concepts*. Artech House, Boston, MA, 2000.
- Swords, S. S., *Technical History of the Beginnings of RADAR*. Peter Peregrinus Ltd., London, 1986.
- Van Trees, H. L., *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory.* J. Wiley, New York, 2002.

Problems

- 1. Compute the range *R* corresponding to echo delays t_0 of 1 ns, 1 µs, 1 ms, and 1 second.
- 2. Compute the time delays for two-way propagation to targets at distances of 100 km, 100 statute miles, and 100 ft.
- 3. Radar is routinely used as one means of measuring the distance to objects in space. For example, it has been used to calculate the orbital parameters and rate of rotation of the planet Jupiter. The distance from Earth to Jupiter varies from 588.5×10^6 to 968.1×10^6 km. What are the minimum and maximum time delays *in minutes* from the time a pulse is transmitted in the direction of Jupiter until the time the echo is received? If pulses are transmitted at a rate of 100 pulses per second, how many pulses are in flight, either on their way to Jupiter or back again, at any given instant?
- 4. Table 1.1 defines the millimeter wave (MMW) band to extend from 40 to 300 GHz. Only certain frequencies in this band are widely used for radar. This is partly due to frequency allocation rules (which frequencies are

allotted to which services), but also due to atmospheric propagation. Based on Fig. 1.3, list two frequencies in the MMW band that might be preferable for radar use, and two that would not be suitable. Explain.

- 5. Compute the bandwidth β needed to achieve range resolutions of 1 m, 1 km, and 100 km. What is the length of a rectangular pulse having this Rayleigh bandwidth (peak-to-first null width of the Fourier transform) for each value of resolution?
- 6. In terms of D_y and λ , what is the peak-to-first null beamwidth (called *Rayleigh beamwidth*) in radians of the antenna pattern for an aperture antenna with constant illumination? Give both the general result, and a small-angle approximation.
- 7. How large must a uniformly illuminated aperture antenna be (value of D_y) in terms of wavelengths so that its 3-dB beamwidth is 1°? What is the estimated gain in decibels of an antenna having azimuth and elevation beamwidths $\theta_3 = \phi_3 = 1^\circ$, based on the approximation in Eq. (1.10)?
- 8. Suppose a police "speed gun" radar has a rectangular antenna. It is desired to have a cross-range resolution ΔCR of 10 ft at a distance of one-quarter mile. What is the required antenna width in inches if the radar frequency is 9.4 GHz? Repeat for 34.4 GHz.
- 9. Continuing problem 8, what is the actual cross-range resolution in feet at each RF if the antenna width is 6 in.?
- 10. Starting from Eq. (1.13) and setting $a_n = 1$, derive Eq. (1.14).
- 11. What is the maximum 3-dB beamwidth θ_3 in degrees such that the approximation for the cross-range resolution, $R\theta_3$, in the last step of Eq. (1.26) has an error of no more than 1 percent?
- 12. Determine the cross-range resolution ΔCR in meters at ranges of 10, 100, and 1000 km for a 3-dB beamwidth $\theta_3 = 3^\circ$.
- 13. Determine the approximate size of a volume resolution cell in cubic meters, ΔV , for R = 20 km, $\Delta R = 100$ m, and $\theta_3 = \phi_3 = 3^\circ$.
- 14. Suppose Eq. (1.31) is modified to consider the magnitude-squared of the signal-plus-noise data:

$$z = \sum_{n=0}^{N-1} \left| A e^{j(2\pi f_D n + \phi)} + w[n] \right|^2$$

Show explicitly that z cannot be expressed as the sum of a signal-only and a noise-only term.

The remaining problems relate to topics covered in <u>Appendix B</u>.

15. What is the Nyquist sampling rate (minimum rate to avoid aliasing) for the signal x(t) having the spectrum X(F) shown in the figure below? Sketch a block diagram of a system for recovering a new signal \hat{x} from samples of x(t) taken at the Nyquist rate, such that the spectrum $\hat{X}(F)$ has the same shape as X(F), but is centered at F = 0.



16. In some cases, the spectrum replication property of sampling can be used as a substitute for demodulation. Given a signal $x_a(t)$ with the spectrum shown, what is the lowest sampling rate that will ensure both no aliasing of the spectrum, and that one of the spectrum replicas is centered at F = 0?



- 17. How many bits are required in an A/D converter to provide a dynamic range of at least 40 dB? What is the expected SQNR with this number of bits, assuming k = 3?
- 18. Numerical values of spatial frequency differ greatly from the usual temporal frequency values. What is the spatial frequency in cycles per meter for a 1-GHz electromagnetic wave?
- 19. What is the instantaneous frequency in hertz of the waveform $x(t) = \exp[j \cdot \exp(-\alpha t)]$?
- 20. Determine a phase function $\psi(t)$ such that the instantaneous frequency in hertz of the waveform $x(t) = \cos[\psi(t)], -\tau/2 \le t \le \tau/2$ seconds, sweeps linearly from $-\beta/2$ Hz to $+\beta/2$ Hz.
- 21. Suppose signal x_1 is 30 dB greater in power than signal x_2 . What is the ratio of their power in linear units (i.e., not in dB)? What is the ratio of the corresponding voltages?

 $\frac{1}{c} = 2.99792458 \times 10^8$ m/s in a vacuum. A value of $c = 3 \times 10^8$ m/s is normally used except where very high accuracy is required.

 $\frac{2}{2}$ Systems exhibiting good or poor resolution are commonly referred to as high- or low-resolution systems, respectively. Since better resolution means a smaller numerical value, in this text the terms "fine" and "coarse" are used instead.

 $\frac{3}{2}$ Gordon Moore's famous 1965 prediction was that the number of transistors on an integrated circuit would double every 18 to 24 months. This prediction has held remarkably true for nearly 40 years, enabling the computing and networking revolutions that began in earnest in the 1980s.

 $\frac{4}{2}$ This is analogous to the use of the two-argument atan2() function instead of the single-argument atan() function in many programming languages such as FORTRAN or C.

 $\frac{5}{2}$ Although these formalizations are not needed for the discussions in this text and are therefore avoided for simplicity, it is worthwhile to note that the complex signal in Eq. (1.21) is the *analytic signal* associated with the real-valued signal of Eq. (1.17). The imaginary part of Eq. (1.21) is the *Hilbert transform* of the real part.

 $\frac{6}{2}$ The effects of unequal signal strength and noise on resolution are considered in (Mir and Wilkinson, 2008).

CHAPTER 2 Signal Models

2.1 Components of a Radar Signal

While a radar transmits a controlled, well-defined signal, the signal measured at the receiver output in response is the superposition of several major components, none of them entirely under the control of the designer. The major components are the *target*, *clutter*, *noise*, and, in some cases, *jamming*. These signals are sometimes subdivided further. For instance, clutter can be separated into ground clutter and weather clutter (such as rain), while jamming can be separated into active jamming (noise transmitters) and passive jamming (such as chaff clouds). Signal processing is applied to this composite signal; the goal is to extract useful information regarding the presence of targets and their characteristics, or to form a radar image. Noise and jamming are interference signals; they degrade the ability to detect targets and measure their position and velocity. Clutter may be interference in some cases, such as when detecting aircraft, or may be the desired signal itself, as with a ground imaging radar. The effectiveness of the signal processing is measured by the improvement it provides in the various figures of merit, such as detection probability, signalto-interference ratio (SIR), or angle accuracy.

It was shown in <u>Chap. 1</u> that conventional pulsed radars transmit narrowband, bandpass signals. Transmitted energy is maximized by restricting amplitude modulation to on-off pulsing; phase modulation is used to expand the instantaneous bandwidth when needed to improve resolution. Thus, an individual transmitted radar pulse can be written as

$$\overline{x}(t) = a(t)\sin[2\pi F_t t + \theta(t)]$$
(2.1)

where a(t) is the constant amplitude pulse envelope, F_t is the radar carrier frequency, and $\theta(t)$ may be a constant or may represent phase modulation of the pulse. It will usually be assumed that a(t) is an ideal, square pulse envelope of amplitude A and duration τ seconds. The instantaneous power of this signal is just $P_s = A^2/2$. The signal at the receiver output will be a combination of echoes of $\overline{x}(t)$ from targets and clutter, noise, and possibly jamming.

Because the target and clutter components are delayed echoes of the transmitted pulse, they are also narrowband signals, although their amplitude and phase modulation will in general be altered, e.g., by propagation loss and Doppler shift. Receiver noise appears as an additive random signal. Thus, the

received signal resulting from a single pulse echoing from a scatterer at range $R_0 = ct_0/2$ can be modeled as

$$\overline{y}(t) = k \cdot a(t - t_0)e^{j[2\pi F_t(t - t_0) + \theta(t - t_0) + \phi(t)]} + n(t)$$
(2.2)

where n(t) = receiver noise

- $k = \frac{\text{echo amplitude factor due to propagation losses and target reflectivity}$
- $\phi(t)$ = echo phase modulation due to target interaction

The important parameters of $\overline{y}(t)$ are the delay time t_0 , the echo component amplitude $k \cdot |a(t)|$ and its power relative to the noise component, and the echo phase modulation function $\theta(t - t_0) + \phi(t)$. These characteristics are used to estimate target range, scattering strength, and radial velocity, suppress jamming and clutter, form images, and so forth.

The amplitude and phase modulation functions also determine the range resolution ΔR of a measurement. For example, $\Delta R = c\tau/2$ if $\theta(t)$ is a constant and $\overline{x}(t)$ is a simple constant-frequency pulse of length τ seconds. Resolution in angle and cross range is determined by the 3-dB width of the antenna pattern in a nonimaging radar.

In order to design good signal processing algorithms, good models of the signals to be processed are needed. In this chapter, an understanding of common radar signal characteristics pertinent to signal processing is developed by presenting models of the effect of the scattering process on the amplitude, phase, and frequency properties of radar measurements. While deterministic models suffice for simple scatterers, it will be seen that complicated real targets require statistical descriptions of the scattering process.

2.2 Amplitude

2.2.1 Simple Point Target Radar Range Equation

The *radar range equation* (Richards et al., 2010; Skolnik, 2001) is a deterministic model that relates received echo power to transmitted power in terms of a variety of system design parameters. It is a fundamental relation used for basic system design and analysis. Since the received signals are narrowband pulses of the form of Eq. (2.2), the received power P_r estimated by the range equation can be directly related to the received pulse amplitude.

To derive the range equation, assume that an isotropic radiating element transmits a waveform of power P_t watts into a lossless medium. Because the transmission is isotropic and no power is lost in the medium, the power density at a range R is the total power P_t divided by the surface area of a sphere of

Isotropic transmitted power density =
$$\frac{P_t}{4\pi R^2}$$
 W/m² (2.3)

Instead of isotropic radiators, real radars use directive antennas to focus the outgoing energy. As described in <u>Chap. 1</u>, the antenna gain G is the ratio of maximum power density to isotropic density. Thus, in the direction of maximum radiation intensity, the power density at range R becomes

Peak transmitted power density =
$$Q_t = \frac{P_t G}{4\pi R^2}$$
 W/m²
(2.4)

This is the power density incident upon the target if it is aligned with the antenna's axis of maximum gain.

When the electromagnetic wave with power density given by Eq. (2.4) is incident upon a single discrete scattering object, or *point target*, at range *R* the incident energy is scattered in various directions; some of it may also be absorbed by the scatterer itself. In particular, some of the incident power is reradiated toward the radar, or *backscattered*. Imagine that the target collects all of the energy incident upon a collector of area σ square meters and reradiates it isotropically. The reradiated power is then

Backscattered power =
$$P_b = \frac{P_t G \sigma}{4\pi R^2} W$$
 (2.5)

The quantity σ is called the *radar cross section* (RCS) of the target. One important fact about RCS is that σ is *not* equal to the physical cross-sectional area of the target; it is an equivalent area that can be used to relate incident power density at the target to the reflected power density that results at the receiver. RCS will be discussed further in <u>Sec. 2.2.3</u>.

Because RCS is defined under the assumption that the backscattered power is reradiated isotropically, the density of the backscattered power at a range R is found by dividing the power of Eq. (2.5) by the surface area of a sphere of radius R as was done in Eq. (2.3), giving the backscattered power density at the radar receiver as

Backscattered power density =
$$Q_b = \frac{P_t G \sigma}{(4\pi)^2 R^4} W/m^2$$
(2.6)

If the effective aperture size of the radar antenna is A_e square meters, the total

backscattered power collected by the receiving antenna will be

Received power =
$$P_r = \frac{P_t G A_e \sigma}{(4\pi)^2 R^4} W$$
 (2.7)

It was shown in <u>Chap. 1</u> that the effective aperture of an antenna is related to its gain and operating wavelength according to $A_e = \lambda^2 G/4\pi$. Thus

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4} W$$
(2.8)

Equation (2.8) describes the power that would be received if an ideal radar operated in free space and used no signal processing techniques to improve sensitivity. Various additional loss and gain factors are customarily added to the formula to account for a variety of additional considerations. For example, losses incurred in various components such as the duplexers, power dividers, waveguide, and radome (a protective covering over the antenna), and propagation effects not found in free space propagation, can be lumped into a *system loss factor* L_s that reduces the received power. System losses are typically in the range of 3 to 10 dB but can vary widely. One of the most important loss factors, particularly at X band and higher frequencies, is atmospheric attenuation $L_a(R)$. Unlike system losses, atmospheric losses are a function of range. If the one-way loss in decibels per kilometer of Fig. 1.3 is denoted by α , the loss in decibels for a target at range R meters (not kilometers) is

$$L_a(R) (dB) = 2\alpha (R/1000) = \alpha R/500 dB$$
(2.9)

In linear units, the loss is therefore

$$L_a(R) = 10^{\alpha R/5000} \tag{2.10}$$

Atmospheric loss can be inconsequential at 10 GHz and moderate ranges, or tens of decibels at 60 GHz and a range of a few kilometers. (This is the reason why 60 GHz is not a popular radar frequency.) This example also shows that, like system losses, atmospheric loss is a strong function of radar frequency.

Incorporating atmospheric and system losses in Eq. (2.8) finally gives

$$P_r = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 L_s L_a(R)} \quad W$$

Equation (2.11) is one simple form of the *radar range equation*. It relates received echo power to fundamental radar system and target parameters such as transmitted power, operating frequency, and antenna gain; radar cross section; and range. Because the power of the radar signal is proportional to the square of the electric field amplitude, the range equation also serves as a model of the amplitude of the target and clutter components of the signal. Note that all variables in Eq. (2.11) are in linear units, not decibels, even though several of the parameters are often specified in decibels; frequent examples include the atmospheric losses, antenna gain, and RCS. Also note that P_r is instantaneous, not average, received power. Finally, realize that for a scatterer at range R, the backscattered EM wave will be received with a time delay of 2R/c seconds after transmission.

As an example, consider an X-band (10-GHz) radar with a peak transmitted power of 1 kW and a pencil beam antenna with a 1° beamwidth, and suppose an echo is received from a jumbo jet aircraft with an RCS of 100 m² at a range of 10 km. The received power can be determined using Eq. (2.11). The antenna gain can be estimated from Eq. (1.10) to be G = 26,000/(1)(1) = 26,000 = 44 dB. The wavelength is $\lambda = c/F = 3 \times 10^8/10 \times 10^9 = 3 \times 10^{-2}$ m = 3 cm. Assuming atmospheric and system losses are negligible, the received power is

$$P_r = \frac{(1000)(26,000)^2(0.03)^2(100)}{(4\pi)^3(10,000)^4} = 3.07 \times 10^{-9} \quad W$$
(2.12)

Even though this example is a large target at short range, the received power is only 3.07 nW, nearly 12 orders of magnitude less than the transmitted power! Nonetheless, this signal level is adequate for reliable detection in many cases. This example illustrates the huge dynamic ranges observed in radar between transmitted and received signal powers.

An important consequence of Eq. (2.11) is that for a point target, the received power decreases as the fourth power of range from the radar to the target. Thus, the ability to detect a target of a given radar cross section decreases rapidly with range. Range can be increased by increasing transmitted power, but because of the R^4 dependence, the power must be raised by a factor of 16 (12 dB) just to double the detection range. Alternatively, the antenna gain can be increased by a factor of 4 (6 dB), implying an increase in antenna area by a factor of 4. On the other hand, designers of "stealth" aircraft and other target vehicles must reduce the RCS σ by a factor of 16 in order to halve the range at which they can be detected by a given radar system.

The range equation is a fundamental radar system design and analysis tool. More elaborate or specialized versions of the equation can be formulated to show the effect of other variables, such as pulse length, *intermediate frequency* (IF) bandwidth, or signal processing gains. Several such variations are given in Richards et al. (2010). The range equation also provides the basis for calibrating a radar system. If the system power, gain, and losses are carefully characterized, then the expected received power of echoes from test targets of known RCS can be computed. Calibration tables equating receiver voltage observed due to those same echoes to incident power density can then be constructed.

Signal processing techniques can increase the effective received power, and therefore increase the obtainable range. The effect of each technique on received power is discussed as they are introduced in later chapters.

2.2.2 Distributed Target Forms of the Range Equation

Not all scattering phenomena can be modeled as a reflection from a single point scatterer. Ground clutter, for example, is best modeled as distributed scattering from a surface, while meteorological phenomena such as rain or hail are modeled as distributed scattering from a three-dimensional volume. The radar range equation can be rederived in a generalized way that accommodates all three cases.

Equation (2.3) is still applicable as a starting point. To consider distributed scatterers, and because the gain of the antenna varies with azimuth and elevation angle, Eq. (2.4) must be replaced with an equation that accounts for the effect of the antenna power pattern $P(\theta, \phi)$ on the power density radiated in a particular direction (θ, ϕ) :

$$Q_t(\theta,\phi) = \frac{P_t P(\theta,\phi)}{4\pi R^2}$$
(2.13)

Assume that the antenna boresight corresponds to $\theta = \phi = 0$. The antenna boresight is normally the axis of maximum gain so that P(0, 0) = G.

Now consider the scattering from an incremental volume dV located at range and angle coordinates (R, θ, ϕ) . Suppose the incremental RCS of the volume element is $d\sigma$ square meters, and that $d\sigma$ in general varies with position in space. The incremental backscattered power from dV is

$$dP_b(\theta,\phi) = \frac{P_t P(\theta,\phi) d\sigma(R,\theta,\phi)}{4\pi R^2}$$
(2.14)

As before, $d\sigma$ is defined such that it is assumed this power is reradiated isotropically, and then collected by the antenna effective aperture, adjusted for the angle of arrival. After substituting for effective aperture and accounting for losses, this results in an incremental received power of

$$dP_r = \frac{P_t P^2(\theta, \phi) \lambda^2 \, d\sigma \left(R, \theta, \phi\right)}{(4\pi)^3 R^4 L_s L_a(R)}$$
(2.15)

Again, this power is received 2R/c seconds after transmission. The total received power is obtained by integrating over all space to obtain a *generalized* radar range equation

$$P_r = \frac{P_t \lambda^2}{(4\pi)^3 L_s} \int_V \frac{P^2(\theta, \phi)}{R^4 L_a(R)} d\sigma(R, \theta, \phi)$$
(2.16)

In Eq. (2.16), the volume of integration V is all of three-dimensional space. However, the backscattered energy from all ranges does not arrive simultaneously at the radar. As discussed in Sec. 1.4.2, only scatterers within a single range resolution cell of extent ΔR contribute significantly to the radar receiver output at any given instant. Thus, a more appropriate form of the generalized radar range equation gives the received power as a function of time,

$$P_r\left(t_0 = \frac{2R_0}{c}\right) = \frac{P_t \lambda^2}{(4\pi)^3 L_s} \int_{\Delta R,\Omega} \frac{P^2(\theta,\phi)}{R^4 L_a(R)} d\sigma(R,\theta,\phi)$$
(2.17)

where ΔR is the range interval of the resolution cell centered at range R_0 and Ω represents integration over the angular coordinates.

By integrating power, it is being assumed that the backscatter from each volume element adds *noncoherently* rather than *coherently*. This means that the power of the composite electromagnetic wave formed from the backscatter of two or more scattering centers is the sum of the individual powers, as opposed to the voltage (electric field amplitude) being the sum of the individual amplitudes, in which case the power would be the square of the voltage sum. Noncoherent addition occurs when the phases of the individual contributors are random and uncorrelated with one another, as opposed to the coherent case when they are in phase. This issue will be revisited in <u>Sec. 2.7</u>.

The general result of Eq. (2.17) is more useful if evaluated for the special cases of point, volume, and area scatterers. Beginning with the point scatterer, the differential RCS in the resolution cell volume is represented by a Dirac impulse function of weight σ :

$$d\sigma(R,\theta,\phi) = \sigma\delta_D(R-R_0,\theta-\theta_0,\phi-\phi_0)dV \text{ (point scatterer)}$$
(2.18)

Using Eq. (2.18) in Eq. (2.17) gives the range equation for a point target at

 $(R_0,\theta_0,\phi_0),$

$$P_{r}(t_{0}) = \frac{P_{t}P^{2}(\theta_{0},\phi_{0})\lambda^{2}\sigma}{(4\pi)^{3}R_{0}^{4}L_{s}L_{a}(R_{0})} \quad \text{(point target)}$$
(2.19)

If the point scatterer is located on the antenna boresight $\theta_0 = \phi_0 = 0$, $P(\theta_0, \phi_0) = G$ and Eq. (2.19) is identical to Eq. (2.11).

Next consider the *volume scattering* case where the RCS seen by the radar is presumed to be due to a distribution of scatterers evenly distributed throughout the volume, rather than associated with a single point. In this case, σ is expressed in terms of RCS per cubic meter, or *volume reflectivity*, denoted as η . The units of reflectivity are m²/m³ = m⁻¹. The RCS of a differential volume element dV is then

$$d\sigma = \eta \, dV = \eta \, R^2 dR \, d\Omega \, \text{(volume scatterer)}$$
(2.20)

where $d\Omega$ is a differential solid angle element. The range equation becomes

$$P_r(t_0) = \frac{P_t \lambda^2 \eta}{(4\pi)^3 L_s} \int_{\Delta R,\Omega} \frac{P^2(\theta,\phi)}{R^2 L_a(R)} dR \, d\Omega$$
(2.21)

If it is assumed that atmospheric loss is slowly varying over the extent of a range resolution cell, then $L_a(R)$ can be replaced by $L_a(R_0)$, where R_0 is the center of the range resolution cell, and removed from the integral. The integral over range that remains is

$$\int_{R_0 - \Delta R/2}^{R_0 + \Delta R/2} \left(\frac{dR}{R^2}\right) = \frac{\Delta R}{R_0^2 - (\Delta R/2)^2} \approx \frac{\Delta R}{R_0^2}$$
(2.22)

provided the range resolution is small compared to the absolute range, which is usually the case. Using Eq. (2.22) in Eq. (2.21) gives

$$P_{r}(t_{0}) = \frac{P_{t}\lambda^{2}\eta \,\Delta R}{(4\pi)^{3}R_{0}^{2}L_{s}L_{a}(R_{0})} \int_{\Omega} P^{2}(\theta,\phi)d\Omega$$
(2.23)

Integration over the angular coordinates requires knowledge of the antenna pattern. One common approximate model of the mainlobe of many antennas is a Gaussian function (Sauvageot, 1992). It can be shown that a good approximation

to the integral in Eq. (2.23) over the cross-range variables for the Gaussian case is (Probert-Jones, 1962)

$$\iint_{\Omega} P^2(\theta,\phi) \sin\phi \, d\theta \, d\phi \approx \frac{\pi \theta_3 \phi_3}{8 \ln 2} G^2 = 0.57 \theta_3 \phi_3 G^2 \tag{2.24}$$

where θ_3 and ϕ_3 are the 3-dB beamwidths in azimuth and elevation. For firstorder calculations, the much simpler assumption is frequently made that the antenna power pattern $P(\theta, \phi)$ is a constant equal to the gain G over the 3-dB beamwidths and zero elsewhere, so that the integral reduces to $G^2\theta_3\phi_3$, a value 2.5 dB higher than that of Eq. (2.24). Using this approximation, Eq. (2.23) reduces to the range equation for volume scatterers:

$$P_r(t_0) = \frac{P_t G^2 \lambda^2 \eta \Delta R \theta_3 \phi_3}{(4\pi)^3 R_0^2 L_s L_a(R_0)} \quad \text{(volume scatterers)}$$
(2.25)

Unlike the point scatterer case described by Eq. (2.11) or (2.19), the received power in the volume scattering case of Eq. (2.25) decreases only as R^2 instead of R^4 . The reason is that the size of the radar resolution cell, which determines the extent of the scatterers contributing to the received power at any one instant, increases as R^2 due to the spreading of the antenna beam in angle at longer ranges.

Finally, the *area scattering* case will be considered. This model is used for the RCS of electromagnetic scattering from the ground, forest, ocean, and other surfaces. The area scattering case must further be divided into two subcases depending on whether the range extent of the scatterers contributing to the echo is limited by the antenna elevation beamwidth or by the range resolution.

First assume that the scattering surface is represented by a flat plane¹ and consider the extent of the mainlobe on the surface. The cross-range extent is simply $R_0\theta_3$, where R_0 is the nominal range to the center of the illuminated area. To estimate the down-range extent, consider Fig. 2.1 that shows the boresight vector intersecting the scattering plane at a *grazing angle* of δ radians. The extent of the beam "footprint" in the down-range dimension is therefore $R_0\phi_3/\sin$ δ meters. Now suppose a pulse of range resolution ΔR is transmitted as shown in Fig. 2.2. Regardless of the antenna footprint, the range extent of scatterers within the resolution cell, and therefore backscattering energy at any instant, is $\Delta R/\cos\delta$ meters.



FIGURE 2.1 Projection of elevation beamwidth onto a horizontal plane at a slant range R_0 and grazing angle δ .



FIGURE 2.2 Projection of range resolution onto a horizontal plane at a slant range R_0 and grazing angle δ .

Scatterers will not contribute significantly to the received signal unless

they are both illuminated (so that there is some backscatter) and within the mainlobe of the antenna (so that their backscatter is not significantly attenuated). Consequently, the effective downrange extent of the resolution cell is the lesser of the range resolution and the elevation beamwidth as each is projected onto the scattering surface. Depending on the relative values of range, range resolution, and grazing angle, either could be the limiting factor. If the range resolution limits the effective extent, the resolution cell is said to be *pulse limited*; if the mainlobe extent is the limiting factor, it is said to be *beam limited*. These two cases are shown in Fig. 2.3. The boundary between the two cases is obtained by equating the pulse length and elevation beam extents as projected onto the ground plane to see which is shorter:



(b)

FIGURE 2.3 Relative geometry of antenna footprint and pulse envelope: (a) beam-limited case and (b) pulse-limited case.

Beam limited:
$$\frac{\Delta R}{R_0} \tan \delta > \phi_3$$

Pulse limited: $\frac{\Delta R}{R_0} \tan \delta < \phi_3$
(2.26)

In area scattering the differential RCS is proportional to the differential area of the scattering surface and can be represented as

$$d\sigma = \sigma^{0}(\delta) \cdot \delta_{D}(R - R_{0})dA$$
(2.27)

where σ^0 (called "sigma nought") is the *area reflectivity* in m²/m² and is therefore dimensionless. The area reflectivity of many surface types is a significant function of the grazing angle δ . The generalized range equation [Eq. (2.17)] becomes

$$P_r(t_0) = \frac{P_t \lambda^2 \sigma^0}{(4\pi)^3 R_0^4 L_s L_a(R_0)} \int_{\Delta A} P^2(\theta, \phi) dA$$
(2.28)

where ΔA is the illuminated area at range R_0 .

If the illuminated area is beam limited, applying the geometry of Fig. 2.3*a* to the differential scattering element at range R_0 shows that the area contributing to the backscatter at one instant is $R^2\phi_3\theta_3/\sin\delta$. Thus, a differential area contributing to the received power is of the form

$$dA = \frac{R_0^2}{\sin\delta} d\theta \, d\phi \qquad \text{(beam-limited case)}$$
(2.29)

Applying this to Eq. (2.28) and again using the constant-gain approximation to the antenna 3-dB beamwidth gives the beam-limited range equation for area scatterers:

$$P_r(t_0) = \frac{P_t G^2 \lambda^2 \phi_3 \theta_3 \sigma^0}{(4\pi)^3 R_0^2 L_s L_a(R_0) \sin \delta}$$
(area scatterers, beam-limited case)
(2.30)

If the illuminated area is pulse limited, the geometry of Fig. 2.3b shows that the area contributing to the backscatter at one instant is $R\theta_3\Delta R/\cos\delta$. The differential contribution is thus

$$dA = \frac{R_0 \Delta R}{\cos \delta} d\theta$$
 (pulse-limited case) (2.31)

The first-order approximation of constant gain over the mainlobe can be used again, though the integral over ϕ is now limited to the range that covers the extent of the pulse on the ground. Equation (2.28) becomes

$$P_r(t_0) = \frac{P_t G^2 \lambda^2 \sigma^0 \Delta R \theta_3}{(4\pi)^3 R_0^3 L_s L_a(R_0) \cos\delta}$$
(area scatterers, pulse-limited case)
(2.32)

Note that power varies as R^{-2} in the beam-limited case because, as with the volume scattering, the resolution cell size grows in both cross-range and down-range extent with increasing range. In the pulse-limited case, power varies as R^{-3} because the resolution cell extent increases in only the cross-range dimension with increasing range.

If the range of interest varies by a large amount, there will be significant variation in the grazing angle δ and therefore in both the antenna beam and pulse footprint extents. For instance, for a radar at a constant altitude *h* and a slant range *R* to the ground, $\sin \delta = h/R$. As *R* increases, the beam-limited antenna footprint area will then increase as R^3 instead of R^2 so that the clutter power would be expected to fall only as R^{-1} . However, σ^0 may also vary significantly with grazing angle (see Section 2.3.1). Additional complications occur when *R* increases so much that a radar that was beam limited at a relatively short range and steep grazing angle becomes pulse limited at a longer range and shallower grazing angle, or the grazing angle falls below the "critical angle". Consequently, the received clutter power may fall off at various rates from R^{-1} to R^{-3} or even more rapidly at very shallow angles (Long, 2001; Currie, 2010).

2.2.3 Radar Cross Section

Section 2.2.1 introduced the radar cross section to heuristically account for the amount of power reradiated by the target back toward the radar transmitter. To restate the concept, assume that the incident power density at the target is Q_t and the backscattered power density at the transmitter is Q_b . If that backscattered power density resulted from isotropic radiation from the target, it would have to satisfy

$$Q_b = \frac{P_b}{4\pi R^2}$$

(2.33)

for some total backscattered power P_b . RCS is the *fictional* area over which the transmitted power density Q_t must be intercepted to collect a total power P_b that would account for the received power desnity. In other words, σ must satisfy

$$P_b = \sigma Q_t \tag{2.34}$$

Combining Eqs. (2.33) and (2.34) gives

$$\sigma = 4\pi R^2 \frac{Q_b}{Q_t} \tag{2.35}$$

This definition is usually written in terms of electric field amplitude. Also, in order to make the definition dependent only on the target characteristics, range is eliminated by taking the limit as R tends to infinity. Thus, the formal definition of radar cross section becomes (Knott et al., 1985)

$$\sigma = 4\pi \lim_{R \to \infty} \left[R^2 \frac{|\mathbf{E}^b|^2}{|\mathbf{E}^t|^2} \right]$$
(2.36)

where \mathbf{E}^{b} and \mathbf{E}^{t} are the backscattered and transmitted electric field complex amplitudes, respectively.

The RCS just defined is a single real scalar number. Implicit in the definition is the use of a single polarization of the transmitted wave and a single receiver polarization, usually the same as the transmitted polarization. However, the polarization state of a transverse electromagnetic plane wave is a two-dimensional vector, and therefore two orthogonal polarization basis vectors are required to fully describe the wave. The most common basis choices are linear (horizontal and vertical polarizations) and circular (left and right rotating polarizations). Furthermore, a general target will modify the polarization of an incident wave, so that the energy backscattered from, say, the vertical component of the incident wave may have both vertical and horizontal components. To account fully for polarization effects, RCS must be generalized to the *polarization scattering matrix* (PSM) S, which relates the complex amplitudes of the incident and backscattered fields. For a radar using a linear polarization basis this relation is (Knott et al., 1985; Mott, 1986; Holm, 1987)

$$\begin{bmatrix} \mathbf{E}_{H}^{b} \\ \mathbf{E}_{V}^{b} \end{bmatrix} = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{H}^{t} \\ \mathbf{E}_{V}^{t} \end{bmatrix}$$
$$= \mathbf{S} \begin{bmatrix} \mathbf{E}_{H}^{t} \\ \mathbf{E}_{V}^{t} \end{bmatrix}$$
(2.37)

Instead of a single real number, the target backscattering characteristics are now described by four complex numbers. If the radar transmitted and received, say, only the vertical component, then the RCS σ would be related to S by

$$\sigma = \left| S_{VV} \right|^2 \tag{2.38}$$

Radars can be designed to measure the full complex PSM. Other designs measure the magnitudes but not the phases of the elements of the PSM, or the magnitudes of two of the PSM elements. These *polarimetric* measurements can be used for a variety of target analysis purposes. However, a discussion of polarimetric techniques is beyond the scope of this book. Henceforth, it will be assumed that only a single fixed polarization is transmitted and a single fixed polarization received, and consequently that RCS is described by a scalar, rather than matrix, function. The reader is referred to the works by Holm (1987) and Mott (1986) for discussions of polarimetric radars and polarimetric signal processing.

Typical values of RCS for targets of interest range from 0.01 m² (-20 dB with respect to 1 m², or -20 dBsm) to hundreds of square meters (\geq +20 dBsm). Both larger and smaller values are also observed. Table 2.1 lists representative RCS values for various types of targets.

Target	RCS, m ²	RCS, dBsm
Conventional unmanned winged missile	0.5	-3
Small single-engine aircraft	1	0
Small fighter aircraft or 4-passenger jet	2	3
Large fighter aircraft	6	8
Medium bomber or jet airliner	20	13
Large bomber or jet airliner	40	16
Jumbo jet	100	20
Small open boat	0.02	-17
Small pleasure boat	2	3
Cabin cruiser	10	10
Large ship at zero grazing angle	10,000+	40+
Pickup truck	200	23
Automobile	100	20
Bicycle	2	3
Man	1	0
Bird	0.01	-20
Insect	0.00001	-50

Source: After Skolnik (2001).

TABLE 2.1	Typical RCS	Values at Microv	wave Frequencies

2.2.4 Radar Cross Section for Meteorological Targets

The field of radar meteorology expresses the reflectivity of weather targets such as rain or snow in terms of a normalized factor called the *reflectivity* (here called the *volume reflectivity*) and usually represented with the symbol Z (Sauvageot, 1992; Doviak and Zrnic, 1993). Weather targets are an example of volume clutter. The actual observed echo is the composite backscatter of many raindrops, suspended water particles, hailstones, or snowflakes in the radar's resolution cell.

Suppose the RCS of the *i*th individual scatterer is σ_i and assume noncoherent addition. Then the total RCS of a volume *V* containing *N* such scatterers is $\Sigma \sigma_i$ and the volume reflectivity is

$$\eta = \frac{1}{\Delta V} \sum_{i=1}^{N} \sigma_i$$
(2.39)

Water droplets are often modeled as small conducting spheres. When the

ratio of the sphere radius *a* to the radar wavelength λ is small, specifically $2\pi a/\lambda \ll 1$, the radar cross section associated with the *i*th scatterer can be expressed as

$$\sigma_i = \frac{\pi^5 |K|^2 D_i^6}{\lambda^4} \tag{2.40}$$

where D_i is the drop diameter, usually given in millimeters, and

$$K = \frac{m^2 - 1}{m^2 + 2} \tag{2.41}$$

with *m* the complex index of refraction. The index of refraction is a function of both the temperature and wavelength. However, for wavelengths between 3 and 10 cm [radar frequencies between X band (10 GHz) and C band (3 GHz)] and temperatures between 0 and 20°C, the value of $|K|^2$ is approximately a relatively constant 0.93 for scatterers composed of water and 0.197 for ice. Substituting Eq. (2.40) in Eq. (2.39) gives

$$\eta = \frac{1}{\Delta V} \sum_{i=1}^{N} \frac{\pi^{5} |K|^{2} D_{i}^{6}}{\lambda^{4}} = \frac{\pi^{5} |K|^{2}}{\lambda^{4}} \frac{1}{\Delta V} \sum_{i=1}^{N} D_{i}^{6}$$
(2.42)

Now define the quantity

$$Z \equiv \frac{1}{\Delta V} \sum_{i=1}^{N} D_i^6$$
(2.43)

Z is called the *reflectivity factor* and is usually expressed in units of mm⁶/m³. Due to the large range of values observed for Z, it is commonly expressed on a decibel scale and denoted as dBZ. Using this definition in Eq. (2.42) gives the following expression for the observed RCS

$$\eta = \frac{\pi^5 |K|^2}{\lambda^4} Z \tag{2.44}$$

Thus, given a measured echo power, the radar range equation can be used to estimate η , and then Eq. (2.44) can be used to convert η to Z.

Because it is related only to the volume density and size of scatterers, meteorologists prefer to express radar echo strength in terms of the reflectivity Z

rather than the RCS η . The value of Z can then be related to the amount of water in the air or the precipitation rate. A number of models are used to relate the observed values of Z to rain rates. These models depend on the type of precipitation, e.g., snow versus thunderstorm rain versus orographic² rain. A common model is that of <u>Table 2.2</u>, which shows the six-level equivalence between observed Z values (in dBZ) and rainfall rates used in the U.S. NEXRAD national weather radar system. Very similar scales are used in the commercial "Doppler weather radar" systems familiar to every watcher of television weather forecasts.

Level	Rain Fall Rate, mm/hr	Reflectivity, dBZ	Category
1	0.49 to 2.7	18 to <30	Light mist
2	2.7 to 13.3	30 to <41	Moderate
3	13.3 to 27.3	41 to <46	Heavy
4	27.3 to 48.6	46 to <50	Very heavy
5	48.6 to 133.2	50 to <57	Intense
6	133.2 and greater	57 and above	Extreme

 TABLE 2.2
 Correspondence between dBZ Reflectivity and Rain Rate

It is important to note that the dBZ values in <u>Table 2.2</u> are 10 times the base 10 logarithm of Z in mm⁶/m³. When Z is given in m⁶/m³ = m³, it must be multiplied by 10^{18} to convert it to units of mm⁶/m³ before converting to a decibel scale and using <u>Table 2.2</u>.

2.2.5 Statistical Description of Radar Cross Section

The radar cross section of real targets cannot be effectively modeled as a simple constant. In general, RCS is a complex function of aspect angle, frequency, and polarization, even for relatively simple scatterers. For example, the conducting trihedral corner reflector of Fig. 2.4 is often used as a calibration target in field measurements. Its RCS when viewed along its axis of symmetry (looking "into the corner") can be determined theoretically; it is (Knott et al., 1985)



FIGURE 2.4 Square trihedral corner reflector.

$$\sigma = \frac{12\pi a^4}{\lambda^2} \tag{2.45}$$

Thus the RCS increases with increasing frequency. On the other hand, at least one frequency-and aspect-independent scatterer exists. The RCS of a conducting sphere of radius *a* is a constant πa^2 , provided $a \gg \lambda$. It is independent of aspect angle because of the spherical symmetry.

A simple example of frequency and aspect dependence is the two-scatterer "dumbbell" target of Fig. 2.5. If the nominal range R is much greater than the separation D, the range to the two scatterers is approximately



FIGURE 2.5 Geometry for determining relative RCS of a "dumbbell" target.

$$R_{1,2}(\theta) \approx R \pm \frac{D}{2} \sin \theta \tag{2.46}$$

If the signal $a \cdot \exp(j2\pi Ft)$ is transmitted, the echo from each scatterer will be proportional to $a \cdot \exp[j2\pi F(t-2R_{1,2}(\theta))/c]$. The voltage $\overline{y}(t)$ of the composite

echo is therefore proportional to RCS is proportional to the power of the composite echo. Taking the squared magnitude of Eq. (2.47) and simplifying leads to the result

$$\overline{y}(t) = a e^{j2\pi F(t-2R_1(\theta)/c)} + a e^{j2\pi F(t-2R_2(\theta)/c)}$$

$$= a e^{j2\pi F(t-2R/c)} [e^{-j\pi FD\sin\theta/c} + e^{+j\pi FD\sin\theta/c}]$$

$$= 2a e^{j2\pi F(t-2R/c)} \cos(\pi FD\sin\theta/c)$$
(2.47)

$$\sigma = 4a^2 \left| \cos(\pi FD \sin \theta / c) \right|^2 = 4a^2 \left| \cos(\pi D \sin \theta / \lambda) \right|^2$$
(2.48)

Equation (2.48) shows that the RCS is a periodic function of both radar frequency and aspect angle. The larger the scatterer separation in terms of wavelengths, the more rapidly the RCS varies with angle or frequency. An exact calculation of the variation in RCS of the dumbbell target is plotted in Fig. 2.6 for the case $D = 10\lambda$ and R = 10,000D. The plot has been normalized so that the maximum value corresponds to 0 dB. Notice the multilobed structure as the varying path lengths traversed by the echoes from the two scatterers cause their echoes to shift between constructive and destructive interference. Also note that the maxima at aspect angles of 90° and 270° (the two "end fire" cases) are the broadest, while the maxima at the two "broadside" cases of 0° and 180° are the narrowest. Figure 2.7 plots the same data in a more traditional polar format.


FIGURE 2.6 Relative radar cross section of the "dumbbell" target of <u>Fig. 2.5</u> when $D = 10\lambda$ and R = 10,000D.



FIGURE 2.7 Polar plot of the data of Fig. 2.6.

The relative RCS of a target with multiple scatterers can be computed as a function of θ and λ using a generalization of Eq. (2.47). Suppose there are N scatterers, each with its own RCS σ_i , located at ranges $R_i(\theta)$ from the radar. Note that the ranges R_i vary with aspect angle θ . The complex voltage of the echo will be, to within a proportionality constant

$$\overline{y}(t) = \sum_{i=1}^{N} \sqrt{\sigma_i} e^{j2\pi F(t - 2R_i(\theta)/c)}$$

$$= e^{j2\pi Ft} \sum_{i=1}^{N} \sqrt{\sigma_i} e^{-j4\pi FR_i(\theta)/c}$$

$$= e^{j2\pi Ft} \sum_{i=1}^{N} \sqrt{\sigma_i} e^{-j4\pi R_i(\theta)/\lambda}$$
(2.49)

Each term in Eq. (2.49) represents the echo from a single scatterer. This equation highlights an extremely important characteristic of these echoes. In addition to being scaled by the scatterer reflectivity, each term is also phase-

shifted relative to the carrier by the amount $-4\pi R/\lambda$ radians. As will be seen, this range-dependent phase shift provides the basis for most fundamental coherent radar signal processing operations such as Doppler processing, imaging, and adaptive beamforming. The phase shift is a very sensitive, but also very ambiguous, indicator of range changes since every $\lambda/2$ change in range produces a 2π change in phase.

The RCS σ is proportional to \overline{y}^2 . Define

$$\zeta \equiv \left| \overline{y} \right| = \left| \sum_{i=1}^{N} \sqrt{\sigma_i} e^{-j4\pi R_i(\theta)/\lambda} \right|$$
(2.50)

and

$$\sigma = \zeta^2 = \left| \sum_{i=1}^N \sqrt{\sigma_i} e^{-j4\pi R_i(\theta)/\lambda} \right|^2$$
(2.51)

RCS variations like those of Fig. 2.6 become very complicated for complex targets having many scatterers of varying individual RCS. Figure 2.8 shows a "target" consisting of 50 point scatterers randomly distributed within a rectangle 5 m wide and 10 m long. The RCS of each individual point scatterer is a constant, $\sigma_i = 1.0$. Figure 2.9 shows the relative RCS, computed at 0.2° increments using Eq. (2.51), which results when this target is viewed 10 km from its center at a frequency of 10 GHz. The dynamic range is similar to that of the simple dumbbell target, but the lobing structure is much more complicated.³



FIGURE 2.8 Random distribution of 50 scatterers used to obtain <u>Fig. 2.9</u>. See text for additional details.



FIGURE 2.9 Relative RCS of the complex target of <u>Fig. 2.8</u> at a range of 10 km and radar frequency of 10 GHz.

The complicated variation of RCS with radar frequency and target aspect angle observed for even moderately complex targets leads to the use of a statistical description for radar cross section (Levanon, 1988; Nathanson, 1991; Skolnik, 2001). This means that the RCS σ of the scatterers within a single resolution cell is considered to be a random variable with a specified *probability density function* (PDF). The mean or median RCS is typically used for radar range equation calculations, but the full PDF is needed for detection probability calculations, as will be seen in <u>Chap. 6</u>.

One of a variety of PDFs is used to describe the statistical behavior of the RCS for different targets. Consider first a target consisting of a large number of individual scatterers (similar to that of Fig. 2.8), each with its own individual but fixed RCS and randomly distributed in space. Because of its high sensitivity to small range changes, the phase of the echoes from the various scatterers can be assumed to be a random variable distributed uniformly on $(0, 2\pi]$. Under these circumstances, the central limit theorem guarantees that the real and imaginary parts of the composite echo can each be assumed to be independent, zero mean Gaussian random variables with the same variance, say α^2 (Papoulis and Pillai, 2001; Beckmann and Spizzichino, 1963). In this case, the squared-

magnitude σ has an exponential PDF:

$$p_{\sigma}(\sigma) = \begin{cases} \frac{1}{\overline{\sigma}} \exp\left[\frac{-\sigma}{\overline{\sigma}}\right] & \sigma \ge 0\\ 0 & \sigma < 0 \end{cases}$$
(2.52)

where $\bar{\sigma} = 2\alpha^2$ is the mean value of the RCS σ . The corresponding amplitude ζ has a Rayleigh PDF:

$$p_{\zeta}(\zeta) = \begin{cases} \frac{2\zeta}{\overline{\sigma}} \exp\left[\frac{-\zeta^2}{\overline{\sigma}}\right] & \zeta \ge 0\\ 0 & \zeta < 0 \end{cases}$$
(2.53)

The phase of the complex echo will be uniformly distributed over $(0,2\pi]$.

While the exponential model for RCS is only strictly accurate in the limit of a very large number of scatterers, in practice it can be a good model for a target having as few as 10 or 20 significant scatterers. Figure 2.10 compares a histogram of the RCS values from Fig. 2.9 to an exponential curve of the form of Eq. (2.52) having the same mean $\bar{\sigma}$. Even though only 50 scatterers are used, the fit of the total RCS histogram to the Rayleigh/exponential distribution is quite good. This same effect is observed when the randomly distributed scatterers also have random individual cross sections drawn from the same Gaussian distribution, a somewhat more general and plausible situation than the fixed-RCS case.



FIGURE 2.10 Histogram of RCS data of <u>Fig. 2.9</u> and exponential PDF with the same mean.

Instead of an ensemble of equal-strength scatterers, some radar targets are better modeled as consisting of one or a few dominant scatterers contributing most of the RCS, modified by the contributions of a number of smaller scatterers. Many PDFs have been advocated and used to model these targets. <u>Table 2.3</u> summarizes several of the more common models for the variability of RCS with aspect angle and radar frequency for targets with and without dominant scatterers.⁴ The mean value $\bar{\sigma}$ of RCS is given for each case in which the PDF is not written explicitly in terms of $\bar{\sigma}$. The variance var(σ) is also given for each case. Other common PDFs not included in <u>Table 2.3</u> are the non-central chi-square and K distributions. Additional information on these PDFs and their relationships, including more conventional forms and in some cases the characteristic functions, is given in <u>App. A</u>. Additional discussion of some of these PDFs in the context of clutter modeling is also given in <u>Sec. 2.3.4</u>.

Model Name	PDF for RCS σ	Comment		
One-Parameter PDF	ŝ			
Nonfluctuating, Marcum, Swerling 0, or Swerling 5	$p_{\sigma}(\sigma) = \delta_D(\sigma - \overline{\sigma})$ $var(\sigma) = 0$	Constant echo power, e.g., calibration sphere or perfectly stationary target with no radar or target motion.		
Exponential, chi-square of degree 2	$p_{\sigma}(\sigma) = \frac{1}{\overline{\sigma}} \exp\left[\frac{-\sigma}{\overline{\sigma}}\right]$ $\operatorname{var}(\sigma) = \overline{\sigma}^{2}$	Many scatterers, randomly distributed, none dominant. Used in Swerling case 1 and 2 models.		
Chi-square of degree 4	$p_{\sigma}(\sigma) = \frac{4\sigma}{\overline{\sigma}} \exp\left[\frac{-2\sigma}{\overline{\sigma}}\right]$ $\operatorname{var}(\sigma) = \overline{\sigma}^{2}/2$	Approximation to case of many small scatterers + one dominant. Used in Swerling case 3 and 4 models.		
Two-Parameter PDF	s			
Chi-square of degree 2 <i>m</i> , Weinstock	$p_{\sigma}(\sigma) = \frac{m}{\Gamma(m)\overline{\sigma}} \left[\frac{m\sigma}{\overline{\sigma}} \right]^{m-1} \exp\left[\frac{-m\sigma}{\overline{\sigma}} \right]$ $\operatorname{var}(\sigma) = \overline{\sigma}^{2}/m$	Generalization of the two preceding cases. Weinstock cases correspond to $0.6 \le 2m \le 4$. Higher degrees correspond to presence of a more dominant single scatterer.		
Noncentral chi-square of degree 2	$p_{\sigma}(\sigma) = \frac{1}{\overline{\sigma}} (1+a^2) \exp\left[-a^2 - \frac{\sigma}{\overline{\sigma}} (1+a^2)\right]$ $\times I_0 \left[2a\sqrt{(1+a^2)(\sigma/\overline{\sigma})} \right]$ $\operatorname{var}(\sigma) = \frac{(1+2a^2)}{(1+a^2)^2} \overline{\sigma}^2$	Exact solution for one dominant scatterer plus many small ones. Corresponds to Rice amplitude PDF. Ratio of dominant RCS to sum of small RCS is a^2 . $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero.		
Weibull	$p_{\sigma}(\sigma) = CB\sigma^{C-1} \exp \left[-B\sigma^{C}\right]$ $\bar{\sigma} = \Gamma(1+1/C)B^{-1/C}$ $\operatorname{var}(\sigma) = B^{-2/C}[\Gamma(1+2/C) - \Gamma^{2}(1+1/C)]$	Empirical fit to many measured targe and clutter distributions. Can have longer "tail" than previous cases. Not readily expressible in terms of d		
Log-normal	$p_{\sigma}(\sigma) = \frac{1}{\sqrt{2\pi s\sigma}} \exp\left[-\ln^2(\sigma/\sigma_m)/2s^2\right]$ $\bar{\sigma} = \sigma_m \exp(s^2/2)$ $\operatorname{var}(\sigma) = \sigma_m^2 \exp(s^2)[\exp(s^2) - 1]$ $= \bar{\sigma}^2[\exp(s^2) - 1]$	Empirical fit to many measured target and clutter distributions. "Tail" is longest of previous cases. σ_m is the median value of σ . Not readily expressible in terms of $\overline{\sigma}$.		

 TABLE 2.3
 Common Statistical Models for Radar Cross Section

The shape of the PDF of RCS directly affects detection performance, as will be seen in Chap. 6. Figure 2.11*a* compares the exponential, fourth-degree chi-square, second-degree noncentral chi-square, Weibull, and log-normal density functions when all have an RCS variance of 0.5. The exponential distribution then necessarily has a mean of 0.5. The fourth-degree chi-square necessarily has a mean of 1.0, and the parameters of the remaining density functions have been chosen to give them a mean of 1.0 as well. Figure 2.11*b* repeats the same data on a semilogarithmic scale so that the behavior of the PDF "tails" is more evident. Note that the Weibull and second-degree non-central chi-square distributions are very similar for this choice of parameters. The chi-square is also similar, but has a somewhat less extensive tail to the distribution.

The log-normal has both the narrowest peak and the longest tail of any of the distributions shown for this choice of parameters. Unlike all of the others, the exponential does not have a distinct peak near the mean RCS. Each of the others does have a distinct peak, making them suitable for distributions with one or a few dominant scatterers.



FIGURE 2.11 Comparison of five models for the probability density function of radar cross section: (a) linear scale, (b) log scale. See text for additional details.

One fundamental difference among the various RCS models of Table 2.3 is

whether the probability density function has one or two free parameters. The nonfluctuating, exponential, and all chi-square (once the order is stated) are all *one-parameter distributions*. The one parameter in the form given earlier is the mean RCS, $\bar{\sigma}$. The non-central chi-square, Weibull, and log-normal are *two-parameter distributions*, as is the chi-square with variable degree. The parameters are $\bar{\sigma}$ and a^2 for the non-central chi-square, $\bar{\sigma}$ and *m* for the variable-order chi-square, *B* and *C* for the Weibull, and σ_m and *s* for the log-normal in the forms given. For a one-parameter distribution, estimating the mean is sufficient to characterize the complete PDF. For the two-parameter case estimates of two parameters, usually the variance and either the mean or median, must be computed to characterize the PDF. This distinction is important in the design of automatic detection algorithms in Chap. <u>6</u>.

Most radar analysis and measurement programs emphasize RCS measurements, which are proportional to received power. Sometimes ζ , the corresponding voltage, is of interest, particularly for use in simulations where Eq. (2.50) is used explicitly to model the composite echo from a multiple scatterer target. The probability density function for the voltage is then required in order to properly model the probabilistic variations of the complex sum. The PDF of $|\zeta|$ is easily derived from the PDF of σ using basic results of random variables (Papoulis and Pillai, 2001). Because RCS is nonnegative, the transformation⁵

$$\zeta = \sqrt{\sigma} \tag{2.54}$$

from RCS to voltage has only one real solution for σ , namely $\sigma = \zeta^2$. It then follows that the PDF of ζ is given by

$$p_{\zeta}(\zeta) = \frac{p_{\sigma}(\zeta^2)}{d\zeta/d\sigma}$$
$$= 2\zeta p_{\sigma}(\zeta^2)$$
(2.55)

Equation (2.55) can be used to write the voltage PDFs by inspection from Table 2.3. The results, given in Table 2.4, are expressed in terms of the parameters of the corresponding RCS distribution from Table 2.3. Additional information is given in App. A.

CS Model Name PDF for Voltage ζ		
$p_{\zeta}(\zeta) = \delta_D(\zeta - \sqrt{\overline{\sigma}})$ $\overline{\zeta} = \sqrt{\overline{\sigma}}, \text{var}(\zeta) = 0$	Also nonfluctuating model.	
tayleigh, central hi of degree 2 $p_{\zeta}(\zeta) = \frac{2\zeta}{\overline{\sigma}} \exp\left[\frac{-\zeta^2}{\overline{\sigma}}\right]$ $\overline{\zeta} = \frac{1}{2}\sqrt{\pi\overline{\sigma}}, \operatorname{var}(\zeta) = \overline{\sigma}(1 - \pi/4)$		
$p_{\zeta}(\zeta) = \frac{8\zeta^3}{\overline{\sigma}^2} \exp\left[\frac{-2\zeta^2}{\overline{\sigma}}\right]$ $\overline{\zeta} = \frac{3}{4}\sqrt{\pi\overline{\sigma}}, \operatorname{var}(\zeta) = \left(1 - \frac{9}{32}\pi\right)\overline{\sigma}$	Corresponding RCS has fourth-degree chi-square distribution. Voltage equivalent of Swerling 3 and 4 models.	
$p_{\zeta}(\zeta) = \frac{2\zeta m}{\Gamma(m)\overline{\sigma}} \left(\frac{m\zeta^2}{\overline{\sigma}}\right)^{m-1} \exp\left[-\frac{m\zeta^2}{\overline{\sigma}}\right]$ $= \sqrt{\overline{\sigma}}$	Voltage equivalent of Weinstock RCS models.	
$\zeta = \sqrt{\frac{m}{m}},$ $\operatorname{var}(\zeta) = \overline{\sigma} \left\{ 1 - \frac{1}{m} [\Gamma(m+0.5)/\Gamma(m)]^2 \right\}$		
$p_{\zeta}(\zeta) = \frac{2\zeta(1+a^2)}{\overline{\sigma}} \exp\left[-a^2 - \frac{\zeta^2}{\overline{\sigma}}(1+a^2)\right]$ $\times I_0\left(2a\zeta\sqrt{(1+a^2)/\overline{\sigma}}\right)$ $\overline{\zeta} = \frac{1}{2}\sqrt{\frac{\pi\overline{\sigma}}{1+a^2}}c_1^{-a^2}F_1[1.5, 1; a^2]$	Voltage equivalent of noncentral chi-square of degree 2. $_1F_1(x)$ is the confluent hypergeometric function, also called Kummer's function.	
$\operatorname{var}(\zeta) = \left(\frac{\overline{\sigma}}{1+a^2}\right) e^{-a^2} \\ \times \left[{}_1F_1(2,1;a^2) - \frac{\pi}{4}e^{-a^2}{}_1F_1^2(1.5,1;a^2)\right]$		
$p_{\zeta}(\zeta) = 2CB\zeta^{2C-1} \exp[-B\zeta^{2C}]$ $\overline{\zeta} = \Gamma(1 + 1/2C)B^{-1/2C}$ $\operatorname{var}(\zeta) = B^{-1/2C}[\Gamma(1 + 1/C) - \Gamma^{2}(1 + 1/2C)]$	Also Weibull, one parameter changed: $C \rightarrow 2C$.	
$p_{\zeta}(\zeta) = \frac{2}{\sqrt{2\pi}s\zeta} \exp\left[-2\ln^2\left(\zeta/\sqrt{\sigma_m}\right)^2/s^2\right]$ $\overline{\zeta} = \sqrt{\sigma_m} \exp(s^2/8)$ $\operatorname{var}(\zeta) = \sigma_m \exp(s^2/4)[\exp(s^2/4) - 1]$		
	PDF for Voltage ζ $p_{\zeta}(\zeta) = \delta_{D}(\zeta - \sqrt{\overline{\sigma}})$ $\overline{\zeta} = \sqrt{\overline{\sigma}}, \operatorname{var}(\zeta) = 0$ $p_{\zeta}(\zeta) = \frac{2\zeta}{\overline{\sigma}} \exp\left[\frac{-\zeta^{2}}{\overline{\sigma}}\right]$ $\overline{\zeta} = \frac{1}{2}\sqrt{\pi\overline{\sigma}}, \operatorname{var}(\zeta) = \overline{\sigma}(1 - \pi/4)$ $p_{\zeta}(\zeta) = \frac{8\zeta^{3}}{\overline{\sigma}^{2}} \exp\left[\frac{-2\zeta^{2}}{\overline{\sigma}}\right]$ $\overline{\zeta} = \frac{3}{4}\sqrt{\pi\overline{\sigma}}, \operatorname{var}(\zeta) = \left(1 - \frac{9}{32}\pi\right)\overline{\sigma}$ $p_{\zeta}(\zeta) = \frac{2\zeta m}{\Gamma(m)\overline{\sigma}} \left(\frac{m\zeta^{2}}{\overline{\sigma}}\right)^{m-1} \exp\left[-\frac{m\zeta^{2}}{\overline{\sigma}}\right]$ $\overline{\zeta} = \sqrt{\frac{\overline{\sigma}}{m}}, \operatorname{var}(\zeta) = \overline{\sigma}\left\{1 - \frac{1}{m}[\Gamma(m + 0.5)/\Gamma(m)]^{2}\right\}$ $p_{\zeta}(\zeta) = \frac{2\zeta(1 + a^{2})}{\overline{\sigma}} \exp\left[-a^{2} - \frac{\zeta^{2}}{\overline{\sigma}}(1 + a^{2})\right]$ $\times I_{0}\left(2a\zeta\sqrt{(1 + a^{2})/\overline{\sigma}}\right)$ $\overline{\zeta} = \frac{1}{2}\sqrt{\frac{\pi\overline{\sigma}}{1 + a^{2}}} e_{1}^{a^{2}} I_{1}^{a} I_{1}^{c} I_{1}^{c}$	

TABLE 2.4 Voltage Distributions Corresponding to Common StatisticalModels of Radar Cross Section

As has been seen, the RCS of a complex target varies with both transmitted frequency and aspect angle. Another important characteristic of a target's signature is the *decorrelation interval* in time, frequency, and angle. This is the change in time, frequency, or angle required to cause the echo amplitude to

decorrelate to a specified degree. If a rigid target such as a building is illuminated with a series of identical radar pulses and there is no motion between the radar and target, one expects the same received complex voltage y from each pulse (ignoring receiver noise). If motion between the two is allowed, however, the relative path length between the radar and the various scatterers comprising the target will change, causing the composite echo amplitude to fluctuate as in Fig. 2.9. Thus, for rigid targets, decorrelation of the RCS is induced by changes in range and aspect angle. On the other hand, if natural clutter such as the ocean surface or a stand of trees is illuminated, the signature will decorrelate even if the radar and target do not move relative to each other. In this case the decorrelation is caused by the "internal motion" of the clutter, such as the wave motion on the sea surface or the blowing leaves and limbs of the trees. The rate of decorrelation is influenced by factors external to the radar such as wind speed. Range or aspect changes also induce decorrelation of clutter signatures.

Although the behavior of real targets can be quite complex, a useful estimate of the change in frequency or angle required to decorrelate a target or clutter patch can be obtained by the following simple argument. Consider a target consisting of a uniform line array of point scatterers tilted at an angle θ with respect to the antenna boresight and separated by Δx from one another, as shown in Fig. 2.12. Assume an odd number 2M + 1 of scatterers indexed from -M to +M as shown. The total target extent is then $L = (2M + 1)\Delta x$. If the nominal distance to the radar R_0 is much larger than the target extent, $R_0 \gg L$, then the incremental distance an EM plane wave must travel from one scatterer to the next is $\Delta x \cdot \sin \theta$. If the target is illuminated with the waveform $A\exp(j\Omega t)$, the received signal will be



FIGURE 2.12 Geometry for calculation of RCS decorrelation interval in frequency and aspect angle.

$$\overline{y}(t) = \sum_{n=-M}^{M} A e^{j(\Omega t - 2(R_0 + n\Delta x \sin \theta)/c)}$$
$$= A e^{j\Omega(t - 2R_0/c)} \sum_{n=-M}^{M} e^{-j4\pi n\Delta x \sin \theta F/c}$$
(2.56)

To simplify the notation, define

$$K_{\theta} = \frac{2\pi F}{c} \sin \theta = \frac{2\pi}{\lambda} \sin \theta, \qquad \alpha = 2\Delta x$$
(2.57)

 K_{θ} is a spatial frequency in rads/m (see <u>App. B</u>). Then $\overline{y}(t)$ can be considered as a function $\overline{y}(t; K_{\theta})$ of both t and K_{θ} . The correlation in the variable K_{θ} , which includes both aspect angle and radar frequency, is of interest. Note that $\overline{y}(t; K_{\theta})$ is periodic in K_{θ} with period $2\pi/\alpha$. The deterministic autocorrelation of $\overline{y}(t; K_{\theta})$ as a function of the lag ΔK_{θ} is therefore

$$s_{\overline{y}}(\Delta K_{\theta}) = \int_{-\pi/\alpha}^{\pi/\alpha} \overline{y}(t; K_{\theta}) \overline{y}^{*}(t; K_{\theta} + \Delta K_{\theta}) dK_{\theta}$$
$$= \int_{-\pi/\alpha}^{\pi/\alpha} \left[Ae^{j\Omega(t - 2R_{0}/c)} \sum_{n=-M}^{M} e^{-j\alpha K_{\theta}n} \right] \left[A^{*}e^{-j\Omega(t - 2R_{0}/c)} \sum_{l=-M}^{M} e^{+j\alpha(K_{\theta} + \Delta K_{\theta})l} \right] dK_{\theta}$$
(2.58)

The complex exponential terms outside the summations cancel. Interchanging integration and summation and collecting terms then gives

$$s_{\overline{y}}(\Delta K_{\theta}) = |A|^{2} \sum_{l=-M}^{M} e^{+j\alpha\Delta K_{\theta}l} \sum_{n=-M}^{M} \left[\int_{-\pi/\alpha}^{\pi/\alpha} e^{-j\alpha(n-l)K_{\theta}} dK_{\theta} \right]$$
(2.59)

A change of variables $K'_{\theta} = \alpha K_{\theta}$ makes it clear that the integral has the form of the inverse discrete-time Fourier transform of a constant spectrum $S(K_{\theta}) = 2\pi/\alpha$. Therefore, the integral is just the discrete impulse function $(2\theta/\alpha)\delta[n-l]$. Using this fact reduces Eq. (2.59) to a single summation over *l* that can be evaluated to give

$$s_{\overline{y}}(\Delta K_{\theta}) = \frac{2\pi |A|^2}{\alpha} \frac{\sin[\alpha(2M+1)\Delta K_{\theta}/2]}{\sin[\alpha\Delta K_{\theta}/2]} = \frac{\pi |A|^2}{\Delta x} \frac{\sin[L \cdot \Delta K_{\theta}]}{\sin[\Delta x \cdot \Delta K_{\theta}]}$$
(2.60)

The decorrelation interval can now be determined by evaluating Eq. (2.60) to find the value of ΔK_{θ} which reduces s_y to a given level. This value of ΔK_{θ} can then be converted into equivalent changes in frequency or aspect angle.

One criterion is to choose the value of ΔK_{θ} corresponding to the first zero of the correlation function, which occurs when the argument of the numerator equals π . The resulting value is the Rayleigh width of the autocorrelation function. Using Eq. (2.61) and recalling that $L = (2M + 1) \Delta x$ gives

$$\pi = L \cdot \Delta K_{\theta} \Longrightarrow \Delta K_{\theta} = \frac{\pi}{L}$$
(2.61)

Recall that $K_{\theta} = (2\pi/c)F \sin\theta$. The total differential of *z* is then $dK_{\theta} = (2\pi/c) \cdot [\sin\theta \cdot dF + F\cos\theta \cdot d\theta]$. To determine the decorrelation interval in angle for a fixed radar frequency, set dF = 0 to give $dK\theta = (2\pi/c) \cdot F\cos\theta \cdot d\theta$, so that $\Delta K_{\theta} \approx (2\pi/c) \cdot F\cos\theta \cdot \Delta\theta$. Similarly, the frequency step required to decorrelate the target is obtained by fixing the aspect angle θ so that $d\theta = 0$, leading to $\Delta K_{\theta} \approx (2\pi/c) \cdot \sin\theta \cdot \Delta F$. Combining these relations with Eq. (2.61) then gives the

desired result for the change in angle or frequency required to decorrelate the echo amplitude:

$$\Delta \theta = \frac{c}{2F \cdot L \cos \theta}, \qquad \Delta F = \frac{c}{2L \sin \theta}$$
(2.62)

Note that $L \cos\theta$ is the projection of the target extent orthogonal to the radar boresight, while $L\sin\theta$ is the projection along the radar boresight. Thus, the decorrelation interval in aspect angle is driven by the width of the target as viewed from the radar, while the interval in frequency is driven by the depth. A more general pair of expressions that can be applied to more arbitrary manyscatterer targets is then

$$\Delta \theta = \frac{c}{2F \cdot L_w}, \qquad \Delta F = \frac{c}{2L_d}$$
(2.63)

where L_w and L_d are the target width and depth, respectively, as viewed from the radar.

As an example, consider a target the size of an automobile, about 5 m long. At L band (1 GHz), the target signature can be expected to decorrelate in $(3 \times 10^8)/(2 \times 5 \times 10^9) = 30$ mrad of aspect angle rotation relative to a broadside view (so the width is 5 m), about 1.7°, while at W band (95 GHz), this is reduced to only 0.018°. The frequency step required for decorrelation from a head-on aspect (depth of 5 m) is 30 MHz. This result does not depend on the nominal transmitted frequency.

As another example, Fig. 2.13a shows the autocorrelation function in angle for many-scatterer targets similar to that of Fig. 2.8, using only the data for aspect angles over a range $\pm 3^{\circ}$. Each of the two autocorrelation functions shown is the average of the autocorrelations of 20 different random targets, $\frac{6}{2}$ each having 20 randomly placed scatterers in a 5 m by 10 m box. The black curve is the autocorrelation of the data around a nominal boresight orthogonal to the 5 m side of the target, while the gray curve is the autocorrelation of the data viewed from the 10-m side. These look angles correspond to viewing the target nominally from the right and from the top in Fig. 2.8. At F = 10 GHz, the expected decorrelation interval in angle when viewed from the right is 0.34°; while when viewed from the top it is 0.17° . These expected decorrelation intervals are marked by the vertical dashed lines in Fig. 2.13a. In both cases, the first minimum of the correlation function occurs the predicted amount of change in the aspect angle. Figure 2.13b shows the average autocorrelation function in frequency over 30 similar random targets. The autocorrelation in this simulation does not have a distinct minimum, but the predicted decorrelation intervals closely approximate the first zero crossing.



FIGURE 2.13 Average autocorrelation function for many-scatterer targets: (a) angle autocorrelation functions, (b) frequency autocorrelation functions. See text for details.

Figure 2.9 demonstrated that viewing a complex target from a sufficiently

different aspect angle will decorrelate the RCS, i.e., result in a significantly different measured value. Figure 2.14 illustrates the ability of frequency agility to force RCS variations. A 20 scatterer, 5 m by 10 m random target was observed from a fixed aspect angle of about 54°, making its effective depth approximately $10\sin(54^\circ) = 8.1$ m. If the same RF frequency was used for each pulse, the RCS and thus received power would be exactly the same on each pulse. However, in this case the RF frequency was increased by 18.5 MHz [calculated from Eq. (2.63)] from one pulse to the next, starting at 10.0 GHz. The resulting relative RCS measurements vary by 38 dB, a factor of about 6300.



FIGURE 2.14 Variation in RCS due to frequency agility for a constant viewing angle. See text for details.

It will be seen in <u>Chap. 6</u> that in certain cases detection performance is improved when successive target measurements are uncorrelated. For this reason, some radars use a technique called *frequency agility* to force decorrelation of successive measurements (Ray, 1966). In this process, the radar frequency is increased by ΔF Hz or more between successive pulses, where ΔF is given by Eq. (2.63), ensuring that the target echo decorrelates from one pulse to the next. Once the desired number of uncorrelated measurements is obtained, the cycle of increasing frequencies is repeated for the next set of measurements.

Equation (2.63) is based on a highly simplified target model and an assumption about what constitutes the correlation interval. A different definition, for example defining the interval by the point at which the correlation function first drops to 1/2 or 1/e of its peak, would result in a smaller estimate of the required change in angle or frequency to decorrelate the target. Also,

many radars operate on the magnitude-squared of the echo amplitude, rather than the magnitude as has been assumed in this derivation. A square law detector produces a correlation function proportional to the square of Eq. (2.60) (Birkmeier and Wallace, 1963). The first zero therefore occurs at the same value of ΔK_{θ} , and the previous conclusions still apply. However, if a different definition of the correlation interval is used (such as the 50 percent decorrelation point), the required change in ΔK_{θ} is less for the square law than for the linear detector.

2.2.6 Target Fluctuation Models

It is common in radar detection algorithms to make a detection decision based not on one, but on a set of *N* noncoherently combined measurements from a given resolution cell. One way such a set of measurements can arise, and possibly the original motivation for this model, is based on the operation of a ground-based surveillance radar. Consider a radar with an antenna that rotates at a constant angular velocity Ω radians per second with an azimuth beamwidth of θ radians and a pulse repetition frequency of *PRF* pulses per second (hertz). Suppose that a target is present at a particular location. The geometry is shown in Fig. 2.15*a*. Assume that significant returns are received only when the target is in the antenna mainlobe. Every complete 360° scan of the antenna results in a new set of $N = (\theta/\Omega)PRF$ mainbeam pulses containing an echo of the target as the beam scans past. Consequently, it would seem to make sense to integrate the measurements from the same range bin over *N* successive pulses in an attempt to improve signal-to-noise ratio before performing a detection test. Early radars could only do this noncoherently.



FIGURE 2.15 Sample scenarios for collection of multiple noncoherently related measurements: (*a*) rotating surveillance antenna with noncoherent radar, (*b*) multiple CPIs with a coherent radar.

This is not the only way a set of related measurements can arise. Many modern systems are designed to transmit a rapid burst of M pulses at a constant PRF, often with the antenna staring in a fixed or nearly fixed direction, forming a coherent processing interval (CPI) of data. As will be seen in Chap. 4, the Mpulse burst is a common waveform well suited to Doppler measurements and interference suppression. The radar may then repeat the measurement, collecting a series of N CPIs in the same look direction. Successive CPIs may share the same radar parameters, or the radar may change the PRF, the waveform, or the RF (frequency agility). The data from the same range bin within a single CPI are usually combined coherently, for instance to form a Doppler spectrum for that range bin. However, combining data across CPIs must generally be done noncoherently. An example would be noncoherent integration of Nmeasurements of the same range-Doppler resolution cell in the N CPIs as shown in Fig. 2.15b, prior to testing that cell for the presence of a target.

If the target, the radar, or both are moving during the time a set of N measurements is collected, a natural question is whether the target RCS during that time should be considered constant or varying. That is, assuming frequency agility is not used, does the radar-target aspect angle vary so little that the RCS should be modeled as the same random variable during the entire set of N pulses or CPIs? Or is the aspect changing so rapidly that the RCS decorrelates from one pulse or CPI to the next, and so should be modeled as independent random variables from the appropriate PDF? The answer has a significant impact on both the procedure and the results for computing detection probabilities, as will be seen in <u>Chap. 6</u>.

These questions require consideration of the dynamics of the radar-aircraft encounter in light of the decorrelation interval in angle given in Eq. (2.63). As an example, consider the crossing encounter of Fig. 2.16a. Aircraft #1 views aircraft #2 at broadside from a range of 5 km with an X band (10 GHz) radar. Assume aircraft #2 is traveling at 100 m/s and has a length (width as viewed from the radar) of 10 m. Assume that aircraft #1 transmits a burst of M = 10pulses at a 1-kHz PRF. In the resulting 10 ms CPI, aircraft #2 will travel 1 m, resulting in an angular change with respect to aircraft #1 of approximately 1/5000 = 0.2 mrad. From Eq. (2.63), the decorrelation interval in angle is expected to be $(3 \times 10^8)/(2 \cdot 10 \cdot 10 \times 10^9) = 1.5$ mrad. Because the actual angle change within a CPI is less than the angular decorrelation interval, one would expect all the pulses within a CPI to experience essentially the same RCS. Now suppose that the radar transmits a series of pulse bursts, each one starting 100 ms after the previous burst. The angular change between aircraft #1 and #2 from one CPI to the next is then 2 mrad, which is greater than the 1.5 mrad decorrelation interval. Consequently, it is expected that the aircraft RCS during a given CPI will be uncorrelated with that during other CPIs.



FIGURE 2.16 Crossing target scenario: (*a*) encounter geometry, (*b*) target RCS for five 10-pulse CPIs. See text for motion and radar details.

Figure 2.16*b* illustrates these effects using another 10×5 m random complex target model with the radar and motion parameters just described. The RCS observed for the target on each pulse is plotted for five CPIs. Notice that the RCS is nearly constant within each CPI. The greatest intra-CPI variation is only about 0.4 dB. Significantly greater variation is seen from one CPI to the next, with the total range in this example being approximately 9 dB. Consequently, the full set of five CPIs of data could reasonably be modeled by drawing one random value of RCS from an exponential PDF to represent each CPI. The same RCS value is used for all pulses within a CPI. If pulse-to-pulse frequency agility was used with a frequency step exceeding the 30 MHz dictated b y Eq. (2.63), the RCS of each individual pulse would be expected to be uncorrelated with all of the other pulses, though still drawn from the exponential PDF.

Traditionally in radar, when a set of N measurements that are to be noncoherently combined are completely correlated with one another but are uncorrelated with a subsequent, new set of N measurements, they are said to exhibit *scan-to-scan decorrelation*. When each individual measurement is uncorrelated with all of the others, they are said to exhibit *pulse-to-pulse decorrelation*. This terminology probably originates from scenarios like that of Fig. 2.15*a*. The N measurements to be combined are the N pulses obtained from one scan of the radar past the target. If the target moves slowly enough that the Npulses from one scan reflect the same RCS value, but fast enough that by the time the radar scans through a full circle and returns to the target again the next group of N pulses reflects a different RCS value, then the term "scan-to-scan decorrelation" would be very descriptive. If frequency agility were used or radar-target motion and collection time were such that each pulse exhibits a different RCS value, the term "pulse-to-pulse decorrelation" would be apt.

The "scan-to-scan" and "pulse-to-pulse" terminology has a long legacy in performance analysis of radars using noncoherent integration. It is used in much of the classical literature, but in modern coherent radars it often does not relate well to the actual data collection and processing methods used. In the CPI-based data collection protocol of Fig. 2.15*b* and the related numerical example of Fig. 2.16, the data are correlated from one pulse to the next but are uncorrelated from one CPI to the next. However, the intra-CPI data will likely be combined coherently. The noncoherent combination will occur from one CPI to the next, for instance by noncoherently integrating the same range-Doppler bin from each CPI. Because the measurements that are actually integrated noncoherently will be uncorrelated in this example, the appropriate detection analysis results from the literature would be those for "pulse-to-pulse" decorrelation even though the actual data are highly correlated from one pulse to the next!

Another example of this confusion can arise when a series of CPIs represents a single short-term "look" at the target region, often called a dwell, rather than the result of different passes over the same region from a regular scan pattern. In the former case the elapsed time may still be short and the target may not decorrelate between CPIs, while in the latter the timeline would likely be longer and the target would be more likely to decorrelate. If noncoherent integration across CPIs is performed, the data might be best modeled as "scantoscan" decorrelation in the former case because the data values from integrated are expected to be similar, while in the second case they would be different and the appropriate mathematical results would be those for noncoherently integrating dissimilar values, namely the classical "pulse-to-pulse" case.

These terminology concerns can become an issue in detection performance analysis. Many published results use the "scan-to-scan" and "pulse-to-pulse" decorrelation terminology. In interpreting radar literature for modern radars, the reader is cautioned to consider carefully the correlation properties of the measurements that will be noncoherently combined for a single detection decision. The critical point is whether those measurements are expected to be highly correlated, i.e., all approximately the same random variable, or whether they are expected to be highly decorrelated (different random variables). If the measurements are highly correlated, published results on "scan-to-scan" mathematical models are applicable. If they are uncorrelated, "pulse-to-pulse" models are applicable. Newer literature is less likely to use the "scan-to-scan" and "pulse-to-pulse" terminology, obviating this problem over time.

A *target fluctuation model* is a combination of a PDF describing the RCS variation with angle, RF, or other important parameters and a decorrelation model for measurements to be combined noncoherently. Any PDF that models the RCS distribution for the targets and radar of interest could be used.

Examples include any of the PDFs in <u>Table 2.3</u>. For manmade targets, the decorrelation model is usually taken as one of the extremes of either the fully correlated or fully decorrelated models. Analysis carried out using these two models produces bounding results for detection performance. In reality, the noncoherently combined measurements will often be partially correlated. Partial correlation models specified with a pulse-to-pulse correlation coefficient or an autocorrelation function are sometimes given, though this is more common in clutter modeling.

2.2.7 Swerling Models

An extensive body of radar detection theory has been built up using the four Swerling models of target RCS fluctuation and noncoherent integration (Swerling, 1960; Meyer and Mayer, 1973; Nathanson, 1991; Skolnik, 2001). They are formed from the four combinations of two choices for the PDF and two for the correlation properties. The two density functions used are the exponential and the chi-square of degree 4 (see Table 2.3). The exponential model describes the behavior of a complex target consisting of many scatterers, none of which is dominant. The fourth-degree chi-square model targets having many scatterers of similar strength with one dominant scatterer. Although the Rice distribution is the exact PDF for this case, the chi-square is an approximation based on matching the first two moments of the two PDFs (Meyer and Mayer, 1973). These moments match when the RCS of the dominant scatterer is $1 + \sqrt{2} = 2.414$ times that of the sum of the RCS of the small scatterers, so the fourth-degree chi-square model fits best for this case. More generally, a chi-square of degree $2m = 1 + [a^2/(1 + 2a)]$ is a good approximation to a Rice distribution with a ratio of a^2 of the dominant scatterer to the sum of the small scatterers. However, only the specific case of the fourth-degree chi-square is considered a Swerling model.

The Swerling models are denoted as "Swerling 1," "Swerling 2," and so forth. <u>Table 2.5</u> defines the four cases. A nonfluctuating target is sometimes identified as the "Swerling 0" or "Swerling 5" model.

	Correlation of Noncoherently Integrated Measurements		
Probability Density Function of RCS	Correlated	Uncorrelated	
Rayleigh/exponential	Case 1	Case 2	
Chi-square, degree 4	Case 3	Case 4	

 TABLE 2.5
 Swerling Models

Figures 2.17 and 2.18 illustrate the difference in the behavior of two of the

Swerling models. In both cases, the received power from a single point scatterer having a unit mean Swerling RCS is plotted, and in both it is assumed that 10 samples are obtained on each of three scans or CPIs of the radar. Figure 2.17 is a sample Swerling 1 (exponential PDF, fully correlated) series. In contrast, Fig. 2.18 illustrates a Swerling 4 case (fourth-degree chi-square PDF, fully decorrelated) in which each individual sample is independent of the others.



FIGURE 2.17 Three scans or CPIs, each having 10 samples of a unit mean Swerling 1 power sequence.



FIGURE 2.18 Three scans or CPIs, each having 10 samples of a unit mean Swerling 4 power sequence.

2.2.8 Effect of Target Fluctuations on Doppler Spectrum

A common operation in radar signal processing is computing the discrete-time Fourier transform (DTFT) of the data in a particular range bin for one CPI. The DTFT is a coherent combination of measurements, usually over a sufficiently short CPI that the target echo RCS and thus amplitude do not decorrelate significantly. As will be seen in <u>Chap. 4</u>, the series of samples within a CPI for a constant-velocity target will form a discrete-time sinusoid. Thus, the usual model for the DTFT of a target is an aliased sinc function [also called an asinc, dsinc (digital sinc), or Dirichlet function] with its mainlobe centered at the appropriate frequency and with sidelobes that peak 13.2 dB below the mainlobe peak and decay at frequencies further from the mainlobe.

In cases where there are significant RCS fluctuations within the CPI, the amplitude and phase of the target data will vary within the CPI, so that the input to the DTFT is no longer a discrete sinusoid with a constant complex amplitude. Figure 2.19 illustrates the resulting effect on the DTFT. The gray spectrum is that of an unwindowed zero-frequency sinusoid, modeling the returns from 20 pulses echoed from a stationary 10×5 m simulated many-scatterer target viewed at a constant aspect angle. The black line is the spectrum of data

observed using the same target and waveform, but with the aspect angle changing 0.7 mrad per pulse. The total angle change over 20 pulses is then 13.3 mrad, nearly nine times the decorrelation interval of 1.5 mrad. The target echo amplitude and phase will then both fluctuate significantly, raising the sidelobes and smearing the target energy over a wider frequency range, effectively whitening the spectrum significantly.



FIGURE 2.19 Effect of intra-CPI target fluctuations on Doppler spectrum.

2.3 Clutter

In radar the term *clutter* refers to a component of the received signal due to echoes from volume or surface scatterers. Such scatterers include the earth's surface, both terrain and sea; weather echoes (for example, rain clouds); and man-made distributed clutter, such as so-called *chaff* clouds of airborne scatterers, typically made out of lightweight strips of reflecting material. Clutter echoes are sometimes interference and sometimes the desired signal. For instance, synthetic aperture imaging radars are designed to image the earth surface, thus the terrain clutter is the target in a SAR. For an airborne or spaceborne surveillance radar trying to detect moving vehicles on the ground, clutter echo from the surrounding terrain is an interference signal.

From a signal processing point of view, the major concern is how to model clutter echoes. As with man-made targets, clutter is a complex target with many scatterers per resolution cell so that the echoes are highly sensitive to radar parameters and encounter geometry. Like complex targets, clutter is therefore modeled as a random process. In addition to temporal correlation, clutter can also exhibit spatial correlation: the reflectivity samples from adjacent resolution cells may be correlated. Two excellent general references on land and sea clutter phenomenology are Ulaby and Dotson (1989) and Long (2001). A good brief introduction is Currie (2010).

Clutter echoes differ from target echoes in that they will typically exhibit different PDFs, temporal and spatial correlation properties, Doppler characteristics, and power levels. These differences can be exploited to separate target and clutter signals. Means to do so are the principal concern of Chaps. 5 and 9. Clutter differs from noise in two major ways: its power spectrum is not white (i.e., it is correlated interference), and, since it is an echo of the transmitted signal, the received clutter power is affected by such radar and scenario parameters as the antenna gain, transmitted power, and the range from the radar to the terrain. In contrast, noise power is affected by none of those factors, but is affected by the radar receiver noise figure and bandwidth.

2.3.1 Behavior of σ^0

Area clutter (land and sea surface) reflectivity is characterized by its mean or median value of radar cross section, σ^0 (dimensionless), the probability density function of the reflectivity variations, and their correlation in space and time. Many of the same PDFs described in Sec. 2.2.5 are applied to modeling σ^0 as well. Popular examples include the exponential, lognormal, and Weibull distributions.

The area reflectivity σ^0 of terrain observed by the radar is a strong function of terrain type and condition (e.g., surface roughness and moisture), weather (wind speed and direction, precipitation), engagement geometry (especially grazing angle), and radar parameters (wavelength, polarization). Consequently, selection of a PDF is not sufficient to model clutter. It is also necessary to model the dependence of σ^0 on these parameters. Consider land clutter. Values of σ^0 commonly range from -60 to -10 dB. Extensive measurement programs over the years have collected statistics of land clutter under various conditions and resulted in many tabulations of σ^0 for various terrain types and conditions, as well as models for the variation of σ^0 . Figure 2.20 shows one set of representative data for the area reflectivity of desert terrain versus radar frequency and grazing angle. Note that σ^0 generally increases with radar frequency, and decreases at shallower grazing angles. For a given frequency, the variation with grazing angle over the range shown is 20 to 25 dB. For a given grazing angle, the variation across frequency in this example is about 10 dB. Figure 2.21 is one example of the variation in σ^0 versus grazing angle for different terrain types at a fixed frequency, in this case S band. Generally, reflectivity increase with terrain roughness, from the presumably smoother desert terrain to the complex, rough urban terrain.



FIGURE 2.20 Behavior of σ^0 of desert terrain versus radar frequency and grazing angle. (Data from Currie, 2010.)



FIGURE 2.21 Behavior of σ^0 versus terrain type and grazing angle at S band. (Data from Currie, 2010.)

As seen in those figures, σ^0 varies significantly with grazing angle. Generally, it decreases rapidly at very low grazing angles, and increases rapidly at very high grazing angles (radar look direction normal to the clutter surface), with a milder variation in a middle "plateau region." Figure 2.22 is a notional diagram of this behavior. A common model for the behavior of σ^0 over the plateau region is the "constant gamma" model (Long, 2001):



FIGURE 2.22 General behavior of σ^0 with grazing angle for land clutter. (After Long, 2001.)

$$\sigma^0 = \gamma \sin \delta \tag{2.64}$$

where γ is a characteristic of the particular clutter type at the radar frequency and polarization of interest. This model predicts that σ^0 is maximum at normal incidence and becomes vanishingly small as the grazing angle tends to zero. However, it does not adequately reflect the degree of change in σ^0 often observed at near-normal or near-zero incidence angles, and additional models are often used at these two extremes.

Various predictive models for σ^0 as a function of important parameters have been presented in the literature. One well-known example is the Georgia Tech Research Institute (GTRI) model given by

$$\sigma^{0} = A(\delta + C)^{B} \exp\left[\frac{-D}{1 + \frac{\sigma_{h}/10}{\lambda}}\right]$$
(2.65)

where σ_h is the RMS surface roughness and the parameters *A*, *B*, *C*, and *D* depend on the clutter type and radar frequency. Sample values for X band are given in <u>Table 2.6</u>.

Parameter	Clutter Type						
	Soil/Sand	Grass	Tall Grass, Crops	Trees	Urban	Wet Snow	Dry Snow
Α	0.25	0.023	0.006	0.002	2.0	0.0246	0.195
В	0.83	1.5	1.5	0.64	1.8	1.7	1.7
С	0.0013	0.012	0.012	0.002	0.015	0.0016	0.0016
D	2.3	0	0	0	0	0	0

Source: Adapted from Currie (2010).

 TABLE 2.6
 GTRI Land Clutter Model Parameters for X Band

Models of a similar spirit exist for sea clutter σ^0 . Important parameters for sea clutter reflectivity in addition to frequency, grazing angle, and polarization include wind speed and direction relative to the radar boresight, wave height, and multipath. Details of one representative model developed at GTRI are given in Currie (2010).

2.3.2 Signal-to-Clutter Ratio

In many scenarios, the dominant interference is not noise, but clutter. Consequently, the *signal-to-clutter ratio* (SCR) is often of more importance than the *signal-to-noise ratio* (SNR). The SCR is easily obtained as the ratio of the received target power, given by Eq. (2.11) to the received clutter power, given by Eq. (2.25), (2.30), or (2.32) for the volume clutter, beam-limited area clutter, and pulse-limited area clutter cases, respectively. The resulting equations are

$$SCR = \frac{\sigma}{R^2 \eta \cdot \Delta R \cdot \theta_3 \phi_3} \quad \text{(volume clutter case)}$$
$$= \frac{\sigma \sin \delta}{R^2 \phi_3 \theta_3 \sigma^0} \quad \text{(beam-limited area clutter case)}$$
$$= \frac{\sigma \cos \delta}{R \sigma^0 \cdot \Delta R \cdot \theta_3} \quad \text{(pulse-limited area clutter case)}$$
(2.66)

In each case, such system parameters as the transmitted power and the antenna gain cancel out. This occurs because both the clutter and target signals are echoes of the radar pulse; increasing power or antenna gain increases the strength of both types of echoes equally. Thus, the SCR just becomes the ratio of the target RCS to the total RCS of the contributing clutter.

2.3.3 Temporal and Spatial Correlation of Clutter

Clutter decorrelation in time is induced by internal motion for clutter, such as

tree leaves moving in the wind or waves on the sea surface, and by changes in radar-target geometry for both clutter and targets. Various investigators have experimentally characterized the decorrelation characteristics of clutter echoes due to internal motion, or equivalently, their power spectrum. For example, one model suggested to estimate the power spectrum of the RCS of foliated trees or rain uses a cubic spectrum (Currie, 2010):

$$S_{\sigma}(F) = \frac{A}{1 + (F/F_c)^3}$$
(2.67)

The corner frequency F_c is a function of the wavelength and either wind speed (for trees) or rain rate (for rain). Some sample measured values are given in <u>Table 2.7</u>. A higher corner frequency (wider power spectrum) implies a shorter decorrelation interval (narrower autocorrelation function). Shorter decorrelation times render the clutter signals more like white noise and degrade the effectiveness of some of the clutter suppression techniques of <u>Chap. 5</u>. Notice that for a given weather condition, the clutter decorrelates more rapidly at higher radar frequencies. Figure 2.23 plots additional windblown tree clutter data that also show the decrease in decorrelation time for both increased clutter motion and increased radar frequency.

	Radar Frequency, GHz			
Target	10	35	95	
Rain, 5 mm/h	35	80	140	
Rain, 100 mm/h	70	120	500	
Trees, 6–15 mph wind	9	21	35	

Source: Currie, N. C. "Clutter Characteristics and Effects," chapter 10 in J. L. Eaves and K. E. Reedy (eds.), *Principles of Modern Radar*. Van Nostrand Reinhold, New York, 1987.

TABLE 2.7 Cubic Power Spectrum Corner Frequencies (Hz) for Rain and TreeClutter



FIGURE 2.23 Decorrelation time of windblown tree clutter versus frequency and wind speed. [Data from Currie (2010)].

Another model frequently used to model generic power spectra is the Gaussian given by

$$S_{\sigma}(F) = A \exp\left[-\alpha \left(\frac{F}{F_0}\right)^2\right]$$
(2.68)

The Gaussian model is very commonly used in weather radar, and is the basis of the pulse pair Doppler velocity estimation technique discussed in <u>Chap. 5</u>.

Both the cubic and Gaussian power spectral models can be well matched by a low-order autoregressive (AR, or all-pole) spectrum model of the form (Haykin et al., 1982)

$$S_{\sigma}(F) = \frac{A}{1 + \sum_{k=1}^{N} \alpha_k F^{2k}}$$
(2.69)

Real clutter measured from ground-based radars appears to be well matched using an order N of only two to four. Other studies of clutter measured by airborne radars in a landing scenario indicate that orders up to 10 may be required (Baxa, 1991). The AR clutter spectrum model has the advantage that its parameters can be computed directly from measured data and adapted in real

time using the Levinson-Durbin or similar algorithms (Kay, 1988). Furthermore, the AR parameters can be used to construct optimal adaptive clutter suppression filters, as is seen in <u>Chap. 5</u>. The disadvantage is that the calculations rapidly become computationally intensive as the model order increases.

Another decorrelation model, more recently developed and popular in studies of detection of ground targets from moving platforms, is the Billingsley model. This model represents the correlation properties of windblown tree clutter and other vegetative cover, said to be the "most pervasive" ground clutter (Billingsley, 2001). This model assumes that the clutter temporal power spectrum is the sum of a two-sided decaying exponential function and an impulse at the origin in Doppler frequency space

$$S_{\sigma}(F) = \sigma_{c}^{2} \left[\underbrace{\frac{\alpha}{\alpha+1} \delta_{D}(F)}_{\text{"DC term"}} + \underbrace{\frac{1}{\alpha+1} \left(\frac{\beta\lambda}{4}\right) \exp\left(-\frac{\beta\lambda}{2}|F|\right)}_{\text{"AC term"}} \right]$$
(2.70)

where the parameter α , which established the ratio of the DC to AC components, is a function of both wind and radar frequency, while β , which determines the width of the AC power spectral component, is dependent primarily on wind conditions. The corresponding autocorrelation function is

$$s_{\sigma}(\upsilon) = \sigma_c^2 \left(\frac{\alpha}{\alpha+1} + \frac{1}{\alpha+1} \frac{(\beta\lambda)^2}{(\beta\lambda)^2 + (4\pi\upsilon)^2} \right)$$
(2.71)

Based on extensive measurements, Billingsley proposed empirical formulas for α and β :

$$\alpha = 489.8 \cdot w^{-1.55} F_0^{-1.21} \tag{2.72}$$

$$\beta^{-1} = 0.1048[\log_{10} w + 0.4147]$$
(2.73)

where w is the wind speed in statute miles per hour and F_0 is the radar carrier frequency in GHz.

Note that β and therefore the decorrelation time does not depend on radar frequency, somewhat in conflict with earlier models. Caution is needed in applying Eq. (2.73) due to mixed units. Specifically, w is in statute miles per hour but β is in meters per second.

The "DC term" in Eqs. (2.70) and (2.71) represents a constant, nonrandom component of the clutter echo that is sometimes called a "persistent component"

of the received signal. For such a component to exist, both the amplitude and phase of the reflectivity of the clutter scatterers involved must be constant. Thus, the DC component is attributable to backscatter from elements such as bare ground, rocks, and tree trunks. The AC term accounts for back-scatter from moving elements such as leaves, branches, and blades of grass. Simple autoregressive filters can be used to implement the model in simulations (Mountcastle, 2004).

2.3.4 Compound Models of Radar Cross Section

As is seen in <u>Chap. 6</u>, radar detection performance predictions depend strongly on the details of target and clutter RCS models. Furthermore, it is well known that RCS statistics vary significantly with a host of factors such as geometry, resolution, wavelength, and polarization. Consequently, the development of good statistical RCS models is a very active area of empirical and analytical research. Following are three brief examples of an extension to the basic modeling approach described earlier, all motivated by the complexities of modeling clutter. Because the literature regarding these models is developed primarily in terms of the echo amplitude (voltage) ζ instead of RCS σ or power, the remainder of this section also concentrates on amplitude PDFs.

Some amplitude PDFs are physically motivated, especially the Rayleigh (exponential RCS) model (which follows from a central limit theorem argument) and the Rice or Rician model (which corresponds to a Rayleigh model with an additional dominant scattering source). Others, such as the lognormal or Weibull, have been developed empirically by fitting distributions to measured data. One attempt to provide a physical justification for a non-Rayleigh model abandons the single-PDF approach, instead assuming that the random variable representing echo amplitude can be written as the product of two independent random variables, $\zeta = x \cdot y$. The PDF of ζ can then be represented in a Bayesian formulation as

$$p_{\zeta}(\zeta) = p_x(x)p_{\zeta|x}(\zeta|x)$$
(2.74)

This model has been used to describe sea clutter (Jakeman and Pusey, 1976; Ward, 1981). The random variable *x* is identified with a slowly decorrelating component having a voltage distribution following a central chi-square of degree 2m with $m \ge 2.5$. This component is introduced to account for "bunching" of scatterers due to ocean swell structure and radar geometry, and represents variation in the mean of the amplitude over time. The distribution $p_{\zeta|x}(\zeta|x)$ is assumed to represent the composite of a large number of independent scatterers. Its amplitude distribution is therefore Rayleigh. The resulting overall PDF $p_{\zeta}(\zeta)$ can be shown to be the *K* distribution, which is given by

$$p_{\zeta}(\zeta) = \begin{cases} \frac{4c}{\Gamma(a)} (c\zeta)^a K_{a-1}(2c\zeta), & \zeta \ge 0\\ 0, & \zeta < 0 \end{cases}$$

$$(2.75)$$

where $K_{a-1}(\cdot)$ is the modified Bessel function of the second kind and order a-1 and $c = b\sqrt{\pi/4}$. Thus, the product formulation suggests that modulation of a standard Rayleigh variable by a central chi-distributed geometric term can account for observed sea clutter distributions. Additional information on the K distribution is given in <u>App. A</u>.

More recent research has begun to bridge the gap between the physics of scattering and the apparent success of compound clutter models of the type promoted by Ward and Jakeman and Pusey. Sangston summarizes the work on extensions of the "many scatterer" physical model that leads to the Rayleigh distribution (Sangston, 1994). Specifically, consider the model of Eq. (2.50), but let the number of scatterers N be a random variable instead of a fixed constant. This representation is referred to as a *number fluctuations* model. Depending on the choice of the statistics of the number N of scatterers contributing to the return at any given time, this modified version of Eq. (2.50) can result in K, Weibull, gamma, Nakagami-m, or any of a number of other distributions in the class of so-called Rayleigh mixtures.

Much of the work in compound RCS models has been performed in the context of sea clutter analysis, and empirical sea clutter data have often been observed to exhibit non-Rayleigh statistics such as Weibull, K, and log-normal distributions. The number fluctuation model is intuitively appealing in this case because it can be related to the physical behavior of waves. Specifically, scattering theory suggests that the principal scatterers on the ocean surface are the small capillary waves, as opposed to the large swells. These small scattering centers tend to cluster near the crest of the swells, with fewer of them in between. In other words, they are nonuniformly distributed over the sea surface. Consequently, a radar illuminating the sea will receive echoes from a variable number N of scatterers as the crests of the swells move into and out of a given resolution cell. By summing echoes from a variable number of scatterers, the number fluctuation model predicts the Weibull and K distributions and provides a link between a phenomenological model of sea scatter and these empirically observed statistics.

All of the statistical models described in Sec. 2.2.5 apply to the scattering observed from a single resolution cell. That is, they represent the variations in RCS observed by measuring the same region of physical space multiple times, for example by transmitting multiple pulses in the same direction and measuring the received power at the same delay after each transmission. Another use of the product model of Eq. (2.74) is to describe the spatial variation of clutter

reflectivity. If the scene being viewed by the radar is nonhomogenous, then the characteristics of the RCS observed in one resolution cell might vary significantly from those of another. For example, the dominant clutter observed by a scanning radar at a coastal site might be an urban area in one look direction and the sea in another. Another example occurs when scattered rain cells occupy only part of the scanned region, so that some resolution cells contain rain while others are clear.

This situation can be modeled by letting the slowly decorrelating term x in the product model represent spatial variations in the local mean of the received voltage. If the PDF of x is log-normal with a large variance and the PDF of ζ conditioned on x is gamma distributed (which includes Rayleigh as a special case), then the overall PDF of the product ζx has a lognormal distribution (Lewinski, 1983). Consequently, the product model implies that lognormal variations of the local mean from one resolution cell to another could account for the log-normal variation often used to model ground clutter returns. A similar argument can be used to justify the log-normal model for target RCS by modeling the variation of RCS with aspect angle as a log-normal process.

2.4 Noise Model and Signal-to-Noise Ratio

The echo signal received from a target or clutter inevitably competes with noise. There are two sources of noise: that received through the antenna from external sources, and that generated in the radar receiver itself.

External noise is a strong function of the direction in which the radar antenna is pointed. The primary contributor is the sun. If the antenna is directed toward the night sky and there are no interfering microwave sources, the primary source is *galactic* (also called *cosmic*) *noise*. Internal noise sources include *thermal noise* (also called *Johnson noise*) due to ohmic losses, *shot noise* and *partition noise* due to the quantum nature of electric current, and *flicker noise* due to surface leakage effects in conducting and semiconducting devices (Carlson, 1976).

Of these various sources, thermal noise is normally dominant. The theories of statistical and quantum mechanics dictate that the thermal noise voltage in an electronic circuit is a zero-mean Gaussian random process (Curlander and McDonough, 1991). The mean energy of the random process is kT/2 joules, where *T* is the temperature of the noise source in kelvins (absolute temperature) and $k = 1.38 \times 10^{-23}$ J/K is *Boltzmann's constant*. The power spectrum $S_n(F)$ of the thermal noise delivered to a matched load is

$$S_n(F) = \frac{hF}{\exp(hF/kT) - 1} \quad W/Hz$$
(2.76)

where $h = 6.6254 \times 10^{-34}$ J/s is *Planck's constant*. If $hF/kT \ll 1$, a series approximation gives $\exp(hF/kT) \approx 1 + hF/kT$ so that Eq. (2.76) reduces to the white noise spectrum

$$S_n(F) = kT \quad W/Hz \tag{2.77}$$

Note that Eq. (2.77), when integrated over frequency, implies infinite power in the white noise process. In reality, however, the noise is not white [Eq. (2.76)] and, in any event, it is observed in any real system only over a finite bandwidth. For frequencies below 100 GHz, the approximation of Eq. (2.77) requires the equivalent noise temperature *T*' (to be defined below) to be larger than about 50 K, which is almost always the case. Consequently, thermal noise has a white power spectrum. For many practical systems it is reasonable to choose the temperature of the system to be the "standard" temperature $T_0 =$ 290 K = 62.3°F so that $kT_0 \approx 4 \times 10^{-21}$ W/Hz.

In a coherent radar receiver, the noise present at the front end of the system contributes noise to both the I and Q channels after the quadrature demodulation. The I and Q channel noises are both zero-mean Gaussian random processes with equal power. Since the total noise spectral density is kT W/Hz, the noise density in each channel individually is kT/2 W/Hz. Furthermore, if the power spectrum of the input noise is white, then the I and Q noise processes are uncorrelated and their power spectra are also white. Since the I and Q noise processes are Gaussian and uncorrelated, it follows that they are also independent (Papoulis and Pillai, 2001). Finally, since the I and Q signals are independent zero-mean Gaussian processes, it also follows that the magnitude of the complex signal I + jQ is Rayleigh distributed, the magnitude-squared is exponentially distributed, and the phase angle tan⁻¹ (Q/I) is uniformly distributed over (0, 2π].

The bandwidths of the various components of a receiver vary, but the narrowest bandwidth is generally approximately equal to the bandwidth of the transmitted pulse. If the receiver contains any component of narrower bandwidth signal, energy will be lost, reducing sensitivity. If the most narrowband component has a bandwidth appreciably wider than the pulse bandwidth, the signal will have to compete against more noise power than necessary, again reducing sensitivity. Thus for the purpose of noise power calculation, the frequency response of the receiver can be approximated as a bandpass filter centered at the transmit frequency with a bandwidth equal to the waveform bandwidth.

Real filters do not have perfectly rectangular passbands. For analyzing noise power the *noise-equivalent bandwidth* β_n of a filter described by the transfer function H(F) is used. Figure 2.24 illustrates the concept. The noise equivalent bandwidth is the width an ideal rectangular filter with gain equal to the peak gain of the actual filter must have so that the area under the two squared

frequency responses are equal. This condition guarantees that given a white noise input, both filters exhibit the same output noise power. Thus



FIGURE 2.24 Illustration of the concept of noise equivalent bandwidth of a filter.

$$\beta_n = \frac{\int_{-\infty}^{\infty} |H(F)|^2 dF}{\max[|H(F)|^2]} = \frac{1}{G_s} \int_{-\infty}^{\infty} |H(F)|^2 dF$$
(2.78)

where the receiver power gain G_s is defined as the maximum gain of $|H(F)|^2$. The total noise power N present at the output of the filter H(F) is then given by

$$N = \int_{-\infty}^{\infty} |H(F)|^2 S_n(F) dF$$

= $kT \int_{-\infty}^{\infty} |H(F)|^2 dF$
= $kT \beta_n G_s$ (2.79)

White noise passed through a filter H(F) is no longer white, but instead has the power spectrum $|H(F)|^2$. If $|H(F)|^2$ is approximated as a rectangular filter of two-sided bandwidth β_n Hz, the autocorrelation function of the noise at the filter output is approximately a sinc function with its first zero at lag $1/\beta_n$ seconds. However, it will be seen in <u>Chap. 3</u> that the receiver output is normally sampled at intervals of approximately $1/\beta_n$ seconds. Consequently, the noise component of the successive receiver output samples are still uncorrelated with one another.

The power spectral density of white noise at the output of any source or circuit can be described as the product of Boltzmann's constant and some equivalent temperature T', mimicking the simple formulation of Eq. (2.77). Source noise power is usually referenced to the input of a system so that the power gain G_s (or loss if $G_s < 1$) of the system must also be taken into account. That is, if the observed output power spectral density (still assumed white over
the receiver bandwidth) at the output of a receiver is some value S_n , then an equivalent temperature T of the noise source at its input is defined to be

$$T' \equiv \frac{S_n}{kG_s}$$
(2.80)

so that $S_n = kT'G_s$ and the total noise power is

$$N = kT'\beta_n G_s \tag{2.81}$$

The total output noise power at the receiver output is the primary quantity of interest. In a radar system, the contributors to this noise include the external noise, the intrinsic $kT_0\beta_n$ thermal noise, and additional thermal noise due to losses in the antenna structure and nonideal receivers. Detailed noise analyses assign individual equivalent noise temperatures to each stage in the system; a good introductory description is given in Curlander and McDonough (1991). When considering the system as a whole, it is common to express the total output noise power as the sum of the power that would be observed due to the minimum noise density kT_0 at the input and a second term that accounts for the additional noise due to the nonideal system

$$N = kT_0\beta_nG_s + kT_e\beta_nG_s$$
(2.82)

In this equation, G_s is now the power gain of the complete receiver system, including antenna loss effects. The equivalent temperature *Te* used to account for noise above the theoretical minimum is called the *effective temperature* of the system.

The noise temperature description of noise power is most useful for lownoise receivers. An alternative description common in radar is based on the idea of *noise figure* F_n , which is the ratio of the actual noise power at the output of a system to the minimum power $kT_0\beta_nG_s$ (Skolnik, 2001). As with noise temperatures, various noise figures can be defined to include the effects of just the receiver, or of the entire antenna and receiver system, and so forth. Here, the term *noise figure* used without qualification will mean the noise figure of the complete receiver system, so that

$$F_n = \frac{N}{kT_0\beta_n G_{\rm s}}$$
(2.83)

Equation (2.83) shows that knowledge of the noise equivalent bandwidth, gain,

and noise figure of the receiver system are sufficient to calculate the output noise power using $N = kT_0\beta_nF_nG_s$. It also follows from using Eq. (2.82) in Eq. (2.83) that $Te = (F_n - 1)T_0$. Typical noise figures for radars can be as low as 2 or 3 dB, and as high as 10 dB or more. Corresponding effective temperatures range from about 170 K to over 2600 K.

In Sec. 2.2, the term "radar range equation" was applied to Eqs. (2.11), (2.25), (2.30), and (2.32). These expressions described the echo power received by the radar given various system and propagation conditions. As will be seen in Chap. 6, the detection performance of a radar depends not on the received power per se but on the SNR at the point of detection. Equation (2.83) can be used to convert the power range equations to SNR range equations.

To illustrate, consider the point target range equation [Eq. (2.11)], which expresses the power P_r of the signal available at the input to the receiver. The signal power at the output will be $P_0 = G_s P_r$ provided the signal bandwidth is entirely contained within the receiver bandwidth B_n . From Eq. (2.83), the output noise power is $N_o = kT_0\beta_nF_nG_s$. The SNR is therefore

$$\chi = \frac{P_o}{N_o} = \frac{G_s P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 L_s L_a(R)} \cdot \frac{1}{k T_0 \beta_n F_n G_s} = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 R^4 k T_0 \beta_n F_n L_s L_a(R)}$$
(2.84)

The last expression in Eq. (2.84) gives the SNR in terms of transmitter and receiver characteristics, target RCS, range, and loss factors. Modifications of Eqs. (2.25), (2.30), and (2.32) for volume and area scatterers to express them in terms of signal to noise ratio are obtained in the same manner by simply including the quantity $kT_0\beta_nF_n$ in their denominators.

Equation (2.84) represents the SNR at the receiver output, but prior to any signal processing. The point of most of the techniques discussed in this text is to increase the SNR above that value through signal processing means so as to obtain better detection, measurement, and imaging results. The impact of signal processing on the SNR can be modeled by simply adding a signal processing gain term G_{sp} to the range equation:

$$\chi = \frac{P_t G^2 \lambda^2 \sigma G_{sp}}{(4\pi)^3 R^4 k T_0 \beta_n F_n L_s L_a(R)}$$
(2.85)

In ensuing chapters, G_{sp} will be expressed in terms of the parameters of specific

techniques such as matched filtering and Doppler processing.

Like Eq. (2.11), Eq. (2.85) is also often called the radar range equation. In the remainder of this text, the term "range equation" or "radar range equation" usually refers to the SNR form of Eq. (2.85) and its analogues for volume and area scatterers.

2.5 Jamming

Jamming refers to intentional interference directed at the radar system from a hostile emitter. Jamming is an example of *electronic countermeasures* (ECM) or *electronic attack* (EA). As noted earlier, the purpose of most radar signal processing is to improve the SIR of the data so as to improve detection, tracking, and imaging performance. The purpose of many jamming techniques is just the opposite: to reduce the SIR so that the radar performance is degraded.

The most basic form of jamming is simple noise jamming. A hostile emitter directs an amplified noise waveform at the victim radar, essentially increasing the noise level out of the receiver. If the noise power spectrum fills the entire radar receiver bandwidth, then the noise out of the receiver will appear like any other white noise process and is modeled in the same way. More advanced forms of noise jamming use various amplitude and frequency modulations. Instead of noise, other jamming techniques use waveforms designed to mimic target echoes and fool the radar into detecting and tracking nonexistent targets.

Even a limited discussion of ECM is outside the scope of this text, due both to the breadth of the topic and the limited amount of material publishable in the open literature. The reader is referred to Lothes et al. (1990) for a good general reference on jamming signals in radar.

2.6 Frequency Models: The Doppler Shift

2.6.1 Doppler Shift

If the radar and scatterer are not at rest with respect to each other, the frequency F_r of the received echo will differ from the transmitted frequency F_t due to the Doppler effect. Doppler shifts can be used to advantage to detect echoes from moving targets in the presence of much stronger echoes from clutter or to drastically improve cross-range resolution. Uncompensated Doppler shifts can also have harmful effects, particularly a loss of sensitivity for some types of waveforms. Thus, characterization and measurement of Doppler shifts is an important topic in radar.

Consider an arbitrary waveform x(t), pulsed or not, transmitted by a monostatic radar. The waveform is reflected from a perfectly conducting target at an arbitrarily time-varying range R(t). For instance, a constant-range target

would have R(t) equal to a fixed R_0 meters, while a constant-velocity target would have $R(t) = R_0 - vt$ meters.² It makes no difference whether the radar, the target, or both are moving such that the range between the two is R(t), so it can be assumed without loss of generality that the radar is stationary and the target is moving, and that all measurements are made in the frame of reference of the radar. Under these conditions the received signal can be shown to be (Cooper, 1980; Gray and Addison, 2003)

$$\overline{y}(t) = -k \cdot [1 - 2h(t)] \,\overline{x} \, [2h(t) - t]$$
(2.86)

where k absorbs all radar range equation amplitude factors and h(t) is the function that satisfies

$$h(t) + \frac{1}{c}R[h(t)] = t$$
(2.87)

The dot over h(t) in Eq. (2.86) denotes the time derivative. The minus sign (180° phase shift) is required by the boundary conditions at a perfectly conducting surface. The function h(t), which has units of seconds, is the time at which a wave must have been launched in order to intercept the moving target at time *t* and range R(t). For example, if R(t) is a constant R_0 , then $h(t) = t - R_0/c$.

For instantaneous velocities R(t) that are a small fraction of the speed of light (virtually always the case as will be discussed shortly), the "quasi-stationary" assumption is commonly made. This holds that the range change during the short flight of any particular point in the waveform from the transmitter to the target is negligible. Then $R[h(t)] \approx R(t)$ so that (Cooper, 1980)

$$h(t) \approx t - \frac{1}{c}R(t)$$

$$\overline{y}(t) \approx -k \cdot \left[1 - 2\dot{h}(t)\right]\overline{x}\left[t - \frac{2R(t)}{c}\right] = -k\left[\frac{2\dot{R}(t)}{c} - 1\right]\overline{x}\left[t - \frac{2R(t)}{c}\right]$$

$$\approx k \cdot \overline{x}\left[t - \frac{2R(t)}{c}\right]$$
(2.88)

The last step also uses the assumption $\hat{R}(t) \ll c$. This result is exact when the target is stationary, $R(t) = R_0$. Then $h(t) = t - R_0/c$ exactly and $\overline{y}(t) = k \cdot \overline{x}[t - 2R_0/c]$ exactly.

The case of a constant-velocity target is of special interest. Returning to the exact result of Eqs. (2.86) and (2.87), let $R(t) = R_0 - vt$ and define $\beta_v \equiv v/c$. It is easy to show that

$$h(t) = \frac{1}{1 - \beta_v} \left(t - \frac{R_0}{c} \right)$$
$$[1 - 2\dot{h}(t)] = -\frac{1 + \beta_v}{1 - \beta_v} \equiv -\alpha_v$$
$$(2.89)$$

so that (Gleiser, 1979; DiFranco and Rubin, 2004; Peebles, 1998)

$$\overline{y}(t) = k\alpha_v \cdot \overline{x} \left(\alpha_v \left(t - \frac{2R_0}{(1+\beta_v)c} \right) \right)$$
(2.90)

If the transmitted waveform is a standard RF pulse

$$\overline{x}(t) = A(t) \exp[j(2\pi F_t t + \varphi_0)]$$
(2.91)

where A(t) is the pulse envelope, the received echo waveform will be

$$\overline{y}(t) = k\alpha_v \cdot A\left(\alpha_v \left(t - \frac{2R_0}{(1+\beta_v)c}\right)\right) \exp\left[j\left(2\pi F_t \alpha_v \left(t - \frac{2R_0}{(1+\beta_v)c}\right) + \varphi_0\right)\right]$$
$$= k\alpha_v \cdot A\left(\alpha_v t - \frac{2R_0}{(1-\beta_v)c}\right) \exp\left[j\left(2\pi\alpha_v F_t t - \frac{4\pi R_0}{(1-\beta_v)\lambda} + \varphi_0\right)\right]$$
(2.92)

Inspection of Eq. (2.92) reveals several characteristics of the received signal. Its frequency is $\alpha_v F_t$ Hz. The *change* in frequency is the Doppler shift F_D :

$$F_D = \alpha_v F_t - F_t = (\alpha_v - 1)F_t = \frac{2v}{(1 - \beta_v)\lambda} \quad \text{Hz}$$
(2.93)

The Doppler shift is positive for approaching targets (v > 0) and negative for receding targets as expected. The phase of the received signal is decreased by

$$\Delta \varphi = -\frac{4\pi R_0}{(1 - \beta_v)\lambda} \text{ radians}$$
(2.94)

The waveform is scaled in time by the factor α_{ν} . For example, for an approaching target, $\alpha_{\nu} > 1$ and a transmitted pulse is shortened by a factor of α_{ν}

on reception; for a receding target it is lengthened by a factor of α_v . The compression (expansion) of the pulse in time results in an expansion (contraction) of the pulse bandwidth by the factor α_v due to the reciprocal spreading property of Fourier transforms. Finally, the amplitude of the waveform is scaled by the factor α_v (in addition to the range equation effects), a consequence of conservation of energy when the time scale is altered.

It is virtually always the case in radar that the ratio $|\beta_v| = |v/c|$ is very small. For example, a car traveling at 60 mph (26.82 m/s) has a ratio |v/c| of 8.94 × 10⁻⁸; an aircraft at Mach 1 (about 340.3 m/s at sea level) has $|v/c| = 1.13 \times 10^{-6}$; and even a low-earth orbit (LEO) satellite with a velocity of 7800 m/s has $|v/c| = 2.6 \times 10^{-5}$. Expand each of the terms $1/(1 \pm \beta_v)$ and $\alpha_v = (1 + \beta_v)/(1 - \beta_v)$ in a binomial series and retain terms only to first order in β_v :

$$\frac{1}{1\pm\beta_{v}} = 1\mp\beta_{v} + \beta_{v}^{2}\mp\beta_{v}^{3} + \dots \approx 1\mp\beta_{v}$$

$$\alpha_{v} = \frac{1+\beta_{v}}{1-\beta_{v}} = (1+\beta_{v})\left(\frac{1}{1-\beta_{v}}\right) = (1+\beta_{v})(1+\beta_{v}+\beta_{v}^{2}+\beta_{v}^{3}+\dots) \approx 1+2\beta_{v}$$
(2.95)

Equation (2.90) and the sinusoidal pulse special case of Eq. (2.92) then reduce to

$$\overline{y}(t) \approx k(1+2\beta_v) \cdot \overline{x} \left[(1+2\beta_v)t - (1+\beta_v)\frac{2R_0}{c} \right]$$

$$= k(1+2\beta_v) \cdot A \left[(1+2\beta_v)t - \frac{2(1+\beta_v)R_0}{c} \right] \exp \left[j \left(2\pi (1+2\beta_v)F_t t - \frac{(1+\beta_v)4\pi R_0}{\lambda} + \varphi_0 \right) \right]$$
(sinusoidal pulse)
(2.96)

The echoed pulse length $\tau' = \tau/\alpha \approx (1 - 2\beta_v)\tau$. This small change in the pulse envelope duration of $2\alpha_v\tau$ seconds is insignificant and can be ignored. The amplitude factor of $\alpha \approx (1 + 2\beta_v)$ is certainly negligible compared to range equation effects and can also be ignored. The change in delay from $2R_0/c$ to 2(1 + β_v)R_0/c represents a percentage change of β_v in the delay and is also usually insignificant, though for a system with fine range resolution at long enough ranges the error could become a significant fraction of a range resolution cell. However, β_v cannot be neglected in the phase term because the factor of $4\pi\beta_vR_0/\lambda$ will frequently be a large fraction or even a multiple of π . With these three approximations to the envelope term and amplitude, the Doppler shift effects on the sinusoidal pulse of Eq. (2.96) reduce to

$$\overline{y}(t) \approx k \cdot A\left(t - \frac{2R_0}{c}\right) \exp\left[j\left(2\pi(F_t + 2\beta_v F_t)t - (1 + \beta_v)\frac{4\pi R_0}{\lambda} + \varphi_0\right)\right]$$
$$= k \cdot A\left(t - \frac{2R_0}{c}\right) \exp\left(-j(1 + \beta_v)\frac{4\pi}{\lambda}R_0\right) \exp\left[+j2\pi\left(\frac{2v}{\lambda}\right)t\right] \exp\left[j(2\pi F_t t + \varphi_0)\right]$$
(2.97)

The key result is that, to an excellent approximation, the pulse echoed from a constant-velocity target exhibits a Doppler shift of $2vF_t/c = 2v/\lambda$ Hz and a phase shift of $-(1 + \beta_v)4\pi/\lambda R_0$ radians.

The numerical values of Doppler shift are small compared to the RF frequencies. <u>Table 2.8</u> gives the magnitude of the Doppler shift corresponding to a velocity of 1 m/s at various radar frequencies. The Mach 1 aircraft observed with the L band radar would cause a Doppler shift of only 2.27 kHz in the 1 GHz carrier frequency.

Band	Frequency, GHz	Doppler Shift (Hz) for $v = 1 \text{ m/s}$
L	1	6.67
С	6	40.0
Х	10	66.7
Ka	35	233
W	95	633

 TABLE 2.8
 Doppler Shift Resulting from a Velocity of 1 m/s

For a monostatic radar and a constant-velocity target, the observed Doppler shift is proportional to the component of velocity in the direction of the radar, called the *radial velocity*. If the angle between the velocity vector relative to the radar of a target traveling at v meters per second and the vector from the radar position to the target position (sometimes called the *cone angle*) is ψ , the radial velocity is $v \cdot \cos\psi$ meters per second. The geometry is illustrated in two dimensions in Fig. 2.25. The magnitude of the Doppler shift is maximum when the target is traveling directly toward or away from the radar. The Doppler shift is zero, regardless of the target velocity, when the target is crossing orthogonally to the radar boresight.



FIGURE 2.25 Doppler shift is determined by the radial component of relative velocity between the target and radar.

Equations (2.86) and (2.87) can be solved for the exact behavior of other regular patterns of radar-target motion as well. The solution for constant acceleration is given in Gray and Addison (2003). Even where a closed form solution for h(t) is difficult or impossible to find, it can still be developed using an iterative approximation approach.

2.6.2 The Stop-and-Hop Approximation and Phase History

The quasi-stationary assumption of Eq. (2.88) provides a simplified but very useful model of reflection of a radar pulse from a target moving relative to the radar. Applying it to the pulsed waveform $A(t)\exp[j(2\pi F_t t + \phi_0)]$ and using the same envelope approximations employed to obtain Eq. (2.97) gives

$$\overline{y}(t) = k \cdot A\left(t - \frac{2R(t)}{c}\right) \exp\left[j\left(2\pi F_t\left(t - \frac{2R(t)}{c}\right) + \varphi_0\right)\right]$$
$$\approx k \cdot A\left(t - \frac{2R_0}{c}\right) \exp\left(-j\frac{4\pi}{\lambda}R(t)\right) \exp[j(2\pi F_t t + \varphi_0)]$$
(2.98)

where R_0 is the initial range at the time of pulse transmission. Equation (2.98) states that the echo is received with a time delay corresponding to the range at the beginning of the pulse transmission but with a phase modulation related to the time variation in range. This is the "stop" part of the *stop-and-hop* assumption common in radar analysis: the envelope of the echo appears as if the target motion effectively stopped while the pulse was in transit. The "hop"

portion will be discussed shortly.

Equation (2.98) adequately describes not only constant but also timevarying Doppler frequency shifts. If the target moves relative to the radar at constant velocity, $R(t) = R_0 - vt$,

$$\overline{y}(t) = k \cdot A\left(t - \frac{2R_0}{c}\right) \exp\left(-j\frac{4\pi}{\lambda}R_0\right) \exp\left[+j2\pi\left(\frac{2v}{\lambda}\right)t\right] \exp\left[j(2\pi F_t t + \varphi_0)\right]$$
(2.99)

Equation (2.99) is identical to the second line of Eq. (2.97), with the exception that the constant phase shift is $-4\pi R_0/\lambda$ instead of $-(1 + \beta_v)4\pi R_0/\lambda$ radians. This difference in the constant phase shift does not affect the magnitude or Doppler shift of the echo and can be ignored. Thus the analysis approach of Eq. (2.88) is consistent with the earlier results in all important respects.

For a more interesting example of the use of Eq. (2.88), consider Fig. 2.25 again. Let the radar be located at (x, y) coordinates $(x_r = 0, y_r = 0)$ with its antenna aimed in the +y direction, and let the coordinates of the target aircraft be $(x_t = vt, y_t = R_0)$. This means that the target aircraft is on the radar boresight at a range R_0 at time t = 0 and is crossing orthogonal to the radar line of sight at a velocity v meters per second. The range between radar and aircraft is

$$R(t) = \sqrt{R_0^2 + (vt)^2} = R_0 \sqrt{1 + \left(\frac{vt}{R_0}\right)^2}$$
(2.100)

While it is possible to work with Eq. (2.100) directly, it is common to expand the square root in a power series:

$$R(t) = R_0 \left[1 + \frac{1}{2} \left(\frac{vt}{R_0} \right)^2 - \frac{3}{8} \left(\frac{vt}{R_0} \right)^4 - \cdots \right]$$
(2.101)

In evaluating this expression, the range of t that must be considered may be limited by any of several factors, such as the time the target is within the radar main beam or the coherent processing interval duration over which pulses will be collected for subsequent processing.

Assume that the distance traveled by the target within this time of interest is much less than the nominal range R_0 so that higher-order terms in (vt/R_0) can be neglected:

$$R(t) \approx R_0 + \left(\frac{v^2}{2R_0}\right) t^2$$

(2.102)

Equation (2.102) shows that the range is approximately a quadratic function of time for the crossing target scenario of Fig. 2.25. Using this truncated series in Eq. (2.98) gives

$$\overline{y}(t) \approx k \cdot A\left(t - \frac{2R_0}{c}\right) \exp\left(-j\frac{4\pi}{\lambda}R_0\right) \exp\left[-j2\pi\left(\frac{v^2}{R_0\lambda}\right)t^2\right] \exp(j2\pi F_t t + \varphi_0)$$
(2.103)

All of the terms are the same as in the constant-velocity case of Eq. (2.99) except for the middle exponential. Recall that instantaneous frequency is proportional to the time derivative of phase. The quadratic phase function therefore represents a Doppler frequency shift that varies linearly with time due to the changing radar-target geometry:

$$F_D(t) = \frac{1}{2\pi} \frac{d}{dt} \left[-2\pi \left(\frac{v^2}{R_0 \lambda} \right) t^2 \right] = -\frac{2v^2}{R_0 \lambda} t$$
(2.104)

As the target aircraft approaches from the left in Fig. 2.25 (t < 0) the instantaneous Doppler shift is positive. When the aircraft is abreast of the radar (t = 0) the Doppler shift is zero because the radial component of velocity is zero. Finally, as the aircraft passes by the radar (t > 0) the Doppler shift becomes negative, as would be expected for a receding target. This quadratic range case is important in synthetic aperture radar and will be revisited in Chap. <u>8</u>.

The exponential term $\exp(-j4\pi R(t)/\lambda)$ in Eq. (2.98) is called the *phase* history of the received signal. This terminology is applied both to the complex exponential and to just its phase function $(-4\pi R(t)/\lambda)$. The phase history encodes the variation of the range between the target and radar during the data collection time. For the constant-velocity example [Eq. (2.99)], the phase history is a linear function of time corresponding to a constant frequency sinusoid, i.e., a constant Doppler shift. For the crossing target example of Eq. (2.103), it is a quadratic function of time, producing a Doppler shift sinusoid having a frequency that varies linearly with time. Other radar-target motions will produce other functional forms for the phase history.

More generally, the term phase history can refer to the variation of phase (or the corresponding complex exponential) in any dimension of the radar data. Two other common uses are to describe the fast-time phase function of a frequency- or phase-modulated waveform or the spatial phase variation across the face of an array antenna at a fixed time. As will be seen, the phase history is central to radar signal processing. The design of many important radar signal processing operations depends critically on accurately modeling or estimating the phase history of the collected data. Examples include pulse compression, adaptive interference cancellation, and imaging.

2.6.3 Measuring Doppler Shift: Spatial Doppler

The Doppler shifts observed in radar are too small to be measured from a single pulse echo in most cases. In Chap. 7 it will be seen that a lower bound on the standard deviation of the error in measuring the frequency of a complex sinusoid with unknown amplitude, frequency, and phase using a discrete Fourier transform (DFT) and an observation of length T_{obs} seconds at an integrated SNR in the DFT of χ is $\sigma_F = \sqrt{6/(2\pi)^2 \chi T_{obs}^2}$ Hz. Applying this to measuring Doppler, this value must be much less than the Doppler shift if that shift is to be measured with reasonable precision, leading to a requirement that $T_{obs} \gg \sqrt{6/(2\pi)^2 \chi F_D^2}$. Even for a rather high Doppler shift of 10 kHz and a very good SNR of 30 dB (χ = 1000), T_{obs} must be much larger than 123 μ s. To measure the Doppler shift with a single pulse would therefore require pulse lengths greater than 1 ms, much longer than the sub-millisecond (usually less than 100 μ s) pulse lengths typically used. For a 1-kHz Doppler shift and 20-dB SNR, a pulse longer than 10 ms would be needed. For this reason, most radars do not measure Doppler shift on an intrapulse basis, although a few designed for very high speed targets (satellites and missiles) and using very long pulses can do so.

The long observation time needed can be obtained by using multiple pulses. Suppose a series of M distinct pulses of duration τ are transmitted beginning at times $t_m = mT$, where T is the pulse repetition interval (PRI). The *m*th transmitted pulse and received echo (using the quasi-stationary assumption) are

$$\overline{x}_{m}(t) = A(t - mT) \exp\left[j(2\pi F_{t}t + \phi_{0})\right]$$
(2.105)

$$\overline{y}_{m}(t) \approx k \cdot A \left[t - mT - \frac{2R(mT)}{c} \right] \exp \left\{ j \left[2\pi F_{t} \left(t - \frac{2R(t)}{c} \right) + \varphi_{0} \right] \right\}$$
(2.106)

After demodulation, the baseband received signal is

$$y_m(t) = k' \cdot A\left(t - mT - \frac{2R(mT)}{c}\right) \exp\left[-j\frac{4\pi}{\lambda}R(t)\right]$$
(2.107)

where k' includes the $\exp(-j\phi_0)$ term. Assume each baseband pulse echo is sampled $2R_s/c$ seconds after transmission, corresponding to a range R_s . Also

assume a target is present within the range bin corresponding to that sample time for the entire data collection time of mT seconds, meaning that R(t) remains in the range interval $[R_s-c\tau/2, R_s]$.[§] The *m*th sample in this range bin is then

$$y_{m}\left(mT + \frac{2R_{s}}{c}\right) = k' \cdot A\left(\frac{2}{c}[R_{s} - R(mT)]\right) \exp\left[-j\frac{4\pi}{\lambda}R\left(mT + \frac{2R_{s}}{c}\right)\right]$$
$$= \hat{k} \cdot \exp\left[-j\frac{4\pi}{\lambda}R\left(mT + \frac{2R_{s}}{c}\right)\right]$$
$$\equiv y[m]$$
(2.108)

The constant \hat{k} combines k' and the amplitude of the sampled pulse envelope $A(\cdot)$. The series of sampled echoes y[m] forms the slow-time series of samples for that range bin, as will be described in <u>Chap. 3</u>.

The "stop" assumption applied in Eq. (2.98), when used across a series of pulses as in (2.107), is called the *stop-and-hop* approximation. Relative to the radar, the target is assumed to "stop" at the time of each pulse transmission at the corresponding range R(mT) and then "hop" to the range at the next pulse transmission time, rather than moving continuously.

Consider again a constant velocity target, $R(t) = R_0 - vt$. The slow-time data series becomes

$$y[m] = \hat{k} \cdot \exp\left(-j\frac{4\pi}{\lambda} \left(R_0 - v \cdot \left(mT + \frac{2R_s}{c}\right)\right)\right)$$
$$= \hat{k} \cdot \exp\left[-j\frac{4\pi}{\lambda} (R_0 - 2\beta_v R_s)\right] \exp\left[+j2\pi \left(\frac{2v}{\lambda}\right)mT\right]$$
(2.109)

The first exponential in Eq. (2.109) is a constant phase shift for all of the slowtime samples y[m] and is of little consequence. The second exponential is a discrete complex sinusoid with normalized frequency $2vT/\lambda$ cycles/sample, corresponding to the expected Doppler frequency of $2v/\lambda$ Hz. Thus, the phase history obtained from a moving target using a series of pulses provides a way to measure the Doppler shift with good precision by observing the signal over an observation time much longer than that of a single pulse.

The manifestation of the target Doppler shift in the slow-time phase history is sometimes referred to as *spatial Doppler*. This terminology emphasizes the fact that the Doppler shift is measured not from intrapulse frequency changes, but rather from the change of phase of the echoes at a given range bin over a series of pulses. Because of the inability to measure intrapulse Doppler frequency shifts in most systems, the term "Doppler processing" in radar usually refers to sensing and processing this spatial Doppler information.

2.7 Spatial Models

Previous sections have dealt with models of Doppler shift and the received power (both mean value and statistical fluctuations) of radar echoes from a single resolution cell. In this section, the variation in received complex voltage or power as a function of the spatial dimensions of range and angle will be considered. It will be seen that the observed complex voltage can be viewed as the output of a linear filter with the weighted variation in reflectivity over range or angle as its input. A similar result holds for power when the reflectivity field has a random phase variation. These relationships will lay the groundwork for an analysis of data sampling requirements and range and angle resolution in subsequent chapters.

2.7.1 Coherent Scattering

Consider a stationary pulsed radar. At time zero it transmits the equivalent complex signal

$$\overline{x}(t) = \sqrt{P_t} x(t) \exp[j(2\pi F_t t + \varphi_0)]$$
(2.110)

Assume that $\overline{x}(t)$ has unit amplitude so that the transmitted signal amplitude is represented by the term $\sqrt{P_t}$. This signal echoes off a differential scatterer of cross section $d\sigma(R, \theta, \phi)$ at coordinates (R, θ, ϕ) . The *baseband complex reflectivity* or just *reflectivity* of the differential scatterer is, from Eq. (2.50), $d\zeta(R, \theta, \phi) \exp[j\psi(R, \theta, \phi)]$ so that $d\sigma = |d\zeta|^2$. The term involving ψ accounts for a possible constant phase shift on reflection at the scatterer surface. The antenna is assumed to be mechanically scanning⁹ in either or both angle coordinates with one-way voltage pattern $E(\theta,\phi)$ so that at the time of transmission it is steered in the direction (θ_0,ϕ_0) . Then analogously to Eq. (2.16), the differential received voltage is

$$d\overline{y}(\theta_0, \phi_0, t; R, \theta, \phi) = \sqrt{\frac{P_t \lambda^2}{(4\pi)^3 R^4 L_s L_a(R)}} E^2(\theta - \theta_0, \phi - \phi_0) \cdot d\zeta(R, \theta, \phi)$$
$$\times \exp[j\psi(R, \theta, \phi)] x \left(t - \frac{2R}{c}\right) \exp\left[j2\pi F_t \left(t - \frac{2R}{c} + \phi_0\right)\right]$$
(2.111)

where $E(\theta,\phi)$ is the one-way antenna voltage pattern. Equation (2.110) can be simplified by separating the reflectivity terms and the terms which depend on spatial location and collapsing all of the other system-dependent amplitude terms into a single constant $A_r = \sqrt{P_t \lambda^2 / (4\pi)^3 L_s}$. The term $d\zeta \exp(j\psi)$ is termed the *baseband complex reflectivity* or just *reflectivity* of the differential scatterer and will be denoted as $d\rho$. Making these substitutions and removing the carrier term with coherent demodulation leaves only the baseband complex received voltage dy for the single differential scatterer

$$dy(\theta_0, \phi_0, t; R, \theta, \phi) = A_r d\rho(R, \theta, \phi) \left[\frac{E^2(\theta - \theta_0, \phi - \phi_0)}{\sqrt{L_a(R)}R^2} x \left(t - \frac{2R}{c} \right) \exp\left(-j\frac{4\pi}{\lambda}R\right) \right]$$
(2.112)

Equation (2.112) gives the contribution to the received echo voltage of a differential scatterer element at coordinates (R, θ, ϕ) . The total received voltage is obtained by integrating these differential contributions over all space:

$$y(\theta_0, \phi_0, t) = \int_{\phi = -\pi/2}^{\pi/2} \int_{\theta = -\pi}^{\pi} \int_{R=0}^{\infty} dy(\theta_0, \phi_0, t; R, \theta, \phi)$$
(2.113)

Equation (2.113) is a coherent scattering model: the differential scatterers are assumed to add as complex voltages. This is most appropriate for reflectivity fields characterized by relatively static configurations of scatterers, e.g., man-made vehicle and urban areas. The case where the scatterers are not static is considered in Sec. 2.7.4.

Now write $d\rho(R, \theta, \phi) = \rho(R, \theta, \phi) \cdot dV = \rho(R, \theta, \phi) R^2 \cos\phi dR d\theta d\phi$ to obtain

$$y(\theta_{0},\phi_{0},t) = A_{r} \int_{\phi=-\pi/2}^{\pi/2} \int_{\theta=-\pi}^{\pi} \int_{R=0}^{\infty} \left\{ \frac{\exp[-j(4\pi/\lambda)R]}{\sqrt{L_{a}(R)}} \rho(R,\theta,\phi) \right\} \times \left[E^{2}(\theta-\theta_{0},\phi-\phi_{0})x\left(t-\frac{2R}{c}\right) \cos\phi dR \, d\theta \, d\phi \right]$$

$$(2.114)$$

Define the *effective reflectivity* ρ' to include the attenuation due to atmospheric loss, the phase rotation due to two-way propagation range, and the $\cos\phi$ term of the differential volume element:

$$\rho'(R,\theta,\phi) = \frac{\exp[-j(4\pi/\lambda)R]}{\sqrt{L_a(R)}}\rho(R,\theta,\phi)\cos\phi$$
(2.115)

Applying Eq. (2.115) to Eq. (2.114), the received signal is seen to be similar to

a *three-dimensional convolution* of the effective reflectivity with a convolution kernel comprising the antenna two-way voltage pattern in the angle coordinates and the pulse modulation function in the range coordinate. Specifically,

$$y(\theta_0, \phi_0, t) \approx A_r \,\rho' \,(ct/2, \theta_0, \phi_0) *_t *_\theta *_\phi \left[E^2(-\theta_0, -\phi_0) \,x(t)\right]$$
(2.116)

where the symbols $*_t$, $*_{\theta}$, and $*_{\phi}$ denote convolution over the indicated coordinate. Now assume the antenna pattern is symmetric in the two angular coordinates, as is often the case; rescale the time variable to units of range; and replace θ_0 and ϕ_0 with general angular variables θ and ϕ . These substitutions finally give

$$y(\theta, \phi, R) \approx A_r \rho'(R, \theta, \phi) *_R *_{\theta} *_{\phi} \left[E^2(\theta, \phi) x\left(\frac{2R}{c}\right) \right]$$
(2.117)

Equations (2.116) and (2.117) are stated as approximate convolutions because of the finite integration limits in the angular variables, which arise due to the periodicity in angle of the antenna pattern and scene reflectivity. A full discussion of a spherical convolution-like equation similar to Eq. (2.114), including development of the Fourier transform relations, is given in Baddour (2010). Nonetheless, like a linear convolution, Eq. (2.117) computes the output at a given point in space as a local average of the reflectivity distribution, weighted by the antenna pattern and waveform. For most antennas and pulses, these patterns concentrate most of their energy in a relatively small finite region defined by the mainlobe for the antenna pattern and the pulse duration for the waveform. Consequently, the output signal can be expected to behave like a true linear convolution.

The convolutional model of Eq. (2.117) is an important result. Its significance is that it allows interpretation of the measured data as the result of a linear filtering process, so that Fourier transform relations between $y(\theta, \phi, R)$, $\rho'(R, \theta, \phi)$, $E^2(\theta, \phi)$, and x(t) can be established and applied to model signal properties, determine sampling rates, and so forth. For example, the range resolution of the measured reflectivity function is seen to be limited by the pulse duration. (In Chap. 4 it will be seen that the introduction of matched filtering will significantly change this statement.) Similarly, for a conventional scanning radar, the angular resolution will be determined by the antenna beamwidth. (In Chap. 8 it will be seen that the introduction of synthetic aperture techniques also significantly changes this statement.) It also follows from the filtering action of x(t) and $E^2(\theta, \phi)$ that the bandwidth of the measured reflectivity function in range and angle is limited by the bandwidth of the waveform modulation function and antenna power pattern. This observation will be used in Chap. 3 to

determine the range and angle sampling requirements.

2.7.2 Variation with Angle

Now consider the variation in reflectivity with angle for a fixed range, say R_0 . Define the range-averaged effective reflectivity

$$\hat{\rho}(\theta,\phi;R_0) = \int_R x \left[\frac{2}{c}(R_0 - R)\right] \rho'(R,\theta,\phi) dR$$
$$= \left[\rho'(R,\theta,\phi) *_R x \left(\frac{2R}{c}\right)\right]_{R=R_0}$$
(2.118)

This is the reflectivity variation in angle, taking into account the range averaging at each angle due to the finite pulse length. Note that in the limit of very fine range resolution, i.e., if the pulse modulation $x(2R/c) \rightarrow \delta_D(R - R_0)$, then $\hat{\rho}(\theta, \phi, R_0) \rightarrow \rho'(R_0, \theta, \phi)$, that is, the "range-averaged" reflectivity would exactly equal the effective reflectivity evaluated at the range of interest R_0 .

Applying <u>Eq. (2.118)</u> to <u>Eq. (2.117)</u> gives

$$y(\theta,\phi;R_0) = A_r \int_{\xi=-\pi/2}^{\pi/2} \int_{\zeta=-\pi}^{\pi} E^2(\zeta-\theta,\xi-\phi)\hat{\rho}(\zeta,\xi;R_0)d\zeta\,d\xi$$
$$\approx \hat{\rho}(\theta,\phi;R_0) *_{\theta} *_{\phi} E^2(\theta,\phi)$$
(2.119)

where again symmetry of the antenna pattern has been assumed in the second line. Equation (2.119) is a special case of Eq. (2.117) showing that the complex voltage at the output of a coherent receiver for a fixed range and a scanning antenna is approximately the convolution in the angle dimensions of the range-averaged effective reflectivity function evaluated at the range R_0 , $\hat{\rho}(\theta, \phi; R_0)$ with the antenna two-way voltage pattern $E^2(\theta, \phi)$.

As mentioned earlier, the interpretation of Eq. (2.119) as a linear convolution is an approximation. Suppose that the elevation angle ϕ is fixed, and consider only the variation in azimuth angle θ . Because the integration is over a full 2π radians and the integrand is periodic in θ with period 2π , the integration over azimuth is a circular convolution of periodic functions.

This would not appear to be the case if instead θ is fixed and ϕ varies because the integrand is over a range of only π radians. However, one could equally well replace Eq. (2.119) as

$$y(\theta, \phi; R_0) = A_r \int_{\xi = -\pi}^{\pi} \int_{\zeta = -\pi/2}^{\pi/2} E^2(\zeta - \theta, \xi - \phi)\hat{\rho}(\zeta, \xi; R_0) d\zeta d\xi$$

= $\hat{\rho}(\theta, \phi; R_0) *_{\theta} *_{\phi} E^2(\theta, \phi)$ (2.120)

For fixed azimuth, this is now a circular convolution of periodic functions in elevation. Taken together, the two integrals over the angular variables implement a two-dimensional weighting and averaging over the (θ, ϕ) space. So long as the antenna beamwidths are small compared to 2π , this circular convolution will closely approximate a linear convolution in the vicinity of (θ, ϕ) .

Figure 2.26 illustrates intuitively in one angle dimension the process described by Eq. (2.119). Assume that the elevation angle is fixed at $\phi = 0^{\circ}$ and consider only the azimuth variation. An array of ideal point scatterers is illuminated by a radar that scans in azimuth across the target field. The response to any one scatterer is maximum when the radar boresight is aimed at that scatterer; as the radar boresight moves away, the strength of the echo declines because less energy is directed to the scatterer on transmission, and the antenna is also less sensitive to echoes from directions other than the boresight on reception. For an isolated scatterer, the amplitude of the coherent baseband received signal $y(\theta, 0; R_0)$ at the receiver output will be proportional to $E^2(\theta,$ 0). Thus, a graph of the received signal mimics the antenna two-way azimuth voltage pattern.



FIGURE 2.26 When scanning past an array of point scatterers, the receiver

output is a superposition of replicas of the antenna pattern.

Assuming a linear receiver so that superposition applies, the response to two closely spaced point scatterers is proportional to two replicas of the antenna pattern, overlapped and added to get a composite response. If the two scatterers are close enough together, the individual responses are not resolved, but instead blur together into a single peak as illustrated in Fig. 2.26. The details of the combined response depend on the relative phase of the two individual responses; they may combine in or out of phase, yielding significantly different composites. However, the separation at which scatterers are consistently resolved regardless of relative phase clearly depends on the antenna pattern $E^2(\theta, 0)$, and in particular on the mainlobe beamwidth.

Because of the approximately linear convolution relation of Eq. (2.119) the spatial Fourier transform of the observed signal is approximately the input spatial Fourier transform multiplied by the Fourier transform of the antenna pattern. Practical antenna patterns have lowpass spectra. Equation (2.121) gives the ideal two-way azimuth voltage patterns for circular and rectangular apertures of width D (Balanis, 2005):

$$E^{2}(\theta, 0) = \left[\frac{J_{1}(\pi D \sin \theta / \lambda)}{\pi D \sin \theta / \lambda}\right]^{2} \quad \text{(circular aperture)}$$
$$E^{2}(\theta, 0) = \left[\frac{\sin(\pi D \sin \theta / \lambda)}{\pi D \sin \theta / \lambda}\right]^{2} \quad \text{(rectangular aperture)} \quad (2.121)$$

)

Figure 2.27 plots these patterns on a decibel scale for the case $D = 40\lambda$.



FIGURE 2.27 Two-way antenna voltage patterns for ideal, uniformly

illuminated circular and rectangular apertures.

The corresponding spatial spectra are shown in Fig. 2.28. For the rectangular aperture, it is a triangle function with a support of twice the aperture width. The reason is easy to see: the one-way voltage pattern is just the inverse Fourier transform of the aperture function, which for uniform illumination is a rectangular pulse of the width of the aperture. When that pattern is squared to get the two-way pattern, the Fourier transform of the squared pattern is the self-convolution of the Fourier transform of the unsquared pattern. Thus, the rectangular aperture function is convolved with itself to give a triangle of twice the aperture width. The spectrum for the circular aperture has the same width but is somewhat smoother.



FIGURE 2.28 Spatial spectra corresponding to the antenna patterns of Fig. 2.27.

The spatial spectra of these idealized, but typical, antenna patterns are lowpass functions. Thus, the upper frequencies in the spatial spectrum of the observed data will be strongly attenuated and in fact effectively removed. Since resolution is proportional to bandwidth, Eq. (2.121) and Fig. 2.28 show that the antenna pattern reduces resolution because it has a strongly lowpass spatial spectrum.

2.7.3 Variation with Range

A development similar to that in Sec. 2.7.2 can be carried out to specialize Eq. (2.117) for the variation of received voltage in the range dimension along the boresight look direction (θ_0, ϕ_0) .¹⁰ First, interchange the order of integration in Eq. (2.114) so that the outer integral is over range. Next, define the new quantity

$$\tilde{\rho}(R;\theta_0,\phi_0) = \iint_{\theta,\phi} E^2(\theta-\theta_0,\phi-\phi_0)\rho'(R,\theta,\phi)d\theta d\phi$$
$$= \rho'(R,\theta,\phi) *_{\theta} *_{\phi} E^2(\theta,\phi) |_{\theta=\theta_0,\phi=\phi_0}$$
(2.122)

This is the reflectivity variation in range, taking into account the azimuth and elevation averaging at each range due to the nonideal antenna power pattern. Note that in the limit as the antenna power pattern tends to the ideal $E^2(\theta, \phi) \rightarrow G\delta_D(\theta, \phi)$, then $\tilde{\rho}(R;m) \rightarrow \rho'(R, \theta_m, \phi_m)$, that is, the "angle-averaged" reflectivity exactly equals the effective reflectivity along the antenna look direction, as expected.

Applying Eq. (2.122) to Eq. (2.117) leaves (Munson and Visentin, 1989)

$$y\left(\frac{2R}{c};\theta_{0},\phi_{0}\right) = A_{r}\tilde{\rho}(R;\theta_{0},\phi_{0})*_{R}\left[x\left(\frac{2R}{c}\right)\right]$$
$$= A_{r}\int_{R'=0}^{\infty} x\left[\frac{2}{c}(R-R')\right]\tilde{\rho}(R';\theta_{0},\phi_{0})dR'$$
(2.123)

or an equivalent equation, using time units instead of range units

$$y(t;\theta_0,\phi_0) = A_r \tilde{\rho} \left(\frac{ct}{2};\theta_0,\phi_0\right) *_t [x(t)]$$
$$= A_r \int_{t'=0}^{\infty} x(t-t') \tilde{\rho} \left(\frac{ct'}{2};\theta_0,\phi_0\right) dt'$$
(2.124)

Equation (2.123) or (2.124) shows that the complex voltage at the output of a coherent receiver versus time for a given antenna look direction is the convolution in the range dimension of the angle-averaged effective reflectivity function in that look direction, $\tilde{\rho}(R; \theta_0, \phi_0)$, with the waveform modulation function x(t).

2.7.4 Noncoherent Scattering

Equation (2.114) and its approximate form (2.117) assume coherent addition of individual differential scatterer echoes; that is, the complex amplitude (magnitude and phase) of the total response is the complex sum of the differential complex echoes. For distributed area or volume clutter contributing very large numbers of scatterers with essentially random phases such as rain or natural ground clutter (grass, trees, water, etc.), it is more useful to model the scatterer reflectivity as having a random phase with either a random or nonrandom magnitude. The total received signal is then also a random variable,

and the expected value of the received power becomes of interest.

From Eqs. (2.112) and (2.113), the power of the integrated received voltage can be written

$$\begin{aligned} \left| y(\theta_{0},\phi_{0},t) \right|^{2} &= \left\{ A_{r} \int_{V} \frac{\exp[-j(4\pi/\lambda)R_{1}]}{R_{1}^{2}\sqrt{L_{a}(R_{1})}} \rho(R_{1},\theta_{1},\phi_{1})E^{2}(\theta_{1}-\theta_{0},\phi_{1}-\phi_{0})x\left(t-\frac{2R_{1}}{c}\right)dV_{1} \right\} \\ &\times \left\{ A_{r} \int_{V} \frac{\exp[+j(4\pi/\lambda)R_{2}]}{R_{2}^{2}\sqrt{L_{a}(R_{2})}} \rho^{*}(R_{2},\theta_{2},\phi_{2})[E^{2}(\theta_{2}-\theta_{0},\phi_{2}-\phi_{0})]^{*}x^{*}\left(t-\frac{2R_{2}}{c}\right)dV_{2} \right\} \end{aligned}$$

$$(2.125)$$

The subscripts "1" and "2" distinguish the spatial variables in the two integrals.

Returning to the expression of the reflectivity distribution $\rho(R, \theta, \phi)$ as the product of its phase term $\exp[j\psi(R, \theta, \phi)]$ and its amplitude $\zeta(R, \theta, \phi)$, model ψ as uniformly distributed over $(0, 2\pi]$ and white in all three spatial variables, while ζ may be random or deterministic; if random, it is statistically independent of ψ . The autocorrelation function of ρ is then $s\rho(R, \theta, \phi) = |\zeta(R, \theta, \phi)|^2 \delta_D(R) \delta_D(\phi)$ and the mean received power becomes

$$\overline{\left|y(\theta_{0},\phi_{0},t)\right|^{2}} = A_{r}^{2} \int_{V} \frac{1}{R^{4}L_{a}(R)} \left|\rho(R,\theta,\phi)\right|^{2} \left|E^{2}(\theta-\theta_{0},\phi-\phi_{0})\right|^{2} \left|x\left(t-\frac{2R}{c}\right)\right|^{2} dV$$
$$= A_{r}^{2} \int_{V} \frac{1}{R^{4}} \left|\rho'(R,\theta,\phi)\right|^{2} \left|E^{2}(\theta-\theta_{0},\phi-\phi_{0})\right|^{2} \left|x\left(t-\frac{2R}{c}\right)\right|^{2} dV$$
(2.126)

Again assuming a symmetric antenna pattern, this becomes finally

$$\left|\overline{y(\theta_0,\phi_0,t)}\right|^2 \approx A_r^2 \frac{\left|\rho'(R,\theta,\phi)\right|^2}{R^2} *_R *_\theta *_\phi \left[\left|E^2(\theta,\phi)\right|^2 \cdot \left|x\left(\frac{2R}{c}\right)\right|^2\right]$$
(2.127)

Equation (2.127) is the noncoherent equivalent of (2.117). It shows the received power in the case of noncoherent scattering to be the 3D convolution of the squared reflectivity, weighted by R^{-2} , convolved with the two-way antenna power pattern and the waveform power envelope. Thus, the received power still obeys a convolutional model, but with differences in the range dependence. Similar results are developed for weather clutter in Sec. 4.4 of Doviak and Zrnic (1993).

2.7.5 Projections

The range-averaged reflectivity $\hat{\rho}(\theta, \phi; R_0)$ of Eq. (2.118) and the angle-averaged reflectivity $\tilde{\rho}(R; \theta_0, \phi_0)$ of Eq. (2.122) are examples of *projections*. In each case, the three-dimensional reflectivity is reduced in dimension by integrating over one or more dimensions. The range-averaged reflectivity was reduced to a two-dimensional function by integrating over range, while the angle-averaged reflectivity was reduced to a one-dimensional function by integrating over both angle coordinates.

The idea of projections, particularly the angle-averaged projection $\tilde{\rho}(R;\theta_0,\phi_0)$, will be important in deriving the polar format spotlight SAR algorithm in Chap. 8. The projections that will be needed are integrals over straight lines or planar surfaces. The averaging in Eq. (2.122) is over the surface of a sphere. However, for small beamwidths only a region of θ_3 radians in azimuth and ϕ_3 radians in elevation contributes significantly to the integral, and at long ranges this limited region is nearly planar.

2.7.6 Multipath

The convolutional model of the measured range profile is based on the assumption of superposition of backscattered fields and a one-to-one mapping of echo arrival time to range, $t \rightarrow R = ct/2$. The superposition of electric fields is a valid assumption, but the mapping of time to range may not be. To illustrate, consider Fig. 2.29, which diagrams two phenomena that violate this assumption. Figure 2.29*a* illustrates the problem of *multipath*, in which echoes from the same target arrive at the radar receiver via two different paths. The first is the direct path of total length $2R_0$. The second is the "multipath" or "ground bounce" path with length $R_0 + R_1 + R_2 > 2R_0$. Though not shown, it is also possible for a portion of the transmitted wave to arrive at the target via the ground bounce and be scattered back along both paths, meaning that there may also be an echo with a time delay corresponding to a two-way path length of $2(R_0 + R_1 + R_2)$. Consequently, one scatterer may produce echoes at three different apparent ranges if multipath is present. Whether these appear as distinct echoes depends on the relationship between the path length difference and the pulse length. The ground bounce echoes are often, but not always, significantly attenuated with respect to the direct path echo. The degree of attenuation depends on the bistatic scattering characteristics of the surface, the antenna pattern characteristics (because the multipath bounce is not on the peak of the mainlobe) and the problem geometry. As the range between target and radar varies, the path length difference also varies, so that the direct and multipath bounces may alternately add in and out of phase, provided the path length difference is such that the two received echoes overlap. Multipath is generally most significant for targets located at low altitude over a good reflecting surface such as a relatively smooth terrain or calm ocean and at long range, so that the grazing angles involved are small.



FIGURE 2.29 Illustration of two scattering phenomena which violate the one-toone mapping of time to range: (*a*) multipath, (*b*) multiple bounce.

Figure 2.29*b* illustrates the effect of *multiple bounce* echoes in a situation involving two scatterers. A portion of the energy reflected from the more distant scatterer bounces off the nearer scatterer, then reflects a second time off the distant scatterer and returns to the radar. Obviously additional multiple bounces are also possible. For the situation sketched, three apparent echoes will again result, with the third due to a phantom scatterer $2\Delta R$ behind the second actual scatterer. As with multipath, the amplitude of multiple bounce echoes often falls off rapidly, and the same considerations of in- and out-of-phase superposition apply.

These possible differences in the measured and actual reflectivity distributions do not mean that range profile measurements are not useful. They do mean that in situations where significant multipath or multiple bounce phenomena are possible, the range profiles must be interpreted with care.

2.8 Spectral Model

There is one more interpretation of the received radar signal that proves useful in subsequent chapters. The preceding two sections have emphasized linear filtering models of the spatial reflectivity distribution as observed through the received complex baseband signals. However, it was pointed out previously that radar cross section is a function of, among many other things, the radar frequency. Thus, it is useful to investigate the significance of the radar transmitted frequency F_t on the reflectivity measurements.

To understand the role of transmitted frequency, it is necessary to deal with the radar signal while it is still at the radar frequency F_t . If the development from Eq. (2.110) to Eq. (2.124) is repeated without demodulating the signals to baseband and the range variation is considered, the RF version of Eq. (2.124) can be obtained:

$$\overline{y}(t;\theta_0,\phi_0) = A_r \int_{t=0}^{\infty} x(t-t') \widetilde{\rho}\left(\frac{ct'}{2};\theta_0,\phi_0\right) \exp\left(j2\pi F_t t'\right) dt'$$
$$= A_r \left[\widetilde{\rho}\left(\frac{ct}{2};\theta_0,\phi_0\right) \exp\left[j(2\pi F_t t+\varphi)\right] \right] *_t [x(t)]$$
(2.128)

Now consider the Fourier transform of $\overline{y}(t; \theta_0, \phi_0)$ with respect to the time (range) variable *t*. Using simple properties of Fourier transforms gives

$$\overline{Y}(F;\theta_0,\phi_0) = \frac{2A_r}{c} e^{j\varphi} X(F) \tilde{P}\left[\frac{2(F-F_t)}{c}\right]$$
(2.129)

Figure 2.30 provides a pictorial interpretation of this equation under the assumption that the transmitted waveform x(t) is a narrowband waveform. In this case,



FIGURE 2.30 Pictorial interpretation of Eq. (2.130), illustrating the spectral sampling effect of a narrowband radar pulse.

$$\overline{Y}(F;\theta_0,\phi_0) \approx \frac{2A_r}{c} e^{j\varphi} \,\,\tilde{P}\left(\frac{-2F_t}{c}\right) X(F)$$
(2.130)

so that the amplitude of the spectrum of the received pulse, and therefore of the pulse itself, is proportional to the amplitude of the spectrum of the angle averaged range profile, evaluated at the transmitted frequency. Since it is the complex spectrum that appears in Eq. (2.124), both the amplitude and phase of the returned pulse are affected by the amplitude and phase of the reflectivity spectral sample. Equation (2.130) shows that a narrowband radar pulse can be interpreted as measuring a frequency sample of the spectrum of the angle-averaged reflectivity range variation.

Another case of interest occurs when x(t) is a wideband pulse of bandwidth β Hz. For some waveforms such as the linear frequency-modulated pulse, the magnitude of the spectrum X(F) is approximately a rectangle as shown in Fig. 2.31. The spectrum of the receiver waveform $\overline{y}(t; \theta_0, \phi_0)$ is then approximately that of the angle-averaged range profile over the bandwidth of the pulse, modified by the phase of the pulse spectrum:



FIGURE 2.31 Pictorial interpretation of Eq. (2.131) illustrating the spectral windowing effect of a wideband radar pulse.

$$\overline{Y}(F;\theta_0,\phi_0) \approx \frac{2A_r}{c} \widetilde{P}\left[\frac{2(F-F_t)}{c}\right] \exp\{j \cdot \arg[X(F)]\} \qquad -\frac{\beta}{2} \le F \le +\frac{\beta}{2}$$
(2.131)

In other words, the pulse spectrum acts as a window selecting a portion of the spectrum of the angle-averaged reflectivity. This result will be useful in understanding the use of linear FM and other modulated waveforms to achieve fine range resolution in <u>Chap. 4</u>.

2.9 Summary

An understanding of the nature of the signals of interest is prerequisite to the design of successful signal processing systems. This chapter reviews the most

common signal models used in designing and analyzing radar signal processors. It has been seen that multiple views of the radar echo are used: its variation in amplitude, space, time, and frequency, and deterministic and statistical interpretations of these variations.

Radar signal modeling traditionally focuses most strongly on amplitude models, that is, on radar cross section. RCS is viewed as a deterministic quantity, predictable in principle through the use of Maxwell's equations if the scattering is modeled accurately enough. The radar range equation in its many forms (only a very small subset of which has been introduced here) is the radar engineer's most fundamental tool for estimating received signal amplitude or, conversely, determining required system characteristics such as transmitted power or antenna gain.

The radar system is a measuring instrument, used to observe the variation of RCS in space. Its pulse function (modulation and carrier term) and antenna power pattern determine its measurement characteristics, which in turn determine the achievable resolution and required sampling rates. The effect of the radar measurement system on the spatial variation of observed RCS is well modeled by the convolution of the combined pulse-and-antenna pattern measurement kernel with the three-dimensional reflectivity function. This important observation means that the tools of linear systems analysis can be brought to bear to help analyze and understand the performance of radar systems. The carrier frequency, in combination with any Doppler shifts, determines what portion of the reflectivity frequency spectrum is sampled by the pulse. This observation reinforces the need for frequency domain analyses of radar measurements. Linear systems and frequency domain viewpoints are relied on heavily throughout the remainder of the book.

Even though RCS is a deterministic quantity, its sensitivity to radar frequency, aspect angle, and range coupled with the complexity of typical targets results in very complex behavior of observed amplitude measurements. Statistical models are used to describe this complexity. A variety of statistical models, comprising both probability density functions and correlation properties, have gained acceptance for various scenarios and form the basis for much analysis, particularly in calculations of probabilities of detection and false alarm, two of the most important radar performance measures.

References

Baddour, N., "Operational and Convolution Properties of Three-Dimensional Fourier Transforms in Spherical Polar Coordinates," J. Optical Society of America, vol. 27, no. 3, pp. 2144–2155, Oct. 2010.

Balanis, C. A., *Antenna Theory*, 3d ed. Harper & Row, New York, 2005. Baxa, E. G., Jr., "Airborne Pulsed Doppler Radar Detection of Low-Altitude Windshear—A Signal Processing Problem," *Digital Signal Processing*, vol. 1, no. 4, pp. 186–197, Oct. 1991.

- Beckmann, P., and A. Spizzichino, *The Scattering of Electromagnetic Waves* from Rough Surfaces. MacMillan, New York, 1963.
- Billingsley, J. B., Radar Clutter. Artech House, Norwood, MA, 2001.
- Birkmeier, W. P., and N. D. Wallace, *AIEE Transactions on Communication Electronics*, vol. 81, pp. 571–575, Jan. 1963.
- Carlson, A. B., Communication Systems. McGraw-Hill, New York, 1976.
- Cooper, J., "Scattering of Electromagnetic Fields by a Moving Boundary: The One-Dimensional Case," IEEE *Transactions on Antennas & Propagation*, vol. AP-28, no. 6, pp. 791–795, Nov. 1980.
- Curlander, J. C., and R. N. McDonough, *Synthetic Aperture Radar*. Wiley, New York, 1991.
- Currie, N. C., "Characteristics of Clutter," Chap. 5 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.
- DiFranco, J. V., and W. L. Rubin, *Radar Detection*, Appendix A. SciTech Publishing, Raleigh, NC, 2004.
- Doviak, D. S., and R. J. Zrnic, *Doppler Radar and Weather Observations*, 2d ed. Academic Press, San Diego, CA, 1993.
- Gleiser, R. J., "Doppler Shift for a Radar Echo," *Amer. J. Physics*, vol. 47, no. 8, pp. 735–737, Aug. 1979.
- Gray, J. E., and S. R. Addison, "Effect of Nonuniform Target Motion on Radar Backscattered Waveforms," *IEE Proceedings Radar, Sonar, Navigation*, vol. 150, no. 4, pp. 262–270, Aug. 2003.
- Haykin, S., B. W Currie, and S. B. Kesler, "Maximum Entropy Spectral Analysis of Radar Clutter," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 953–962, Sep. 1982.
- Holm, W. A., "MMW Radar Signal Processing Techniques," Chap. 6 in N. C. Currie and C. E. Brown (eds.), *Principles and Applications of Millimeter-Wave Radar*. Artech House, Boston, MA, 1987.
- Jakeman, E., and P. N. Pusey "A Model for Non-Rayleigh Sea Echo," *IEEE Transactions on Antennas and Propagation*, vol. 24, no. 6, pp. 806–814, Nov. 1976.
- Kay, S. M., *Modern Spectral Estimation*. Prentice Hall, Englewood Cliffs, NJ, 1988.
- Knott, E. F., J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section*. Artech House, Boston, MA, 1985.
- Levanon, N., Radar Principles. Wiley, New York, 1988.
- Lewinski, D. J., "Nonstationary Probabilistic Target and Clutter Scattering Models," *IEEE Transactions on Antennas and Propagation*, vol. AP-31, no. 3, pp. 490–498, May 1983.
- Long, M. W., Radar Reflectivity of Land and Sea, 3d ed. Artech House,

Boston, MA, 2001.

- Lothes, R. N., M. B. Szymanski, and R. G. Wiley, *Radar Vulnerability to Jamming*. Artech House, Boston, MA, 1990.
- Meyer, D. P., and H. A. Mayer, *Radar Target Detection*. Academic Press, New York, 1973.
- Mott, H., Polarization in Antennas and Radar. Wiley, New York, 1986.
- Mountcastle, P. D., "New Implementation of the Billingsley Clutter Model for GMTI Data Cube Generation," *Proceedings of IEEE 2004 Radar Conference*, pp. 398–401, April 2004.
- Munson, D. C., and R. L. Visentin, "A Signal Processing View of Strip-Mapping Synthetic Aperture Radar," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 12, pp. 2131–2147, 1989.
- Nathanson, F. E., *Radar Design Principles*, 2d ed. McGraw-Hill, New York, 1991.
- Papoulis, A., and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. McGraw-Hill, New York, 2001.
- Peebles, Jr., P. Z., Radar Principles, pp. 12-17. Wiley, New York, 1998.
- Probert-Jones, J. R., "The Radar Equation in Meteorology," *Quarterly Journal* of the Royal Meteorological Society, vol. 88, pp. 485–495, 1962.
- Ray, H., "Improving Radar Range and Angle Detection with Frequency Agility," *Microwave Journal*, p. 64ff, May 1966.
- Richards. M. A., J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.
- Sangston, K. J., and K. R. Gerlach, "Coherent Detection of Radar Targets in a Non-Gaussian Background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-30, no. 2, pp. 330–340, April 1994.

Sauvageot, H., Radar Meteorology, Artech House, Boston, MA, 1992.

- Skolnik, M. I., *Introduction to Radar Systems*, 3d. ed. McGraw-Hill, New York, 2001.
- Swerling, P., "Probability of Detection for Fluctuating Targets," *IRE Transactions on Information Theory*, vol. IT-6, pp. 269–308, April 1960.
- Ulaby, F. T., and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*. Artech House, Norwood, MA, 1989.
- Ward, K. D., "Compound Representation of High Resolution Sea Clutter," *Electronics Letters*, vol. 17, no. 16, pp. 561–563, Aug. 6, 1981.

Problems

1. Find the received power P_r expected from a radar and target having the following parameters: RF frequency = 95 GHz (W band), transmitted power = 100 W, antenna beamwidth = 2° in azimuth and 5° in elevation, system losses = 5 dB, target range = 3 km, target RCS = 20 m². Use Fig.

<u>1.3</u> and assume operation near sea level to estimate atmospheric losses. What is the ratio in decibels of the received power to the transmitted power?

- 2. By how many dB will the received power be reduced (not the absolute power received, but the *reduction* in power) in problem 1 if the weather changes from clear to a heavy rain of 25 mm/h? Estimate the needed parameters from Fig. 1.4.
- 3. Suppose the parameters of a radar are such that the power of the echo from a particular target is just detectable at a range of 50 miles. If the target RCS is reduced by 10 dB, what will be the new detection range? By how many dB must the RCS be reduced to reduce the detection range to 5 miles?
- 4. According to the Smithsonian Air and Space Museum, the RCS of a B-52 bomber is about 1000 m², while that of a B-2 stealth bomber is 10⁻⁶ m². If a given radar system could detect the B-52 at a range of 100 km, at what range could the same radar system detect the B-2 stealth bomber? Assume that atmospheric losses are negligible. If the B-2 flies at 550 mph, how much warning time would the radar give?
- 5. The example around Eq. (2.12) in the text calculated a received power $P_r = 3.07 \times 10^{-9}$ W at a range of 10 km. If the radar noise figure is 10 dB and the receiver noise equivalent bandwidth is 10 MHz, what is the expected signal-to-noise ratio in dB for the same target and range? Assume the receiver is at the standard temperature $T_0 = 290$ K.
- 6. Consider a "millimeter wave (MMW) seeker," which is a small radar typically used on a small missile. The RF frequency is 95 GHz and the 3 dB beamwidth is 1° in both azimuth and elevation. The range resolution ΔR is 5 m. The grazing angle of the antenna beam with respect to the ground is 20°, and the slant range to the ground, i.e., along the line of sight of the antenna, is 5 km. The terrain has a reflectivity of $\sigma^0 = -10$ dB. Is the range resolution of this system beam limited or pulse limited? What is the approximate area of the resolution cell on the ground? What is the total RCS σ of the resolution cell?
- 7. A radar is attempting to detect a point target in the presence of ground clutter. The parameters of the radar and its environment are such that the SNR at a range of R = 10 km is 30 dB, while the clutter-to-noise ratio (CNR) at the same range is 20 dB. The detection performance at this range is "clutter limited" because the clutter is the dominant interference. Assuming pulse-limited clutter interference, at what range will the SNR and CNR be equal? Equivalently, at what range will the signal-to-clutter ratio (SCR) equal 1 (0 dB)?
- 8. Consider the WSR-88D "NEXRAD" weather radar used by the U.S.

National Weather Service. It is an S-band (3 GHz) system with elevation and azimuth beamwidths of 0.88° and pulse length in its "short pulse" mode of $\tau = 1.57 \text{ ms}$. Suppose the radar measures the RCS in a region of a rain storm 50 km away as $\sigma = 20 \text{ m}^2$. Assume the simplified model of the resolution cell volume used to get Eq. (2.25), namely $\Delta V = (\Delta R)(R\theta_3)(R\phi_3)$ $= \Delta R \cdot R^2 \theta_3 \phi_3$. What are the range, azimuth, and elevation resolutions, and the resolution cell volume ΔV ? What is the volume reflectivity η ? Using Eq. (2.44), compute the value of the meteorological reflectivity factor Z in mm⁶/m³. Using Table 2.2, how hard is it raining? *Note:* The value of Z that results from Eq. (2.44) will be in m⁶/m³ = m³ if wavelength is in meters. Convert this to units of mm⁶/m³ by multiplying by 10¹⁸ before converting to a decibel scale to get dBZ.

- 9. In terms of wavelengths, by how much must the range between the radar and a scatterer change in order for the received echo phase to change by 180°? How far is this at RF frequencies of 1 GHz (L band), 10 GHz (X band), and 95 GHz (MMW band)?
- 10. The fourth degree chi-square PDF used to model the case of one dominant scatterer with many small scatterers is an approximation to the exact model for this case, which is the non-central chi-square PDF of degree 2. Both PDFs are listed in Table 2.3 along with the formulas for their variances. Show that when they both have the same mean $\bar{\sigma}$, their variances will also be the same if the non-central chi-square parameter $a^2 = 1 + \sqrt{2}$.
- 11. Suppose a target was modeled as consisting of one large scatterer and many small ones, but that the ratio a^2 of the large scatterer RCS to the sum of the small scatterer RCS values is 1 (instead of $1+\sqrt{2}$ as assumed by the fourth-degree chi-square model). Assuming the means of the two distributions are the same, what degree 2m should be chosen for the chi-square so as to match the variance of the non-central chi-square of degree 2 (See Table 2.3.) Repeat for $a^2 = 10$. Note: *m* does not have to be an integer.
- 12. Part of the significance of choosing the probability density function used to model target RCS (or clutter or other interference) is that the differences in the "tails" of the PDF can have a significant impact on the probability of observing relatively large signal values, sometimes called signal "spikes." Recall that the probability that a random variable *x* described by a PDF $p_x(x)$ exceeds some value *T* is given by

$$P\{x > T\} = \int_{T}^{+\infty} p_x(x) dx$$

Consider a set of RCS data with a mean value (linear scale) of 1.0. Compute the probability that the RCS σ is greater than 2 when an

exponential PDF is a good model for the RCS statistics, and again when a fourth-degree chi-square is a good model for the statistics. What is the ratio of the exponential PDF value to the chi-square PDF value? Repeat for $\sigma > 10$.

- 13. A ground-based airport surveillance radar has an antenna that rotates at 10 rpm (revolutions per minute). The 3-dB azimuth bandwidth of the antenna is 3°. Assume the PRF is chosen to give an unambiguous range of 150 km. How many pulses will be transmitted during the time a given target is within the antenna mainbeam during a single rotation? (This is the number of "hits" the radar gets on the target on each rotation.)
- 14. Consider a complex target with dimensions of 2 m by 4 m. What is the maximum change in aspect angle needed to decorrelate the target RCS, assuming the radar frequency remains fixed at F = 3 GHz? Repeat for F = 35 GHz.
- 15. For the same target considered in the previous problem, what is the maximum frequency step needed to decorrelate the target RCS, regardless of radar-target aspect angle?
- 16. A stationary, fixed-frequency radar collects several CPIs of data, each consisting of several pulses, from a stationary complex target. What is an appropriate correlation model for this data within a CPI? What is the appropriate model from one CPI to the next? Justify the answers.
- 17. Consider a radar at a fixed height above the ground of *h* m, with the boresight intercepting the ground at a slant range of *R* m. Show that in this scenario, the constant-gamma model for σ^0 given by Eq. (2.64) implies that the received clutter power will be proportional to R^{-2} if the clutter echo is beam-limited. What will be the proportionality if σ^0 is independent of grazing angle?
- 18. Use the GTRI clutter model of Eq. (2.65) with $\lambda = 3$ cm and the X-band data in Table 2.6 to compute and plot or sketch on a single graph the variation in σ^0 in dBsm for grass, trees, and urban clutter as the grazing angle δ varies from 10° to 70°. Assume $\sigma_h = 10\lambda$.
- 19. Show that the GTRI model predicts that σ^0 becomes independent of surface roughness when $\sigma_h \ll 10\lambda$.
- 20. Consider the four radar system parameters of antenna azimuth plane size, antenna elevation plane size, transmitter power, and pulse length. Which of these affect the signal-to-clutter ratio (SCR) in a beam-limited area clutter scenario? Repeat for a pulse-limited area clutter scenario. Justify the answers.
- 21. Use the Billingsley clutter model to estimate the time lag required for the

AC term only of the clutter to decorrelate to 10 percent of its power (maximum correlation value) for wind speeds of 5, 15, and 25 mph. Assume $F_0 = 1$ GHz. Repeat for 10 GHz.

- 22. At the standard temperature of $T_0 = 290$ K, at what frequency in hertz does $S_n(F)$ of Eq. (2.76) fall to 3 dB below its value at F = 0 Hz? It may be necessary to find the answer numerically, but a small number of terms in the series approximation to the exponential can be used to develop a good initial estimate.
- 23. What is the effective noise temperature in kelvins for a system with a noise figure of 3 dB? Repeat for 6 dB and 10 dB.
- 24. Consider an aircraft that has a radial velocity of Mach 2 (about 660 m/s) away from a stationary L-band (1 GHz) radar. What is the difference in the magnitude of the Doppler shift calculated from Eq. (2.93) and the approximation resulting from Eq. (2.97)? *Note*: On some calculators, the difference may be too small for the calculator precision. If this is the case, try using MATLAB[®] or a similar computational system to calculate the difference; use format long in MATLAB[®] to display more decimal places than the default. The difference can also be estimated by considering the largest term dropped from the series expansion [Eq. (2.95)].
- 25. Suppose two aircraft are flying straight and level at the same altitude. At a particular instant one is traveling due north at 100 m/s, while the other is flying directly at the first but in the southwesterly direction. What is the radial velocity between the two aircraft at that instant? What is the Doppler shift in hertz, including the sign, at that instant?
- 26. A stationary radar with a rotating antenna (typical of an airport approach radar, for instance) observes an aircraft moving through its airspace in a straight line at a speed of 200 mph. The aircraft approaches from the east, flies directly in front of the radar, and continues to the west. Sketch the general behavior of the radial velocity of the target relative to the radar as it flies from east to west through the airspace. Label significant values.
- 27. Derive an expression for the maximum radial velocity of a constant-radial velocity target such that the total range migration in a CPI of *M* pulses collected with a PRI of *T* seconds is less than the range resolution $c\tau/2$. The answer will be in terms of *M*, *T*, and τ . What is this velocity for a relatively long (for Doppler processing purposes) CPI of 30 pulses at T = 1 ms and a pulse length of 10 μ s?
- 28. In Eqs. (2.100) to (2.104) the time-varying range between a moving radar and a stationary target was computed. A series expansion was used to approximate the range as a quadratic function (specifically, a parabola),

and then the instantaneous Doppler shift was shown to be a linear function of time. What kind of conic section curve describes the range variation if the approximation to the square root is *not* made? Derive the formula for the instantaneous Doppler shift in this case.

- 29. Consider again the series approximation to range referred to in the previous problem. Find the maximum absolute value of *t* such that the magnitude of the dropped fourth-order term in *t* in the series approximation [see Eq. (2.101)] is less than 10 percent of the magnitude of the retained second-order term. (This condition is a limit on the amount of data that can be collected while still using the approximation to the range.) Give the numerical value of the maximum allowable *t* when v = 100 m/s and $R_0 = 10$ km.
- 30. Suppose the reflectivity distribution $\rho(\theta, \phi, R)$ consists of a single isolated point scatterer at coordinates (θ_p, ϕ_p, R_t) , i.e., $\rho(\theta, \phi, R) = \rho_t \delta_D(\theta \theta_t) \delta_D(\phi \phi_t) \delta_D(R R_t)$. Determine $y(\theta_0, \phi_0, R_0)$ of Eq. (2.114). What determines the shape of this function in the azimuth (θ) dimension for fixed ϕ and t? Repeat for the elevation and fast time dimensions with the other two variables fixed.
- 31. The first zero of the function $J_1(x)$ occurs at $x \approx 3.8317$. What is the ratio of the Rayleigh azimuth beamwidth of a circular aperture of diameter *D* with uniform illumination to the azimuth beamwidth of a rectangular antenna of the same width *D*, also with uniform illumination? Figure 2.27 can be used as an approximate check on the result.
- 32. Consider a scatterer at elevation *h* above the ground plane, and an incoming EM plane wave at a grazing angle of δ radians. What is the difference in path lengths between the "single bounce" direct reflection (path #1 in the adjoining figure) and the "double bounce" multipath reflection (path #2), as a function of δ and *h*? If *h* = 50 m and the radar slant range resolution is ΔR = 20 m, will the double-bounce echo appear in the same range bin as the single-bounce echo?



33. Continuing with the previous problem, assume the range bin spacing is greater than the path length difference so that both echoes combine in a

single range bin. Develop a formula for the phase difference between the two echoes in terms of δ , h, and the wavelength λ . If the amplitudes of the direct and double-bounce reflections are the same (very unlikely), at what values of h will there be a maximum total signal amplitude? At what values of h will there be a null in the signal amplitude?

34. δR in the figure below is the maximum path length difference between a constantrange path of length *R* over an integration angle θ and a straightline path of equal range *R* at the center of the integration angle. Find the maximum integration angle θ such that $\delta R \le \lambda/8$. The answer will depend on *R*. A path length difference of this amount would lead to an echo phase change of $-4\pi\delta R/\lambda = -\pi/2$ radians, often used as an upper bound on the tolerable phase error in various calculations.



 $\frac{5}{2}$ Because the power P of a real sinusoid is related to its amplitude A according to $P = A^2/2$ instead of just P

 $= A^2$ as with a complex sinusoid, some authors present a slightly different form for the voltage distributions.

 $\frac{6}{5}$ The autocorrelation function for any one many-scatterer target can vary significantly from the results predicted by this analysis and the average shapes seen in Fig. 2.13; the expected value of the autocorrelation must be estimated by averaging over many similar targets.

^{$\frac{1}{2}$} It is probably more common to define a constant-velocity target so that positive *v* corresponds to increasing range, but the preference here is to define *v* so that a positive *v* gives a positive Doppler shift.

 $[\]frac{1}{2}$ This ignores earth curvature effects that are significant in very long range or spaceborne radars. See Nathanson (1991) or Skolnik (2001) for additional details.

 $^{^{2}}$ A form of rain that occurs when moist air is lifted over an obstacle such as a mountain range, cooling as it rises and condensing into rainfall.

 $[\]frac{3}{2}$ RCS also varies similarly with angle over the full 3D sphere, not just in a 2D plane.

 $[\]frac{4}{2}$ An unfortunate tendency in radar terminology is to use the name of the PDF of amplitude when actually discussing the PDF of power. The most common example is to refer to a "Rayleigh RCS" target, or some similar term. A target with an exponential PDF for RCS is almost certainly meant when this occurs. Caution is advised to be certain as to whether a given PDF models amplitude or power.

 $\frac{8}{8}$ Movement of a target across multiple range bins during the series of pulses due to high rates of radartarget motion is known as *range migration*. It is much more common in imaging radar due to their much longer observation times, and so is discussed in <u>Chap. 8</u>.

 $\frac{9}{2}$ The results in this section must be modified for an electronically scanned antenna, for which the antenna pattern is a function of the scan angles.

 $\frac{10}{10}$ The analysis can be carried out equally easily for an off-boresight look direction. The only difference is to substitute an antenna gain value other than the peak gain *G*.

CHAPTER 3

Pulsed Radar Data Acquisition

As has been seen, radar measures the spatial distribution of reflectivity in the three-dimensional spherical coordinate system of range, azimuth angle, and elevation angle. Pulsed radars do this by emitting a series of individual pulses and recording the received voltage as a function of time, equivalent to range. Modern pulsed radars use coherent receivers so that the measured voltage is complex valued. They also record and process the data digitally. As with any digital data acquisition system, the selection of sampling rates and quantization strategies are crucial design decisions, affecting signal fidelity, resolution, aliasing, and noise properties, as well as processor memory and computational requirements.

3.1 Acquiring and Organizing Pulsed Radar Data

3.1.1 One Pulse: Fast Time

Suppose a radar transmits a single pulse of length τ seconds. The leading edge of the pulse is emitted at time t = 0. As discussed in <u>Chap. 2</u>, the echo power at the receiver due to clutter and targets will decay with range or time, typically at rates between R^{-1} and R^{-4} , while the noise power generated within the receiver will be constant. Figure 3.1 is a notional illustration of this behavior. Depending on the goals of the particular radar mode of operation, the radar will measure the received power over some interval in range, say from R_1 to R_2 . The interval $R_2 - R_1$ is called the *range swath* or the *range window* R_w . The range to the beginning of the range swath, R_1 , may be influenced by a number of factors. For example, for an airborne downlooking radar it might equal the altitude of the radar, since no clutter echoes could occur at a shorter range. In a ground imaging mode, it would likely be determined by the range to the nearest edge of the antenna mainbeam. Similarly, the end of the range swath, R_2 , might be set in different radars by the far edge of the mainbeam on the ground, or by the maximum expected detection range for targets of interest. Another constraint on R_2 is the unambiguous range, to be discussed shortly.


FIGURE 3.1 Signal level versus range, and range swath or window R_w .

In a monostatic radar, the radar receiver must be isolated from the antenna during pulse transmission so as to avoid damaging the sensitive receiver circuits with leakage of the high-power transmitted signal. Consequently, the receiver is off for the first τ seconds after pulse transmission begins, so the *minimum range* from which a full echo can be received is $R_{\min} = c\tau/2$ m. In practice, R_{\min} may be somewhat larger to allow for the finite switching time needed to reconnect the receiver to the antenna after pulse transmission, and in some environments to protect against particularly strong *near-in clutter*. Any clutter or target scatterer closer to the radar than R_{\min} will produce an echo that arrives in part during the initial τ seconds after transmission. That portion of the pulse echo will therefore not be seen at the receiver. A pulse that is not received in whole or part because it arrives during the time the receiver is isolated is said to be *eclipsed*.

The received signal is demodulated using a coherent receiver as described in <u>Chap. 1</u>. The resulting complex-valued baseband signal is sampled at a high rate, typically in the range of hundreds of kilohertz to a few tens of megahertz and sometimes higher. To implement the desired range swath, sampling begins at time $t_1 = 2R_1/c$ after pulse transmission and ends at time $t_2 + \tau = 2R_2/c + \tau$. The additional τ seconds at the end of the sampling period are needed to capture the end of the echo of the pulse from the far edge of the swath. The resulting samples are stored in a digital memory as shown in Fig. 3.2*a*, where each cube represents a single baseband sample. A set of *L* samples from a single transmitted pulse are referred to as *range bins*, *range gates*, *range cells*, or *fast-time samples*.¹ The phases of these complex samples are the fast-time phase history of the pulse echo data.



FIGURE 3.2 Organization of received data: (*a*) vector of fast-time samples for one pulse, (*b*) matrix of fast-time/slow-time samples for one CPI.

How rapidly should one sample the echo from a single received pulse, i.e., what should the spacing of the range bins be? The Nyquist theorem states that the sampling rate F_s should equal or exceed the bandwidth of the received signal (see App. B). In Chap. 2 it was shown that the received signal in the range dimension can be modeled as the convolution of the range reflectivity function and the modulation function x(t) of the transmitted waveform. The spectrum of the received signal is thus the product of the spectra of the range reflectivity function and the modulation function. This means that the bandwidth of the received fast-time signal will be limited by the bandwidth of the transmitted pulse.

As shown in <u>App. B</u>, the spectrum of the simple complex exponential pulse of frequency F_0 Hz is a sinc function in the frequency domain centered at F_0 Hz. This spectrum is not strictly bandlimited; however, the 3-dB bandwidth β_3 is $0.89/\tau$ Hz, both the Rayleigh bandwidth β_r and 4 dB bandwidth β_4 are $1/\tau$ Hz, and the null-to-null bandwidth β_{nn} is $2/\tau$ Hz. These approximate bandwidth measures are shown in Fig. 3.3 with $F_0 = 0$ for convenience.



FIGURE 3.3 Four definitions of bandwidth for the spectrum of a simple rectangular pulse of duration τ .

Since the sinc spectrum is decidedly not bandlimited, a Nyquist bandwidth cannot be unambiguously defined for the simple pulse. An approximate bandwidth can be defined as the two-sided frequency interval beyond which the spectrum amplitude is "insignificant," but this approach is not very useful in this case because of the slow decay of the sinc function. For example, a criterion of 40 dB reduction in the spectrum from its peak gives an approximate Nyquist bandwidth of about 66 times the 3-dB bandwidth. In radar, 3-dB bandwidths are commonly used; thus "the bandwidth" of the simple pulse is often defined to be $\beta_3 = 0.89/\tau$ Hz. More conservative definitions use the Rayleigh bandwidth of $1/\tau$ Hz and the null-to-null bandwidth $\beta_{nn} = 2/\tau$ Hz, which encompass 78 percent and 91 percent of the total spectrum energy, respectively. The Rayleigh bandwidth β_r is used here to estimate appropriate range sampling rates, i.e., the sampling rate $F_s = \beta_r$, so the sampling interval in fast time is $T_s = 1/\beta_s = 1/\beta_r$. The corresponding range bin spacing is

$$\Delta R_s = \frac{cT_s}{2} = \frac{c}{2\beta_r} \quad \text{meters}$$
(3.1)

In practice, the fast time signal is often sampled at some margin above the Nyquist rate. This compensates both for the transition band of receiver antialiasing filters and for some of the non-bandlimited nature of common pulse waveforms. Sampling rate margins of 20 to 50 percent are common.

It will be seen in <u>Chap. 4</u> that pulses are often phase modulated in order to increase their bandwidth. The pulse spectrum is then no longer a sinc function.

In fact, many phase modulated pulses are designed to have a spectrum that is approximately constant magnitude (but with complicated phase characteristics) over some desired bandwidth β , where β is much larger than the simple pulse bandwidth of approximately $1/\tau$. Thus, an idealized model of the spectrum of the ideal received phase-modulated radar pulse after translation to baseband is

$$Y(F) = \begin{cases} A \exp[j\Phi(F)] & |F| < \frac{\beta}{2} \\ 0 & |F| > \frac{\beta}{2} \end{cases}$$

$$(3.2)$$

where $\Phi(F)$ is some phase function. Figure 3.4 shows an example, the magnitude spectrum of a linear frequency-modulated or "chirp" waveform with a time-bandwidth product of 100; this waveform is studied in <u>Chap. 4</u>. On the normalized frequency scale shown, the spectrum is approximately rectangular with support $f \in (-0.5, +0.5)$, corresponding to $\pm \beta/2$ Hz. This case offers a relatively unambiguous definition of the bandwidth of the pulse (namely, β Hz), making application of the Nyquist criterion to range sampling straightforward.



FIGURE 3.4 Magnitude of the Fourier transform of a linear FM "chirp" waveform having a time-bandwidth product of 100.

3.1.2 Multiple Pulses: Slow Time and the CPI

The radar transmits not just a single pulse, but a periodic series of pulses. In some cases (e.g., a rotating weather or surveillance radar), the pulse series may be continuously ongoing, but in many cases it is organized into groups of M

pulses. The time between pulses is denoted as the *pulse repetition interval* (PRI) or *inter-pulse period* (IPP) and denoted as *T*. Its inverse is the *pulse repetition frequency* (PRF).² The PRF may range from a few hundred pulses per second (also called, casually, hertz) to tens and sometimes a few hundreds of kilohertz.

The vectors of *L* fast-time samples collected for each of the *M* pulses are typically organized into a two-dimensional matrix y[l, m] as shown in Fig. 3.2*b*. The pulse number dimension is called the *slow time* axis. The time required to collect this data is simply *MT* seconds. If a coherent series of pulses was used that time is called the *coherent processing interval* (CPI). The term CPI is used to refer both to the matrix of data and the time required to collect it. While there are exceptions, a CPI of data is usually collected using a constant PRI, constant radar frequency (RF), and the same pulse waveform for all pulses in the CPI.

Although the data for a single CPI is collected by columns (pulses), once it is stored in memory it may be accessed in any fashion. In Fig. 3.2*b* the fourth range bin for each pulse is shaded gray. This row of samples in the data matrix is the *slow-time signal* for that range bin. These samples represent the echo received after the same delay from the time of transmission for successive pulses. Assuming the antenna boresight is not moving significantly from pulse to pulse, these samples represent the reflectivity from the same range and angle, i.e., the same region in three-dimensional space, measured with a sampling interval equal to the pulse repetition interval PRI. The slow-time sampling frequency is therefore the PRF.

How should the PRF be chosen? The PRF affects, and is affected by, many aspects of the radar and environment. As was seen in the discussion of spatial Doppler in <u>Chap. 2</u>, the slow-time phase history reflects the Doppler components in the received signal. One criterion for choosing the PRF is to avoid aliasing of the spectrum replicas so as to preserve the information in the Doppler spectrum for subsequent processing such as pulse Doppler target detection or synthetic aperture imaging. Thus, the Nyquist requirement in slow time is that the PRF be at least as large as the slow-time signal bandwidth.

A nonzero Doppler bandwidth results from two sources: intrinsic motion of the scatterers in the area being measured, and motion of the radar platform. If the area being measured is a target in the conventional sense of a man-made vehicle or object, its intrinsic motion is simply the motion of the vehicle.³ If it is clutter, then intrinsic motion can be due to wind blowing the leaves of trees or blades of grass, waves on the ocean, falling and swirling rain, air-conditioning fans on tops of buildings, and so forth. For instance, the Doppler power spectrum corner frequencies in <u>Table 2.7</u> imply an intrinsic Doppler spread on the order of 0.5 to 1.0 m/s for rain at X band. The intrinsic Doppler spread of moving man-made objects can be much larger. Consider an urban clutter scene where a stationary radar observes automobile traffic with a maximum speed of 55 mph both toward and away from the radar. The radar therefore sees targets

with a velocity spread of 110 mph, or about 50 m/s. For a more extreme example, consider a moving radar installed on one of two subsonic (200 m/s) jet aircraft flying in opposite directions. As they approach, the closing rate is 400 m/s; once they pass, they separate at 400 m/s. The change in velocities observed by the radar on one of the aircraft over time is 800 m/s.

A moving radar can also induce a spread in the Doppler bandwidth of stationary objects in the beam. This is most relevant in air-to-ground radars. Figure 3.5 illustrates in two dimensions an approach to estimating the Doppler bandwidth of a patch of terrain induced by radar platform motion. The 3-dB radar beamwidth is θ_3 radians. Recall from Chap. 2 that the Doppler shift for a radar moving at velocity *v* with its boresight squinted ψ radians off the velocity vector is



FIGURE 3.5 Geometry for estimating the Doppler bandwidth component due to radar platform motion.

$$F_D = \frac{2v}{\lambda} \cos \psi \quad \text{Hz}$$
(3.3)

Now consider three point scatterers P1, P2, and P3, each at the same range from the radar. P1 and P3 are at the 3-dB edges of the antenna beam, while P2 is on boresight. Because all three are at the same range, the received echo at a delay corresponding to that range is the superposition of the echoes from all three scatterers. However, each is at a slightly different angle with respect to the aircraft velocity vector. P2 is on the boresight at the squint angle of ψ , but P1 and P3 are at $\psi \pm \theta_3/2$ radians. The *difference* in the Doppler shift of the echoes from P1 and P3 is then

$$\beta_D = \frac{2v}{\lambda} [\cos(\psi - \theta_3/2) - \cos(\psi + \theta_3/2)]$$
$$= \frac{4v}{\lambda} \sin\left(\frac{\theta_3}{2}\right) \sin\psi$$
(3.4)

Many radar antenna beamwidths are small, typically less than 5°. Applying a small angle approximation to the $\sin(\theta_3/2)$ term in Eq. (3.4) gives a simple expression for Doppler bandwidth due to platform motion

$$\beta_D \approx \frac{2v\theta_3}{\lambda}\sin\psi$$
 Hz (3.5)

As can be seen from Fig. 3.5, Eq. (3.4) or (3.5) assumes the radar is squinted sufficiently that the main beam does not include the velocity vector, that is, $|\psi| > \theta_3/2$. If the radar is forward looking or nearly so, then the $\cos(\psi - \theta_3/2)$ term in Eq. (3.4), which represents the largest Doppler shift in the mainbeam, is replaced by 1. A more complete expression for the platform motion-induced Doppler bandwidth is therefore

$$\beta_{D} \approx \begin{cases} \frac{2v\theta_{3}}{\lambda}\sin\psi & \text{Hz, } |\psi| > \frac{\theta_{3}}{2} \text{ (squinted)} \\ \frac{2v}{\lambda}[1-\cos(\psi+\theta_{3}/2)] & \text{Hz, } |\psi| < \frac{\theta_{3}}{2} \text{ (forward looking)} \end{cases}$$

$$(3.6)$$

For example, an L band (1 GHz) side-looking ($\psi = 90^{\circ}$) radar with a beamwidth of 3° traveling at 100 m/s will induce $\beta_D \approx 35$ Hz, while an X band (10 GHz) side-looking radar with a 1° beam flying at 200 m/s will induce $\beta_D \approx 233$ Hz. The same two radars in a forward-looking configuration induce negligible Doppler bandwidths of only 0.9 Hz and 0.5 Hz, respectively. Thus, while absolute Doppler shift due to platform motion is highest for a forward-looking system, the Doppler bandwidth spread is highest for a side-looking system.

In the previous example, the radar was viewing a patch of ground and the Doppler bandwidth observed by a stationary radar would be 0 Hz. The nonzero Doppler bandwidth β_D is entirely due to the motion of the observing radar, not to the characteristics of the target scene itself. The total Doppler bandwidth observed is approximately the sum of the bandwidth induced by platform motion [Eq. (3.6)] and the intrinsic bandwidth of the scene being measured. The PRF should be chosen equal to or greater than this value if possible to meet the Nyquist sampling criterion for the slow-time signal.

Although the Doppler spectrum of the illuminated terrain is both shifted

according to Eq. (3.3) and broadened according to Eq. (3.6) by relative motion between the terrain and platform, the shift in center frequency is not relevant to selection of the PRF; only the bandwidth determines the Nyquist rate. Also, note that antenna patterns are not strictly limited in angular extent, and therefore the motion-induced Doppler spectrum is not strictly bandlimited to the value based on the 3-dB bandwidth given in Eq. (3.6). Nonetheless, Eq. (3.6) provides a good basis for estimating motion-induced bandwidth.

3.1.3 Doppler and Range Ambiguities

For any sampled time signal, the sampling frequency determines the aliasing interval of the DTFT of that signal. The DTFT of the slow-time signal in a given range bin is the Doppler spectrum. The PRF is the slow-time sampling rate and therefore determines the aliasing interval of the Doppler spectrum. For reasons to be seen in <u>Chap. 5</u>, this interval is called the *blind Doppler shift*, denoted F_{Db} . The equivalent radial velocity interval is called the *blind velocity* v_b . They are given by

$$F_{Db} = PRF = \frac{1}{T} \text{ Hz}$$

$$v_b = \frac{\lambda}{2} PRF = \frac{\lambda}{2T} \text{ m/s}$$
(3.7)

Because the Doppler or radial velocity spectrum is usually plotted on the interval $[-F_{Db}/2, +F_{Db}/2)$ or $[-v_b/2, +v_b/2)$, the quantities $F_{Db}/2$ and $v_b/2$ are sometimes called the unambiguous Doppler or velocity, F_{Dua} or v_{ua} . Care must be taken to determine if a stated value refers to the full unambiguous interval F_{Db} or v_b , or to the frequency (velocity) cutoff point $\pm F_{Dua}$ or $\pm v_{ua}$.

Because the Doppler spectrum measured with a pulse burst waveform is periodic in frequency, targets having a Doppler shift F_D (radial velocity v) outside of the unambiguous Doppler or radial velocity interval will be aliased into that interval. The apparent (aliased) Doppler shift F_{Da} or radial velocity v_a will satisfy

$$F_D = F_{Da} + n \cdot F_{Db}$$

$$v = v_a + n \cdot v_b$$
(3.8)

where the integer *n* is chosen such that F_{Da} and v_a fall in the unambiguous interval.

As a numerical example, consider a 10-GHz radar viewing a target with a radial velocity v of +100 m/s. The Doppler shift will be $F_D = 6.67$ kHz. If the radar collects a CPI of data using a PRF of 3 kHz and obtains the Doppler

spectrum via a DTFT of the slow time signal in the appropriate range bin, the blind Doppler and velocity intervals are $F_{Db} = 3$ kHz and $v_b = 45$ m/s, respectively. The unambiguous Doppler shift and velocity are then 1.5 kHz and 22.5 m/s. The DTFT will be interpreted as covering [-1.5, +1.5) kHz and [-22.5, +22.5) m/s. The target's Doppler shift and velocity are outside of these intervals and will therefore be aliased into them. With n = 2 in Eq. (3.8), it is seen that the apparent Doppler and velocity are $F_{Da} = 0.67$ kHz and $v_a = +10$ m/s.

The PRF also sets a limit on the range swath. In particular, the maximum range from which the end of the echo of one pulse can be received before the next pulse is transmitted is the *unambiguous range* R_{ua} . It must satisfy $2R_{ua}/c + \tau < T$, so

$$R_{ua} = \frac{c}{2}(T - \tau) = \frac{c}{2} \left(\frac{1}{PRF} - \tau\right)$$
$$\approx \frac{cT}{2} = \frac{c}{2PRF}$$
(3.9)

The approximation in the second line applies when $T \gg \tau$ and is often cited as the definition of R_{ua} . The far edge of the sampled range swath (R_2 in Fig. 3.1) must be limited to R_{ua} m or less so the full pulse echo from a scatterer at R_2 is received before the next pulse is transmitted.

In general, it is desirable to be able to choose the PRF to provide the desired unambiguous range and Doppler bandwidth simultaneously. However, unambiguous range increases with decreasing PRF, while the blind Doppler interval increases with increasing PRF. Equations (3.7) and (3.9) can be combined to make this explicit:

$$R_{ua} F_{Db} = \frac{c}{2}, \quad R_{ua} v_b = \frac{\lambda c}{4}$$
 (3.10)

Because the right-hand side of either version of Eq. (3.10) is a constant for a given RF, increasing R_{ua} requires decreasing F_{Db} or v_b . As a result, in many situations it is not possible to simultaneously obtain the desired unambiguous velocity interval and unambiguous range with a single PRF.

To illustrate the effect of the PRF on range measurements more fully, consider the idealized signals of Fig. 3.6*a*. In the first line, a pulse is transmitted at time t = 0. It is assumed that there are two targets present at ranges R_1 and R_2 , and that the PRF is such that the unambiguous range falls between these two. The target echoes occur $2R_1/c$ and $2R_2/c$ seconds after transmission, as shown. Now suppose a second pulse is transmitted at a PRI $T = 2R_{ua}/c$. Assume $\tau \ll T$

so that the approximation $R_{ua} \approx cT/2$ from Eq. (3.9) can be used for simplicity. The same target echo profile is repeated, simply delayed by *T* seconds as shown on the second line. The next line continues this behavior for a third pulse and any subsequent pulses.



FIGURE 3.6 Illustration of range ambiguous target returns: (*a*) pattern of received data for three pulses, (*b*) total received signal.

Figure 3.6*b* shows the resulting total return observed by the radar. In the first PRI (0 < t < T), only one target echo is observed because target #2 is past the unambiguous range. In the second PRI (T < t < 2T) two target echoes are observed: target #1 from pulse #2, and target #2 from pulse #1. This pattern repeats in the third and subsequent PRIs until the pulse train ends. If the radar receives detectable echoes from ranges up to *N* times the unambiguous range, the pattern of returns observed after each pulse will reach steady state in the *N*th PRI. In Fig. 3.6, it reaches steady state in the second PRI.

Once steady state is achieved, each pulse appears to result in two detections at the apparent ranges R_{1a} and R_{2a} as shown in Fig. 3.6b. For target #1, the apparent range is the actual range. However, target #2 was beyond R_{ua} and so aliases to the apparent range $R_{2a} = R_2 - R_{ua}$. In general, if the radar is sensitive enough to detect targets beyond the unambiguous range for a given PRF, the apparent ranges will be ambiguous. In particular, given a detection at an apparent range R_a , the target's actual range could be any value R_0 that satisfies

$$R_0 = R_a + n \cdot R_{ua} \tag{3.11}$$

and is within the plausible maximum detection range of the radar. Note that n = 0 for target #1 and n = 1 for target #2 in the example of Fig. 3.6. Techniques to deal with range and Doppler ambiguities are discussed in <u>Chap. 5</u>.

Figure 3.7 illustrates the structure of the CPI of data that would result from this example. Suppose the unambiguous range corresponds to seven range bins.⁴ Assume the ranges to targets #1 and #2 correspond to the sixth and eleventh range bins. During PRI #1, only the first target is detected, so the first fast-time column of data has only one detection in range bin #6. Target #2 will alias to range bin 11-7 = 4, so the second and subsequent pulses will show detections in range bins #4 and #6.



FIGURE 3.7 Steady-state range ambiguous return for the scenario of Fig. 3.6.

This example also illustrates the existence of a start-up transient in processing when the returns are range ambiguous. If the scenario and PRF are such that targets can be detected at ranges of at least $(N-1)R_{ua}$ but no further than NR_{ua} (N = 2 in the example), the received signal will not achieve steady state until the *N*th PRI (second PRI in this example).

Range aliasing also affects clutter returns. If the clutter echo power is above the receiver noise levels at ranges exceeding R_{ua} , the clutter echoes will also fold over, so that the clutter competing with targets in the steady state may actually be the combined clutter of several range ambiguity intervals. Also like targets, the clutter level versus range in a given PRI will not reach steady state until multiple pulses have been transmitted if the clutter is range ambiguous. If the clutter reaches steady state in the *N*th PRI, the first N - 1 pulses are often called *clutter fill* pulses. Processing of this nonstationary data generally gives degraded results; it is often better to discard the data from the clutter fill pulses and use only the data received after steady state is achieved. Conversely, if it is desired to have a certain number of steady state PRIs for subsequent processing, the number of pulses is often augmented by the necessary number of clutter fill pulses.

It is tempting to conclude that range ambiguities could be resolved by observing whether or not a target detection appears in all of the pulses. A target detection missing from the first n pulses in a CPI suggests that the actual range is the apparent range plus n times the unambiguous range. This idea will work if detection algorithms are applied to the fast time data for each pulse separately and the SNR is high enough that the probability of missed detections is small. However, it is rare to use a CPI of data in this manner. More commonly, the SNR of the single-pulse data is not adequate for reliable detection so that it is not known whether the target is absent in the first n pulses. Instead, the slow time data will be coherently or noncoherently integrated in order to obtain an adequate SNR.

In addition to creating the possibility of range ambiguities, the use of multiple pulses also aggravates the eclipsing phenomenon. A target at any integer multiple of the range $R_{ua} \approx cT/2$, corresponding to time delays that are integer multiples of the PRI *T*, will produce an echo that arrives as the next pulse is being transmitted. During this interval, the receiver will again be isolated, so the target echo will be eclipsed. Targets at other time delays within the interval $(nT - \tau, nT + \tau)$ for any integer *n* will be partially eclipsed. Thus, the pulse burst creates a series of *blind zones* in range or time delay. Targets in these blind zones will be difficult or impossible to detect, even when they have adequate SNR. Techniques to overcome this limitation are discussed in <u>Chap. 5</u>.

3.1.4 Multiple Channels: The Datacube

Some radars, but by no means all, have antennas that provide multiple simultaneous outputs. The most obvious example is a system using a phased array antenna with multiple subarrays, each having its own receiver, or even with one receiver per array element in some cases. Each receiver will generate a matrix of data like that of Fig. 3.2b for every pulse burst. The complete set of data y[l, m, n] from all N channels is called a *datacube* and is illustrated in Fig. 3.8. The third dimension is often referred to as the receiver channel or phase center dimension. Another type of system that generates a datacube uses a monopulse antenna, common in some types of tracking radars. A monopulse antenna has three output channels and so generates a datacube having N = 3 layers. Radar data is often explicitly organized in the processor memory in a datacube format, i.e., as a three-dimensional structure of complex-valued data.



FIGURE 3.8 Datacube illustrating one CPI of data from a multichannel pulsed radar.

How large is a datacube? The number of samples in each dimension are determined by the characteristics of the desired radar measurements. The number *L* of range samples is simply the length of the range swath divided by the range bin spacing, $L = R_w/\Delta R_s$. The swath length is determined by mission requirements, while the range bin spacing is determined primarily by the waveform bandwidth as seen in Eq. (3.1). Both may vary significantly for the same radar as it switches between various operating modes with different search ranges and range resolutions.

One important determinant of the number of pulses M in a CPI is the desired Doppler resolution (equivalently, velocity resolution). The Doppler spectrum is the DTFT of the slow-time data. The duration of the slow-time signal is the CPI length of MT seconds. The Doppler resolution will therefore be on the order of $\Delta F_D = 1/MT$,⁵ giving the required number of pulses as $M = 1/\Delta F_D T = PRF/\Delta F_D$. Thus, M depends on the *PRF* as well as the Doppler resolution, and can vary widely. In pulse-Doppler processing for basic target detection and tracking, M is frequently a small number of tens of pulses. However, in fine-resolution imaging it can be hundreds or even thousands of pulses.

For a multichannel receiver, the number N of channels is more difficult to characterize. A phased array antenna with a receiver per element may have hundreds or thousands of phase centers, each constituting a receiver channel. A subarrayed architecture may have many fewer, perhaps ranging from as little as three or four to a few tens. A monopulse antenna has three phase centers. The antenna type, size, and architecture all significantly influence N.

The datacube view of a CPI of data from a multichannel pulsed radar provides a good conceptual model for understanding most digital radar signal processing operations. Many of the basic radar signal processing operations considered in the remainder of this text correspond to processing onedimensional subvectors or two-dimensional submatrices of the datacube in various dimensions. Figure 3.9 illustrates these relationships. The particular operations depicted are discussed in upcoming chapters. For example, pulse compression (see Chap. 4) is implemented as a one-dimensional convolution on a single vector in the fast-time (range) dimension. Pulse compression can be performed independently on each such range vector for every pulse and receiver channel.



FIGURE 3.9 Correspondence between key radar signal processing functions and operations on the radar datacube.

3.1.5 Dwells

A *dwell*, sometimes also called the *time on target*, is another term for a radar data collection interval. Like the term CPI, a dwell can refer both to a time interval and to the data collected within that time. It is defined in the IEEE standard for radar definitions as "a data acquisition interval during which the data is usually processed together for detection or measurement" (IEEE, 2008). Consider a rotating radar like that shown in Fig. 2.15*a* and imagine that the beamwidth is 3° and the antenna scans at a rate of 60° per second. A point target would be in the beam for 50 ms during each scan. If the radar has a PRF of 2000 pulses per second, a target would be illuminated with 100 pulses during the traversal of the mainbeam. Since it is known that the echo of a target, if present, would be present in 100 successive pulses, it might be sensible to integrate 100 pulses at a time for SNR improvement before performing a detection test. In this scenario, the dwell time would be the full 50 ms. If only 50 pulses at a time were integrated, the dwell time would be considered to be 25 ms.

The idea of a dwell is not limited to mechanically scanning radars. Consider a pulsed Doppler radar with an electronically scanned antenna pointed in some particular direction. The radar emits a CPI of data consisting of 20 pulses at a 2-kHz PRF. The CPI is then 10 ms. Suppose the radar collects three such CPIs while illuminating the same region, with 50 ms from the start of one CPI to the start of the next. The total data collection time of 110 ms from the beginning of the first CPI to the end of the third would be the dwell time for the radar in that look direction.

The terms "dwell" and "CPI" are sometimes used synonymously, but the preceding example illustrates that they are not the same. For a coherent radar that organizes its data into CPIs, a dwell can correspond to one or more CPIs. For rotating and similar mechanically scanned radars, a dwell is usually the time it takes to scan across a point target.

3.2 Sampling the Doppler Spectrum

Selecting a value for the pulse repetition frequency determines the sampling rate for the slow-time signal. The frequency spectrum of the slow-time signal is traditionally called the Doppler spectrum, because the nonzero frequency components are due to the spatial Doppler effect arising from the relative motion between the radar and target scene. Doppler processing, which is the analysis or modification of the information about the target scene contained in the Doppler spectrum, will be the subject of <u>Chap. 5</u>. Doppler processing will sometimes be performed directly in the slow-time domain, that is, directly on the time signal represented by a row of v[l, m]; but frequently the spectrum of each row will be explicitly calculated. In a digital processor, this must be done with a discrete Fourier transform (DFT) or other discrete spectral analysis technique. In this section, it is assumed that the spectrum is computed using conventional DFT techniques; no nonlinear spectral estimation methods or other alternatives are considered. The question then arises as to how closely successive samples of the computed Doppler spectrum should be spaced, i.e., what should be the Doppler sampling interval?

3.2.1 The Nyquist Rate in Doppler

The Nyquist criterion concept can be applied to sampling in frequency as well as the more usual application to sampling in time. The result will be a frequency sampling rate that is dependent on a "bandwidth" in the time domain.

The Nyquist sampling rate in the frequency domain can be determined by reviewing the relation between the sampled Doppler spectrum and the slow-time signal. Let a single finite duration slow-time signal (one row of y[l, m]) be denoted as $y_s[m]$, $0 \le m \le M - 1$. The DTFT of $y_s[m]$ is (Oppenheim and Schafer, 2010)

$$Y_{s}(\omega) \equiv \sum_{m=-\infty}^{\infty} y_{s}[m]e^{-j\omega m} \qquad \omega \in (-\pi, \pi)$$

 $Y_s(\omega)$ is a function of a continuous frequency variable, despite the fact that the signal $y_s[m]$ is discrete. Furthermore, it is periodic in ω with period 2π radians per sample.

Consider the *K*-point discrete spectrum $Y_s[k]$ formed by sampling $Y_s(\omega)$ at *K* evenly spaced points along the interval $[0, 2\pi)$,

$$Y_s[k] = Y_s\left(\frac{2\pi k}{K}\right) \qquad k \in [0, K-1]$$
(3.13)

Interpret $Y_s[k]$ as a *K*-point DFT. To find the relation between it and the original signal $y_s[m]$, compute its inverse DFT:

$$\hat{y}_{s}[m] = \frac{1}{K} \sum_{k=0}^{K-1} Y_{s} \left(\frac{2\pi k}{K} \right) e^{j2\pi mk/K} \qquad m \in [0, K-1]$$

$$= \frac{1}{K} \sum_{k=0}^{K-1} \sum_{p=-\infty}^{\infty} y_{s}[p] e^{-j2\pi pk/K} e^{j2\pi mk/K}$$

$$= \sum_{p=-\infty}^{\infty} y_{s}[p] \left[\frac{1}{K} \sum_{k=0}^{K-1} e^{j2\pi (m-p)k/K} \right]$$
(3.14)

The inner sum can be evaluated as

$$\frac{1}{K} \sum_{k=0}^{K-1} e^{j2\pi(m-p)k/K} = \sum_{q=-\infty}^{\infty} \delta\left[m - p - qK\right]$$
(3.15)

where δ [·] is the discrete-time unit impulse function.⁶ Substituting Eq. (3.15) back in Eq. (3.14) gives

$$\hat{y}_{s}[m] = \sum_{q=-\infty}^{\infty} y_{s}[m-qK] \qquad m \in [0, K-1]$$
(3.16)

Although a finite length signal $y_s[m]$ was assumed in the previous analysis, the result also holds for infinite-length signals.

Equation (3.16) shows that, in the dual of time domain sampling behavior, sampling the frequency spectrum replicates the signal in the time domain with a period proportional to the frequency sampling rate. Specifically, if the slow-time signal spectrum is computed at *K* frequency points, the time domain signal

obtained by an inverse DFT of those frequency samples is the original slowtime signal replicated at intervals of *K* samples. Recall that $y_s[m]$ is confined to the interval [0, M-1]. If $K \ge M$, $y_s[m-qK] = 0$ in the interval of interest [0, K-1] for all $q \ne 0$, so that $\hat{y}_s[m] = y_s[m]$. This is the usual case where the DFT size is at least as long as the data sequence size.

The Nyquist rate for sampling in the frequency dimension is now apparent. If $K \ge M$ the original slow-time signal $y_s[m]$ is not aliased by the frequency domain sampling operation. Consequently, it can be recovered from the replicated signal $\hat{y}_s[m]$ implied by the sampled spectrum by simply excising the principal period $m \in [0, K - 1]$. This is the equivalent of the lowpass filter required to reconstruct a sampled time-domain signal; the time and frequency domains have simply been reversed in this discussion of sampling in the frequency domain.

Since $K \ge M$, the frequency domain sampling interval ω_s must satisfy

$$\omega_{\rm s} \le \frac{2\pi}{M}$$
 rad (3.17)

The corresponding Nyquist sampling rate in the frequency domain is $\frac{1}{2}$

$$K \ge M$$
 samples per Doppler spectrum period (3.18)

Thus, the width of the signal's region of support (i.e., its length or "bandwidth") in the time domain of M samples plays the same role for sampling in frequency as does the width of a signal's region of support in frequency (its actual bandwidth) for sampling in time.

In some systems the number of Doppler samples computed is less than the number of data samples available, i.e., K < M. This can occur if only a limited number of spectrum samples are required by the system design. In early digital radar processors, it was more likely motivated by the difficulty of implementing a larger DFT at radar data rates, a problem mostly obviated by computing advances enabled by Moore's law. One way to compute a *K*-point DFT from an *M*-point sequence when K < M is to simply retain only *K* data samples and compute their *K*-point DFT. This is not desirable when there are M > K samples available for two reasons. First, the DTFT of the *K* samples used is not the same as that of the full *M*-point sequence, so the DFT will give us samples of a reduced-resolution DTFT. Second, by not using all *M* available samples, the signal-to-noise ratio (SNR) of the calculated spectrum is reduced because only *K* samples instead of all *M* available samples are coherently integrated by the DFT. It is rarely a good idea to discard measured data if the highest possible measurement quality is desired.

If the Doppler spectrum samples are still to be equal to samples of the

DTFT of $y_s[m]$ in this case, Eqs. (3.13) to (3.16) imply that it is necessary to form a new, reduced-length *K*-point sequence $\hat{y}_s[m]$ from the slow-time data sequence $y_s[m]$ by aliasing it according to Eq. (3.16). This operation, depicted pictorially in Fig. 3.10, is sometimes called *data turning*. It maximizes the SNR of the Doppler spectrum samples by using all of the available samples, and is in fact used in some older operational radars.



FIGURE 3.10 Illustration of the "zero padding" and "data turning" operations: (*a*) original 12-point data sequence, (*b*) zero-padded to 16 points for use in a 16-point DFT, (*c*) data turning to create an aliased 8-point sequence shown in (*d*) for use in an 8-point DFT.

3.2.2 Straddle Loss

The previous section established the Nyquist sampling rate in Doppler frequency. When actually computing the sampled spectrum, whether by the DFT or other means, one would like to be confident that the sampled spectrum captures all of the important features of the underlying DTFT. For example, if the DTFT exhibits significant peaks, it is hoped that one of the spectral samples will fall on or very near that peak so that the sampled spectrum captures this feature.

An appropriate signal model to consider this issue is a pure complex sinusoid, corresponding for example to a target moving at constant velocity relative to the radar over the observation interval and therefore exhibiting a constant Doppler shift. Thus, the slow-time signal $y_s[m]$ is modeled as

$$y_{s}[m] = Ae^{j\omega_{D}m}$$
 $m \in [0, M-1]$ (3.19)

where ω_D is the Doppler frequency shift in normalized radian frequency units. The DTFT of $y_s[m]$ is

$$Y_{s}(\omega) = A \frac{\sin[(\omega - \omega_{D})M/2]}{\sin[(\omega - \omega_{D})/2]} \exp[-j(\frac{M-1}{2})(\omega - \omega_{D})] \qquad \omega \in (-\pi, \pi]$$
(3.20)

That is, $Y_s(\omega)$ is a so-called *digital sinc*, *aliased sinc* (*asinc*), or *Dirichlet* function, circularly shifted in the frequency domain so that its peak occurs at $\omega = \omega_D$. An example is shown in Fig. 3.11 for the case $\omega_D = \pi/2$ (corresponding to $f_D = \omega_D / 2\pi = 0.25$) and M = 20. Significant features of this DTFT include the peak amplitude and frequency, the mainlobe bandwidth, and the sidelobe structure. In particular, the *M*-point DTFT of a pure complex sinusoid of amplitude *A* has a peak value of *MA*, with the peak sidelobe about 13.2 dB below the peak. The 3-dB width of the mainlobe in normalized frequency units is $\beta_3 = 0.89/M$ cycles per sample, the Rayleigh width is $\beta_r = 1/M$ cycles per sample. These metrics are illustrated in Fig. 3.11.



FIGURE 3.11 The magnitude of the DTFT of a sampled pure complex sinusoid of 20 samples length, normalized frequency 0.25 cycles per sample, and amplitude 1.

The DFT computes per samples of this spectrum at normalized frequencies $2\pi k/K$ rads/sample. Figure 3.12 shows the result when K = M and the sinusoid

frequency exactly equals one of the DFT frequencies, that is, $\omega_D = 2\pi k_0/K$ for some k_0 ($k_0 = 5$ and K = 20 in this example, corresponding to $\omega_D = \pi/2$ rads per sample). One DFT sample falls on the peak of the asinc function, while all of the others fall on its zeroes, so that the DFT becomes an impulse function. This could be viewed as an ideal measurement, since the discrete spectrum indicates a single sinusoid at the correct frequency and nothing else; but it does not reveal the mainlobe width or sidelobe structure of the underlying DTFT.



FIGURE 3.12 The 20-point DFT of a sampled pure complex sinusoid of 20 samples length, normalized frequency 0.25 cycles per sample, and amplitude 1. The dotted line shows the underlying DTFT of the same data from Fig. 3.11.

More importantly, the good result of Fig. 3.12 depends critically on the actual sinusoid frequency exactly matching one of the DFT sample frequencies. If this is not the case, the DFT samples will fall somewhere on the asinc function other than the peak and zeros. Figure 3.13 shows the result when the example of Figs. 3.11 and 3.12 is modified by changing the normalized frequency from 0.25 to 0.275 (equivalently, changing ω_D to 0.55π), exactly halfway between two DFT sample frequencies. Now a pair of DFT samples straddle the actual underlying peak of the asinc function, while the other samples fall near the sidelobe peaks. Even though the underlying asinc function is identical in shape in both cases, differing only by a half-bin shift on the frequency axis, the effect on the apparent spectrum measured by the DFT is dramatic: a broadened and attenuated mainlobe, and the appearance of significant sidelobes where before there apparently were none.



FIGURE 3.13 Same as Fig. 3.12 except for a frequency shift of the sinusoid by one-half DFT bin to a normalized frequency of 0.275 cycles per sample.

Because the DFT sample frequencies straddle the true peak of the underlying DTFT, the apparent peak amplitude of the spectrum in Fig. 3.13 is about 13, whereas the peak amplitude of the underlying DTFT (and thus of the DFT in Fig. 3.12) is 20. This reduction in measured peak signal amplitude is called a *straddle loss* (because the samples straddle the true peak location).⁸

One obvious way to reduce straddle loss is to sample the Doppler frequency axis more densely, i.e., to choose the number of spectrum samples K > M. The resulting samples are more closely spaced so that the maximum amount by which a sample frequency can miss the peak frequency of the DTFT is reduced, thus reducing the straddle loss. Figure 3.14 continues the example of Fig. 3.13, but with the sampling density doubled to 2*M* samples per Doppler spectrum period (40 samples in this case), and then to 12.8*M* samples per spectrum measured by the DFT to begin to resemble the underlying asinc of the DTFT even at as little as 2*M* samples per period. At 12.8*M* samples per spectrum period, the DFT gives an excellent representation of the details of the underlying DTFT.



FIGURE 3.14 Continuation of the example of Fig. 3.13: (*a*) 40-point DFT of the 20-point sinusoid of normalized frequency 0.275, (*b*) 256-point DFT of the same sequence.

The off-peak sampling loss (straddle loss) for a sinusoidal signal can be limited to a specified value, at least for this idealized signal, by appropriate choice of the spectrum sampling rate *K*. For example, the loss can be limited to 3 dB or less by choosing *K* such that the interval $2\pi/K$ between samples does not exceed the 3 dB width of the asinc function. The 3-dB width can be found by considering just the magnitude of Eq. (3.20) with $\omega = 0$ for convenience. The peak value of the asinc function is *MA*, thus it is necessary to find the value ω_3 o f ω such that the asinc function has the value $MA/\sqrt{2}$. This is best done numerically. The answer is a strong function of *M* for small *M* but rapidly approaches an asymptotic value of $\omega_3 = 2.79/M$ for $M \ge 10$. It follows that the 3 dB width of the asinc function is $\Delta \omega = 5.58/M$ radians.

The sampling interval for a rate of *K* samples per period is $2\pi/K$ radians. Equating this to the 3-dB width and solving gives the sampling rate required to limit off-peak sampling attenuation to 3 dB in the Doppler spectrum in terms of the Nyquist rate of *M* samples per Doppler spectrum period,

$$K \ge \frac{2\pi}{5.58}M = 1.13M$$
 samples per Doppler spectrum period (3.21)

which is 13 percent higher than the Nyquist sampling rate in Doppler. If the offpeak sampling loss is to be kept significantly less than 3 dB, the Doppler spectrum must be oversampled still more.

The analysis leading to Eq. (3.21) can be repeated for any specified level of tolerable straddle loss. Figure 3.15 shows the worst-case straddle loss as a function of the oversampling factor κ (i.e., $K = \kappa M$) for the case M = 100. Both undersampled ($\kappa < 1$) and oversampled ($\kappa > 1$) cases are shown. The loss is

somewhat less for very short duration sequences (M < 10) but varies little for larger M.



FIGURE 3.15 Maximum off-peak Doppler spectrum sampling loss for a sinusoidal slow-time signal sampled at κM samples per Doppler spectrum period.

3.3 Sampling in the Spatial and Angle Dimensions

As discussed earlier, two distinct types of spatial sampling are of concern in a radar system. One type concerns the design of phased array antennas. A phased array samples the incoming wavefront at the individual array element locations. Thus, the spacing of these elements must be chosen to adequately sample the wavefront for any incidence angle. The second concerns beam steering. Mechanically or electronically steered antennas can change the pointing direction of their antenna beam. As the beam is scanned to search or map a region in space, a decision must be made as to how far it is permissible to scan before another pulse (or burst of pulses) must be emitted by the radar so that the external environment is adequately sampled. The next two subsections address these questions.

3.3.1 Spatial Array Sampling

<u>Chapter 1</u> introduced the concept of spatial frequency and wavenumber. Consider a uniform linear array with element spacing *d*, as shown in Fig. 3.16. The wavenumber (spatial radian frequency) of an RF signal with wavelength λ impinging on the array antenna from a direction of arrival θ radians off the normal to the array, as shown in the figure, is



FIGURE 3.16 Geometry of a uniform linear array antenna.

$$K_x = \frac{2\pi}{\lambda} \sin\theta$$
 rads/m (3.22)

The equivalent spatial frequency in cyclical units is just

$$F_x = \frac{1}{\lambda} \sin\theta$$
 cycles/m (3.23)

The angle of arrival θ can vary between -90° and $+90^{\circ}$, so the spatial frequency bandwidth becomes

$$\beta_x = \frac{1}{\lambda} \sin\left(\frac{\pi}{2}\right) - \frac{1}{\lambda} \sin\left(-\frac{\pi}{2}\right) = \frac{2}{\lambda} \qquad \text{cycles/m}$$
(3.24)

It follows immediately by the Nyquist criterion that the required spatial sampling interval is

$$d \le \frac{1}{\beta_x} = \frac{\lambda}{2}$$
 meters (3.25)

Thus, the elements of the array should be spaced no more than $\lambda/2$ meters apart to avoid aliasing of spatial frequencies.⁹

Many practical arrays, particularly large wideband systems, employ more complicated architectures effectively having two spatial sampling intervals. The complete antenna array is broken into a relatively small number of subarrays, each of which is populated with elements obeying the Nyquist spacing of Eq. (3.25). Since the multielement subarrays are necessarily separated by multiples

of d, they contribute a term to the complete antenna pattern that does exhibit spatial aliasing. As a result, the overall antenna pattern can exhibit aliasing in some circumstances, depending on whether phase or time delay steering is used for individual elements and across subarrays, the steering direction of the array, and the bandwidth of the radar waveform. An introduction to these issues is given in Bailey (2010).

3.3.2 Sampling in Angle

Consider a steerable or scanning antenna, whether mechanically steered (typically a parabolic dish or slotted flat-plate array, and others) or electronically steered (phased array), with a 3-dB beamwidth θ_3 radians. Each pulse transmitted samples the reflectivity of the environment in the direction in which the antenna is pointed. If a region in angular (elevation and azimuth) space is to be searched, the question arises: how densely in angle must the space be sampled? That is, how much can the antenna be steered before another pulse should be transmitted? Smaller angular sampling intervals provide a better representation of the search volume, but also require more pulses and therefore more time to search a given volume. Since the antenna boresight, one intuitively expects that to adequately sample the reflectivity of the scene scanned by the antenna, it will be necessary to make a new measurement every time it scans by some angle on the order of θ_3 . The Nyquist criterion can be applied to this spatial sampling problem to quantify this expectation.

It was seen in <u>Chap. 2</u> [Eq. (2.119)] that the observed reflectivity in angle for a constant range is the convolution of the range-averaged reflectivity with the two-way antenna voltage pattern. An equivalent expression in just one angle dimension for simplicity, say azimuth, is

$$y(\theta; R_0) = \hat{\rho}(\theta; R_0)_{*\theta} E^2(\theta)$$
$$= A_r \int_{-\pi}^{\pi} E^2(\zeta - \theta) \hat{\rho}(\zeta; R_0) d\zeta$$
(3.26)

where $y(\theta; R_0)$ is the complex coherent receiver output as a function of azimuth angle θ at range R_0 , $\hat{\rho}(\theta; R_0)$ is the range-averaged reflectivity evaluated at range R_0 and $E^2(\theta)$ is the two-way voltage pattern in the angular dimension θ . It follows that the Fourier transform in the angle dimension of y is the product of the Fourier transforms of the antenna pattern and the range-averaged reflectivity.

Taking the pattern of the ideal rectangular aperture as representative, it was seen in <u>Chap. 1</u> that the two-way antenna voltage pattern is

$$E^{2}(\theta) = \left[\frac{\sin(\pi D \sin\theta/\lambda)}{\pi D \sin\theta/\lambda}\right]^{2}$$

Defining $s = \sin\theta$ and $\alpha = D/\lambda$, Eq. (3.27) can be rewritten as

$$E^{2}(s) = \left[\frac{\sin(\pi\alpha s)}{\pi\alpha s}\right]^{2}$$
(3.28)

which is a sinc-squared function. It follows immediately that its Fourier transform is a triangle function in the normalized variable (x/λ) , where x is the spatial dimension of the antenna aperture (Bracewell, 1999). This function is illustrated in Fig. 3.17.



FIGURE 3.17 Fourier transform of the two-way antenna voltage pattern for an ideal rectangular antenna aperture with uniform illumination.

Because the Fourier transform of the antenna pattern has a width of 2α , the Nyquist sampling interval in *s* must be

$$T_{\rm s} \le \frac{1}{2\alpha} \tag{3.29}$$

Recall that $s = \sin\theta$. To convert T_s into a sampling interval in θ , consider the differential $ds = \cos\theta \ d\theta$, so that $d\theta = ds/\cos\theta$. Thus, a small interval T_s in s corresponds approximately to an interval $T_{\theta} = T_s/\cos\theta$ in θ . The minimum value for T_{θ} occurs when $\theta = 0$ so that $T_{\theta} = T_s$. Thus, the sampling interval in angle becomes (using $\alpha = D/\lambda$ for the second step)

$$T_{\theta} \le \frac{1}{2\alpha} = \frac{\lambda}{2D} \tag{3.30}$$

This is the Nyquist sampling interval in angle for a rectangular aperture of size D with uniform illumination.

As a final step, this result can be expressed in terms of 3-dB beamwidths.

The 3-dB beam-width of an aperture antenna is of the form (Balanis, 2005)

$$\theta_3 = k \frac{\lambda}{D} = \frac{k}{\alpha}$$
 rad (3.31)

For the uniformly illuminated case, k = 0.89. Combining Eqs. (3.30) and (3.31) gives

$$T_{\theta} \le \frac{\theta_3}{2k} \tag{3.32}$$

For k = 0.89, this gives a Nyquist sampling rate of 0.56 times the 3-dB beamwidth, or 1.8 samples per 3-dB beamwidth. In practice, many systems sample in angle at approximately one sample per 3-dB beamwidth. The search space is then undersampled in angle, at least according to the Nyquist criterion.

While derived for the uniformly illuminated aperture, these results apply to all aperture antennas. For a finite aperture of size D, different antenna patterns (for instance, with lower sidelobes at the expense of a wider mainlobe) are obtained by changing the aperture illumination function, typically by tapering it in a manner similar to windowing operations in signal processing. The Fourier transform of these antenna power patterns will still be the autocorrelation of the corresponding illumination function. Since the illumination function still has finite support, its autocorrelation will still be limited to a width of 2α in s, as shown in Fig. 3.17; only the detailed shape of the function will change. Thus, Eq. (3.32) applies for any finite aperture antenna. The difference is that the factor k will be different for different illumination functions. Lower sidelobe antennas will have values of k in the range of approximately 1.4 to 2.0, giving corresponding Nyquist sampling rates on the order of 2.8 to four samples per 3-dB beamwidth for low sidelobe antennas.

For a rotating radar, the angular sampling rate of Eq. (3.31) implies a lower bound on PRF. Suppose the rate of rotation is Ω_0 radians per second. In order that successive pulses be transmitted in directions differing by no more than the T_{θ} of (3.31), the PRI and PRF must satisfy

$$PRI \le \frac{(\theta_3/2k)}{\Omega_0} = \frac{\theta_3}{2k\Omega_0} \implies PRF \ge \frac{2k\Omega_0}{\theta_3}$$
(3.33)

Equations (3.33) and (3.9) illustrate a conflict between volume coverage and search rate in a rotating search radar. For a given antenna design, θ_3 and k are fixed. Then, increasing the sweep rate Ω_0 will increase the volume search rate, but will also require an increased PRF; but a higher PRF reduces the unambiguous range, reducing the volume that can be searched without ambiguities.

3.4 I/Q Imbalance and Digital I/Q

In <u>Chap. 1</u>, it was shown that the output of a quadrature receiver given a realvalued bandpass signal as input is the same as would be obtained by using the equivalent analytic (one-sided spectrum) complex signal with complex demodulation by the signal $\exp(-j\Omega_0 t)$. In other words, the quadrature receiver acts to select the upper band of the bandpass signal and shift it to baseband. Any system that accomplishes this same result can be used to derive the in-phase and quadrature signals needed for further signal processing.

The quadrature receiver could, in principle, be implemented entirely digitally. The input signal would be converted to a digital signal after the low-noise amplifier. The mixing operations would be replaced by multiplications, and the analog lowpass filters by digital filters. This is not done in practice because a straightforward implementation would require the A/D converter to operate at about twice the carrier frequency rather than twice the information bandwidth of the signal (specifically, $2F_0 + \beta$ rather than just β samples per second), a technologically unreasonable requirement. On the other hand, the conventional analog quadrature receiver also has technological limitations, as mentioned briefly in <u>Chap. 1</u>. Correct operation assumes that the two channels are perfectly matched in delay and gain across the frequency band of interest, there are no DC biases in either channel, and the two reference oscillators are exactly 90° out of phase. In this section, the effect of I/Q imbalances is investigated, and then two digital I/Q receiver structures that combat imbalance errors are described.

3.4.1 I/Q Imbalance and Offset

Figure 1.9 describing the conventional quadrature receiver is repeated below as Fig. 3.18, but with the addition of an amplitude mismatch factor $(1 + \varepsilon)$, a phase mismatch ϕ , and DC offsets γ and κ in the in-phase (I) and quadrature (Q) channels, respectively. Take the I channel as the gain and phase reference without loss of generality, so the gain and phase errors are placed entirely in the Q (upper) channel. As shown in the figure, the introduction of these errors is reflected as an undesired gain and phase shift in the Q channel output, along with the DC offset in each channel. For processing, the I and Q channel outputs are combined as usual into a single complex signal, x(t) = I(t) + jQ(t). In the absence of mismatch errors, $x(t) = A \exp[j\theta(t)]$. How are the mismatch errors manifested in x(t)?



FIGURE 3.18 Conventional coherent receiver of <u>Fig. 1.9</u> with amplitude and phase mismatch errors and DC offsets.

Inspection of Fig. 3.18 gives

$$\begin{aligned} x(t) &= A\cos\theta + \gamma + j[A(1+\varepsilon)\sin(\theta-\phi) + \kappa] \\ &= A\{[1-j(1+\varepsilon)\sin\phi]\cos\theta + j(1+\varepsilon)\cos\phi\sin\theta\} + (\gamma+j\kappa) \\ &\equiv A(\alpha\cos\theta + j\beta\sin\theta) + (\gamma+j\kappa) \end{aligned}$$
(3.34)

where the time dependence of $\theta(t)$ has been dropped to simplify the notation slightly. Note that the constant α is complex but β is not, and that $\alpha = \beta = 1$ in the absence of gain and phase errors, i.e., if $\varepsilon = \phi = 0$. Using the identities

$$\alpha = \frac{\alpha + \beta}{2} + \frac{\alpha - \beta}{2}$$
$$\beta = \frac{\alpha + \beta}{2} - \frac{\alpha - \beta}{2}$$
(3.35)

in Eq. (3.34) and collecting terms of equal amplitude gives

$$x(t) = A \left[\frac{\alpha + \beta}{2} (\cos \theta + j \sin \theta) + \frac{\alpha - \beta}{2} (\cos \theta - j \sin \theta) \right] + (\gamma + j\kappa)$$
$$= A \left[\frac{\alpha + \beta}{2} \exp(+j\theta) + \frac{\alpha - \beta}{2} \exp(-j\theta) \right] + (\gamma + j\kappa)$$
(3.36)

Equation (3.36) shows that in the presence of amplitude or phase errors, the complex signal x(t) will not only contain the desired signal component (with a slightly modified amplitude) $A\exp[j\theta(t)]$, but also an *image* component with a different amplitude and a conjugated phase function, as well as a complex DC term. The image component is an error resulting from the amplitude and phase

mismatches; the DC component is the direct result of the individual channel DC offsets.

Recall that the phase function $\exp[j\theta(t)]$ can represent phase modulation of the radar waveform, the effect of the environment on the waveform (such as a phase shift due to spatial Doppler), or both. In the case of a spatial Doppler phase shift, $\theta(t)$ on the *m*th pulse will be of the form $\omega_D m$ for some normalized Doppler radian frequency ω_D . The image component will then have a phase shift of the form $-\omega_D m$. Thus, over a series of *M* pulses, the mismatches will give rise to a false signal at the negative of each actual Doppler frequency component in addition to the desired signal. Furthermore, the DC component is equivalent to a false signal at a Doppler shift of zero, i.e., clutter or a stationary target.

As another example, suppose θ (*t*) represents the intentional quadratic phase modulation used to construct a linear FM chirp signal, θ (*t*) = αt^2 (see Chap. 4 for details). Then the image component will have a phase modulation of $-\alpha t^2$, which represents a linear frequency modulation (FM) signal with a slope opposite to the transmitted pulse. This signal will not be properly compressed by the matched filter, instead causing an apparent increase in the noise floor (Sinsky and Wang, 1974).

To judge the significance of the gain and phase mismatch errors, consider the ratio P_r of the power in the image component relative to that in the desired component. From Eq. (3.36), this is

$$P_r = \frac{\left|(\alpha - \beta)/2\right|^2}{\left|(\alpha + \beta)/2\right|^2}$$
$$= \frac{\left[1 - (1 + \varepsilon)\cos\phi\right]^2 + \left[(1 + \varepsilon)\sin\phi\right]^2}{\left[1 + (1 + \varepsilon)\cos\phi\right]^2 + \left[(1 + \varepsilon)\sin\phi\right]^2}$$
(3.37)

Figure 3.19 illustrates the value of P_r as a function of the phase and amplitude imbalance.



FIGURE 3.19 Relative power of I/Q mismatch-induced signal image as a function of amount of phase and amplitude mismatch.

It is also useful to consider simplifications of Eq. (3.37) for the cases of small amplitude mismatch only and small phase mismatch only. First consider the case of small amplitude mismatch only ($\varepsilon \ll 1$), so that $\phi = 0$. Then

$$P_r = \frac{\left[1 - (1 + \varepsilon)\right]^2}{\left[1 + (1 + \varepsilon)\right]^2} = \frac{\varepsilon^2}{(2 + \varepsilon)^2}$$
$$\approx \frac{\varepsilon^2}{4}$$
(3.38)

Amplitude mismatch is often specified in decibels, as is the relative power of the image signal component. A mismatch of *k* dB implies that $20 \log_{10}(1 + \varepsilon) = k$. Substituting this relation in Eq. (3.38) and expressing the result in decibels gives

$$P_r (dB) = 20 \log_{10} (10^{k/20} - 1) - 6.02 \quad (k \text{ in } dB)$$

(3.39)

For example, an amplitude mismatch of 0.1 dB gives rise to an image component 44.7 dB below the desired component of x(t).

A result similar to Eq. (3.38) holds for the case of small phase mismatch

only, that is, $\varepsilon = 0$ and $\phi \ll 1$. In this case Eq. (3.37) reduces to

$$P_{r} = \frac{[1 - \cos\phi]^{2} + \sin^{2}\phi}{[1 + \cos\phi]^{2} + \sin^{2}\phi} = \frac{1 - \cos\phi}{1 + \cos\phi}$$
$$\approx \frac{\frac{1}{2}\phi^{2}}{2 + \frac{1}{2}\phi^{2}} \approx \frac{\phi^{2}}{4}$$
(3.40)

where the second line is obtained using the small angle approximation $\cos \phi \approx 1 - \frac{\phi^2}{2}$. Note that ϕ is in radians. The relative power of the image component in decibels is then

$$P_r(dB) = 20\log_{10}(\phi) - 6.02 \tag{3.41}$$

As an example, a phase mismatch of 1° gives an image component approximately 41.2 dB below the desired response.

3.4.2 Correcting I/Q Errors

As shown in Fig. 3.18, the I and Q signals in the presence of mismatch can be modeled as

$$I = A\cos\theta + \gamma \qquad Q = A(1+\varepsilon)\sin(\theta-\phi) + \kappa \tag{3.42}$$

where the dependence on time *t* continues to be suppressed to simplify the notation. The desired in-phase signal *I* is $A\cos\theta$, and in the quadrature channel is $A\sin\theta$. Is it possible to recover the desired outputs from the available measurements of Eq. (3.42)?

Consider forming a new I' and Q' as a linear combination of the measured I and Q. Specifically, require that $I' = A\cos\theta$ and $Q' = A\sin\theta$. Although it is straightforward to solve the general problem, it is obvious that the DC offsets should simply be subtracted, and then a linear combination of the zero-offset data formed

$$\begin{bmatrix} I'\\Q' \end{bmatrix} = \begin{bmatrix} A\cos\theta\\A\sin\theta \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12}\\a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} I\\Q \end{bmatrix} - \begin{pmatrix} \gamma\\\kappa \end{bmatrix}$$
$$= \begin{bmatrix} a_{11} & a_{12}\\a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} A\cos\theta\\A(1+\varepsilon)\sin(\theta-\phi) \end{bmatrix}$$
(3.43)

By inspection, $a_{11} = 1$ and $a_{12} = 0$. The remaining equation is

$$Q' = A\sin\theta = a_{21}I + a_{22}Q$$

= $a_{21}A\cos\theta + a_{22}A(1+\varepsilon)\sin(\theta - \phi)$
(3.44)

Applying a trigonometric identity for $\sin(\theta - \phi)$ and equating terms in $\sin\theta$ and $\cos\theta$ on both sides of Eq. (3.44) leads to the following solution for a_{21} and a_{22} :

$$a_{21} = \tan\phi \qquad a_{22} = \frac{1}{(1+\varepsilon)\cos\phi}$$
(3.45)

Using Eq. (3.45) in Eq. (3.43) gives the final transformation required

$$\begin{bmatrix} I'\\Q' \end{bmatrix} = \begin{bmatrix} 1 & 0\\\tan\phi & 1/(1+\varepsilon)\cos\phi \end{bmatrix} \begin{bmatrix} I\\Q \end{bmatrix} - \begin{pmatrix} \gamma\\\kappa \end{bmatrix}$$
(3.46)

Once the I/Q errors ε , ϕ , γ , and κ are determined, Eq. (3.46) can be used to compute a new value Q' for the quadrature channel sample for each measured I-Q sample pair. The difficulty, of course, is in actually determining the errors; the correction is then easy. The errors are generally estimated by injecting a known *pilot signal*, usually a pure sinusoid, into the receiver and observing the outputs. Details for one specific technique to estimate gain and phase errors are given in Churchill et al. (1981); that paper also derives limits to mismatch correction (and thus to image suppression) caused by noise, which introduces errors into the estimates of ε and ϕ .

A second method for eliminating I/Q error is based on the idea of transmitting multiple pulses, stepping the starting phase of each pulse through a series of evenly spaced values, and then integrating the measured returns. To see how this technique works, suppose the input signal in Fig. 3.18 is changed to $A \sin \left[\Omega t + \theta \left(t\right) + k(2\pi/N)\right]$ for some fixed integer N and variable integer k; i.e., the pulse is one of a series of N pulses, where the initial phase is increased by $2\pi/N$ radians on each successive pulse. The extra phase shift propagates to the output signals

$$I_{k} = A \cos \left[\theta(t) + k \frac{2\pi}{N} \right] + \gamma$$

$$Q_{k} = A(1+\varepsilon) \sin \left[\theta(t) + k \frac{2\pi}{N} - \phi \right] + \kappa$$
(3.47)

for k = 0, 1, ..., N - 1. The development leading to Eq. (3.36) can be repeated to obtain the complex signal for this case, which is (still suppressing the *t*

dependence of θ)

$$x_{k}(t) = A \left[\frac{\alpha + \beta}{2} e^{j(\theta + 2\pi k/N)} + \frac{\alpha - \beta}{2} e^{-j(\theta + 2\pi k/N)} \right] + (\gamma + j\kappa)$$
(3.48)

Now coherently integrate the N pulses x_k to form a single composite measurement, applying a counter phase rotation to each to realign their phases:

$$\begin{aligned} x(t) &= \frac{1}{N} \sum_{k=0}^{N-1} x_k(t) e^{-jk2\pi/N} \\ &= \frac{A}{N} \frac{\alpha + \beta}{2} \sum_{k=0}^{N-1} e^{j\theta} + \frac{A}{N} \frac{\alpha - \beta}{2} \sum_{k=0}^{N-1} e^{-j\theta} e^{-j4\pi k/N} + \frac{(\gamma + j\kappa)}{N} \sum_{k=0}^{N-1} e^{-j2\pi k/N} \\ &= A \frac{\alpha + \beta}{2} e^{j\theta} + A \frac{\alpha - \beta}{2} e^{-j\theta} \sum_{k=0}^{N-1} e^{-j4\pi k/N} + \frac{(\gamma + j\kappa)}{N} \sum_{k=0}^{N-1} e^{-j2\pi k/N} \end{aligned}$$
(3.49)

The summations in the middle and last terms of Eq. (3.49) can be evaluated in closed form to give

$$\sum_{k=0}^{N-1} e^{-j4\pi k/N} = \begin{cases} N, & N=1,2\\ 0, & N \ge 3 \end{cases}$$
(3.50)

and

$$\sum_{k=0}^{N-1} e^{-j2\pi k/N} = \begin{cases} 1, & N=1\\ 0, & N \ge 2 \end{cases}$$
(3.51)

so that

$$x(t) = A \frac{\alpha + \beta}{2} e^{j\theta} \qquad (N \ge 3)$$
(3.52)

Thus, as long as at least three pulses are used, the process of rotating the transmitted phase, compensating the received measurements, and integrating will suppress both the undesired image component and the DC component!

The algebraic correction technique of Eq. (3.46) is applied to individual I/Q sample pairs, requiring two real multiplies and three real additions per time sample (assuming the correction coefficients have been precomputed). The

major advantage of this technique is that it can be applied individually to each pulse of data. Its major disadvantage is that it requires the transmitter/receiver control and analog hardware be augmented to allow pilot signal insertion for determining the correction coefficients. The pilot signal operation is performed relatively infrequently on the assumption that ε and ϕ vary only slowly.

The phase rotation and integration technique of Eqs. (3.47) to (3.49), in contrast, requires integration of at least three pulses with the transmitted phase adjusted for each pulse. Thus, the technique requires both high-speed transmitter phase control and more time to complete a measurement since multiple pulses must be collected. The increase in required time implies also an assumption that the scene being measured does not vary during the time required for the multiple pulses; decorrelation of the scene degrades the effectiveness of the technique. This method also places a heavier load on the signal processor, since the integration requires N complex multiplies and N-1 complex additions per time sample, or a total of 4N real multiplies and 4N - 2 real additions, with $N \ge 3$. However, the integration method has one very important advantage: it does not require knowledge of any of the errors ε , ϕ , γ , and κ . It also has the side benefit that the integration of multiple pulses increases the signal-to-noise ratio of the final result x(t). Given these considerations, it is often used in instrumentation systems at fixed site installations, such as turntable RCS measurement facilities. In these systems, N is often on the order of 16 to 64, and may even be as high as 65,536 (64K) in some cases.

Note also that Eqs. (3.47) to (3.49) implicitly assume that the phase modulation $\theta(t)$ is the same for each pulse $x_k(t)$. If $\theta(t)$ represents waveform modulation (e.g., a linear FM chirp), this will be true; but if $\theta(t)$ contains a term representing environmental phase modulation, for example due to Doppler shift, then the technique assumes that the appropriate component of $\theta(t)$ is the same on each of the pulses integrated. This is the case for stationary targets (assuming the radar is also stationary). For constant Doppler targets, the frequency implied by $\theta(t)$ will be the same from pulse to pulse, but the absolute phase will change in general, so that the target response does not integrate properly. For accelerating targets, the assumption will fail entirely. The phase rotation and integration technique is therefore most appropriate for stationary or nearly stationary (over *N* PRIs) targets. The algebraic technique does not have this limitation, since it operates on individual pulses only.

3.4.3 Digital I/Q

Digital I/Q or *digital IF* is the name given to a collection of techniques that form the I and Q signals digitally in order to overcome the channel matching limitations of analog receivers. A number of variations have been described in the literature. In general, they all share two characteristics. First, they use analog mixing and filtering to shift the single real-valued input signal to a low intermediate frequency prior to A/D conversion, greatly relaxing the A/D speed

requirements compared to RF sampling. Furthermore, the intermediate frequency (IF) is chosen so that required complex multiplications by functions of the form $\exp(j\omega_0 n)$ reduce to particularly simple forms. Second, they use a combination of digital filtering and down sampling to obtain a final output consisting only of the desired sideband of the original spectrum, sampled at or near the appropriate Nyquist rate of β complex samples per second. Two approaches are briefly described here.

The first method, which is particularly elegant, is described in (Rader, 1984.) The RF signal is assumed to have a bandpass spectrum with an information bandwidth of β Hz. Figure 3.20 is a block diagram of the system, and Fig. 3.21 sketches the signal spectrum at various points in the system. The first step is an analog frequency shifting operation that translates this spectrum to a low IF of β Hz. The bandpass filter rejects the double frequency terms created by the mixer. The spectrum is therefore bandlimited to $\pm \beta/2$ Hz, so the Nyquist rate is 3β samples per second. However, for reasons that will become clear shortly a higher sampling rate of 4β samples per second is used, giving a discrete-time signal with the spectrum shown in Fig. 3.21*c*.



FIGURE 3.20 Architecture of Rader's system for digital generation of in-phase and quadrature signals. (After Rader, 1984.)


FIGURE 3.21 Spectra corresponding to successive signals in the digital I/Q system of Fig. 3.20. (*a*) Spectrum of bandpass input signal with information bandwidth β Hz, (*b*) result of translation to an IF frequency also equal to β , (*c*) one period of spectrum on normalized frequency scale after A/D conversion, (*d*) only the upper sideband remains after filtering, (*e*) a replica of the upper sideband is centered at DC after decimation. (After Rader, 1984.)

Recall that the goal of quadrature demodulation is to select one sideband of the bandpass signal and translate it to baseband. Assume that the upper sideband is to be retained. The next step is therefore to filter the real signal $\tilde{x}[n]$ to eliminate the lower sideband. Since the resulting spectrum will not be Hermitian, the output signal must be complex; this is shown in Fig. 3.20 as a one-input, two-output filter. The required frequency response is clear from the spectrum diagrams in Fig. 3.21; it is

$$H(\omega) = \begin{cases} 1, & \frac{\pi}{4} < \omega < \frac{3\pi}{4} \\ 0, & -\frac{3\pi}{4} < \omega < -\frac{\pi}{4} \\ \text{don't care, otherwise} \end{cases}$$
(3.53)

This asymmetric filter frequency response corresponds to a complex-valued impulse response, giving rise to the complex output from the single real input.

While Eq. (3.53) states that the value of $H(\omega)$ around DC is unconstrained, in fact it should be close to zero. The filter will then also suppress any DC component in the signal (not sketched in Fig. 3.21) that may have been introduced by nonideal mixing in the first analog frequency translation. Thus, this digital I/Q architecture also makes it easier to suppress mixer bias terms. This would not be possible if the spectrum had been translated to the lowest possible IF frequency, namely β Hz, since there would then be no region of the spectrum around DC that did not contain signal components of interest.

A particularly efficient design for realizing the filter H(z) as a pair of loworder recursive filters is based on the mathematics of phase-splitting networks; details are given in Rader (1984). However, the particular design of the filters is not central to the architecture of the approach.

The final step is to translate the remaining spectral sideband, centered at $\omega_0 = \pi/2$, to baseband and to reduce the sampling rate from 4β to the final Nyquist rate β . This can be accomplished by multiplying the complex filter output \hat{x} by the sequence $exp(-j\pi n/2) = (-j)n$ and then simply discarding three of every four samples. Because of the special form of the multipliers, the complex multiplications could be implemented simply with sign changes and interchanges of real and imaginary parts, rather than with actual complex multiplications. This is a consequence of having selected the original sampling rate to be 4β instead of 3β .

However, this multiplication is not shown in Fig. 3.20 because, in fact, it is not necessary at all. The spectrum of the decimator output y[n] is related to the spectrum of \hat{x} according to (Oppenheim and Schafer, 2010)

$$Y(\omega) = \frac{1}{4} \sum_{k=0}^{3} \hat{X}\left(\frac{\omega - 2\pi k}{4}\right) = \frac{1}{4} \sum_{k=0}^{3} \hat{X}\left(\frac{\omega}{4} - k\frac{\pi}{2}\right)$$
(3.54)

Equation (3.54) states that the decimation process causes the spectrum to replicate at intervals of $\pi/2$ radians. Since the nonzero portion of the spectrum is bandlimited to $\pi/2$ radians, these replications abut but do not alias; furthermore, since the spectrum prior to decimation is centered at $\omega = \pi/2$, one of the

replications (k = 3, specifically) is centered at $\omega = 2\pi$ radians. The periodicity of the spectrum of a discrete-time signal therefore guarantees that there is a replica centered at $\omega = 0$ as well; this replica is the final desired spectrum. Thus, the real and imaginary outputs of the decimator are the desired I and Q signals. Another digital I/Q system that uses the spectrum replicating properties of decimation to advantage is described in Rice and Wu (1982).

The success of the decimation operation in eliminating the need for a final complex frequency translation depended on the proper relationship between the bandwidth and center frequency of the signal, and the decimation factor. This is the major reason for choosing the IF to be β instead of $\beta/2$ (or some other permissible value), and the sampling frequency as 4β instead of 3β (or some other value).

Rader's digital I/Q architecture has reduced the number of analog signal channels from two to one, making the issues of oscillator quadrature and gain and phase matching completely moot, while also providing a natural opportunity to filter out DC biases introduced by the remaining analog mixer. Furthermore, the two A/D converters required at the output of the conventional quadrature receiver to enable subsequent digital processing have been reduced to one. There are two major costs to these improvements. The first is an increase by a factor of four in the A/D converter speed requirement, from β samples per second for conventional baseband sampling to 4β samples per second for Rader's system; this may be difficult at radar signal bandwidths. The second is the introduction of the need for high-rate digital filtering, which is computationally expensive (although Rader's efficient filter design lessens this cost).

Figures 3.22 and 3.23 sketch the processor conceptual architecture and the relevant signals of the second digital I/Q architecture (Shaw and Pohlig, 1995). In this case, analog frequency translation is used to shift the signal spectrum to a lower IF than used by Rader, namely 0.625β . The signal is then A/D converted at a rate of 2.5β samples per second, resulting in the signal $\frac{\pi}{n}$ having a spectrum centered at $\omega = \pi/2$ as shown in Fig. 3.23. An explicit complex modulation by $exp(+j\pi n/2) = j^n$ then shifts one of the sidebands, in this case the lower one, to baseband, resulting in the spectrum shown in Fig. 3.23*d*. Clearly $\overline{x}[n]$ is complex as a result of this complex modulation.



FIGURE 3.22 Conceptual architecture of Lincoln Laboratory system for digital

generation of in-phase and quadrature signals. (After Shaw and Pohlig, 1995.)



FIGURE 3.23 Spectra corresponding to successive signals in the digital I/Q system of Fig. 3.22. (*a*) Spectrum of bandpass input signal with information bandwidth of β Hz, (*b*) result of translation to an IF frequency of 0.625β , (*c*) one period of spectrum on normalized frequency scale after A/D conversion, (*d*) digital complex modulation centers the lower sideband at DC, (*e*) only the lower sideband remains after lowpass filtering, (*f*) decimation by two reduces the sampling rate to 1.25β . (After Shaw and Pohlig, 1995.)

The next step is to lowpass filter $\overline{x[n]}$ to remove the upper sideband, leaving only the baseband portion of the spectrum. A 16-point finite impulse

response (FIR) digital filter is used for this task in Shaw and Pohligh (1995). Once the lowpass filtering is completed, the spectrum is nonzero only for $\omega \in (-0.4 \pi, +0.4\pi)$. The sampling rate is then reduced by a factor of two by discarding every other output sample. The final result is the desired digital I and Q signals, sampled at a rate of 1.25β samples per second.

As with Rader's system, the computational complexity is actually reduced by taking advantage of the properties of decimation and FIR filters. The decimation is performed immediately after the A/D conversion by splitting the data into even- and odd-numbered sample streams. The complex modulation by j^n , which implies both sign changes and real/imaginary interchanges, then reduces only to sign changes on every other sample in each channel, and the 16point FIR filters are replaced with 8-point FIR filters in each channel without any reduction in filtering quality.

A significant advantage of this system over Rader's is that the A/D converter must operate at only 2.5 times the signal information bandwidth, rather than four times the bandwidth. This is an important savings at high radar bandwidths. There are three disadvantages. The first is that the lower IF and sampling rate require sharper transitions in the digital filter, therefore increasing the filter order necessary to achieve a given stopband suppression and thus the computational complexity of the filter. The second is the requirement for an explicit multiplication by *jn*. Although this reduces to switching and sign changes, it nonetheless represents extra processing. Third, the final sampling rate exceeds the signal Nyquist rate by 25 percent, whereas in Rader's system it equaled the Nyquist rate. This increases the computational load by 25 percent over the minimum necessary throughout the remainder of the digital processing. This may not be a problem in practice. Sampling rates are usually set somewhat above Nyquist rates anyway to provide a margin of safety, since real signals are never perfectly bandlimited.

Two other details merit mention. It may appear that modulating the sideband to baseband before filtering eliminates the possibility of using the digital filter to suppress DC bias errors from the analog mixer. However, that same modulation will move any DC term contributed by the mixer to $\omega = \pi/2$, where it can still be removed by the lowpass filter. Finally, in the Rader system the I and Q signals were derived from the upper sideband of the original bandpass signal, while in the Shaw and Pohlig system the lower sideband was used. Because the original signal was real valued, its spectrum was Hermitian, and consequently the spectra of the complex outputs of the two systems, say $Y_1(\omega)$ and $Y_2(\omega)$, are related according to $Y_2(\omega) = Y_1^* (-\omega)$ so that $y_2[n] = y_1^*[n]$. Thus the I outputs of the two systems are (ideally) identical, while the Q outputs differ in sign. Clearly, either system could be modified to use the opposite sideband.

References

Bailey, C. D., "Radar Antennas," Chap. 9 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.

Balanis, C. A., Antenna Theory, 3d ed. Harper & Row, New York, 2005.

Bracewell, R. N., *The Fourier Transform and Its Applications*, 3d ed. McGraw-Hill, New York, 1999.

Churchill, F. E., G. W. Ogar, and B. J. Thompson, "The Correction of *I* and *Q* Errors in a Coherent Processor," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, no. 1, pp. 131–137, Jan. 1981. See also "Corrections to 'The Correction of *I* and *Q* Errors in a Coherent Processor," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 17, no. 2, p. 312, Mar. 1981.

IEEE Standard Radar Definitions. IEEE Standard 686-2008, May 21, 2008, New York.

Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing*, 3d ed. Prentice Hall, Englewood Cliffs, NJ, 2010.

Rader, C. M., "A Simple Method for Sampling In-Phase and Quadrature Components," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 20, no. 6, pp. 821–824, Nov. 1984.

- Rice, D. W., and K. H. Wu, "Quadrature Sampling with High Dynamic Range," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 18, no. 4, pp. 736–739, Nov. 1982.
- Shaw, G. A., and S. C. Pohlig, "I/Q Baseband Demodulation in the RASSP SAR Benchmark," Project Report RASSP-4, Massachusetts Institute of Technology Lincoln Laboratory, Aug. 24, 1995.
- Sinsky, A. I., and P. C. P. Wang, "Error Analysis of a Quadrature Coherent Detector Processor," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 10, no. 6, pp. 880–883, Nov. 1974.

Problems

- 1. Compute the minimum range for a radar using pulse lengths of 1 ns, 1 μ s, and 1 ms.
- 2. A radar transmits a series of 10 μ s long pulses at a PRI of 100 μ s. Determine the maximum and minimum target range such that at least a portion of the echo from one pulse will arrive back at the receiver during the transmission of the next pulse. Targets in this range interval will be completely or partially eclipsed. What target range produces a completely eclipsed echo?
- 3. Suppose a radar has a pulse length of 100 ns. What is the Rayleigh

bandwidth of the pulse spectrum, in Hz? What is the 3 dB bandwidth in Hz?

- 4. Consider two signals: a square pulse $x_1(t)$ of length τ seconds, and a triangle $x_2(t)$ of length 2τ seconds obtained by convolving the square pulse with itself, $x_2(t) = x_1(t) * x_1(t)$. What is the relationship between the two spectra $X_1(F)$ and $X_2(F)$? Determine the relationship between the 3-dB bandwidths of $X_1(F)$ and $X_2(F)$. How does this compare to the relationship between the Rayleigh bandwidths of $X_1(F)$ and $X_2(F)$?
- 5. Consider two RF pulses at frequencies of 5.0 GHz and 5.01 GHz. Assume two pulses are resolvable in frequency if their center frequencies are separated by at least the Rayleigh resolution of the individual pulses. What is the minimum pulse length required so that the two pulses could be resolved in frequency?
- 6. A finite pulse train waveform is composed of 20 pulses, each of 10 μ s length and separated by a PRI of 1 ms. What is the coherent processing interval for this waveform?
- 7. Consider an X-band (10 GHz) radar on a aircraft traveling at 100 m/s. Assume the 3-dB azimuth beamwidth of the antenna is 3°. Compute the Doppler shift F_D of a scatterer on the antenna boresight and the Doppler bandwidth β_D across the beam for squint angles of $\psi = 0^\circ$, 30°, 60°, and 90°.
- 8. Consider a radar with a PRF of 5 kHz. What is the maximum unambiguous range, R_{ua} , of this radar in km? If a target is located at a range of 50 miles, how many pulses will the radar have transmitted before the first echo from the target arrives? What will be the apparent range of the target in kilometers?
- 9. Assuming a sidelooking radar ($\psi = 90^{\circ}$) and Nyquist sampling in slow time, determine the relationship between the maximum unambiguous range R_{ua} and the antenna beamwidth θ_{ax} . Ignore eclipsing. Using $\theta_{ax} \approx \lambda/D_{az}$, determine the relationship between R_{ua} and the antenna azimuth dimension D_{az} .
- 10. Suppose a target is at a range of 10 km from a C-band (5 GHz) radar and has a radial velocity of +50 m/s with respect to the radar. Determine whether the target is ambiguous in range and velocity, and the apparent range and velocity R_a and v_a for PRFs of 1, 10, and 100 kHz. The unambiguous velocity interval is considered to be $[-v_{ua}/2, +v_{ua}/2]$.
- 11. Suppose a C-band (5 GHz) radar has a fast-time sampling rate of 2M samples per second and a PRF of 5000 pulses per second. A CPI of data is to be collected that will cover a range interval of 30 km and support velocity resolution of 10 m/s. What are the dimensions of the data matrix

for one CPI of data?

- 12. There is sometimes a concern as to whether a target will stay in the same range bin while during the time it takes to collect a CPI of data. Sometimes this is an issue, but often it is not. Using the PRF and number of pulses in the CPI from the previous problem, what is the total duration of the CPI in seconds? Consider a target moving at 100 m/s (about 224 mph). How far does the target move during one CPI, and how does that compare to the range bin size? What is the minimum target velocity in m/s such that the target would move more than one range bin in range during the CPI?
- 13. Suppose N samples of time-domain data are collected at a sampling rate of F_s samples per second. The K-point DFT of the data is computed. Depending on the relative values of K and N, zero padding or data turning is used as required. Develop a formula for the spacing of the DFT bins in hertz.
- 14. Consider a sequence of 20 slow-time data samples collected at a PRF of 2 kHz. If a 1000-point DFT of this sequence is computed, what is the spacing between DFT frequency samples in hertz?
- 15. Derive a condition on the DFT size K similar to that of Eq. (3.21) for a maximum straddle loss of 1 dB. The result will depend on the value of M. Instead of solving the appropriate equations numerically, use the first two terms of the Taylor series for sin(x) to get a closed-form result. Figure 3.15 can be used as an approximate check on the result for the case M = 100.
- 16. Consider a search radar at 1 GHz (L band) with a rotating D = 10 m dish antenna. Suppose the beamwidth is $\theta_3 = 2^\circ$. What is the antenna parameter k in Eq. (3.31)? What is the Nyquist sampling rate in degrees for this antenna? If the antenna rotates at a rate of one revolution every 6 seconds, what is the PRF required to achieve this angular sampling rate?
- 17. Compute the relative power ratio P_r in the image component of the output of an I/Q receiver when there is a simultaneous mismatch of 0.1 dB in gain and 1° in phase. Express the answer in dB. Use Fig. 3.19 to check the answer.
- 18. Consider a digital I/Q architecture similar to Rader's. Starting with the original signal spectrum of Fig. 3.21*a*, assume that the signal is demodulated from the original center frequency F_0 to an IF of β Hz. What will be the minimum required sampling rate F_s of the real-valued data? Assuming this value for F_s and also that the sampling rate is reduced to the minimum possible without aliasing in the last step, sketch the complete set of spectra from the original analog spectrum X(F) to the final discrete time spectrum $Y(\omega)$, similar to Fig. 3.21. Also show the required frequency response $H(\omega)$ of the digital filter, assuming the upper sideband is the one

that is retained. Discuss the spectrum recentering step in going from the equivalent of Fig. 3.21*d* to the equivalent of Fig. 3.21*e*: can demodulation by decimation be used and, if not, why not? If multiplication by complex exponentials is used, do the multipliers assume a simplified form? Summarize the pros and cons of this system versus Rader's system.

 $\frac{4}{2}$ Seven range bins is unrealistically few in most situations, but is used here for ease of illustration.

 $\frac{5}{2}$ The precise value depends on the definition of mainlobe width that is used, and whether or not weighting for sidelobe control is employed.

 $\frac{6}{2}$ Not to be confused with the "Dirac delta" impulse function $\delta_D(\cdot)$ used in continuous-time analysis.

 $\frac{8}{5}$ Straddle loss is also called *scallop loss* by some authors, e.g., Harris (1978).

 $\frac{9}{2}$ This same result is often derived in antenna literature by requiring that the antenna pattern not contain *grating lobes*, which are replicas of the antenna pattern caused by sampling of the aperture of a phased array antenna by the elements.

¹ The terms *resolution bins* and *resolution cells* are also sometimes used synonymously with range bins, and in many cases are synonymous; but the sampling interval in range does not always equal the range resolution, so caution should be used in interpreting the latter two terms.

 $[\]frac{2}{2}$ In this text, the abbreviations "PRF" and "PRI" are used both as acronyms and as mathematical variables. When used as an acronym, they are not italicized (PRF, PRI); when used as mathematical variables, they are italicized (*PRF*, *PRI*).

 $[\]frac{3}{2}$ Note that the velocity of portions of a vehicle may differ from the nominal velocity of the vehicle as a whole. The Doppler frequency corresponding to the nominal velocity is frequently termed the *skin return*. However, the lug nuts on the wheels on a wheeled vehicle have velocities ranging from zero to twice the nominal velocity of the vehicle. Some laser radars even attempt to measure the Doppler shifts due to vehicle vibration. The spectrum of Doppler shifts from different portions of a vehicle is sometimes called a *microDoppler spectrum*.

² Using a DFT size *K* that is strictly greater than the number of time samples *M* is often referred to as *zero padding*. This term is a relic of traditional software for computing the fast Fourier transform, which often returned the vector of frequency samples in the same variable used to input the time samples. When K > M, the input vector was "padded" to the required length of *K* by adding *K*–*M* zeros at the end. The term has nothing to do with the mathematics of the DFT.

CHAPTER 4 Radar Waveforms

4.1 Introduction

A radar transmits a waveform typically modeled as

$$\overline{x}(t) = a(t)\sin[\Omega t + \theta(t)]$$

(4.1)

The term Ω in the argument of the sine function is the carrier radar frequency (RF) in radians per second. The term a(t) represents amplitude modulation of the RF carrier; in a pulsed radar, this is typically just a rectangular function that pulses the waveform on and off. The term $\theta(t)$ models any phase or frequency modulation of the carrier. It can be zero, a nonzero constant, or a nontrivial function. The overbar on $\overline{x}(t)$ denotes that the signal is on a carrier, i.e., it has not yet been demodulated. Figure 4.1 illustrates three example waveform types common in pulsed radar. The simple pulse is simply a constant-amplitude burst at the RF frequency. The frequency of the *linear frequency modulated* (LFM) pulse increases at a constant rate during the time the pulse is on. LFM pulses can also have decreasing frequency during the pulse. The third example is a binary phase-coded pulse. In this waveform, the frequency is constant but the absolute phase of the waveform changes from zero to π radians several times within the pulse. That is, the value of $\theta(t)$ changes between the constants zero and π at specific times within the pulse.



FIGURE 4.1 Examples of common pulsed radar waveforms: (*a*) simple pulse, (*b*) linear frequency modulated (LFM) pulse, (*c*) binary phase-coded pulse.

As discussed in <u>Chap. 1</u>, the real-valued waveform of <u>Eq. (4.1)</u> is more conveniently modeled by its complex equivalent

$$\overline{x}(t) = a(t)e^{j[\Omega t + \theta(t)]}$$
(4.2)

The portion of $\overline{x}(t)$ other than the carrier term, or equivalently the complex baseband signal after demodulation, is called the *complex envelope* of the waveform

$$x(t) = a(t)e^{j\theta(t)}$$

$$(4.3)$$

It is this function that describes the amplitude and phase or frequency modulation applied to the RF carrier and is considered to be "the waveform" in this chapter.

Radar waveforms can be characterized in several ways. Perhaps first is whether the waveform is *continuous wave* (CW) or pulsed; sometimes variations such as "interrupted CW" are defined as well. Pulsed waveforms can be defined based on a single pulse, or "the waveform" can be considered to be a multipulse burst. Both pulsed and CW waveforms can be further categorized based on the presence or absence of frequency or phase modulation. If present, the modulation may be intrapulse (applied to individual pulses), interpulse (applied across the pulses of a multipulse waveform), or both. Phase modulation can be biphase (two possible states) or polyphase (more than two phase states); frequency modulation can be linear or nonlinear. Intrapulse amplitude modulation may be used, but usually is not.

The choice of waveform directly determines or is a major contributor to several fundamental radar system performance metrics. These include the *signal-to-noise ratio* (SNR) χ , the range resolution ΔR , the Doppler (velocity) resolution ΔF_D (Δv), ambiguities in range and Doppler, range and Doppler sidelobes, and range-Doppler coupling. These metrics are determined by such waveform attributes as the pulse duration, bandwidth, amplitude, and phase or frequency modulation. While all of these metrics are discussed, the primary emphasis is on SNR, range resolution, and Doppler resolution because these are the most fundamental drivers in choosing the waveform. As an example, the simple pulse of Fig. 4.1*a* has a duration of τ seconds and an amplitude of *A* volts. The SNR will prove to be proportional to the waveform energy, which is the product $A^2\tau$ of its power and duration. The range resolution of $c\tau/2$ is proportional to the pulse duration. It will be shown shortly that both the waveform bandwidth and the Doppler resolution of the simple pulse are inversely proportional to the pulse length.

Two classic references on radar waveforms are Cook and Bernfeld (1993) and Rihaczek (1996). Most radar system books cover the fundamentals of radar waveforms (e.g., Nathanson, 1991; Peebles, 1998). A complete modern reference on radar waveforms is Levanon and Mozeson (2004), while good brief surveys of basic and advanced waveforms are in Keel (2010) and Keel and Baden (2012), respectively. In addition to covering the mainstream waveforms such as pulse bursts and LFM, this text covers the many developments in phase codes in recent decades. Another new text that focuses more on advanced waveforms and emerging applications is Gini et al. (2012).

4.2 The Waveform Matched Filter

4.2.1 The Matched Filter

So far, it has been implicitly assumed that the overall frequency response of the radar receiver is a bandpass characteristic with a bandwidth equal to or greater than that of the transmitted signal. Equivalently, once the carrier is demodulated out, the effective frequency response is a lowpass filter with a bandwidth equal to that of the complex envelope. It will be shown in <u>Chap. 6</u> that detection performance improves with increasing SNR. Thus, it is reasonable to ask what overall receiver frequency response $H(\Omega)$ will maximize the SNR.

To answer this question, note that the spectrum of the receiver output, y(t),

will be $Y(\Omega) = H(\Omega)X(\Omega)$, where $X(\Omega)$ is the spectrum of the waveform (and thus, except for a phase shift due to an overall delay, of a received target echo). Consider maximizing the SNR at a specific time T_M . The power of the signal component of the output at that instant is

$$\left|y(T_{M})\right|^{2} = \left|\frac{1}{2\pi}\int_{-\infty}^{\infty} X(\Omega)H(\Omega)e^{j\Omega T_{M}}d\Omega\right|^{2}$$

$$(4.4)$$

To determine the output noise power, consider the case where the interference is white noise with power spectral density $\sigma_w^2 W/Hz$. The noise power spectral density at the output of the receiver will be $\sigma_w^2 |H(\Omega)|^2 W/Hz$. The total output noise power is then

$$n_p = \frac{\sigma_w^2}{2\pi} \int_{-\infty}^{\infty} \left| H(\Omega) \right|^2 d\Omega$$
(4.5)

and the SNR measured at time T_M is

$$\chi = \frac{\left|y(T_{M})\right|^{2}}{n_{p}} = \frac{\left|(1/2\pi)\int_{-\infty}^{\infty}X(\Omega)H(\Omega)e^{j\Omega T_{M}}\,d\Omega\right|^{2}}{(\sigma_{w}^{2}/2\pi)\int_{-\infty}^{\infty}\left|H(\Omega)\right|^{2}d\Omega}$$
(4.6)

Clearly, χ depends on the receiver frequency response. The choice of $H(\Omega)$ that will maximize χ can be determined via the Schwarz inequality. One of many forms of the Schwarz inequality is

$$\left|\int A(\Omega)B(\Omega)d\Omega\right|^{2} \leq \left\{\int \left|A(\Omega)\right|^{2}d\Omega\right\} \left\{\int \left|B(\Omega)\right|^{2}d\Omega\right\}$$
(4.7)

with equality if and only if $B(\Omega) = \alpha A^*(\Omega)$, with α any arbitrary constant. Applying Eq. (4.7) to the numerator of Eq. (4.6) give the upper bound on SNR as

$$\chi \leq \frac{(1/2\pi)^2 \int_{-\infty}^{\infty} \left| X(\Omega) e^{j\Omega T_M} \right|^2 d\Omega \int_{-\infty}^{\infty} \left| H(\Omega) \right|^2 d\Omega}{(\sigma_w^2/2\pi) \int_{-\infty}^{\infty} \left| H(\Omega) \right|^2 d\Omega}$$
(4.8)

The SNR is maximized when

$$H(\Omega) = \alpha X^*(\Omega) e^{-j\Omega T_M} \text{ or}$$
$$h(t) = \alpha x^*(T_M - t)$$
(4.9)

This particular choice of the receiver filter frequency or impulse response is called the *matched filter*, because the response is "matched" to the signal waveform. Thus, the waveform and the receiver filter needed to maximize the output SNR are a matched pair. If the radar changes waveforms, it must also change the receiver filter response in order to stay in a matched condition. The impulse response of the matched filter is obtained by time-reversing and conjugating the complex waveform. The gain constant α is often set equal to unity; it has no impact on the achievable SNR, as seen later in this chapter. The time T_M at which the SNR is maximized is arbitrary. However, $T_M \ge \tau$ is required for h(t) to be causal.

Given an input signal x'(t) consisting of both target and noise components, the output of the matched filter is given by the convolution

$$y(t) = \int_{-\infty}^{\infty} x'(s)h(t-s)ds$$
$$= \alpha \int_{-\infty}^{\infty} x'(s)x^*(s+T_M-t)ds$$
(4.10)

(4.11)

The second line of Eq. (4.10) is recognized as the cross-correlation of the target-plus-noise signal x'(t) with the transmitted waveform x(t), evaluated at lag $T_M - t$. Thus, the matched filter implements a correlator with the transmitted waveform as the reference signal.

It is useful to determine the maximum value of SNR achieved by the matched filter. Using $H(\Omega) = \alpha X^*(\Omega) \exp(-j\Omega T_M)$ in Eq. (4.6)

$$\chi = \frac{\left| (1/2\pi) \int_{-\infty}^{\infty} X(\Omega) [\alpha X^*(\Omega) e^{-j\Omega T_M}] e^{j\Omega T_M} d\Omega \right|^2}{(\sigma_w^2/2\pi) \int_{-\infty}^{\infty} |\alpha X^*(\Omega) e^{-j\Omega T_M}|^2 d\Omega}$$
$$= \frac{\left| (1/2\pi) \alpha \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega \right|^2}{|\alpha|^2 \sigma_w^2/2\pi \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega}$$
$$= \frac{1}{2\pi \sigma_w^2} \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega$$

The energy in the signal x(t) is

$$E = \int_{-\infty}^{\infty} \left| x(t) \right|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| X(\Omega) \right|^2 d\Omega$$
(4.12)

where the second step follows from Parseval's relation. Using Eq. (4.12) in Eq. (4.11) gives

$$\chi = \frac{1}{2\pi\sigma_w^2} \int_{-\infty}^{\infty} |X(\Omega)|^2 d\Omega = \frac{E}{\sigma_w^2}$$
(4.13)

Equation (4.13) states the remarkable result that the maximum achievable SNR depends only on the energy of the waveform and not on other details such as its modulation. Two waveforms having the same energy will produce the same maximum SNR, provided each is processed through its own matched filter.

Although it is the ratio of the peak signal component power to the noise power, the SNR of Eq. (4.13) is called the *energy SNR* because the peak signal power at the matched filter output equals the energy of the transmitted signal. To see this, note that the peak signal component at the matched filter output is given by Eq. (4.10) with $t = T_M$

$$y(T_M) = \int x(s) \,\alpha x^*(s) \, ds = \alpha E \tag{4.14}$$

Also, the duration of the signal component of the matched filter output is exactly 2τ seconds, since it is the convolution of the τ -second pulse with the τ -second matched filter impulse response.

The previous results can be generalized to develop a filter that maximizes output *signal-to-interference ratio* (SIR) when the interference power spectrum is not white. In radar, this is useful for example in cases where the dominant interference is clutter, which generally has a colored power spectrum. The result can be expressed as a two-stage filtering operation. The first stage is a *whitening filter* that converts the interference power spectrum to a flat spectrum (and also modifies the signal spectrum in the process); the second stage is then a conventional matched filter as described earlier, but designed for the now-modified signal spectrum. Details are given by Kay (1998).

4.2.2 Matched Filter for the Simple Pulse

To illustrate the previous ideas, consider a simple pulse of duration τ :

$$x(t) = \begin{cases} 1, & 0 \le t \le \tau \\ 0, & \text{otherwise} \end{cases}$$

(4.15)

The corresponding matched filter impulse response is

$$h(t) = \alpha x^* (T_M - t)$$

$$= \begin{cases} \alpha, & T_M - \tau \le t \le T_M \\ 0, & \text{otherwise} \end{cases}$$
(4.16)

where $T_M > \tau$ for causality. Because x(t) is a much simpler function than its Fourier transform (a sinc function), it is easier to work with the correlation interpretation of Eq. (4.10) to compute the output. Figure 4.2 illustrates the two terms in the integrand, helping to establish the regions of integration. Part *a* of the figure shows that



FIGURE 4.2 Convolution of simple pulse and its matched filter: (a) $T_M - \tau \le t \le T_M$, (b) $T_M \le t \le T_M + \tau$.

$$y(t) = \begin{cases} 0, & t < T_M - \tau \\ \int_0^{t - T_M + \tau} (1)(\alpha) \, ds, & T_M - \tau \le t \le T_M \end{cases}$$
(4.17)

while part b is useful in identifying the next two regions

$$y(t) = \begin{cases} \int_{t-T_M}^{\tau} (1)(\alpha) \, ds, & T_M \le t \le T_M + \tau \\ 0, & t > T_M + \tau \end{cases}$$
(4.18)

The result is

$$y(t) = \begin{cases} \alpha [t - (T_M - \tau)], & T_M - \tau \le t \le T_M \\ \alpha [(T_M + \tau) - t], & T_M \le t \le T_M + \tau \\ 0, & \text{otherwise} \end{cases}$$
(4.19)

This result is illustrated in Fig. 4.3. The matched filter output is a triangle function of duration 2τ seconds with its peak at $t = T_M$ as expected. The peak value is $\alpha\tau$; since the energy of the unit amplitude pulse is just τ , the peak value equals αE as predicted.



FIGURE 4.3 Matched filter output for a simple pulse.

The noise power at the output of the matched filter is

$$n_{p} = \frac{\sigma_{w}^{2}}{2\pi} \int_{-\infty}^{\infty} |H(\Omega)|^{2} d\Omega$$

= $\sigma_{w}^{2} \int_{-\infty}^{\infty} |h(t)|^{2} dt$ (Parseval's relation)
= $\sigma_{w}^{2} |\alpha|^{2} \tau$ (4.20)

The SNR is therefore

$$\chi = \frac{|\alpha\tau|^2}{\sigma_w^2 |\alpha|^2 \tau} = \frac{\tau}{\sigma_w^2} = \frac{E}{\sigma_w^2}$$
(4.21)

consistent with Eq. (4.13). Note that the gain α has no effect on the SNR.

4.2.3 All-Range Matched Filtering

The matched filter was designed to maximize the output SNR at a particular time instant T_M . This raises several questions. How should T_M be chosen, and how can the range of a target be related to the resulting output? What happens if the

received signal contains echoes from multiple targets at different ranges?

Start by choosing $T_M = \tau$, the minimum value that results in a causal matched filter. Now suppose the input to the matched filter is the echo from a target at an unknown range R_0 , corresponding to a time delay $t_0 = 2R_0/c$. The signal component of the output of the matched filter will be

$$y(t) = \int_{-\infty}^{\infty} x(s - t_0) \alpha x^* (s + \tau - t) \, ds \tag{4.22}$$

This is just the correlation of the received, delayed echo and the matched filter impulse response. The output waveform will again be a triangle with its peak at correlation lag zero. This occurs when $s - t_0 = s + \tau - t$, or $t = t_0 + \tau$. The matched filter output will appear as in Fig. 4.4. The peak will occur at time $t_{\text{peak}} = t_0 + \tau$, corresponding to the actual delay to the target plus the delay of the causal matched filter. The target range can be easily determined from observation of the matched filter output as $R_0 = c(t_{\text{peak}} - \tau)/2$.



FIGURE 4.4 Output of the matched filter for a target at range $R_0 = ct_0/2$.

This discussion shows that the matched filter parameter T_M can be chosen arbitrarily (typically as $T_M = \tau$). Once T_M is known, the range of a target can be determined by detecting the time at which a peak occurs at the matched filter output, subtracting T_M to get the delay to the target and back, and converting to units of range. Thus, a single choice of T_M allows detection of targets at all ranges. One simply samples the matched filter output at a series of fast-time sample instants t_k ; if a peak occurs at time t_k , it corresponds to a target at range $c(t_k - T_M)/2$. If the received signal contains echoes from multiple targets at different ranges, by superposition the matched filter output will contain multiple copies of the single-pulse triangle response, one centered at the time delay (plus filter delay) of each of the various targets.

4.2.4 Straddle Loss

In modern practice, matched filtering is carried out digitally so that y(t) is sampled at some fast-time sampling rate $F_s = 1/T_s$. Typically F_s equals or is slightly greater than the waveform bandwidth β . The range sample spacing is

then $cT_s/2$ meters. In general, targets do not arrange themselves precisely at ranges corresponding to the range samples. The receiver then will not sample the matched filter output precisely at its peak. The result is a reduction in the measured signal amplitude and therefore an SNR loss.

This is exactly the issue of straddle loss that was discussed in <u>Chap. 3</u> with regard to the DFT of frequency domain data. In either case, the finite sampling rate allows the processor to "miss" the peak response, whether it is the matched filter output in fast time or the spectrum of a slow-time signal. Straddle loss also arises in angular sampling with scanning antennas. In any of these cases it can be reduced with higher sampling rates or various interpolation methods. Consideration of these methods is deferred to the discussion of pulse Doppler analysis in <u>Chap. 5</u> and the analyses of time delay, frequency, and angle estimation in <u>Chap. 7</u>. All of the methods there can be applied to the fast-time straddle loss for the various waveforms in this chapter.

4.2.5 Range Resolution of the Matched Filter

By determining the range separation that would result in nonoverlapping echoes, it was shown in <u>Chap. 1</u> that the range resolution achieved by a simple pulse of duration τ seconds is $c\tau/2$ meters. When a matched filter is used, the output due to each scatterer is now 2τ seconds long, but is also triangular rather than rectangular in shape. Does the longer matched filter output result in a larger value of range resolution?

Before considering this question, it is useful to recall that the demodulated echo from a scatterer at range R_0 meters has not only a delay of $t_0 = 2R_0/c$ seconds, but also an overall phase shift of $exp[j(-4\pi/\lambda)R_0]$ radians.¹ A change of only $\lambda/4$ in range will cause a change of 180° in the received echo phase. Two overlapping target responses may therefore add either constructively or destructively in phase, and small deviations in their spacing can result in large changes in the composite response. Consider two targets at ranges $ct_0/2$ and $ct_0/2 + c\tau/2$ and assume τ is such that the two matched filter responses add in phase. Then the composite response at the matched filter output is a flat-topped trapezoid as shown in Fig. 4.5a. Clearly, if the separation between the two scatterers increases, a dip will begin to develop in the composite response, even when the separation is such that they remain in phase. If the separation decreases, the in-phase response will still be a trapezoid, but with a higher peak and a shorter flat region as the responses overlap more. Because any increase in separation will result in a dip between the two responses, the separation of $c\tau/2$ meters is still considered to be the range resolution of the matched filter output. Thus, using a matched filter does not degrade the range resolution. To reinforce this further, recall that the definition of the Rayleigh resolution is the peak-tofirst null distance. Inspection of Fig. 4.3 shows that $c\tau/2$ is also the Rayleigh resolution of the simple pulse matched filter output.



FIGURE 4.5 Composite matched filter response due to two scatterers separated by $c\tau/2$ meters: (*a*) target responses in phase, (*b*) target responses 180° out of phase.

Scatterers that are closer together than the Rayleigh resolution may still be resolved if the spacing is such that the individual responses add out of phase. Figure 4.5b illustrates the case where the two responses differ in phase by 180°. Destructive interference in the region of overlap causes a deep null in the composite response. However, this null is very sensitive to the fine spacing of the scatterers and cannot be relied on to resolve two targets.

4.3 Matched Filtering of Moving Targets

Suppose a simple pulse is transmitted, x(t) = 1, $0 \le t \le \tau$, and it echoes from a target moving toward the radar with a radial velocity of v meters per second. After demodulation, the received waveform (ignoring the overall time delay) will be $x'(t) = x(t)\exp(j\Omega_D t)$, with $\Omega_D = 4\pi v/\lambda$. Because the echo is different from x(t), a filter matched to x(t) will *not* be matched to x'(t). If the target velocity is known, the matched filter for x'(t) can be constructed:

$$h(t) = \alpha x^{\prime *} (-t) = \alpha x^{*} (-t) e^{+j\Omega_{D}t}$$
(4.23)

The frequency response of this matched filter is

$$H(\Omega) = \alpha \int_{-\infty}^{\infty} x^* (-t) e^{+j\Omega_D t} e^{-j\Omega t} dt, \qquad t' = -t$$
$$= \alpha \left[\int_{-\infty}^{\infty} x(t') e^{-j(\Omega - \Omega_D)t'} dt' \right]^*$$
$$= \alpha X^* (\Omega - \Omega_D)$$
(4.24)

Thus, the matched filter for x'(t) can be obtained by simply shifting the center frequency of the matched filter for x(t) to the expected Doppler shift.

A more interesting situation occurs when the velocity is not known in advance so that the receiver filter is not matched to the target Doppler shift. More generally, suppose the filter is matched to some Doppler shift Ω_i radians per second but the actual Doppler shift of the echo is Ω_D . Choosing $T_M = 0$ for simplicity, the matched filter output will be zero for $|t| > \tau$. For $0 \le t \le \tau$ the response is

$$y(t) = \alpha \int_{t}^{t} \exp(j\Omega_{D}s) \exp[-j\Omega_{i}(s-t)] ds$$
(4.25)

If the filter is in fact matched to the actual Doppler shift, $\Omega_i = \Omega_D$, the output becomes

$$y(t) = \alpha \exp(j\Omega_D t) \int_t^\tau (1) ds$$

= $\alpha e^{j\Omega_D t} (\tau - t), \quad 0 \le t \le \tau$
(4.26)

The analysis is similar for negative $t, -\tau \le t \le 0$. The complete result is

$$y(t) = \begin{cases} \alpha e^{j\Omega_D t} (\tau - |t|), & -\tau \le t \le \tau \\ 0, & \text{otherwise} \end{cases}$$

$$(4.27)$$

Thus, |y(t)| is the usual triangular function, peaking as expected at t = 0.

If there is a Doppler mismatch, $\Omega_i \neq \Omega_D$, the response at the expected peak time t = 0 is

$$\begin{aligned} y(t)|_{t=0} &= \alpha \int_0^\tau \exp(j\Omega_D s) \exp(-j\Omega_i s) \, ds \\ &= \alpha \int_0^\tau \exp[+j(\Omega_D - \Omega_i)s] \, ds \\ &= \frac{\alpha}{j(\Omega_D - \Omega_i)} \exp[j(\Omega_D - \Omega_i)s] \Big|_0^\tau \end{aligned}$$

$$(4.28)$$

Defining $\Omega_{\text{diff}} \equiv \Omega_D - \Omega_i$

$$\left| y(0) \right| = \left| \frac{2\alpha \sin(\Omega_{\text{diff}} \tau/2)}{\Omega_{\text{diff}}} \right|$$
(4.29)

Equation (4.29) is plotted in Fig. 4.6. The first zero of this sinc function occurs at $F_{diff} = 1/\tau$ Hz.² Relatively small Doppler mismatches ($F_{diff} \ll 1/\tau$) will cause only slight reductions in the matched filter output peak amplitude. Large mismatches, however, can cause very substantial reductions.



FIGURE 4.6 Effect of Doppler mismatch on matched filter response at expected peak time.

The effect of Doppler mismatch can be either good or bad. If targets are moving and the velocities are unknown, mismatch will cause reductions in observed peaks and, if severe enough, may prevent detection. The signal processor must either estimate the target Doppler so that the matched filter can be adjusted or construct matched filters for a number of different possible Doppler frequencies and observe the output of each to search for targets. On the other hand, if the goal is to be selective in responding only to targets of a particular Doppler shift, it is desirable to have a matched filter that suppresses targets at other Doppler shifts.

From Fig. 4.6, it is clear that the Rayleigh resolution of the Doppler mismatch response is $1/\tau$ Hz. The resolution in velocity is therefore $\lambda/2\tau$ meters

per second. For typical pulse lengths, these are fairly large values. For example, a 10 μ s pulse would exhibit a Rayleigh resolution in Doppler of 100 kHz, or in velocity at X band (10 GHz) of 1500 m/s. Many systems do not observe such high Doppler shifts, so Doppler mismatch effects are insignificant and targets cannot be resolved in Doppler on a single pulse. If finer Doppler resolution is desired, a very long pulse may be needed. For example, velocity resolution of 1 m/s at X band requires a 15-ms pulse. The range resolution is then a very poor 2250 km. This conflict between good range resolution and good Doppler resolution can be resolved using a pulse burst waveform, which will be addressed in Sec. 4.5.

4.4 The Ambiguity Function

4.4.1 Definition and Properties of the Ambiguity Function

In the preceding sections, the matched filter response for the simple pulse waveform has been analyzed to show its behavior both in time and in response to Doppler mismatches. The *ambiguity function* (AF) is an analytical tool for waveform design and analysis that succinctly characterizes the behavior of a waveform paired with its matched filter. The AF is useful for examining resolution, sidelobe behavior, and ambiguities in both range and Doppler for a given waveform, as well as phenomena such as range-Doppler coupling (introduced in Sec. 4.6.4).

Consider the output of a matched filter for a waveform x(t) when the input is a Doppler-shifted response $x(t)\exp(j2\pi F_D t)$. Also assume that the filter has unit gain ($\alpha = 1$) and is designed to peak at $T_M = 0$; this merely means that the time axis at the filter output is relative to the expected peak output time for the range of the target. The filter output will be

$$y(t; F_D) = \int_{-\infty}^{\infty} x(s) \exp(j2\pi F_D s) x^* (s-t) \, ds$$

$$\equiv \hat{A}(t, F_D)$$
(4.30)

which is defined as the *complex ambiguity function* $\hat{A}(t, F_D)$. An equivalent definition can be given in terms of the signal spectrum by applying basic Fourier transform properties:

$$\hat{A}(t, F_D) = \int_{-\infty}^{\infty} X^*(F) X(F - F_D) \exp(j2\pi F t) \, dF$$
(4.31)

The *ambiguity function*³ is defined as the magnitude of $\hat{A}(t, F_D)$,

$$A(t, F_D) \equiv \left| \hat{A}(t, F_D) \right|$$

$$(4.32)$$

It is a function of two variables: the time delay relative to the expected matched filter peak output, and the mismatch between that Doppler shift for which the filter was designed, and that which is actually received. For example, the AF evaluated at time t = 0 corresponds to the output of the actual matched filter at time $t = 2R_0/c + \tau$ for a target at range R_0 . The particular form of the AF is determined entirely by the complex waveform x(t).

Three properties of the ambiguity function are of immediate interest. The first states that if the waveform has energy E, then

$$|A(t, F_D)| \le |A(0, 0)| = E$$

(4.33)

Thus, when the filter is matched in Doppler to the echo and is sampled at a delay corresponding to the target range, the response will be maximum. If the filter is not matched or is sampled at a different delay, then the response will be less than or equal to (usually less than) the maximum. The second property states that total area under any ambiguity function is constant and is given by

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(t, F_D)|^2 dt dF_D = E^2$$
(4.34)

This conservation of energy statement implies that, in the design of waveforms, one cannot remove energy from one portion of the ambiguity surface without placing it somewhere else; it can only be moved around on the ambiguity surface. The third property is a symmetry relation:

$$A(t, F_{\rm D}) = A(-t, -F_{\rm D})$$
(4.35)

In order to prove the first property, start with the square of Eq. (4.32)

$$|A(t, F_D)|^2 = \left| \int_{-\infty}^{\infty} x(s) x^*(s-t) \exp(j2\pi F_D s) \, ds \right|^2$$
(4.36)

Applying the Schwartz inequality to Eq. (4.36) yields

$$|A(t, F_D)|^2 \le \int_{-\infty}^{\infty} |x(s)|^2 ds \int_{-\infty}^{\infty} |x^*(s-t)\exp(j2\pi F_D s)|^2 ds$$
$$= \int_{-\infty}^{\infty} |x(s)|^2 ds \int_{-\infty}^{\infty} |x^*(s-t)|^2 ds$$
(4.37)

Each integral is just the energy E in x(t), so that

$$\left|A(t, F_D)\right|^2 \le E^2 \tag{4.38}$$

The equality holds only if $x(s) = x(s - t)\exp(-j2\pi F_D s)$ for all *s*, which occurs if and only if $t = F_D = 0$. Making these substitutions in Eq. (4.38) gives the equality in Eq. (4.33).

The proof of the second property starts by defining the complex conjugate of the complex ambiguity function, where

$$\hat{A}^{*}(t, F_{D}) = \int_{-\infty}^{\infty} x^{*}(s) x (s-t) \exp(-j2\pi F_{D}s) ds$$
$$= \int_{-\infty}^{\infty} X(F) X^{*}(F - F_{D}) \exp(-j2\pi Ft) dF$$
(4.39)

The squared magnitude of the ambiguity function can then be written as

$$|A(t, F_D)|^2 = \hat{A}(t, F_D)\hat{A}^*(t, F_D)$$

= $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(s)x^*(s-t)X(F)X^*(F-F_D)\exp[j2\pi(F_Ds-Ft)]ds dF$
(4.40)

The total energy in the ambiguity surface is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| A(t, F_D) \right|^2 dt \, dF_D = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(s) x^*(s-t) X(F) X^*(F-F_D) \\ \times \exp[j2\pi(F_D s - Ft)] \, ds \, dF \, dt \, dF_D$$

$$(4.41)$$

Isolating those terms integrated over t and F_D yields the following two relationships:

$$\int_{-\infty}^{\infty} x^* (s-t) \exp(-j2\pi Ft) \, dt = \exp(-j2\pi Fs) X^*(F)$$
(4.42)

$$\int_{-\infty}^{\infty} X^* (F - F_D) \exp(j2\pi F_D s) \, dF_D = \exp(j2\pi F s) x^*(s)$$
(4.43)

Substituting these into Eq. (4.41) yields

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |A(t, F_D)|^2 dt dF_D = (1/2\pi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(s) X^*(F) X(F) x^*(s) ds dF$$
$$= \left\{ \int_{-\infty}^{\infty} |x(s)|^2 ds \right\} \left\{ 1/2\pi \int_{-\infty}^{\infty} |X(F)|^2 dF \right\}$$
(4.44)

The first integral on the right-hand side of Eq. (4.44) is just the energy *E* of the pulse measured in the time domain; the second is, by Parseval's theorem, also the energy. Thus

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| A(t, F_D) \right|^2 dt \, dF_D = E^2$$
(4.45)

The symmetry property can be proved by substituting -t and $-F_D$ for t and F_D , respectively, in the definition in Eq. (4.30)

$$\hat{A}(-t, -F_D) = \int_{-\infty}^{\infty} x(s) \exp(-j2\pi F_D s) x^*(s+t) \, ds$$
(4.46)

Now make the change of variables s' = s + t to get

$$\hat{A}(-t, -F_D) = \int_{-\infty}^{\infty} x(s'-t) \exp(-j2\pi F_D(s'-t)) x^*(s') \, ds'$$

= $\exp(j2\pi F_D t) \int_{-\infty}^{\infty} x(s'-t) \exp(-j2\pi F_D s') x^*(s') \, ds'$
= $\exp(j2\pi F_D t) \hat{A}^*(t, F_D)$ (4.47)

Since $A(t, F_D) \equiv |\hat{A}(t, F_D)|$, Eq. (4.35) follows immediately.

It is reasonable to ask what would be an ideal ambiguity function. The answer varies depending on the intent of the system design, but a commonly cited goal is the "thumbtack" ambiguity function of Fig. 4.7, which features a single central peak with the remaining energy spread uniformly throughout the delay-Doppler plane. The narrow central peak implies good resolution in both range and Doppler. The lack of any secondary peak implies that there will be no range or Doppler ambiguities. The uniform plateau suggests low and uniform sidelobes, minimizing target masking effects. All of these features are beneficial for a system designed to make fine-resolution measurements of targets in range and Doppler or to perform radar imaging. On the other hand, a waveform intended to be used for target search might be preferred to be more tolerant of Doppler mismatch so that the Doppler shift of targets whose velocity is not yet known does not prevent their detection due to a weak response at the matched filter output. Thus, what is "ideal" in the way of an ambiguity function depends on the use to which the waveform will be put.



FIGURE 4.7 "Thumbtack" ambiguity function.

4.4.2 Ambiguity Function of the Simple Pulse

As a first example of an ambiguity function, consider a simple pulse centered on the origin and normalized to have unit energy (E = 1) for convenience

$$x(t) = \frac{1}{\sqrt{\tau}}, \quad \frac{-\tau}{2} \le t \le \frac{\tau}{2}$$
(4.48)

Applying Eq. (4.30) gives for t > 0

$$\hat{A}(t, F_{D}) = \int_{-\tau/2+t}^{\tau/2} \frac{1}{\tau} \exp(j2\pi F_{D}s) \, ds$$

$$= \frac{\exp\left[j2\pi F_{D}\tau/2\right] - \exp\left[j2\pi F_{D}(-\tau/2+t)\right]}{\tau j 2\pi F_{D}}$$

$$= \frac{1}{\tau j 2\pi F_{D}} e^{j2\pi F_{D}t/2} \left\{ \exp\left[j2\pi F_{D}\left(\frac{\tau}{2} - \frac{t}{2}\right)\right] - \exp\left[-j2\pi F_{D}\left(\frac{\tau}{2} - \frac{t}{2}\right)\right] \right\}$$

$$(4.49)$$

The ambiguity function for t > 0 is the magnitude of Eq. (4.49)

$$A(t, F_D) = \left| \hat{A}(t, F_D) \right| = \left| \frac{\sin(\pi F_D(\tau - t))}{\tau \pi F_D} \right|, \quad 0 \le t \le \tau$$
(4.50)

Repeating the derivation for t < 0 gives a similar result, but with the quantity ($\tau - t$) replaced by ($\tau + t$). The complete AF of the simple pulse is therefore

$$A(t, F_D) = \left| \frac{\sin[\pi F_D(\tau - |t|)]}{\tau \pi F_D} \right|$$
$$= \left(1 - \frac{|t|}{\tau} \right) \left| \frac{\sin[\pi F_D \tau (1 - |t|/\tau)]}{\pi F_D \tau (1 - |t|/\tau)} \right|, \quad -\tau \le t \le \tau$$
(4.51)

Equation (4.51) is plotted in Fig. 4.8 in a three-dimensional surface plot and in Fig. 4.9 as a contour plot, which is often easier to interpret and is therefore used in most cases in the remainder of this chapter. The AF for a simple pulse is a triangular ridge oriented along the delay axis. Doppler mismatches on the order of $1/\tau$ Hz or more drastically reduce and spread the matched filter output peak, as was shown previously.



FIGURE 4.8 Ambiguity function of a unit-energy simple pulse of length τ .



FIGURE 4.9 Contour plot of the simple pulse ambiguity function of Fig. 4.8.

1

The zero-Doppler response A(t, 0) gives the matched filter output when there is no Doppler mismatch. Setting $F_D = 0$ in Eq. (4.51) and using L'Hôpital's rule to resolve the indeterminate form gives

$$A(t,0) = \left| \frac{\pi(\tau - |t|) \cos\left[\pi F_D(\tau - |t|)\right]}{\tau \pi} \right|_{F_D = 0}$$
$$= \frac{\tau - |t|}{\tau}, \quad -\tau \le t \le \tau$$
(4.52)

Similarly, the zero-delay cut $A(0, F_D)$ gives the output of the matched filter at the expected peak time t = 0 as a function of Doppler mismatch. Using t = 0 in Eq. (4.51) immediately gives

$$A(0, F_D) = \left| \frac{\sin(\pi F_D \tau)}{\tau \pi F_D} \right|$$
(4.53)

Equations (4.52) and (4.53) are the expected triangle and sinc functions derived previously. They are illustrated in Fig. 4.10.



FIGURE 4.10 (a) Zero-Doppler cut of simple pulse AF, (b) zero-delay cut.

A Doppler mismatch not only reduces the peak amplitude but, if severe enough, completely alters the shape of the range response of the matched filter. Figure 4.11 shows the effect of varying degrees of Doppler mismatch on the matched filter range response. These curves should be compared to Fig. 4.10*a*. A mismatch of $0.31/\tau$ Hz results in a reduction of about 16 percent in the peak

amplitude, but the peak remains at the correct time delay. A larger shift, for example $0.94/\tau$, not only reduces the maximum output amplitude by 65 percent but eliminates the central peak altogether. By the time the mismatch is several times $1/\tau$, the response becomes completely unstructured. Note that a mismatch of n/τ Hz means that there will be *n* cycles of the Doppler frequency during the pulse duration τ . Also recall that for typical pulse lengths, $1/\tau$ is a large Doppler shift, so that the simple pulse still ranks as a relatively Doppler-tolerant waveform. For instance, if $\tau = 10 \ \mu$ s, a Doppler shift of $0.31/\tau$ is 31 kHz, corresponding at an RF of 10 GHz to a velocity of 465 m/s, or 1040 mph. Even with this very large Doppler mismatch, the simple pulse matched filter output retains its basic shape, correct peak location, and suffers only the 16 percent (1.5 dB) amplitude loss.



FIGURE 4.11 Effect of Doppler mismatch on the range response of the matched filter for the simple pulse.

4.5 The Pulse Burst Waveform

The flip side of the Doppler tolerance of the simple pulse described in the preceding example is that its Doppler resolution is very poor. If the designer wants the radar system to respond to targets only at certain velocities and reject targets at nearby velocities, the simple pulse is not adequate as a waveform. Better frequency resolution requires a longer observation time. The *pulse burst waveform* is one way to meet this requirement. It is defined as

$$x(t) = \sum_{m=0}^{M-1} x_p(t - mT)$$
(4.54)

where $x_p(t) = \text{single pulse of length } \tau$

M = number of pulses in the burst

T = pulse repetition interval

While the constituent pulse $x_p(t)$ can be any single-pulse waveform, for the moment only the simple pulse will be considered. Figure 4.12 illustrates this waveform. The solid line forming the envelope of the sinusoidal pulses is the actual baseband waveform x(t). The train of RF pulses that results when it is impressed upon a carrier is denoted as usual as $\overline{x}(t)$. The total duration *MT* (which includes the dead time after the last pulse) is the *coherent processing interval* (CPI).



FIGURE 4.12 Pulse burst waveform and the resulting train of RF pulses.

4.5.1 Matched Filter for the Pulse Burst Waveform

The matched filter for the pulse burst is (with $\alpha = 1$ and $T_M = 0$)

$$h(t) = x^*(-t) = \sum_{m=0}^{M-1} x_p^*(-t - mT)$$
(4.55)

and the matched filter output, given an echo from a range corresponding to a time delay t_0 , is therefore

$$y(t) = \int_{-\infty}^{\infty} \left\{ \sum_{m=0}^{M-1} x_p(s - t_0 - mT) \right\} \left\{ \sum_{n=0}^{M-1} x_p^*(s - t - nT) \right\} ds$$
$$= \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \int_{-\infty}^{\infty} x_p(s - t_0 - mT) x_p^*(s - t - nT) ds$$
(4.56)

The inner integral is the matched filter output for the constituent simple pulse.

Let $t_0 = 0$ for simplicity; the results can be adjusted for any other delay t_0 by shift invariance. Renaming the simple pulse matched filter output from Eq. (4.19) as $s_p(t)$, Eq. (4.56) becomes

$$y(t) = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} s_p(-t - (m-n)T) = \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} s_p^*(t - (n-m)T)$$
(4.57)

where the symmetry of $s_p(t)$ has been used in the last step. Equation (4.57) states that the matched filter output is a superposition of shifted copies of $s_p(t)$. The double summation can be simplified by noting that all terms that have the same value of (n - m) are identical and can be combined. There are M combinations of m and n such that m - n = 0, namely, all those where m = n. There are M - 1cases where m - n = +1 and another M - 1 cases where m - n = -1. Continuing in this vein gives

$$y(t) = \sum_{m=-(M-1)}^{M-1} (M - |m|) s_p^*(t - mT)$$
(4.58)

The matched filter output for the pulse burst waveform is simply a sum of scaled and shifted replicas of the output of the filter matched to a single constituent pulse.

Since the constituent pulse $x_p(t)$ is of duration τ , $s_p(t)$ is of duration 2τ . If $T > 2\tau$ as is usually the case, none of the replicas of $s_p(t)$ overlap one another. Figure 4.13 illustrates a pulse burst waveform and the corresponding matched filter output for this case and M = 3. The peak output occurs at $t = T_M = 0$:



FIGURE 4.13 (a) Pulse burst waveform, M = 3, (b) matched filter output.

$$y(0) = \sum_{m=-(M-1)}^{M-1} (M - |m|) s_p^*(-mT) = \sum_{m=-(M-1)}^{M-1} (M - |m|) s_p(mT)$$

= $Ms_p(0) = ME_p = E$
(4.59)

where the last step uses $s_p(mT) = 0$ when $T > \tau$. In this equation E_p is the energy in the single pulse $x_p(t)$, while *E* is the energy in the entire *M*-pulse waveform. Note that the peak response is *M* times that achieved with a single pulse of the same amplitude. Recall the radar range equation signal processing gain factor G_{sp} of Eq. (2.85). The increase in the matched filter output peak for the pulse burst waveform represents a coherent signal processing gain of a factor $G_{sp} = M$ that will improve the SNR compared to a single-pulse waveform, aiding detection probability and measurement precision.

4.5.2 Pulse-by-Pulse Processing

The structure of Eq. (4.58) suggests that it is not necessary to construct an explicit matched filter for the entire pulse burst waveform x(t), but rather that the matched filter can be implemented by filtering the data from each individual pulse with the single-pulse matched filter and then combining those outputs. This process, called *pulse-by-pulse* processing, uses separable two-dimensional processing in fast time and slow time. It provides a much more convenient implementation and is consistent with how pulse burst waveforms are processed in real systems.

Define the matched filter impulse response for the individual pulse in the burst, assuming $T_M = 0$

$$h_p(t) = x_p^*(-t)$$
(4.60)

The output from this filter for the *m*th transmitted pulse, assuming a target at some delay t_l , is

$$y_m(t) = x_p(t - t_l - mT) * h_p(t)$$

= $s_p^*(t - t_l - mT) = s_p(-t + t_l + mT), \quad 0 \le m \le M - 1$
(4.61)

Assume that the echo from the individual pulse matched filter for the first pulse (m = 0) is sampled at $t = t_i$; that value will be $y_0(tl) = sp(0)$. Now sample the filter response to each succeeding pulse at the same delay after its transmission

(i.e., sample the same range bin for each pulse). The filter output for pulse *m* is sampled at $t = t_l + mT$, giving $y_m(t_l + mT) = s_p(0)$ again.

If the sample taken at time *tl* after pulse transmission is associated with range bin *l*, the *M* samples so obtained form a discrete constant-valued sequence $y[l, m] = sp(0), 0 \le m \le M - 1$. The discrete-time causal matched filter in the slow-time (*m*) dimension for such a sequence is $h[m] = \alpha y^*[M - 1 - m]$; with $\alpha = 1/s_p(0), h[m] = 1$ for $0 \le m \le M - 1$. The output of this discrete-time matched filter is

$$z[m] = \sum_{r=0}^{M-1} y[l, r]h[m-r]$$

$$= \begin{cases} \sum_{r=0}^{m} y[l, r](1), & 0 \le m \le M-1 \\ \sum_{r=m-M+1}^{M-1} y[l, r](1), & M-1 \le m \le 2(M-1) \end{cases}$$
(4.62)

The peak output will occur when the two functions in the summand completely overlap, which requires m = M - 1; then

$$z[M-1] = \sum_{r=0}^{M-1} y[l,r] = Ms_p(0) = ME_p = E$$
(4.63)

Equation (4.63) indicates that in pulse-by-pulse processing, matched filtering of the slow-time sequence from a given range bin reduces to coherently integrating the slow-time samples in each range bin, and the resulting peak output is identical to that obtained with a whole-waveform continuous matched filter of Eq. (4.55). Figure 4.14 illustrates the row of slow-time samples that are integrated (after matched filtering of the single pulse in fast time) to complete the matched filtering process for the pulse burst. This operation is performed independently for each range bin.



Pulse # = Slow Time

FIGURE 4.14 Slow-time sequence to be integrated for matched filtering of a pulse burst waveform.

4.5.3 Range Ambiguity

Evaluating the pulse burst matched filter output at t = 0 gave the peak output for a target at the time delay t_0 under consideration. Normally $t_0 < T$ and if a peak is observed it will be interpreted as implying the presence of a target at range $R_0 = ct_0/2$ m. However, suppose the data instead contain echoes from a target an additional *T* seconds of delay further away. The received waveform will be unchanged except for a delay of *T* seconds and a reduced amplitude according to the range equation. The amplitude reduction is not pertinent to the discussion and is ignored. By shift invariance, the matched filter output of Eq. (4.58) will also be delayed by *T* seconds

$$y(t) = \sum_{m=-(M-1)}^{M-1} (M - |m|) s_p^* [t - (m+1)T]$$
(4.64)

Now when the matched filter output is evaluated at t = 0 the result is

$$y(0) = \sum_{m=-(M-1)}^{M-1} (M - |m|) s_p^* [-(m+1)T]$$
(4.65)

In this expression (and continuing to assume $T > 2\tau$) only the m = -1 term is nonzero, so that

$$y(0) = (M-1)s_p(0) = (M-1)E_p$$
(4.66)

This equation shows that the output at the sample time is reduced from ME_p to $(M - 1)E_p$. The situation is illustrated in Fig. 4.15 from both the wholewaveform matched filter and pulse-by-pulse viewpoints. From the former viewpoint, a local peak of the matched filter output is sampled, but the global peak is missed because the filter is "tuned" for the wrong delay. The result, while not zero, is a reduced-amplitude sample, reducing SNR. From the latter viewpoint, the echo appears in only M - 1 of the M slow-time samples integrated because it first returns after the sampling window following transmission of the second pulse rather than the first.


FIGURE 4.15 Pulse burst matched filtering when target is range ambiguous by one PRI: (*a*) whole-waveform matched filter output. Compare to Fig. 4.13*b*. (*b*) Slow-time data in pulse-by-pulse processing viewpoint. Compare to Fig. 4.14.

This behavior creates two problems. The reduced amplitude of the target component of the matched filter output reduces the SNR and thus the probability of detecting the target. Assuming the reduced-amplitude response does prove large enough to be detected, the processor will assume the target is at delay t_0 when in fact it is at $t_0 + T$. This phenomenon whereby there is more than one possible range that can be associated with a detection is called a range ambiguity. First discussed in Chap. 3, it is a characteristic of pulse burst waveforms. It is not readily apparent if a peak at the matched filter output is due to a target at the implied range or at that range plus a multiple of the unambiguous range $R_{ua} = cT/2$ meters.

As will be seen in <u>Chap. 5</u>, it is common in some radars to operate at a PRF for which the unambiguous range is less than the maximum detection range, so methods are needed to counter these two problems. Range ambiguities can be resolved using multiple pulse burst waveforms at different PRFs as discussed in <u>Chap. 5</u>. The reduction in matched filter output amplitude and SNR for range-ambiguous targets is countered by noting that it occurs because the pulse burst echo is not fully overlapped with the matched filter reference pulse burst at the output sampling time when the target time delay $t_0 > T$. The solution to this problem is to extend the transmitted waveform. Suppose the radar can be

expected to detect targets at ranges up to $P \cdot R_{ua}$. Extend the transmitted waveform from M to M + P - 1 pulses. The receiver matched filter remains the same M-pulse waveform. Still using $t_0 = 0$ and $T_M = 0$, the matched filter output will be the waveform shown in Fig. 4.16*a*. It indicates that the full integrated target power of $MA^2\tau = ME_p$ will be obtained for a target at the delay T_M (zero in the figure) for which the matched filter is tuned, but also for targets at the ambiguous delays $T_M + pT$, p = 0, ..., P - 1. Evaluating the matched filter output for delays T_M between zero and T allows full-SNR matched filtering of targets at delays up to $T_M + (P - 1)T$, corresponding to ranges up to $P \cdot R_{ua}$ as desired. The pulse-by-pulse viewpoint is shown in Fig. 4.16*b*. A target in the *P*th delay interval, $(P - 1)T \le t_0 < PT$, will produce a response in the appropriate range bin in the *P*th and later slow time samples. By integrating only samples *P* through M + P - 1 in each range bin the design integration gain of M can be achieved for all targets up to ranges $P \cdot R_{ua}$.



FIGURE 4.16 Effect of extending the transmitted waveform with additional pulses when P range ambiguities can occur: (*a*) whole-waveform matched filter output. (*b*) slow-time data in pulse-by-pulse processing viewpoint.

Another issue arises with pulse burst waveforms and clutter echoes.

Suppose the radar can be expected to receive significant clutter returns at ranges up to $P \cdot t_{ua}$, and consider the clutter component of the slow-time signal for a given range bin in the pulse-by-pulse processing viewpoint. When the range bin of interest is sampled at delay $t_0 < T$ after the first pulse is transmitted, only clutter echoes from the corresponding range $ct_0/2$ will be sampled at the receiver. When the range bin is sampled again after the second pulse, the clutter component will include echoes from the second pulse and range $ct_0/2$ as well as from the first pulse and range $c/2(t_0 + T)$. These two contributions represent echo from two physically different patches of clutter scatterers. The first slowtime sample, which includes echoes from only the nearer patch, may differ significantly in power and statistical behavior from the second slow-time sample, which includes echoes from both. The Pth and subsequent slow-time samples will contain contributions from all P contributing range intervals and therefore exhibit the consistent clutter power levels and statistical behavior needed for effective clutter filtering and target detection. Extending the transmitted waveform to M + P - 1 pulses as above therefore allows collection of M steady-state clutter measurements. In <u>Chap. 5</u> these additional pulses will be called "clutter fill" pulses. The first P - 1 slow-time samples will be discarded in each range bin and only the remaining M samples will be used in clutter filtering, coherent integration, and detection processing.

4.5.4 Doppler Response of the Pulse Burst Waveform

To consider the effect of a Doppler mismatch on the pulse burst waveform and its matched filter, consider a target moving toward the radar at velocity v meters per second so that its range is $R_0 - vt$ meters at time t. Assume that the "stopand-hop" approximation is valid and that the target motion does not exceed one range bin over the CPI, that is, $MvT < c\tau/2$; this ensures that all echoes from a given target appear in the same range bin over the course of a CPI. The demodulated echoes will have a phase shift of $-(4\pi/\lambda)R(t) = -(4\pi/\lambda)(R_0 - vt)$. Adopting the pulse-by-pulse processing viewpoint and absorbing the phase $\exp(-j4\pi R_0/\lambda)$ due to the nominal range R_0 into the overall gain, the individual matched filtered outputs for each pulse become

$$y_m(t) = x_p(t - mT) * h_p(t)$$

= $e^{j(4\pi v/\lambda)mT} s_p(-t + mT), \quad 0 \le m \le M - 1$
(4.67)

The corresponding slow-time sequence is

$$y[l, m] = y_m(mT) = e^{j(4\pi v/\lambda)mT} s_p(0), \qquad 0 \le m \le M - 1$$
$$= e^{j\omega_D m} E_p \qquad \qquad (\omega_D = 4\pi vT/\lambda)$$
(4.68)

Integrating the slow-time samples gives

$$\sum_{m=0}^{M-1} y[l, m] = E_p \sum_{m=0}^{M-1} e^{j\omega_D m} \equiv Y[l, \omega_D)$$
$$= E_p \frac{\sin[\omega_D M/2]}{\sin[\omega_D/2]} \exp\left[-j\left(\frac{M-1}{2}\right)\omega_D\right]$$
(4.69)

Equation (4.69) gives the system response to the pulse burst waveform in an arbitrary range bin *l* as a function of the normalized Doppler mismatch ω_D .

This is the familiar asinc function. Figure 4.17 shows the central portion of the magnitude of this function. The zeros occur at intervals of 1/M cycles per sample in normalized frequency; thus, the Rayleigh resolution in Doppler is 1/M cycles per sample or 1/MT Hz. MT is the duration of the entire pulse burst waveform. The Doppler resolution is therefore determined by the duration of the entire waveform instead of the duration of a single pulse. In this manner, the pulse burst waveform achieves much better Doppler resolution than a single pulse of the same duration while maintaining the same range resolution. The cost is the time and energy required to transmit and receive M pulses instead of one and the computational load of integrating M samples in each range bin.



FIGURE 4.17 Central portion of the Doppler mismatch response of the slow-time signal using a pulse burst waveform.

Integrating the slow-time samples of the pulse burst echo corresponds to implementing a matched filter in slow time for a signal with zero Doppler shift;

in this case the expected slow-time signal is simply a constant. A matched filter for a Doppler-shifted pulse burst can be implemented by continuing to use the single-pulse matched filter in fast time and constructing the appropriate slowtime matched filter for the signal expected for a given Doppler shift.

Suppose the normalized Doppler shift of interest is ω_D radians per sample. The expected slow-time signal is then of the form $A\exp(j\omega_D m)$. After conjugation and time-reversal the slow-time matched filter coefficients will be $h[m] = \exp(+j\omega_D m)$. Consider the response of this filter when the actual Doppler shift of the signal is ω . The matched filter peak output occurs when the impulse response and data sequence are fully overlapped, giving

$$Y[l, \omega; \omega_D) = \sum_{m=0}^{M-1} e^{-j\omega_D m} y[l, m]$$

= $E_p \sum_{m=0}^{M-1} e^{-j\omega_D m} e^{j\omega m} = E_p \sum_{m=0}^{M-1} e^{-j(\omega-\omega_D)m}$
= $E_p \frac{\sin[(\omega-\omega_D)M/2]}{\sin[(\omega-\omega_D)/2]} \exp\left[-j\left(\frac{M-1}{2}\right)(\omega-\omega_D)\right]$
(4.70)

which is identical to Eq. (4.69) except that the peak of the asinc function has been shifted to $\omega = \omega_D$ radians per sample.

Note that the first line of Eq. (4.70) is simply the discrete time Fourier transform of the slow-time data sequence. Thus, a matched filter for a pulse burst waveform and a Doppler shift of ω_D radians can be implemented with a single-pulse matched filter in fast time and a DTFT in slow time, evaluated at ω_D . If ω_D is a discrete Fourier transform frequency, i.e., of the form $2\pi k/K$ for some integers k and K, the slow-time matched filter can be implemented with a DFT calculation. It follows that a K-point DFT of the data y[l, m] in the slow-time dimension simultaneously computes the output of K-matched filters, one at each of the DFT frequencies. These frequencies correspond to Doppler shifts of $F_k = k/KT$ hertz or radial velocities $v_k = \lambda k/2KT$ meters per second, k = 0, ..., K - 1. The fast Fourier transform (FFT) algorithm then allows very efficient search of the data for targets at various Doppler shifts by simply applying an FFT to each slow-time row of the data matrix.

4.5.5 Ambiguity Function for the Pulse Burst Waveform

Inserting the definition of the pulse burst waveform of Eq. (4.54) into the definition of the complex ambiguity function of Eq. (4.30) gives

$$\hat{A}(t, F_D) = \int_{-\infty}^{\infty} \left(\sum_{m=0}^{M-1} x_p(s - mT) \right) \left(\sum_{n=0}^{M-1} x_p^*(s - t - nT) \right) e^{j2\pi F_D s} ds$$
$$= \sum_{m=0}^{M-1} \sum_{n=0}^{M-1} \int_{-\infty}^{\infty} x_p(s - mT) x_p^*(s - t - nT) e^{j2\pi F_D s} ds$$
(4.71)

Substitute s' = s - mT

$$\hat{A}(t, F_D) = \sum_{m=0}^{M-1} e^{j2\pi F_D mT} \sum_{n=0}^{M-1} \int_{-\infty}^{\infty} x_p(s') x_p^* (s' - t - nT + mT) e^{j2\pi F_D s'} ds'$$
(4.72)

If the complex ambiguity function of the single simple pulse $x_p(t)$ is denoted as $\hat{A}_p(t, F_D)$, the integral in Eq. (4.72) is $\hat{A}_p(t + (n - m)T, F_D)$. Thus

$$\hat{A}(t, F_D) = \sum_{m=0}^{M-1} e^{j2\pi F_D mT} \sum_{n=0}^{M-1} \hat{A}_p(t - (m - n)T, F_D)$$
(4.73)

The double sum in Eq. (4.73) is somewhat difficult to deal with. Obviously, all combinations of *m* and *n* having the same difference m - n result in the same summand in the second sum, but the dependence of the exponential term on *m* only prevents straightforward combining of all such terms. Defining n' = m - n, it can be shown by simply enumerating all of the combinations that the double summation of some function f[m, n] can be written (Rihaczek, 1996)

$$\sum_{m=0}^{M-1} \sum_{n=0}^{M-1} f[m, n] = \sum_{n'=-(M-1)}^{0} \sum_{m=0}^{M-|n'|-1} f[m, m-n'] + \sum_{n'=1}^{M-1} \sum_{m=0}^{M-|n'|-1} f[m+n', m]$$
(4.74)

Applying the decomposition of Eq. (4.74) to Eq. (4.73) gives

$$\hat{A}(t, F_D) = \sum_{n'=-(M-1)}^{0} \hat{A}_p(t - n'T, F_D) \sum_{m=0}^{M-|n'|-1} e^{j2\pi F_D mT} + \sum_{n'=1}^{M-1} e^{j2\pi F_D n'T} \hat{A}_p(t - n'T, F_D) \sum_{m=0}^{M-|n'|-1} e^{j2\pi F_D mT}$$

$$(4.75)$$

The geometric series that appears in both halves of the right-hand side of this equation sums to

$$\sum_{m=0}^{M-|n'|-1} e^{j2\pi F_D mT} = \exp[j\pi F_D (M-|n'|-1)T] \frac{\sin(\pi F_D (M-|n'|)T)}{\sin(\pi F_D T)}$$
(4.76)

Using this result in Eq. (4.75) and combining the two remaining sums over n' into one while renaming the index of summation as m gives

$$\hat{A}(t, F_D) = \sum_{m=-(M-1)}^{M-1} \hat{A}_p(t - mT, F_D) e^{j\pi F_D(M-1+m)T} \frac{\sin(\pi F_D(M-|m|)T)}{\sin(\pi F_D T)}$$
(4.77)

Equation (4.77) expresses the complex ambiguity function of the coherent pulse train in terms of the complex ambiguity function of its constituent simple pulses and the PRI.

Recall that the support in the delay axis of $\hat{A}_p(t, F_D)$ is $|t| \le \tau$. If $T > 2\tau$, which is almost always the case, the replications of \hat{A}_p in Eq. (4.77) will not overlap and the magnitude of the sum of the terms as *m* varies will be equal to the sum of the magnitude of the individual terms. The ambiguity function of the pulse burst waveform can then be written as

$$A(t, F_D) = \sum_{m=-(M-1)}^{M-1} A_p(t - mT, F_D) \left| \frac{\sin[\pi F_D(M - |m|)T]}{\sin(\pi F_D T)} \right| \qquad (T > 2\tau)$$
(4.78)

To understand this ambiguity function, it is convenient to first look at the zero Doppler and zero delay responses. The zero Doppler response is obtained by setting $F_D = 0$ in Eq. (4.78) and recalling that $A_p(t, 0) = 1 - |t|/\tau$:

$$A(t, 0) = \begin{cases} \sum_{m=-(m-1)}^{M-1} (M - |m|) \left(1 - \frac{|t - mT|}{\tau} \right), & |t - mT| < \tau \\ 0, & \text{elsewhere} \end{cases}$$
(4.79)

Equation (4.79) describes the triangular output of the single-pulse matched filter, repeated every *T* seconds and weighted by an overall triangular function M - |m|. Figure 4.18 illustrates this function for the case M = 5 and $T = 4\tau$. The ambiguity function has been normalized by the signal energy *E* so that it has a maximum value of 1.0. Note that, as with any waveform, the maximum of the AF occurs at t = 0 and the duration is twice the total waveform duration (2*MT* in this case). The local peaks every *T* seconds represent the range ambiguities discussed previously in Sec. 4.5.3 and illustrated in Fig. 4.15*a*. If the transmitted waveform were extended by *P* pulses while the reference waveform

remained M pulses long as discussed above, there would be P consecutive spikes with the full amplitude of 1.0, similar to Fig. 4.16a.



FIGURE 4.18 Zero-Doppler cut of the ambiguity function of a pulse burst. M = 5 pulses, $T = 4\tau$.

The zero delay cut is obtained by setting t = 0 in Eq. (4.78) and recalling that $A_p(0, F_D) = |\sin(\pi F_D \tau)/\pi F_D \tau|$ (assuming a unit energy simple pulse), giving

$$A(0, F_D) = \left| \frac{\sin(\pi F_D \tau)}{\pi F_D \tau} \right| \left| \frac{\sin(\pi F_D MT)}{\sin(\pi F_D T)} \right|$$
(4.80)

The response is an asinc function with a first zero at $F_D = 1/MT$ Hz, repeating with a period of 1/T Hz. This basic behavior is weighted by a more slowly varying true sinc function with its first zero at $1/\tau$ Hz. This structure is evident in Fig. 4.19, which shows a portion of the zero delay cut for the same case with M = 5 and $T = 4\tau$. The 1/T spacing of the principal peaks in the zero-delay response is the blind Doppler shift first defined in Chap. 3.



FIGURE 4.19 Zero-delay cut of the pulse burst ambiguity function with M = 5 and $T = 4\tau$.

Figure 4.20 is a contour plot of a portion of the complete ambiguity function for this waveform. Note the broadening of the response peaks in Doppler when sampling at the range-ambiguous delays such as 4τ and 8τ (corresponding to ± 0.2 and ± 0.4 on the normalized delay scale of the contour plot). This phenomenon is caused by the (M - |m|) term in the asinc term of Eq. (4.78) and reflects the fact that at these range-ambiguous delays fewer than M pulses are contributing to the matched filter local output peak. The reduced observation time results in degraded Doppler resolution. Again, if the transmitted waveform were extended by P pulses, there would be P consecutive range peaks which retained the full Doppler resolution. This plot also illustrates the breakup of the well-defined peaks in delay when the Doppler mismatch reaches $1 / \tau$ hertz (corresponding to 20 on the normalized Doppler scale of the plot).



FIGURE 4.20 A portion of the ambiguity function for the pulse burst waveform with M = 5 and $T = 4\tau$. Positive frequencies only shown.

Figure 4.21 is a diagram of the structure of the central peak of the pulse burst ambiguity function and the first repeated peaks in Doppler and range. This figure summarizes how the various waveform parameters determine the resolution in range and Doppler, the range ambiguity interval, and the blind Doppler interval. The individual pulse length τ is chosen to achieve the desired range resolution ($c\tau/2$ meters). The pulse repetition interval *T* sets the ambiguity interval in range (cT/2 meters) and the blind interval in Doppler (1/T Hz). Finally, once the PRI is chosen, the number of pulses in the burst determines the Doppler resolution (1/MT Hz).



FIGURE 4.21 Relationship between pulse burst waveform parameters and range and Doppler resolution and ambiguities.

4.5.6 The Slow-Time Spectrum and the Periodic Ambiguity Function

It would seem that the DTFT $Y[l, \omega_D)$ of the slow-time sequence y[l, m] should be related to the variation of the complex ambiguity function $\hat{A}(t, F_D)$ in Doppler. The slow-time sequence y[l, m] is obtained by sampling the output of the simple pulse matched filter output $s_p(t)$ at the same delay after transmission on each pulse. If the target motion across the CPI is small compared to the range resolution (i.e., the target moves only a small fraction of a range bin during the CPI), the amplitude of the sample taken on each pulse will be the same. This amplitude will be the maximum value $s_p(0)$ if the sampling time exactly corresponds to the target range; if the sampling time differs from that corresponding to the target range by Δt seconds the measured amplitude on each pulse will be $s_p(\Delta t)$. Thus, the slow-time sequence in a given range bin will have constant amplitude but the ambiguity function of the waveform will determine that amplitude based on the alignment of the target range and the range bin sampling times.

If there is relative motion between the radar and target, there will be a sample-to-sample decrease in the phase of the slow-time samples of the form – $4\pi mvT/\lambda$. If the target is within the first unambiguous range interval the target echo will be present in all *M* slow-time samples for the appropriate range bin and the magnitude of the DTFT will have the $|\sin(\pi F_D MT)/\sin(\pi F_D T)|$ form seen in Eq. (4.80). However, the $|\sin(\pi F_D \tau)/\pi F_D \tau|$ term of the AF Doppler response due to the individual pulse shape will not be observed in the DTFT; rather, this term will weight the overall amplitude of the DTFT. Finally, if the target range

exceeds R_{ua} the target echo will not be present in all of the slow-time samples and the Doppler resolution will degrade in $Y[l, \omega_D)$ in the same manner it did in $\hat{A}(t, F_D)$.

When the transmitted pulse burst waveform is extended by P pulses to provide full integration gain of a factor of M for targets extending over P range ambiguities, the matched filter output maintains its full maximum peak value of ME_p over the P range ambiguities (delay interval 0 to (P - 1)T) of interest as shown in Fig. 4.16*a*. The same result over only that delay interval could be obtained by at least two equivalent calculations: correlation of an M-pulse transmitted waveform with an infinitely extended reference, evaluated over [0, (P - 1)T], or circular correlation of an M-pulse waveform with an M-pulse reference. The *periodic ambiguity function* (PAF) is a modification of the complex AF of Eq. (4.30) that, when applied to a pulse burst waveform, produces the full-gain AF over this delay interval. A typical definition is (Levanon, 2010; Levanon and Mozeson, 2004)

$$PA(t, F_D) = \left| \int_0^{MT} x(s) \exp(j2\pi F_D s) x^*(s-t) \, ds \right|, \quad 0 \le t < P \cdot T$$
(4.81)

A significant property of the PAF is its relation to the AF of the single constituent pulse in the pulse burst when $T > 2\tau$:⁴

$$PA(t, F_D) = A_p(t, F_D) \left| \frac{\sin(\pi F_D MT)}{\sin(\pi F_D T)} \right|, \quad 0 \le t < P \cdot T$$

$$(4.82)$$

That is, the PAF is the AF of the single pulse multiplied by the DTFT of a discrete *M*-sample pulse. This is exactly the DTFT $Y[l, \omega_D)$ that will result from the pulse-by-pulse processing approach as described above.

4.6 Frequency-Modulated Pulse Compression Waveforms

A simple pulse has only two parameters, its amplitude A and its duration τ . The range resolution $c\tau/2$ is directly proportional to τ ; better resolution requires a shorter pulse. Most modern radars operate with the transmitter in saturation. That is, any time the pulse is on, its amplitude is kept at the maximum value of A; amplitude modulation other than on/off switching is not used. The energy in the pulse is then $A^2\tau$. This mode of operation maximizes the pulse energy, which is then also directly proportional to τ . As will be seen in Chaps. 6 and 7, increasing pulse energy improves detection and estimation performance. Thus, improving resolution requires a shorter pulse, while improving detection and

estimation performance requires a longer pulse. The two metrics are coupled in this unfortunate way because there is effectively only one free parameter τ in the design of the simple pulse waveform.

Pulse compression waveforms decouple energy and resolution. Recall that a simple pulse has a Rayleigh bandwidth $\beta = 1/\tau$ Hz and a Rayleigh resolution in time at the matched filter output of τ seconds. Thus, the *time-bandwidth product* (BT product) of the simple pulse is $\tau(1/\tau) = 1$. A pulse compression waveform, in contrast, has a bandwidth β that is much greater than $1/\tau$. Equivalently, it has a duration τ much greater than that of a simple pulse with the same bandwidth, $\tau \gg 1/\beta$. Either condition is equivalent to stating that a pulse compression waveform has a BT product $\beta\tau$ much greater than one.

Pulse compression waveforms are obtained by adding frequency or phase modulation to a simple pulse. There are a vast number of pulse compression waveforms in the literature. In this text, only the most commonly used types will be described. These include linear frequency modulation, biphase codes, and certain polyphase codes. Nonlinear FM will also be briefly introduced. Many other waveforms are described in Levanon and Mozeson (2004) and Keel and Baden (2012).

4.6.1 Linear Frequency Modulation

A linear frequency modulated waveform is defined by

$$x(t) = \cos\left(\pi\frac{\beta}{\tau}t^2\right), \quad 0 \le t \le \tau$$
(4.83)

The complex equivalent is

$$x(t) = e^{j\pi\beta t^2/\tau} = e^{j\theta(t)} \quad 0 \le t \le \tau$$
(4.84)

The instantaneous frequency in hertz of this waveform is the time derivative of the phase function

$$F_i(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} = \frac{\beta}{\tau} t \quad \text{hertz}$$
(4.85)

This function is shown in Fig. 4.22, assuming $\beta > 0$. $F_i(t)$ sweeps linearly across a total bandwidth of β Hz during the τ -second pulse duration. The waveform x(t)[Eq. (4.83), or the real part of Eq. (4.84)] is shown in Fig. 4.23 for $\beta \tau = 50$. The LFM waveform is often called a *chirp* waveform in analogy to the sound of an acoustic sinusoid with a linearly changing frequency. When β is positive the pulse is an *upchirp*; if β is negative it is a *downchirp*. The BT product of the LFM pulse is simply $\beta \tau$; $\beta \tau \gg 1$ if the LFM pulse is to qualify as a pulse compression waveform.



FIGURE 4.22 Instantaneous frequency of an LFM pulse.



FIGURE 4.23 Real-valued LFM upchirp waveform, BT product $\beta \tau = 50$.

Figure 4.24 shows the magnitude spectrum of the LFM waveform for a relatively low BT product case ($\beta \tau = 10$), and again for a higher BT product case ($\beta \tau = 100$). For low BT the spectrum is relatively poorly defined. As the BT product increases the spectrum takes on a more rectangular shape. This is intuitively reasonable: because the sweep is linear, the waveform spreads its energy uniformly across the spectrum.



FIGURE 4.24 Magnitude spectrum of an LFM waveform: (a) $\beta \tau = 10$, (b) $\beta \tau = 100$.

Figure 4.25 shows the output of the matched filter for the same two chirp waveforms. The dotted line superimposed on the output waveform is the output of a matched filter for a simple pulse of the same duration. As always, the total duration of the matched filter output is 2τ seconds. In both cases, the LFM waveform results in a matched filter output with a Rayleigh resolution much narrower than τ . In fact, the Rayleigh resolution is very nearly $1/\beta$ in each case (this will be confirmed shortly), an improvement over the simple pulse by a factor of the time-bandwidth product $\beta\tau$.



FIGURE 4.25 Output of matched filter for an LFM waveform: (a) $\beta \tau = 10$, (b) $\beta \tau = 100$. The dotted line is the output of a matched filter for a simple pulse of the same duration.

Simple and LFM pulses of the same amplitude and duration will have the same peak power at the matched filter output and achieve the same output SNR in accordance with Eq. (4.13). However, for an LFM pulse and a simple pulse of the same amplitude to have the same fast-time Rayleigh resolution at the output of their matched filters, the simple pulse must be shorter than the LFM pulse by the factor $\beta\tau$. The energy in the simple pulse and the SNR achieved are then also less by the factor of $\beta\tau$. In other words, the LFM waveform with

proper matched filtering achieves a signal processing gain of $G_{sp} = \beta \tau$ compared to a simple pulse of the same Rayleigh resolution.

Unlike the simple pulse case, the matched filter output for the LFM pulse exhibits a sidelobe structure. Figure 4.26 expands the central portion of Fig. 4.25*b*, showing the distinctly sinc-like mainlobe and first few sidelobes. This should not be surprising: the waveform spectrum X(F) (Fig. 4.24*b*) is approximately a rectangle of width β Hz. Consequently, the spectrum of the matched filter output, $|X(F)|^2$, will also be approximately a rectangle of width β . The time-domain output of the matched filter is therefore expected to be approximately a sinc function with a Rayleigh resolution of $1/\beta$ seconds.



FIGURE 4.26 Expanded view of central portion of Fig. 4.25b.

To summarize, the LFM waveform enables separate control of pulse energy (through its duration) and range resolution (through its swept bandwidth). The possibility of pulse compression is created by the use of matched filters. The output of the matched filter is not a replica of the transmitted waveform x(t), but of its autocorrelation function $s_x(t)$. Therefore, if a waveform can be designed that has a long duration but a narrowly concentrated autocorrelation, both good range resolution and good energy can be obtained simultaneously. This in turn is accomplished by modulating a long pulse to spread its bandwidth beyond the usual $1/\tau$. Since the spectrum of the autocorrelation function is just the squared magnitude of the waveform spectrum, a spectrum spread over β Hz will tend to produce a filter output with most of its energy concentrated in a mainlobe of about $1/\beta$ seconds duration. The linear FM pulse is the first example of such a waveform, but phase coded waveforms will provide more examples of this approach.

4.6.2 The Principle of Stationary Phase

The Fourier transform of Eq. (4.84) is a relatively complicated result involving the sine integral Si(*F*) (Rihaczek, 1996). A very useful and much simpler approximation can be derived using the *principle of stationary phase* (PSP), an advanced technique in Fourier analysis. The PSP is useful for approximate evaluation of integrals with highly oscillatory integrands; thus, it applies particularly well to Fourier transforms. Write x(t) in amplitude and phase form, $x(t) = A(t) \exp[j\theta(t)]$, and consider its Fourier transform

$$X(\Omega) = \int_{-\infty}^{+\infty} \underbrace{A(t)e^{j\theta(t)}}_{x(t)} e^{-j\Omega t} dt$$
(4.86)

Define the phase $\phi(t, \Omega)$ of the Fourier integral as the combination of the signal phase and the Fourier kernel phase

$$X(\Omega) = \int_{-\infty}^{+\infty} A(t)e^{j\theta(t)}e^{-j\Omega t} dt = \int_{-\infty}^{+\infty} A(t)e^{j[\theta(t)-\Omega t]} dt$$
$$\equiv \int_{-\infty}^{+\infty} A(t)e^{j\phi(t,\Omega)} dt$$
(4.87)

Of course, the exact Fourier transform is known for many signals having relatively simple phase functions $\tau(t)$. The PSP is most useful when the signal phase function and thus the total integral phase $\phi(t, \Omega)$ is continuous but nonlinear or otherwise complicated.

Define a *stationary point* of the integrand as a value of $t = t_0$ such that the first time derivative of the integral phase $\phi'(t_0, \Omega) = 0$. Then the PSP approximation to the spectrum is (Born and Wolf, 1959; Papoulis and Pillai, 2002; Raney, 1992)

$$X(\Omega) \approx \sqrt{\frac{-2\pi}{\phi^{\prime\prime}(t_0,\Omega)}} e^{-j\pi/4} A(t_0) e^{j\phi(t_0,\Omega)}$$

$$(4.88)$$

where $\phi'(t_0, \Omega)$ is the second time derivative of $\phi(t, \Omega)$ evaluated at $t = t_0$. If there are multiple stationary points the spectrum is the sum of such terms for each stationary point. Equation (4.88) states that the magnitude of the spectrum at a given frequency Ω is proportional to the amplitude of the signal envelope at the time that the stationary point occurs and, more importantly, is inversely proportional to the square root of the rate of change of the frequency $\phi'(t_0, \Omega)$ at that time. The PSP also implies that only the stationary points significantly influence $X(\Omega)$.

The PSP can be applied to estimate the spectrum of the LFM waveform.

The waveform is defined as

$$x(t) = A(t)e^{j\alpha t^{2}}, \quad A(t) = \begin{cases} 1, & -\tau / 2 \le t \le +\tau / 2\\ 0, & \text{otherwise} \end{cases}, \quad \alpha \equiv \pi \frac{\beta}{\tau}$$

$$(4.89)$$

Thus

$$X(\Omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\Omega t} dt = \int_{-\infty}^{+\infty} A(t)e^{j(\alpha t^2 - \Omega t)} dt$$
(4.90)

The integrand phase and its derivatives are then

$$\phi(t, \Omega) = \alpha t^{2} - \Omega t$$

$$\phi'(t, \Omega) = 2\alpha t - \Omega$$

$$\phi''(t, \Omega) = 2\alpha$$

(4.91)

The stationary points are found by setting $\phi'(t, \Omega) = 0$ and solving for *t*. In this case, there is only one stationary point:

$$0 = \phi'(t_0, \Omega) \Longrightarrow t_0 = \frac{\Omega}{2\alpha}$$
(4.92)

Inserting Eq. (4.92) into Eq. (4.88) gives

$$X(\Omega) \approx \sqrt{\frac{-2\pi}{\phi''(t_0, \Omega)}} e^{-j\pi/4} A(t_0) e^{j\phi(t_0, \Omega)}$$
$$= \sqrt{\frac{-2\pi}{2\alpha}} e^{-j\pi/4} A\left(\frac{\Omega}{2\alpha}\right) e^{j\left[\alpha(\Omega/2\alpha)^2 - \Omega(\Omega/2\alpha)\right]}$$
$$= j\sqrt{\frac{\pi}{\alpha}} e^{-j\pi/4} A\left(\frac{\Omega}{2\alpha}\right) e^{-j\Omega^2/4\alpha}$$
(4.93)

Recalling the finite support of the signal envelope A(t), the term $A(\Omega/2\alpha)$ becomes (using $\alpha = \pi \beta/\tau$)

$$A\left(\frac{\Omega}{2\alpha}\right) = \begin{cases} 1, & -\tau/2 \le \frac{\Omega}{2\alpha} \le +\tau/2 \\ 0, & \text{otherwise} \end{cases} \Rightarrow$$
$$A\left(\frac{\Omega}{2\alpha}\right) = \begin{cases} 1, & -2\pi\left(\frac{\beta}{2}\right) \le \Omega \le +2\pi\left(\frac{\beta}{2}\right) \\ 0, & \text{otherwise} \end{cases}$$
(4.94)

The final result is

$$X(\Omega) \approx j \sqrt{\frac{\pi}{\alpha}} e^{-j\pi/4} e^{-j\Omega^2/4\alpha}, \qquad -2\pi \left(\frac{\beta}{2}\right) \le \Omega \le +2\pi \left(\frac{\beta}{2}\right)$$
(4.95)

Figure 4.27 compares this approximation with the exact spectrum when $\beta \tau = 100$. Equation (4.95) estimates that $|X(\Omega)|$ is constant over the range $\pm \beta/2$ Hz and is zero outside of this range. This is both intuitively satisfying, since this is exactly the range over which the instantaneous frequency of the LFM pulse sweeps, and consistent with the increasingly rectangular shape of the exact spectrum observed in Fig. 4.23 as the BT product increases. The PSP result also gives an estimate of the phase of the spectrum which, like the temporal phase of the waveform x(t), is seen to be quadratic in frequency.



FIGURE 4.27 Comparison of actual magnitude spectrum and PSP approximation for an LFM pulse with $\beta \tau = 100$.

4.6.3 Ambiguity Function of the LFM Waveform

The ambiguity function of an LFM pulse can be obtained by direct calculation,

similar to the simple pulse, but with a good deal more tedium. An easier way is to introduce the "chirp property" of the ambiguity function and then apply it to the LFM case. Suppose that a waveform x(t) has an ambiguity function $A(t, F_D)$. Create a modified waveform x'(t) by modulating x(t) with a linear FM complex chirp and compute its complex ambiguity function

$$\begin{aligned} x'(t) &\equiv x(t)e^{j\pi\beta t^{2}/\tau} \\ \hat{A}'(t, F_{\rm D}) &= \int_{-\infty}^{\infty} x'(s)x'^{*}(s-t)e^{j2\pi F_{\rm D}s} \, ds \\ &= \int_{-\infty}^{\infty} x(s)e^{j\pi\beta s^{2}/\tau} x^{*}(s-t)e^{-j\pi\beta(s-t)^{2}/\tau}e^{j2\pi F_{\rm D}s} \, ds \\ &= e^{-j\pi\beta t^{2}/\tau} \int_{-\infty}^{\infty} x(s)x^{*}(s-t)e^{j2\pi(F_{\rm D}+\beta t/\tau)s} \, ds \\ &= e^{-j\pi\beta t^{2}/\tau} \hat{A}\bigg(t, F_{\rm D} + \frac{\beta}{\tau}t\bigg) \end{aligned}$$

$$(4.96)$$

Taking the magnitude of $\hat{A}'(t, F_D)$ gives the ambiguity function of the chirp signal in terms of the ambiguity function of the original signal without the chirp

$$A'(t, F_D) = A\left(t, F_D + \frac{\beta}{\tau}t\right)$$
(4.97)

Equation (4.97) states that adding a chirp modulation to a signal skews its ambiguity function in the delay-Doppler plane. Applying this property to the simple pulse AF [Eq. (4.51)] gives the AF of the LFM waveform

$$A(t, F_D) = \left(1 - \frac{|t|}{\tau}\right) \frac{\sin[\pi (F_D + \beta t/\tau)\tau (1 - |t|/\tau)]}{\pi (F_D + \beta t/\tau)\tau (1 - |t|/\tau)}$$
$$= \left(1 - \frac{|t|}{\tau}\right) \frac{\sin[\pi (F_D\tau + \beta t)(1 - |t|/\tau)]}{\pi (F_D\tau + \beta t)(1 - |t|/\tau)}, \quad -\tau \le t \le \tau$$

$$(4.98)$$

Figure 4.28 is a contour plot of the AF of an LFM pulse of duration $\tau = 10 \ \mu s$ and swept bandwidth $\beta = 1$ MHz; thus, the BT product is 10. The AF retains the triangular ridge of the simple pulse but is now skewed in the delay-Doppler plane as predicted by Eq. (4.97).



FIGURE 4.28 Contour plot of the ambiguity function of an LFM waveform with $\beta \tau = 10$.

The zero-Doppler cut of the LFM ambiguity function is the matched filter output when there is no Doppler mismatch:

$$A(t,0) = \left| \frac{\sin[\pi\beta t(1-|t|/\tau)]}{\pi\beta t} \right|, \quad -\tau \le t \le \tau$$

$$(4.99)$$

This function was illustrated for BT products of both 10 and 100 in Fig. 4.25. The Rayleigh resolution of the LFM pulse is obtained by examination of Eq. (4.99). The peak of A(t, 0) occurs at t = 0. The first zero occurs when the argument of the numerator equals π , which occurs when $\beta t(1-|t|/\tau) = 1$. For positive *t*, this becomes

$$\beta t - \frac{\beta t^2}{\tau} = 1 \qquad \Rightarrow \qquad t^2 - \tau t + \tau / \beta = 0 \tag{4.100}$$

The roots of this equation are $t = (\tau \pm \sqrt{\tau^2 - 4\tau/\beta})/2 = \tau (1 \pm \sqrt{1 - 4/\beta\tau})/2$. Since the argument of the square root in the last expression is less than one, taking the negative sign gives the positive root closest to zero and thus the Rayleigh resolution in time. This result can be simplified with the following series

expansion of the square root

$$\sqrt{1-x} = 1 - \frac{x}{2} - \frac{x^2}{8} - \dots \approx 1 - \frac{x}{2} \quad (x \ll 1) \qquad \Rightarrow$$
$$t \approx \frac{\tau}{2} \left[1 - \left(1 - \frac{2}{\beta \tau} \right) \right] \approx \frac{1}{\beta} \quad (\beta \tau \gg 1)$$
(4.101)

Thus, the Rayleigh resolution in time is approximately $1/\beta$ seconds, corresponding to a Rayleigh range resolution ΔR of

$$\Delta R = \frac{c}{2\beta} \quad m \tag{4.102}$$

The zero-delay response is

$$A(0, F_D) = \left| \frac{\sin(\pi F_D \tau)}{\pi F_D \tau} \right|$$
(4.103)

which is simply a standard sinc function. The Doppler resolution of the LFM pulse is the same as that of a simple pulse, namely

$$\Delta F_D = \frac{1}{\tau} \quad \text{Hz} \tag{4.104}$$

Equation (4.103) shows that, like the simple pulse, the Doppler resolution of an LFM pulse is inversely proportional to the pulse length. Furthermore, the energy in the LFM pulse is still $A^2\tau$, directly proportional to the pulse length. Equation (4.102) shows that, *un*like the simple pulse, the range resolution is inversely proportional to the swept bandwidth. The LFM waveform has two parameters, bandwidth and duration, which can now be used to independently control pulse energy and range resolution. The pulse length is chosen (along with the pulse amplitude *A*) to set the desired energy, while the swept bandwidth is chosen to obtain the desired range resolution.

The expression $c/2\beta$ for range resolution is quite general. For instance, the Rayleigh bandwidth of a simple pulse is $\beta = 1/\tau$ Hz; using this in $c/2\beta$ gives $\Delta R = c\tau/2$ as before. While bandwidth and pulse length are directly related in the simple pulse, modulation of the LFM waveform has decoupled them. If $\beta \tau > 1$ for the LFM pulse the range resolution will be better than that of a simple pulse of the same duration by the factor $\beta \tau$. Alternatively, the range resolution of a simple pulse of length τ can be matched by an LFM pulse that is longer (and thus

higher energy, given the same transmitted power) by the factor $\beta\tau$.

4.6.4 Range-Doppler Coupling

The skew in the ambiguity function for the LFM pulse gives rise to an interesting phenomenon. Consider the AF of Eq. (4.98). The peak of this sinc-like function will occur when

$$F_D + \frac{\beta}{\tau}t = 0 \qquad \Rightarrow \qquad t = -\frac{\tau F_D}{\beta}$$
(4.105)

That is, when there is a Doppler mismatch the peak of the matched filter output will not occur at t = 0 as desired. Instead, it will be shifted by an amount proportional to the Doppler shift. Because the target range will be estimated based on the time of occurrence of this peak, a Doppler mismatch will induce an error in measuring range. The corresponding range error will be

$$\delta R = -\frac{c\tau F_D}{2\beta} \tag{4.106}$$

The amplitude of the peak will also be reduced by the factor $(1 - |t|/\tau) = (1 - F_D/\beta)$. Figure 4.29 illustrates the skewed ridge of the LFM ambiguity function and the relationship between Doppler shift and range measurement error.



FIGURE 4.29 Illustration of the effect of range-Doppler coupling on apparent target range.

While an incorrect range measurement is certainly undesirable, range-Doppler coupling is a useful phenomenon in some systems. A simple pulse with duration τ will have a Doppler Rayleigh resolution of $1/\tau$ Hz; targets with Doppler mismatches approaching this value or larger will produce a greatly attenuated output from the matched filter and will likely go undetected. An LFM pulse of the same duration will still produce a significant output peak for a much broader range of Doppler shifts, even though the peak will be mislocated in range. Nonetheless, the target will be more likely to be detected. The LFM waveform is said to be more *Doppler tolerant* than the simple pulse. This makes it a good choice for surveillance applications because a relatively large range of Doppler shifts can be searched with an LFM pulse. The range error can be eliminated, at least for isolated targets, by repeating the measurements with an LFM pulse of the opposite slope, e.g., an upchirp followed by a downchirp. In this case the sign of the range error will be reversed. Averaging the two measurements will give the true range and also allow determination of the Doppler shift.

4.6.5 Stretch Processing

LFM waveforms are often the waveform of choice for exceptionally wideband radar systems where the swept bandwidth β may be hundreds of megahertz or even exceed 1 GHz. Digital processing can be difficult to implement in such systems because the high instantaneous bandwidth of the waveform requires equally high sampling rates in the A/D converter. It is difficult to obtain highquality A/D converters at these rates with wordlengths longer than perhaps 8 bits with current technology; wordlengths at 1 GHz are expected to reach only about 11 bits by 2020 (Jonsson, 2010). In addition, the sheer number of samples generated can be stressing for the signal processor.

Stretch processing is a specialized technique for matched filtering of wideband LFM waveforms. It is also called *deramp processing, deramp on receive, dechirp*, and *one-pass processing*. It is essentially the same as the processing used with linear *frequency-modulated continuous wave* (FMCW) radar. Stretch processing is most appropriate for applications seeking very fine range resolution over relatively short range intervals (called *range windows* or *range swaths*).

Figure 4.30 shows the scenario for analyzing stretch processing. The *central reference point* (CRP) is in the middle of the range window of interest at a range of R_0 meters, corresponding to a time delay of t_0 seconds. Consider a scatterer at range R_b and time delay $t_b = t_0 + \delta t_b$. The problem will be analyzed in terms of differential range or delay relative to the CRP, denoted δR_b and δt_b . The transmitted waveform is the LFM pulse of Eq. (4.84). The echo from the scatterer, with the carrier frequency included, is



FIGURE 4.30 Scenario for stretch processing analysis.

$$\overline{x}(t) = \rho \exp\left[j\pi \frac{\beta}{\tau} (t - t_b)^2\right] \exp[j\Omega(t - t_b)], \quad 0 \le t - t_b \le \tau$$
(4.107)

where ρ is proportional to the scatterer reflectivity. This echo is processed with the modified coherent receiver in complex equivalent form shown in Fig. 4.31. The unique aspects of this stretch receiver are the reference oscillator and the Fourier transform. The oscillator contains a conventional term $\exp(-j\Omega t)$ to remove the carrier. However, it also contains a replica of the transmitted chirp, referenced to the time delay t_0 corresponding to the CRP. The reason for the Fourier transform will be apparent shortly.



FIGURE 4.31 Complex equivalent receiver for stretch processor.

After some algebra and using $t_b = t_0 + \delta t_b$, the output y(t) can be expressed s

$$y(t) = \rho \exp\left(-j\frac{4\pi R_b}{\lambda}\right) \exp\left[-j2\pi \frac{\beta}{\tau} \delta t_b(t-t_0)\right] \exp\left[j\pi \frac{\beta}{\tau} (\delta t_b)^2\right], \quad t_0 \le t - \delta t_b \le t_0 + \tau$$
(4.108)

The phase term that is quadratic in δt_b is a complex constant. In synthetic aperture imaging it is called the *residual video phase* (RVP). The middle complex exponential contains a term that is linear in *t* and therefore represents a constant-frequency complex sinusoid. By inspection, the sinusoid frequency is $F_b = -\beta \delta t_b / \tau$ Hz. F_b is proportional to δt_b and thus to the range of the scatterer relative to the CRP. The differential range can be obtained from the mixer output frequency as

$$\delta R_b = -\frac{cF_b\tau}{2\beta} \tag{4.109}$$

Heuristically, the scatterer produces a constant frequency tone at the output of the stretch receiver because the receiver not only removes the carrier from the LFM echo but also combines it in a mixer with a replica of the LFM with a delay corresponding to the CRP. In the conventional real-signal receiver of Fig. 1.13, the mixer produces sum and difference "beat" frequencies. The sum frequency is removed by a lowpass filter. (This LPF is not needed in the complex representation and is therefore not shown in Fig. 4.31.) The difference frequency is the difference between the instantaneous frequency of the LFM echo and the LFM reference. Since both have the same sweep rate, this beat frequency is a constant.

If there are several scatterers distributed at ranges R_i and delays δt_i , the stretch receiver output is simply the superposition of several terms of the form of Eq. (4.108)

$$y(t) = \sum_{i} \rho_{i} \exp\left(-j\frac{4\pi R_{i}}{\lambda}\right) \exp\left[-j2\pi\frac{\beta}{\tau}\delta t_{i}(t-t_{0})\right] \exp\left[j\pi\frac{\beta}{\tau}(\delta t_{i})^{2}\right]$$
(4.110)

Thus the output of the stretch receiver contains a different beat frequency tone for each scatterer. The reason for the Fourier transform block in Fig. 4.31 is now apparent. Spectral analysis of y(t) can identify the beat frequencies present in the mixer output and therefore the ranges and amplitudes of the scatterers present in the composite echo. Figure 4.32 illustrates the instantaneous frequencies and timing of the signals involved for three scatterers, one in the middle and one at each edge of the scene.



FIGURE 4.32 Instantaneous frequency vs. time for an LFM transmitted pulse and echoes from three scatterers. See text for details.

It is desirable that the reference LFM chirp completely overlap the echo from a scatterer anywhere within the range window. If the range window is $R_w = cT_w/2$ meters long the leading edge of the echo from a scatterer at the nearest range, $R_0 - R_w/2$, will arrive $t_0 - T_w/2$ seconds after transmission as shown in Fig. 4.32. The trailing edge of the echo from the scatterer at the far limit of the range window, $R_0 + R_w/2$, will arrive $t_0 + T_w/2 + \tau$ seconds after transmission. Thus, data from the range window have a total duration of $T_w + \tau$ seconds. To ensure complete overlap of the reference chirp with echoes from any part of the range window the reference chirp must be $T_w + \tau$ seconds long and so will sweep over $(1 + T_w/\tau)\beta$ Hz.

Another issue evident in Fig. 4.32 is *range skew*. This is the phenomenon whereby the beat frequencies for scatterers at different ranges, while all of the same duration (provided the reference chirp is lengthened), start and stop at different times. This complicates weighting of the mixer output prior to spectral analysis for range sidelobe control. If the window is aligned with the beat frequency for the center scatterer response, it will be misaligned with earlier and later scatterer responses. If it is lengthened to cover the full mixer output duration of $T_w + \tau$ seconds, none of the beat frequencies will be weighted by the full window and each will have a different effective window function. In either case, sidelobe suppression will be poor.

This problem can be solved by placing an additional filter between the mixer output and the Fourier transform. Notice that the scatterer at delay δt_b relative to the patch center generates a beat frequency of $-\beta \cdot \delta t_b/\tau$ Hz. What is needed is a filter whose frequency response has unit magnitude for all

frequencies so as not to distort the scatterer amplitudes, but also a group delay⁵ of $-\delta t_b$ seconds at the frequency $-\beta \cdot \delta t_b/\tau$. It can be shown that the required frequency response in analog radian frequency units is $H(\Omega) = \exp(-j\Omega^2 \tau/2\beta_{\Omega})$, which has unit magnitude for all frequencies and a quadratic phase in the frequency domain (see Prob. 15). All of the beat frequencies will be aligned in time at the output of this filter. As an extra benefit, this filter also corrects RVP (Carrara et al., 1995).

The bandwidth of the stretch receiver output can be obtained by considering the difference in beat frequencies for scatterers at the near and far edges of the range window. This gives

$$F_{\text{near}} - F_{\text{far}} = \left[-\frac{\beta}{\tau} \left(-\frac{T_w}{2} \right) \right] - \left[-\frac{\beta}{\tau} \left(+\frac{T_w}{2} \right) \right]$$
$$= \frac{T_w}{\tau} \beta$$
(4.111)

If $T_w < \tau$, the bandwidth at the receiver output is less than the original signal bandwidth β . The mixer output can then be sampled with slower A/D converters and the number of range samples needed to represent the range window data is reduced. Thus, the stretch technique is most effective for systems performing fine range resolution analysis over limited range windows. Also note that while the digital processing rates have been reduced, the analog receiver hardware up through the LFM mixer must still be capable of handling the full instantaneous signal bandwidth.

As an example, consider a 100 µs pulse with a swept bandwidth of 750 MHz, giving a BT product of 75,000. Suppose the desired range window is $R_w = 1.5$ km, corresponding to a sampling window of $T_w = 10$ µs. In a conventional receiver the sampling rate will be 750 megasamples per second. Data from scatterers over the extent of the range window will extend over $T_w + \tau$ seconds, requiring (750 MHz)(10 µs + 100 µs) = 82,500 samples to represent the range window. In contrast, the bandwidth at the output of the stretch receiver will be $(T_w/\tau)\beta = 75$ MHz. The sampled time interval remains the same, so only 8250 samples are required. Restricting the analysis to a delay window one-tenth the length of the pulse and using the stretch technique has resulted in a factor of 10 reduction in both the sampling rate required and the number of samples to be digitally processed.

Stretch processing of linear FM waveforms preserves both the resolution and the range-Doppler coupling properties of conventionally processed LFM. Consider the output of the stretch mixer for a scatterer at differential range δt_b from the central reference point. This signal will be a complex sinusoid at a frequency $F_b = -\beta \cdot \delta t_b / \tau$ Hz observed for a duration of τ seconds. In the absence of windowing, the Fourier transform of this signal will be a sinc function with its peak at F_b and a Rayleigh resolution of $1/\tau$ Hz. The processor will be able to resolve scatterers whose beat frequencies are at least $\Delta F_b = 1/\tau$ Hz apart. The time-delay spacing that gives this frequency separation satisfies

$$\frac{1}{\tau} = \left| \frac{\beta}{\tau} \delta t_b \right| \implies \delta t_b = \frac{1}{\beta}$$
(4.112)

The corresponding range separation is then the usual result for range resolution

$$\delta R_b = \frac{c}{2} \delta t_b = \frac{c}{2\beta}$$
(4.113)

If the reference oscillator sweep is not lengthened as discussed above to fully overlap the echo from scatterers at any location in the range window, the range resolution will be degraded. Specifically, the duration τ' of the beat frequency will be less than τ seconds for scatterers at any delay other than the center of the window due to the incomplete overlap. The Rayleigh resolution of the Fourier transform of that scatterer's beat frequency will increase to a value $1/\tau' > 1/\tau$, causing the range resolution of Eq. (4.113) to increase proportionately. The processing gain will similarly be reduced from the ideal factor of $\beta\tau$.

To consider the effect of Doppler shift on the stretch processor, replace $\overline{x}(t)$ in Eq. (4.107) with

$$\overline{x}(t) = \rho \exp\left[j\pi \frac{\beta}{\tau} (t - t_b)^2\right] \exp(j2\pi F_D t) \exp[j\Omega(t - t_b)], \quad 0 \le t - t_b \le \tau$$
(4.114)

Repeating the previous analysis, Eq. (4.108) becomes

$$y(t) = \rho \exp\left(-j\frac{4\pi R_b}{\lambda}\right) \exp\left[-j2\pi\left(\frac{\beta}{\tau}\delta t_b - F_D\right)t - 2\pi\frac{\beta}{\tau}\delta t_b t_0\right] \exp\left[j\pi\frac{\beta}{\tau}(\delta t_b)^2\right]$$
(4.115)

Equation (4.115) shows that the effect of a Doppler shift is to increase the beat frequency F_b by F_D Hz. Since beat frequency is mapped to differential range by the stretch processor according to $\delta R_b = -cF_b\tau/2\beta$, this implies a measured range shift of

$$\delta R = -\frac{c\tau}{2\beta} F_D \quad \mathrm{m} \tag{4.116}$$

which is the same range-Doppler coupling relationship obtained previously.

Stretch processing and especially Eq. (4.110) will be revisited and extended in <u>Chap. 8</u>, where the technique is central to the polar format algorithm for spotlight synthetic aperture radar imaging. Additional details of stretch processing are given in Keel and Baden (2012).

4.7 Range Sidelobe Control for FM Waveforms

It was seen in the previous section that the output of the LFM matched filter exhibits sidelobes in range (equivalently, delay). These are a consequence of the approximately rectangular LFM matched filter output spectrum, which produces a sinc-like range response. The first range sidelobe is approximately 13 dB below the output peak for moderate-to-high BT products, and about -15 dB for small BT products. Sidelobes this large are unacceptable in many systems that will encounter multiple targets in range due to *target masking*. This phenomenon is shown in Fig. 4.33*a*, where the smaller target is barely visible above the sidelobes of the stronger target despite being separated by approximately sixteen times the Rayleigh resolution. The smaller target could not be reliably detected in this scenario. If the sidelobes could be reduced, this masking effect could be greatly reduced as shown in part *b* of the figure.



FIGURE 4.33 Effect of windowing on target masking: (a) no windowing, (b) Hamming window applied.

For a simple pulse, the matched filter output is a triangle function, which exhibits no sidelobes. Thus, sidelobe reduction is not an issue for that waveform and it will not be further discussed. For the LFM pulse, there are two basic approaches to delay sidelobe reduction: shaping the receiver frequency response, and shaping the waveform spectrum.

4.7.1 Matched Filter Frequency Response Shaping

Recall from finite impulse response (FIR) digital filter design that to reduce

sidelobes of the frequency response of a digital filter, a window function is applied in the time domain to the impulse response. The goal here is to reduce sidelobes in range, corresponding to the time domain, so the analogous approach is to window the receiver frequency response in the frequency domain.

The matched filter frequency response is $H(F) = X^*(F)$, and at least for larger BT products is approximately rectangular. A modified frequency response H'(F) can be obtained by multiplying H(F) by a window function $w(F)^{6}$

$$H'(F) = w(F)H(F) = w(F)X^{*}(F)$$
(4.117)

Figure 4.34*a* shows a Hamming window function overlaid on the matched filter frequency response for an LFM waveform with $\beta \tau = 100$. *H'*(*F*) is the product of these two functions. The resulting impulse response *h'*(*t*) is shown in Fig. 4.34*b*. The response of both the matched filter and the filter of Fig. 4.34*b* to the unwindowed LFM echo are overlaid in Fig. 4.35. The LFM waveform used had $\beta \tau = 100$ and the window cutoff was placed at $\pm \beta/2$ Hz. The peak sidelobe has dropped 23.7 dB, from 13.5 dB below the mainlobe peak in the unwindowed case. This comes at a cost of the mainlobe peak gain dropping 5.35 dB and the Rayleigh time (range) resolution increasing by 93 percent.



FIGURE 4.34 Hamming weighting of the LFM receiver frequency response: (*a*) Hamming window overlaid on matched filter frequency response, (*b*) resulting filter impulse response h'(t).



FIGURE 4.35 Comparison of the receiver filter output with (black curve) and without (gray curve) frequency-domain Hamming weighting of the matched filter. See text for details.

Since the matched frequency response does not have a perfectly sharp cutoff frequency, there is some uncertainty as to where in frequency to place the window cutoff. In the example, the support of the window equals the instantaneous frequency cutoff, which is 0.36 cycles per sample on the normalized frequency scale for this particular sampled LFM waveform. However, this choice cuts off some of the waveform energy in the sidelobes, increasing the mismatched filtering losses. A case could be made for a narrower support so that the window is applied only over the relatively flat portion of the spectrum. This would provide range sidelobes more closely matching those expected for the chosen window but would reduce the effective bandwidth, further degrading the range resolution. A case could also be made for increasing the support to maximize the output energy, but this choice might increase the range sidelobes by "wasting" some of the window shape on the skirts of the LFM spectrum.

Since $H'(F) \neq X^*(F)$, the modified receiver is not matched to the transmitted LFM pulse and therefore the output peak and SNR will be reduced from their maximum values. This effect was evident in Fig. 4.33*b*, where the peak of the dominant target response is several dB lower than the unwindowed case in part *a* of the figure. The losses in output peak amplitude and SNR can be estimated from the window function w(F). In practice, a discrete window w[k] will be applied to a discrete-frequency version of H(F), H[k]. The loss in the peak signal output from the matched filter, called the *loss in processing gain* (LPG), is

$$LPG = \frac{K^2}{\left|\sum_{k=0}^{K-1} w[k]\right|^2}$$
(4.118)

where *K* is the window length. The loss in SNR at the matched filter output is called the *processing loss* (PL) and is

$$PL = \frac{K \sum_{k=0}^{K-1} |w[k]|^2}{\left|\sum_{k=0}^{K-1} w[k]\right|^2}$$
(4.119)

With these definitions *LPG* and *PL* are both greater than one so that the losses in decibels are positive numbers. For a relatively long Hamming window, the LPG is approximately 5.4 dB while the PL is approximately 1.4 dB. Both are weak functions of K and are slightly larger for small K. These formulas are approximate when applied to windowing of the LFM spectrum due to the finite-width transition of the LFM spectrum and the designer's discretion in choosing the cutoff of the window in frequency. In the example above, the LPG is 5.35 dB. Derivation of these formulas is deferred to Chap. 5, where they will arise again in the context of Doppler processing and where the results will be exact.

4.7.2 Matched Filter Impulse Response Shaping

The impulse response of the filter just obtained and illustrated in Fig. 4.34*b* suggests that similar results could have been obtained by windowing the LFM waveform in the time domain. Consider again the signal in Eq. (4.89) having an arbitrary amplitude function A(t) and a quadratic phase function. The PSP approximation to its spectrum was given in Eq. (4.93). The magnitude of the spectrum is proportional to the *time*-domain amplitude:

$$|X(\Omega)| \propto \left| A\left(\frac{\Omega}{2\alpha}\right) \right|$$
(4.120)

If A(t) has finite support on $-\tau/2 \le t \le \tau/2$, it follows that $X(\Omega)$ will have finite support on $-\beta/2 \le F \le \beta/2$ and in that interval $|X(\Omega)|$ has the same shape as the window magnitude |A(t)|. Thus, a Hamming-shaped (for example) spectrum can be obtained by applying a Hamming window to the impulse response h(t)instead of the frequency response H(F). Note that this result is specific to the use of a linear FM waveform.

The output of the resulting filter is overlaid on the matched filter response in <u>Fig. 4.36</u>. It has the same general character as the frequency-domain
weighting result but with some differences in details of the sidelobe structure. The peak is reduced from 60 to 54.64 dB with weighting, a nearly identical LPG of 5.36 dB. The peak sidelobe of the weighted response is 40.7 dB below the corresponding mainlobe peak, 3.5 dB better than the frequency-domain case. The Rayleigh width has increased 97 versus 93 percent in the frequency domain weighted case. This is consistent with the better sidelobe performance of the time-domain case. Additional detail on frequency- and time-domain weighting of LFM waveforms is available in Richards (2006).



FIGURE 4.36 Comparison of the receiver filter output with (black curve) and without (gray curve) time-domain Hamming weighting of the matched filter. See text for details.

4.7.3 Waveform Spectrum Shaping

The principal limitation of the receiver weighting approach to range sidelobe control is that the resulting filter is not matched to the transmitted waveform, resulting in an SNR loss. An alternative approach is to design a modified pulse compression waveform whose matched filter output inherently has lower sidelobes than the standard LFM. The waveform should be designed to have a spectrum shaped like that of a window function with the desired sidelobe behavior. Such a waveform would combine the maximized SNR of a truly matched filter with low sidelobes. There are two common ways to shape the spectrum. Both start with the idea that the LFM spectrum's relatively square shape is the result of a linear sweep rate combined with a constant pulse amplitude, resulting in a fairly uniform distribution of the signal energy across the spectral bandwidth. The spectral energy could be reduced at the edges, giving a "window-shaped" spectrum, by reducing the signal amplitude at the pulse edges while maintaining a constant sweep rate, by using a faster sweep rate at the edges with a constant pulse amplitude so as to spend less time in each spectral interval near the band edges, or both. The technique using variable sweep rates is referred to as *nonlinear FM* (NLFM).

The amplitude modulation technique implies operating the power amplifier at less than full power over the pulse length. This requires more complicated transmitter control but, more importantly, results in a pulse with less than the maximum possible energy for the given pulse length. This technique is not discussed further in this book; see Levanon and Mozeson (2004) for more information.

Two methods that have been proposed for NLFM waveform design are the principle of stationary phase method and empirical techniques. The PSP technique is used to design a temporal phase function from a prototype spectral amplitude function; the instantaneous frequency function is then obtained from the temporal phase. Examples of using this technique for deriving NLFM waveforms from common window functions such as Hamming or Taylor functions are given in Keel and Baden (2012).

One empirically developed design gives the instantaneous frequency function as (Price, 1979)

$$F_{i}(t) = \frac{t}{\tau} \left(\beta_{L} + \beta_{C} \frac{1}{\sqrt{1 - 4t^{2}/\tau^{2}}} \right), \quad |t| \le \frac{\tau}{2}$$
(4.121)

The term $\beta_L t/\tau$ represents a linear FM component, while the term involving β_C is designed to achieve a result that approximates a Chebyshev-shaped (constant sidelobe level) spectrum. Since $F_i(t) = (1/2\pi)(d\theta(t)/dt)$, integrating and scaling this instantaneous frequency function gives the required phase modulation

$$\theta(t) = \frac{\pi\beta_L}{\tau} t^2 - \frac{\pi\beta_C \tau}{2} \sqrt{1 - 4t^2/\tau^2}, \quad |t| \le \frac{\tau}{2}$$
(4.122)

Figure 4.37 illustrates the behavior of the resulting nonlinear FM waveform for the case where $\beta_L \tau = 50$ and $\beta_C \tau = 20$. The waveform is sampled at 10 times the bandwidth of the linear term, $T_s = 1/10\beta_L$. The instantaneous frequency (part *a* of the figure) is nearly linear in the center of the pulse but sweeps much more rapidly near the pulse edges. This reduces the spectral density at the pulse edge, resulting in the spectrum shown in part *c*, which has a window-like tapered shape instead of the usual nearly square LFM spectrum. The resulting matched filter output, shown in part *d*, has most of its sidelobes between -48 and -51 dB with the first sidelobe at -29 dB. In contrast, Fig. 4.38 illustrates the spectrum of the same waveform with β_c set to zero. This results in a linear FM waveform with the usual nearly square spectrum.

are on the same normalized frequency scale. Comparing the spectra of these two waveforms illustrates how the nonlinear term has spread and tapered the LFM spectrum to lower the matched filter sidelobes. The LFM matched filter output has a peak sidelobe of -13.5 dB, decaying approximately as 1/F at higher frequencies. The Rayleigh resolution in time of the NLFM waveform is approximately $0.8/\beta_L$, less than the $1/\beta_L$ value observed for the LFM case but greater than $1/(\beta_L + \beta_C)$.





FIGURE 4.38 FM waveform having same linear component as that of Fig. 4.33, but no nonlinear component: (*a*) magnitude of Fourier spectrum, (*b*) magnitude of matched filter output.

An example of a hybrid technique that combines a similar frequency modulation function with amplitude tapering of the matched filter impulse response is described in De Witte and Griffiths (2004). It is claimed there that the far sidelobes are controlled primarily by the maximum instantaneous frequency, while the near-in sidelobes are controlled by the amplitude weighting.

In addition to the more difficult phase control required, the major drawback of nonlinear FM pulses is their Doppler intolerance. Figure 4.39 shows the matched filter output for the waveform of Fig. 4.37 when a Doppler mismatch of $7/\tau$ Hz is present. While the general sidelobe level remains largely unchanged, the mainlobe is seriously degraded, exhibiting both range-Doppler coupling (a shift of the peak) and severe spreading and ambiguity caused by very high near-in sidelobes. The major advantage of NLFM over linear FM with receiver weighting is that the receiver filter for the NLFM waveform is a matched filter so that lower sidelobes are achieved with is no reduction of the matched filter output peak.



FIGURE 4.39 Output of NLFM matched filter when $F_D = 7/\tau$ Hz.

4.8 The Stepped Frequency Waveform

The LFM waveform increases resolution well beyond that of a simple pulse by sweeping the instantaneous frequency over the desired range β within the pulse. This technique is very effective and very common but does have drawbacks in some systems, particularly those using very large bandwidths on the order of hundreds of megahertz or more. First, the transmitter hardware must be capable of generating the LFM sweep. Second, all of the analog components must be able to support an instantaneous bandwidth of β Hz without introducing distortion. Even if stretch processing is used, the same is true of the receiver components up to and including the dechirp mixer and reference oscillator.

A second issue arises in systems using phase-steered array antennas. Recall from <u>Chap. 1</u> that the antenna pattern of a phase-steered array antenna is determined primarily by the array factor where *d* is the element spacing and the $\{a_n\}$ are the complex weights on each subarray output. The antenna is steered to a particular look direction θ_0 by setting the steering weights a_n according to²

$$E(\theta) = E_0 \sum_{n=0}^{N-1} a_n e^{j(2\pi/\lambda)nd\sin\theta}$$
(4.123)

$$a_n = |a_n| e^{-j(2\pi/\lambda)nd\sin\theta_0}$$
(4.124)

The magnitudes of the weights are chosen to provide the desired sidelobe level. $E(\theta)$ will exhibit a peak at $\theta = \theta_0$; for example, if $|a_n| \equiv 1$, $E(\theta)$ will be an asinc function with its peak at θ_0 . Note that the phases of the required weights $\{a_n\}$ are a function of the wavelength λ . If an LFM pulse is transmitted, the effective wavelength changes during the pulse sweep. If the system is wideband, this wavelength change will be significant and the value of θ at which $E(\theta)$ peaks will change as well. That is, the antenna look direction will actually change during the LFM sweep (see Prob. 18). This undesired frequency steering effect is an additional source of SNR loss.

Stepped frequency waveforms are an alternative technique for obtaining a large bandwidth and thus fine range resolution without requiring intrapulse frequency modulation. A stepped frequency waveform is a pulse burst waveform. Each pulse in the burst is a simple, constant-frequency pulse; however, the RF is changed from one pulse to the next. The most common stepped frequency waveform employs a linear frequency stepping pattern, where the RF of each pulse is increased by ΔF Hz from the preceding pulse. Factoring out the starting RF gives the following baseband waveform

$$x(t) = \sum_{m=0}^{M-1} x_p (t - mT) e^{j2\pi m\Delta F(t - mT)}$$
(4.125)

Figure 4.40 illustrates the linearly stepped frequency waveform.



FIGURE 4.40 Linearly frequency-stepped waveform.

Because only simple pulses are used for each constituent pulse, the instantaneous bandwidth capability of the transmitter and receiver need be only on the order of $1/\tau$ Hz. The total bandwidth of the waveform as a whole is $M \cdot \Delta F$. When used with a phase-steered array antenna, the time between pulses can be used to reset the phase shifters to update the $\{a_n\}$ sequence and maintain a nearly constant steering direction θ_0 as the effective wavelength changes from pulse to pulse. The major disadvantages of this waveform are that it requires a pulse-to-pulse tunable transmitter and receiver, and that M PRIs are required to collect data over the desired bandwidth instead of just one.

The pulse-by-pulse processing viewpoint applied to the constant-frequency pulse burst waveform can be applied again to analyze the matched filter response for the stepped frequency waveform. Suppose the radar is stationary, and a stationary target is located at a range corresponding to a delay $t_l + \delta t$, where δt represents an incremental delay relative to the nominal delay t_l corresponding to range bin *l*. Individual pulses are processed through the simple pulse matched filter as before, producing the output waveforms (assuming $T_M = 0$)

$$y_m(t) = s_p^* [t - (t_l + \delta t) - mT] e^{j2\pi m\Delta F[t - (t_l + \delta t) - mT]}$$
(4.126)

This output is then sampled at $t = t_l + mT$ (that is, t_l seconds after the current pulse was transmitted), corresponding to range $R_l = ct_l/2$. The resulting sample becomes the *l*th coarse range bin sample for the current pulse

$$y[l, m] = y_m(t_l - mT)$$

= $s_p^*(\delta t)e^{j2\pi m\Delta F\delta t}$
(4.127)

Equation (4.127) shows that the slow-time sequence at a fixed coarse range bin l when using a linearly stepped frequency waveform is a discrete time sinusoid. The frequency is proportional to the displacement of the scatterer from the nominal range bin location of $R_l = ct_l/2$ meters. The amplitude of the sequence is weighted by the triangular simple pulse matched filter response evaluated at the incremental delay $s_p(\delta t)$.

Following the earlier discussion of pulse-by-pulse processing for the conventional pulse burst waveform, the slow-time matched filter impulse response for a target located at the nominal delay $t_i + \delta t$ is $h[m] = \exp(-j2\pi m\Delta F\delta t)$. Thus, the matched filter impulse response is different for every value of δt . Consider a DTFT of the slow-time data

$$Y[l, \omega) = \sum_{m=0}^{M-1} y[l, m] e^{-j\omega m} = \sum_{m=0}^{M-1} s_p^*(\delta t) e^{j2\pi m\Delta F \delta t} e^{-j\omega m}$$
$$= s_p^*(\delta t) \sum_{m=0}^{M-1} e^{-j(\omega - 2\pi \Delta F \delta t)m}$$
(4.128)

The summation will yield an asinc function having its peak at $\omega = 2\pi\Delta F\delta t$. Thus, the peak of the DTFT of the slow-time data in a fixed range bin with a linearly stepped frequency waveform provides a measure of the delay of the scatterer relative to the nominal delay t_l . Specifically, if the peak of the DTFT is at $\omega = \omega_p$, the scatterer is at an incremental delay

$$\delta t = \frac{\omega_p}{2\pi\Delta F} = \frac{f_p}{\Delta F} \quad \text{seconds}$$
(4.129)

Note also that the DTFT evaluated at ω_p is the matched filter for the slow-time sequence, so that the data samples are integrated in phase

$$Y[l, \omega_{p}) = \sum_{m=0}^{M-1} s_{p}^{*}(\delta t) e^{j2\pi m\Delta F \delta t} e^{-j\omega_{p}m}$$
$$= s_{p}^{*}(\delta t) \sum_{m=0}^{M-1} (1) = M s_{p}^{*}(\delta t)$$
(4.130)

The factor of *M* is the coherent integration gain from using *M* pulses. If $\delta t = 0$, meaning the matched filter output was sampled at its peak, $Y[l, \omega) = ME_p = E$, the total waveform energy. If $\delta t \neq 0$ the ambiguity function of the individual pulses reduces the amplitude of the slow-time samples by $|s_p(\delta t)|$. This represents a straddle loss.

It follows that applying a *K*-point DFT to the slow-time sequence implements *K* filters, each matched to a different incremental delay δt . Thus, the DFT of the slow-time data within a single range bin for a stepped frequency waveform is a map of echo amplitude versus incremental range within that coarse range bin.

The DTFT of an *M*-point sinusoid has a Rayleigh frequency resolution of $\Delta f = 1/M$ cycles per sample. Using the scaling between *f* and *t* from Eq. (4.129), the corresponding time resolution is $\Delta t = 1/M \cdot \Delta F$ seconds; the range resolution is therefore

$$\Delta R = \frac{c}{2M \cdot \Delta F} = \frac{c}{2\beta} \quad \text{meters}$$
(4.131)

where β is the total stepped bandwidth $M \cdot \Delta F$. Thus, the linearly stepped frequency waveform achieves the same range resolution as a single pulse of bandwidth β . If a *K*-point DFT is used to process the slow-time data the DFT output will provide range measurements at intervals of

$$\delta R = \frac{c}{2K \cdot \Delta F} = \frac{M}{K} \Delta R \tag{4.132}$$

Since $K \ge M$ normally, the DFT output provides echo amplitude samples at intervals equal to or less than the range resolution. This fine-resolution reflectivity map is often called a *high resolution range profile*[§] or just a *range profile*.

The total bandwidth β of the stepped frequency waveform is determined by the desired range resolution. It can be realized by various combinations of the number of frequency steps M and the step size ΔF . To determine how to choose these parameters, note that the DTFT of the slow-time data is periodic in ω with period 2π radians per sample. Because the DTFT peak is at $\omega_p = 2\pi \Delta F \delta t$, the range profile is periodic in δt with period $1/\Delta F$. This periodicity establishes the required coarse range bin spacing. Specifically, avoiding range ambiguities in the range profile requires $c/2\Delta F > L_t$, where L_t is the maximum target length of interest. Once ΔF is chosen, M is selected to span the bandwidth required to provide the desired fine range resolution. The DFT range profile then effectively breaks each relatively large coarse range bin ($c/2\Delta F$ meters) into M fine-resolution range bins ($c/2\beta$ meters) sampled at K points within the coarse range bin. If K = M the range sample spacing equals the range resolution. If K > 1M the range profile is oversampled compared to the resolution by the factor K/M. The pulse length τ is chosen to balance straddle losses and range ambiguities. Recall that the single-pulse matched filter output $s_n(t)$ is 2τ seconds long. Choosing $\tau < 1/2\Delta F$ means that $s_p(t)$ will be no more than $1/\Delta F$ seconds

long so that a scatterer will only influence measurements in one coarse range bin, avoiding range ambiguities. On the other hand, the shorter τ is made, the greater the potential straddle loss for targets located between coarse range samples. A detailed consideration of these tradeoffs is in Keel and Baden (2012).

Details of the Doppler response and ambiguity function of the linearly stepped frequency waveform are available in Levanon and Mozeson (2004). A small central portion of the ambiguity function is shown in Fig. 4.41 for the case M = 8 pulses, PRI $T = 10\tau$, and a frequency step size of $\Delta F = 0.8/\tau$. The resulting bandwidth is $\beta = M \cdot \Delta F = 6.4/\tau$ Hz. The AF displays both the skewed response typical of a linear FM modulation, and the range and Doppler ambiguities typical of pulse burst waveforms. Ambiguities in delay (range) are evident at intervals of T seconds, corresponding to 1/8 = 0.125 on the normalized scale of the figure. The first zero in Doppler of the main ridge occurs at 1/MT Hz, corresponding to 1 on the normalized Doppler scale.



FIGURE 4.41 Contour plot of the central portion of the ambiguity function of a pulse burst waveform. M = 8, $T = 10\tau$, and $\Delta F = 0.8/\tau$.

Figure 4.42*a* further magnifies the delay coordinate of this AF. The delay coordinate now covers the interval $\pm \tau = \pm 0.0125 \ (\pm 1/80)$ on this normalized scale. The zero-delay and zero-Doppler axes are highlighted by the heavier gray lines. The expected Rayleigh resolution in delay is $1/\Delta F = \tau/6.4$, which

becomes 0.002 on this scale. The dotted heavy gray line marks the + 0.002 delay coordinate. It can be seen that this intersects the first null on the zero-Doppler axis of the AF, confirming that the intended resolution is achieved.



FIGURE 4.42 Contour plot of the central portion of the ambiguity function of a pulse burst waveform: M = 8, $T = 10\tau$, (a) $\Delta F = 0.8/\tau$ (same waveform as Fig. 4.41), (b) $\Delta F = 2.5/\tau$.

Choosing $\Delta F > 1/\tau$ allows generation of a wide total bandwidth with fewer pulses and therefore a shorter data collection time, but the resulting undersampling creates aliasing that appears as extra range ambiguities (also called grating lobes). Figure 4.42*b* shows a similar view of another case with $\Delta F \cdot \tau = 2.5$ but the other parameters unchanged. The bandwidth is now 20/ τ Hz so the resolution in delay is correspondingly finer. However, there are now five peaks along the zero-Doppler axis, representing five range ambiguities, within ± 1 pulse length.

4.9 The Stepped Chirp Waveform

The stepped chirp waveform is a stepped frequency waveform that substitutes an LFM constituent pulse for the constant-frequency pulse used above. It can achieve very wideband operation without resorting to stretch processing, thereby avoiding the restriction of short range windows. In addition, it avoids the array frequency steering effects mentioned previously so long as the individual pulse bandwidth is not too large.

The stepped chirp waveform can allow a large frequency step $\Delta F > 1/\tau$ without suffering the aliasing seen in the conventional stepped frequency waveform. Careful design is needed to relate the LFM pulse bandwidth and length to the RF step size in order to achieve effective suppression of the ambiguities. Details and sample parameter sets are given in Levanon and Mozeson (2004). Processing of the waveform requires individually demodulating and matched filtering each individual pulse, and then post-processing the ensemble to construct a new signal with the full bandwidth. This post-processing can be performed in either the time or frequency domain. Details are given in Keel and Baden (2012).

4.10 Phase-Modulated Pulse Compression Waveforms

The second major class of pulse compression waveforms is referred to as *phase* coded waveforms. A phase coded waveform has a constant RF but an absolute phase that is switched between one of two or more fixed values at regular intervals within the pulse length. Such a pulse can be modeled as a collection of N contiguous subpulses $x_n(t)$ of duration τ_c , each with the same frequency but a (possibly) different phase

$$\begin{aligned} x(t) &= \sum_{n=0}^{N-1} x_n (t - n\tau_c) \\ x_n(t) &= \begin{cases} \exp(j\phi_n), & 0 \le t \le \tau_c \\ 0, & \text{elsewhere} \end{cases} \end{aligned}$$

(4.133)

The total pulse length is $\tau = N\tau_c$. Individual subpulses are often referred to as *chips*. Phase coded waveforms are divided into *biphase codes* and *polyphase* codes. A biphase code has only two possible choices for the phase state ϕ_n , typically 0 and π ; a polyphase code has more than two phase states. There are several common subcategories of each. Figure 4.1c was an example of a biphase-coded waveform.

The matched filter output for a phase coded pulse is derived in detail in Levanon and Mozeson (2004); the result is now summarized. Denote the sequence of complex amplitudes of the individual pulse chips $x_n(t)$ of Eq. (4.133) as $\{A_n\} = \{\exp[j\phi_n]\}$. Express the time variable *t* in terms of the chip duration τ_c and an offset η , as $t = k\tau_c + \eta$, $0 \le \eta < \tau_c$. The matched filter output, which is just the autocorrelation of x(t), is

$$y(t) = s_x(t) = y(k\tau_c + \eta) = \left(1 - \frac{\eta}{\tau_c}\right) s_A[k] + \frac{\eta}{\tau_c} s_A[k+1]$$
(4.134)

where $s_A[k]$ is the discrete autocorrelation of the complex amplitude sequence $\{A_n\}$. Equation (4.134) shows that $s_x(t)$ takes on the value $s_A[k]$ at $t = k\tau_c$, and is linearly interpolated (in the complex plane) between adjacent samples. Thus, the matched filter output can be determined by computing the autocorrelation of the amplitude sequence and interpolating between those values. One consequence of this result and the fact that the $\{A_n\}$ have unit magnitude is that the peak value of the autocorrelation will always be $s_x(0) = N$.

4.10.1 Biphase Codes

The most common biphase codes in radar are the *Barker codes*. Barker codes are a specific set of biphase sequences that have a maximum sidelobe magnitude of 1 at the matched filter output and therefore attain an N: 1 ratio of the peak to the highest sidelobe. A low-frequency Barker coded waveform for N = 13 is shown in Fig. 4.43. The phase switches are visible at $\tau = 5\tau_c$, $7\tau_c$, $9\tau_c$, $10\tau_c$, $11\tau_c$, and $12\tau_c$. Because there are only two phase states, the waveform is often represented by a diagram such as the one shown in Fig. 4.44, using either "+" and "–" symbols as shown, or +1 and –1 symbols. Note that biphase codes do not necessarily change phase state at every subpulse transition.



FIGURE 4.43 Barker coded waveform, N = 13.



FIGURE 4.44 Binary sequence describing the Barker code of Fig. 4.43.

Recall that pulse compression waveforms have a bandwidth $\beta \gg 1/\tau$. Because phase coded waveforms are constant frequency, it may not be obvious that their spectrum is spread. However, the discontinuities caused by the phase transitions do spread the signal spectrum. As an example, Fig. 4.45 shows the effect of a single phase switch of 180° on the spectrum of a constant-frequency waveform. While the effect depends on the point in the pulse at which the switch occurs, clearly it significantly spreads the signal energy in frequency. Multiple phase transitions increase this effect: Fig. 4.46 compares the spectra of the 13-bit Barker coded waveform with that of a simple pulse of the same duration. The Rayleigh bandwidth of the Barker spectrum is $\beta = 1/\tau_c$ Hz (about 13 times as wide as that of the simple pulse in this specific case). In addition, the sidelobes of the Barker waveform spectrum decay much more slowly than those of the simple pulse.



FIGURE 4.45 Effect of a single 180° phase switch on the spectrum of a constant-frequency pulse: (*a*) phase switch occurs at $t = \tau/2$, (*b*) phase switch occurs at $t = 3\tau/4$.



FIGURE 4.46 Spectra of a 13-bit Barker coded pulse and a simple pulse of the same length.

Because of the increased bandwidth, the signal processing gain due to pulse compression with the Barker codes in particular, and phase codes in general, is the same factor $G_{sp} = \beta \tau$ that applied for the LFM waveform. This is again because a simple pulse must be shorter than a phase-coded pulse by the factor $\beta \tau$ to achieve the same Rayleigh resolution in time. Note that in the case of phase codes, $\beta \tau$ can also be expressed as τ/τ_c .

One of the major disadvantages of Barker codes is that there are not very many of them. Barker codes have been found only for *N* up to 13. Sample Barker codes of all known lengths are listed in <u>Table 4.1</u>; more than one code exists for some lengths, while none exist for N = 6, 8 - 10, and 12. The table also lists the peak sidelobe level (PSL) relative to the mainlobe peak, which is simply 20 $\log_{10}(1/N)$. Because of the modest lengths, low PSLs are not attainable.

	Code Sequence			
N	+/-Format	Octal	PSL, dB	
2	+	2	-6.0	
2	+.+	3	-6.0	
3	++-	6	-9.5	
4	++-+	15	-12.0	
4	+++-	16	-12.0	
5	+++-+	35	-14.0	
7	++++-	162	-16.9	
11	++++-	3422	-20.8	
13	+++++++-+-+	17465	-22.3	

TABLE 4.1Barker Codes



FIGURE 4.47 Matched filter output for a 13-bit Barker code.

Barker codes have two major disadvantages. The first is that the lack of Barker codes longer than N = 13 limits the degree of sidelobe suppression possible. The second is that they are very Doppler intolerant. A Doppler phase rotation of 360° across the full pulse is more than sufficient to completely break up the structure of the matched filter output, as seen in the 13-bit Barker ambiguity function contour plot of Fig. 4.48. As a consequence, it is common to design Barker coded waveforms to limit the Doppler phase rotation to one-quarter cycle or less, which requires that the maximum expected Doppler shift and target velocity satisfy



FIGURE 4.48 Contour plot of ambiguity function of a 13-bit Barker code.

$$F_{D_{\max}}\tau < \frac{1}{4} \implies v_{\max} < \frac{\lambda}{8\tau}$$

$$(4.135)$$

This constraint limits the Doppler mismatch loss to 1 dB or less.

The limited number and length of Barker codes has led to various techniques for constructing longer biphase codes with good sidelobe properties. *Combined* or *nested* Barker codes form a longer code as the Kronecker product of two shorter Barker codes. If an *N*-bit Barker code sequence is denoted as B_N , an *MN*-bit code can be constructed as $B_M \otimes B_N$. The Kronecker product is simply the B_N code repeated *M* times, with each repetition multiplied by the corresponding element of the B_M code. For example, a 20-bit code can be constructed as the product $B_4 \otimes B_5$

These codes have a peak sidelobe higher than 1. The autocorrelation of the code of Eq. (4.136) is shown in Fig. 4.49. Notice that the magnitude of the peak sidelobes is 5, so that the PSL compared to the autocorrelation peak is only 1/4

instead of the 1/20 that would be obtained if a 20-bit Barker code existed. A code that obtains a PSL = 2 for N = 20 is discussed momentarily.



FIGURE 4.49 Autocorrelation of combined $B_4 \otimes B_5$ code.

Another technique uses pseudorandom noise sequences to generate much longer biphase codes. Pseudorandom sequences have length $N = 2^{P} - 1$ for some P and generally exhibit range sidelobes on the order of $-10 \log_{10}(N)$. For example, the matched filter output for a typical N = 1023 (P = 10) code, shown in Fig. 4.50, has peak sidelobes just above -30 dB.



FIGURE 4.50 Matched filter output for a 1023-bit pseudorandom biphase code.

While the Barker codes are the only biphase codes with a peak sidelobe value of 1 that are known to exist, one can seek longer codes with *minimum* peak sidelobe (MPS) levels for the length of interest. These MPS codes are found by exhaustive search techniques, taking advantage of certain properties of biphase code autocorrelations to prune the search somewhat. As an example, the MPS code for N = 20 has a maximum sidelobe level of 2, giving a PSL ratio of 1/10 instead of the 1/4 obtained by the nested Barker code above. The state of the art is summarized and references given in Keel (2010). The peak sidelobe for MPS codes of lengths 2 through 5, 7, 11, and 13 (Barker codes) is 1; for N =6, 8 - 10, 12, 14 - 21, 25, and 28 is 2; for N = 22 - 24, 26 - 27, 29 - 48, and 51 is 3; for N = 49 - 50 and 52 - 82 is 4; and for N = 83 - 105 is 5. The MPS has not been established at this writing for codes longer than 105. Table 4.2 lists one sample code for the longest code length in each of these sidelobe level regimes; additional sample codes for other lengths are available in the references. It is evident that the sidelobe level in dB improves only very slowly as the code length increases.

Code Length N	Sample Code (Hexadecimal)	Peak Sidelobe Level (PSL)	PSL, dB
13 (Barker)	1F35	1	-22.3
28	8F1112D	2	-22.9
51	0E3F88C89524B	3	-24.6
82	3CB25D380CE3B7765695F	4	-26.2
105	1C6387FF5DA4FA325C895958DC5	5	-26.4

 TABLE 4.2
 Sample Minimum Peak Sidelobe Biphase Codes

4.10.2 Polyphase Codes

Biphase codes, as noted previously, have poor Doppler tolerance. They also suffer from precompression bandlimiting effects. As is shown in Fig. 4.51, the spectrum of a typical biphase code not only exhibits the desired mainlobe spreading, but also a very slow falloff of the far sidelobes. This is a direct consequence of the sharp phase discontinuities. Practical receivers will have a noise-limiting bandpass filter that will bandlimit the biphase waveform spectrum, smoothing the phase transitions. This has the effect of mismatching the received waveform relative to the correlator, reducing the peak gain and widening the mainlobe.



FIGURE 4.51 Matched filter output for a 16-bit Frank code.

Polyphase codes allow arbitrary values for the chip phases ϕ_n . Compared to biphase codes, they can exhibit lower sidelobe levels and greater Doppler tolerance. A number of polyphase codes are in common use. These include "polyphase Barker codes," the Frank codes, and the P1, P2, P3, P4, and P(*n*, *k*) codes. All of these except the polyphase Barker codes are related to LFM or NLFM waveforms. Numerous other polyphase codes have been proposed; many are described in Levanon and Mozeson (2004). The special case of quadriphase codes is described in Keel and Baden (2012).

Frank codes are codes whose length is a square, $N = M^2$ for some *M*. The phase sequence for a Frank code is given by

$$\phi_n = \phi(Mp+q) = \frac{2\pi}{M}pq, \quad p = 0, 1, 2, \dots, M-1, \quad q = 0, 1, 2, \dots, M-1$$
(4.137)

As an example, if M = 4 so N = 16, the sequence of phases becomes

$$\phi_n = \left\{ \underbrace{\underbrace{0 \ 0 \ 0 \ 0}_{p=0} \underbrace{0 \ \frac{\pi}{2} \pi \ \frac{3\pi}{2}}_{p=1} \underbrace{0 \ \pi \ 0 \ \pi}_{p=2} \underbrace{0 \ \frac{3\pi}{2} \pi \ \frac{\pi}{2}}_{p=3} \underbrace{0 \ \frac{3\pi}{2} \pi \ \frac{\pi}{2}}_{p=3} \right\}$$
(4.138)

Figure 4.51 shows the magnitude of the matched filter output for the case $N = 16.^{2}$ Note that while the mainlobe has a local minimum at $t = \tau_{c}$, it does not go to zero at that point as the Barker code autocorrelations do. The largest sidelobe in this example is $\sqrt{2}$, larger than the Barker codes. The sidelobe level is $\sqrt{2}/16$ or -21.1 dB. Figure 4.52 shows the ambiguity function in contour plot form. The main ridge is skewed in the delay-Doppler plane, similar to the range-Doppler coupling of an LFM ambiguity function.



FIGURE 4.52 Contour plot of 16-bit Frank code ambiguity function.

The P3 and P4 codes of length N are given, respectively, by

P3:
$$\phi_n = \begin{cases} \frac{\pi}{N} n^2, & n = 0, 2, ..., N-1 \quad (N \text{ odd}) \\ \frac{\pi}{N} n(n+1), & n = 0, 1, 2, ..., N-1 \quad (N \text{ even}) \end{cases}$$

P4: $\phi_n = \frac{\pi}{N} n^2 - \pi n, \quad n = 0, 1, 2, ..., N-1 \qquad (4.139)$

Unlike the Frank code, these codes can be generated for any length *N*. Figure <u>4.53</u> shows the matched filter output for the N = 20 P3 code, while Fig. <u>4.54</u> shows the corresponding ambiguity function. Again, range-Doppler coupling is evident.



FIGURE 4.53 Matched filter output for a 20-bit P3 code.



FIGURE 4.54 Contour plot of ambiguity function of 20-bit P3 code.

The Frank, P3, and P4 codes all are based on quadratic phase progressions, as is evident from Eqs. (4.137) and (4.139), and are therefore related to LFM waveforms. Figure 4.55 shows the (unwrapped) phase progression of these three codes for the case N = 16. The P3 and P4 codes are truly quadratic, the difference being whether the minimum phase "slope" occurs at the beginning (P3) or the middle (P4) of the waveform. The smallest phase increments, and thus the minimum discontinuities in the actual RF waveform, occur where the phase slope is least. The Frank code uses a piecewise linear approximation to a quadratic phase progression. The phase increment is constant for *M* bits at a time and then increases for the next *M* bits. This can be viewed as a phase code approximation to a stepped-frequency waveform having *M* steps and *M* bits per step (Lewis and Kretschmer, 1986). As a result, the Frank code is less Doppler tolerant than the P3 and P4 codes.



FIGURE 4.55 Unwrapped phase sequences of 16-bit Frank, P3, and P4 codes.

Bandlimiting of the phase-coded waveform prior to matched filtering results in an increase in mainlobe width but a decrease in PSL in codes that have the smallest phase increments in the middle of the codes (Lewis and Kretschmer, 1986; Levanon and Mozeson, 2004). Codes with the largest phase increments near the end exhibit the opposite behavior. Thus, of the three codes shown, the P4 will show the greatest tolerance to precompression bandlimiting in the sense of maintaining or improving its sidelobe level at the matched filter output.

Just as phase codes can be designed based on linear frequency modulation waveforms, they can also be designed based on nonlinear frequency modulation waveforms. A class of codes based on NLFM waveforms designed using the PSP technique mentioned earlier is given in Felhauer (1994). No closed form expression is known for these P(n, k) codes; they must be found numerically. Typical results are very similar to those for the empirical NLFM waveforms described earlier. The effect of Doppler mismatch is similar to that observed in Fig. 4.39. This is an improvement over conventional polyphase codes, which are prone to exhibiting significantly increased sidelobes near the ends of the code and, in many cases, large spurious peaks well above the general sidelobe level. P(n, k) codes also exhibit better tolerance to precompression bandlimiting than do codes based on linear FM, since their spectra are already shaped by the basic NLFM design approach. Their chief disadvantage is the difficulty of their design.

Another approach to reducing spectral sidelobes and thus improving precompression bandlimiting tolerance is the use of quadriphase codes. These codes are obtained from biphase codes by mapping the binary phase progression to a four-phase code using a specified transformation, and also by replacing the rectangular subpulse chips with half-cosine chips of twice the width. Compared to the biphase code, the resulting codes have significantly lower spectral sidelobes, nearly the same autocorrelation sidelobes, but a significant loss of time (range) resolution. Details are given in Keel and Baden (2012).

Polyphase Barker codes are polyphase codes that exhibit a maximum sidelobe peak level of 1. The phases of a length-N code are either unrestricted, or restricted to a Pth root of unity, $\phi_n = 2\pi p_n/P$ for some integers p_n and P and $n \in [0, N-1]$. Figure 4.56 shows the discrete autocorrelation of a polyphase Barker code having N = 51, P = 50, and the following $\{p_n\}$ sequence:



FIGURE 4.56 Autocorrelation of an N = 51 polyphase Barker code.

0, 0, 4, 4, 18, 20, 27, 25, 25, 26, 24, 15, 15, 14, 9, 32, 36, 2, 21, 17, 9, 27, 46, 49, 19, 29, 9, 32, 7, 45, 21, 46, 22, 47, 18, 35, 0, 22, 9, 31, 44, 5, 29, 21, 4, 49, 33, 24, 9, 49, 29

The PSL is -34.2 dB, significantly better than the -24.6 dB for the 51-point MPS code in <u>Table 4.2</u>.

4.10.3 Mismatched Phase Code Filters

The sidelobe structure of phase-coded waveforms can be improved with the use o fmismatched filters, just as is done with stepped frequency and FM waveforms to improve their sidelobe structures. For phase-coded waveforms, this implies correlating the code sequence with another discrete-time sequence, not necessarily restricted in the amplitudes or phases of its coefficients, such that some metric of the sidelobe structure is optimized. Mismatched filters can be designed to minimize the output PSL, minimize the output integrated sidelobe level (ISL) (sum of the squares of all the sidelobe values of the discrete correlation of the code and filter, divided by the square of the peak), or to shape the output sidelobe response, for instance to enforce particularly low near-in sidelobes at the expense of higher distant sidelobes. The filter order L is usually larger than the code length. In many cases the design of the mismatched filter coefficients can be formulated as the solution of a weighted least squares problem, for which many numerical algorithms are available. Other optimization techniques, such as L¹ minimization using convex optimization algorithms, can also be employed.

Figure 4.57 illustrates two examples of mismatched filter design. In both cases, the waveform phase code was the same N = 64 MPS biphase code. The peak sidelobe of the matched filter output for this length is 4, giving a PSL in dB of 20 $\log_{10}(4/64) = -24.1$ dB. The ISL for the matched filter is -6.7 dB. The solid line in the figure is the result of a mismatched filter of length L = 130 designed to minimize the PSL. The filter impulse response is normalized to have the same energy as the code and its matched filter impulse response, namely 64. This filter achieves a PSL of -31.4 dB, an 8.3 dB improvement compared to the matched filter. The ISL is -9.8 dB, an improvement of about 3.1 dB. However, there is now an LPG of 1.11 dB relative to the matched filter.



FIGURE 4.57 Autocorrelation of two L = 130 mismatched filters for the same N = 64 MPS code. Solid line: Optimal PSR filter response. Dotted line: Optimal ISR filter response. (After Keel and Baden, 2012.)

The dotted line is the result of an L = 130 filter designed to minimize the ISL and also normalized to an energy of 64. The PSL is now about -23.1 dB, 1 dB worse than the matched filter and 7.3 dB worse than the minimum-PSL filter. However, the ISL is now -12 dB, 5.3 dB better than the matched filter and 2.2 dB better than the minimum-PSL filter. The LPG is 0.85 dB relative to the matched filter, a 0.26 dB improvement compared to the minimum-PSL filter.

4.11 Costas Frequency Codes

Costas waveforms are a class of pulse compression waveforms having aspects of both phase-coded and stepped frequency pulse burst waveforms (Costas, 1984). A Costas waveform is similar to a polyphase waveform in that it is a single pulse waveform divided into *N* subpulses. It is similar to the linearly stepped frequency waveform in that, rather than maintaining a constant frequency and altering the phase of each subpulse, it alters the subpulse frequencies, stepping through a set of *N* frequencies that differ by ΔF Hz. Unlike the stepped frequency pulse burst, however, the Costas waveform does not step through the frequencies in linear order. The Costas pulse can be expressed as

$$x(t) = \sum_{n=0}^{N-1} x_n (t - \eta \tau_c)$$

$$x_n(t) = \begin{cases} \exp(jc[n] \cdot \Delta F \cdot t), & 0 \le t \le \tau_c \\ 0, & \text{elsewhere} \end{cases}$$
(4.140)

where the sequence c[n] denotes the ordering of the stepped frequencies.

Figure 4.58 shows the frequency sequence for a typical low-order Costas waveform. With proper design of the frequency step sequence, the Costas waveform can be designed to have a more thumbtack-like ambiguity function than the linearly stepped waveform. Figure 4.59 illustrates the ambiguity function of a Costas waveform with N = 15; the frequency step sequence was $c[n] = \{1,7,8,11,3,13,9,14,12,6,5,2,10,0,4\}$. Note the generally low and relatively uniform sidelobe structure throughout the delay-Doppler plane. The construction and properties of Costas waveform are discussed and more examples given in Levanon and Mozeson (2004).



FIGURE 4.58 Frequency sequence for Costas waveform with N = 7.



FIGURE 4.59 Ambiguity function for a Costas waveform with N = 15.

4.12 Continuous Wave Radar

All of the preceding discussion in this chapter, and indeed in virtually this entire text, is centered on pulsed radar. Pulsed radars are capable of very long-range application, can easily measure range and velocity, and can achieve fine-resolution imaging. However, pulsed radars require high peak powers in order to achieve good average power and suffer eclipsing and blind zones.

Continuous wave (CW) radar is another class of radar system that transmits and receives continuously. They do not necessarily require as complex a transceiver as does a pulsed system. Because transmission is continuous, the average power equals the peak power, a situation more amenable to the use of solid-state or other peak-power-limited transmit sources. Solid state sources in particular enable the development of very low-cost radar systems. This also means that good average power can be achieved without high peak powers, which is helpful when a low probability of intercept is desired. Eclipsing does not occur, making CW systems superior for short-range measurements. Furthermore, CW systems do not suffer blind zones in range or velocity. Given these characteristics, CW radar is popular for a variety of low-power, shortrange applications, especially those involving velocity measurements. Common examples include police and sports "speed guns," radar altimeters and fuzes, missile seekers, meteorology, and automotive cruise control and collision avoidance radar. There are also many more complex or unusual applications such as short-range synthetic aperture imaging, RCS measurements, and storage tank level measurements.

Like pulsed radar, CW radar can be operated with different waveforms, many of them analogous to pulse waveforms. These include constant-frequency, linear and nonlinear FM, biphase and polyphase coding, and frequency coding, as well as techniques less common in pulsed radar such as frequency shift keying (FSK), sinusoidal modulation, and noise modulation. The most common CW waveform is the linear FM, usually called FMCW.

An excellent overview of CW radar configurations, design, waveforms, and applications is available in Piper (2014). A comparative discussion of CW, FMCW, and FSK waveforms in the context of automotive radar is given in Rohling and Kronauge (2012).

References

Born, M., and E. Wolf, *Principles of Optics*. Pergamon Press, London, 1959. Carrara, W. G., R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic* Aperture Radar. Artech House, Norwood, MA, 1995.

- Cook, C. E., and M. Bernfeld, *Radar Signals: An Introduction to Theory and Application*. Artech House, London, 1993.
- Costas, J. P., "A Study of a Class of Detection Waveforms Having Nearly Ideal Range-Doppler Ambiguity Properties," *Proceedings of the IEEE*, vol. 72(8), pp. 996–1009, Aug. 1984.
- De Witte, E., and H. D. Griffiths, "Improved Ultra-Low Range Sidelobe Pulse Compression Waveform Design," *Electronics Letters*, vol. 40, no. 22, pp. 1448–1450, 2004.
- Felhauer, T., "Design and Analysis of New P(n, k) Polyphase Pulse Compression Codes," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 865–874, Jul. 1994.
- Gini, F., A. De Maio, and L. Patton (eds.), *Waveform Design and Diversity for Advanced Radar Systems*. Institution of Engineering and Technology (IET), London, 2012.
- Jonsson, B. E., "A Survey of A/D Converter Performance Evolution," *Proceedings 17th IEEE Intl. Conf. Electronics, Circuits, and Systems* (ICECS), pp. 766–769, 2010.
- Kay, S. M., Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory. Prentice Hall, Upper Saddle River, NJ, 1998.
- Keel, B. M., "Fundamentals of Pulse Compression Waveforms," Chap. 20 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.
- Keel, B. M., and J. M. Baden, "Advanced Pulse Compression Waveform Modulations and Techniques," Chap. 2 in W. L. Melvin and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Edison, NJ, 2012.
- Levanon, N., "The Periodic Ambiguity Function—Its Validity and Value," *Proceedings 2010 IEEE International Radar Conference*, pp. 204–208, Arlington, VA, 2010.
- Levanon, N., and E. Mozeson, Radar Signals. Wiley, New York, 2004.
- Lewis, B. L., F. K. Kretshcmer, Jr., and W. W. Shelton, *Aspects of Radar Signal Processing*. Artech House, Canton, MA, 1986.
- Nathanson, F. E., (with J. P. Reilly and M. N. Cohen), *Radar Design Principles*, 2d ed. McGraw-Hill, New York, 1991.
- Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Pearson, Englewood Cliffs, NJ, 2010.
- Papoulis, A., and S. U. Pillai, *The Fourier Integral and Its Applications*, 6th ed. McGraw-Hill, New York, 2002. Peebles, P. Z., Jr., *Radar Principles*. Wiley, New York, 1998.
- Piper, S. O., "Continuous Wave Radar," Chap. 2 in J. A. Scheer and W. L. Melvin (eds.), *Principles of Modern Radar: Radar Applications*. SciTech Publishing. to appear, 2014.

Price, R., "Chebyshev Low Pulse Compression Sidelobes via a Nonlinear FM," URSI National

Radio Science meeting, Seattle, WA, June 18, 1979.

- Raney R. K., "A New and Fundamental Fourier Transform Pair," *Proceedings* of the IEEE 12th International Geoscience & Remote Sensing Symposium (IGARSS' 92), pp. 26–29, 106–107 May, 1992.
- Richards, M. A., "Time and Frequency Domain Weighting of LFM Pulses," unpublished technical note, Sep. 29, 2006. Available at <u>http://www.radarsp.com</u>.
- Rihaczek, A. W., *Principles of High-Resolution Radar*. Artech House, Boston, MA, 1996.
- Rohling, H., and M. Kronauge, "Continuous Waveforms for Automotive Radar Systems," Chap. 7 in F. Gini et al. (eds.), *Waveform Design and Diversity for Advanced Radar Systems*. Institution of Engineering and Technology (IET), London, 2012.

Problems

1. Consider a stationary radar transmitting a simple square pulse (modulation only, not including the carrier term) of duration τ :

$$x(t) = \begin{cases} 1, & 0 \le t \le \tau \\ 0, & \text{otherwise} \end{cases}$$

The receiver uses a causal matched filter with $T_M = \tau$, so $h(t) = x^*(\tau - t)$. The pulse is transmitted with the leading edge being emitted at time t = 0. An echo is received from a stationary target at a range of *R* meters. At what time t_{peak} will the peak output of the matched filter be observed? Show all work.

2. Consider the same pulse and matched filter used in the previous problem. Assume that now the target is at range *R* meters when the pulse hits it, but is moving with a radial velocity toward the radar of $k\lambda/2\tau$ m/s, where *k* is any integer (except $k \neq 0$). The received signal (again, after the carrier is removed) can be modeled as

$$r(t) = x \left(t - \frac{2R}{c} \right) \exp \left(j 2\pi F_D \left(t - \frac{2R}{c} \right) \right)$$

where F_D is the Doppler shift in hertz. Find the output waveform of the causal matched filter. What is its value at $t = 2R/c + \tau$ seconds?

3. Suppose the ambiguity function of some waveform x(t) of duration $\tau = 1$

millisecond is given by

$$A(t, F_D) = \exp\left\{-\left[\left(\frac{t}{3\tau}\right)^2 + (2F_D\tau)^2\right]\right\}$$

(Note: This is not a possible AF because it is not time-limited to $\pm \tau$ seconds duration in the delay coordinate, but it will do for this problem.) Suppose there are two targets in the radar's line of sight, one at R = 10 km and one at R = 10.1 km. Also assume that both have the same RCS and ignore the effect of the small range difference on the received echo power. The radar and the first target are stationary. The second target is traveling toward the radar at 100 m/s. The radar is operating at 1 GHz. What is the Doppler shift of the echo from the second target, in hertz? If the matched filter output is sampled at a time delay corresponding to the range to the first target (= 2 × (10 km)/c = 66.67 µs), the sample will contain contributions from both the first and second targets. Use $A(t, F_D)$ to determine the relative amplitude of the contribution from the second target compared to that of the first target. Express the answer in dB.

- 4. Consider a simple pulse burst waveform with M = 30 pulses, each of 10 µs duration, and a PRI of T = 100 µs. Assuming no weighting functions are used, what are the range resolution, Doppler resolution, unambiguous range, and unambiguous Doppler shift of this waveform?
- 5. Consider a linear FM waveform that sweeps from 9.5 to 10.5 GHz over a pulse length of 20 μ s. What is the bandwidth β ? What is the time-bandwidth product? What will be the Rayleigh resolution (peak to first null) of the matched filter output in meters? What would be the Rayleigh resolution in meters of a square pulse of the same energy (assuming both have the same amplitude)?
- 6. Continuing with the same LFM waveform as in the previous problem, what will be the frequency in hertz of the first zero of the zero delay cut of the ambiguity function? (This will be the Doppler resolution, or Doppler sensitivity, of the pulse.)
- 7. Consider an LFM waveform of bandwidth $\beta = 1$ MHz and pulse length $\tau = 1$ ms. Suppose an echo is received from a target at a true range of 10 km that is Doppler shifted by 1 kHz. What will be the apparent range of the target, i.e., what will be the range corresponding to the time at which the matched filter output peaks?
- 8. Consider an LFM pulse with $\beta = 50$ MHz and $\tau = 1$ msec. Compute the Doppler shift required to displace the matched filter output by three Rayleigh range resolution cells. No windowing for sidelobe control is used. At 10 GHz, compute the radial velocity associated with that value of

Doppler shift. Compute the loss in peak amplitude due to the Doppler shift in dB.

- 9. Suppose a radar uses a simple rectangular pulse of duration τ seconds and processes it through the corresponding matched filter. Assume the matched filter output is sampled at a rate equal to its Rayleigh bandwidth. What is the worst-case straddle loss in dB? Repeat for an LFM waveform with a sufficiently large BT product so that its spectrum is well-approximated by a rectangle of width β Hz. Assume no weighting for sidelobe control is used with either waveform.
- 10. Consider an LFM pulse of duration $\tau = 1$ ms. Suppose that a range window of only 1.5 km extent is of interest, so it is decided to use stretch processing. The range window is centered on a nominal range of 100 km (think of this as "zooming in" on targets in the vicinity of 100 km). A range resolution of 1.5 meters is required. What is the required bandwidth β ? What will be the $\beta \tau$ product of the LFM pulse? What will be the bandwidth of the stretch mixer output?
- 11. Continue with the same scenario and LFM waveform as in the previous problem. Suppose that a beat frequency of 100 kHz is observed at the mixer output. What is the range of the target, relative to the 100 km center of the range window? Ignore any delay in the matched filter.
- 12. Consider a stationary X-band (10 GHz) radar transmitting a $\beta = 500$ MHz LFM waveform and using stretch processing in the receiver. The pulse length is $\tau = 10 \ \mu$ s. A radar is often considered "narrowband" if the percentage bandwidth, defined as β divided by the RF frequency, is less than 10 percent; otherwise it is "wideband." Is this radar narrowband or wideband? What is the expected range resolution in meters?
- 13. Continuing with the same LFM waveform, suppose a Hamming window is applied to the signal at the output of the stretch mixer, before the FFT is performed. What will be the new value for the expected range resolution, based on the Rayleigh definition of resolution? (*Hint*: The peak-to-null width of the DTFT of a Hamming window of length τ seconds is $2/\tau$ Hz; for a rectangular window it is $1/\tau$ Hz.) What bandwidth β would be required to achieve 0.3 m resolution if the Hamming window is used to keep the range sidelobes low?
- 14. Continuing with the same radar and 500 MHz LFM pulse as in the previous two problems, suppose the stretch processor is set up for a nominal range (center of the range window) of $R_0 = 200$ km and a range window of 300 m (200 km ± 150 m). However, suppose the reference LFM signal is only $\tau = 10 \ \mu$ s seconds long, i.e., it is not lengthened to allow for signals arriving from the leading or trailing edges of the range window. The reference

signal is timed to overlap exactly with the echo from a target at range R_0 . What will be the duration at the mixer output of the beat frequency tone between the echo from a scatterer at the leading edge of the range window (200 km – 150 m) and the LFM reference? Assuming a rectangular window (i.e., no Hamming window), what will be the range resolution at the leading edge of the window?

- 15. It was stated that range skew at the output of a stretch processor could be corrected with a filter having the frequency response $H(\Omega) = \exp(-j\Omega^2 \tau/2\beta_{\Omega})$ where $\beta_{\Omega} = 2\pi\beta$ is in radian frequency units. Show that the group delay function $d_g(\Omega)$ of this filter meets the stated requirement, namely $d_g(-\beta_{\Omega} \cdot \delta t_b/\tau) = -\delta t_b$ seconds. Group delay in seconds is defined as $dg(\Omega) = -d\Phi(\Omega)/d\Omega$, where $\Phi(\Omega) = \arg[H(\Omega)]$.
- 16. Assuming a sampling rate of F_s samples per second at the stretch mixer output, convert the analog frequency response $H(\Omega)$ of the previous problem to an equivalent discrete-time frequency response $H(\omega)$. Also give the expression for $H(\omega)$ in the particular case when F_s is chosen to match the stretch mixer output bandwidth of Eq. (4.111).
- 17. Explicitly compute the loss in processing gain LPG and the processing loss PL as a function of K for a triangular window of odd length K + 1 (so K is even) defined according to

$$w[k] = \begin{cases} 2k/K, & 0 \le k \le K/2 \\ 2-2k/K, & K/2 \le k \le K \\ 0, & \text{otherwise} \end{cases}$$

Numerically evaluate the result for K = 4 and K = 20 and give the answers in dB. What are the asymptotic values in dB for *LPG* and *PL* as $K \rightarrow \infty$? The following facts may be useful (be careful about the limits):

$$\sum_{k=1}^{n} k = n(n+1)/2, \quad \sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$$

(*Hint:* sum just the first half of the triangle, then use symmetry to get the sum of the whole function. Be careful not to double-count any samples.)

18. Consider the array steering factor $E(\theta_0)$ of Eq. (4.123) and use the weights given in Eq. (4.124) with $|a_n| = 1$ for all *n*. Assume the phases of the weights are computed for a wavelength λ_0 and steering angle θ_0 , but the waveform bandwidth is approximately 10 percent of the nominal frequency so that the effective wavelength varies over the range of $(1 \pm 0.05)\lambda_0$. Derive an equation that gives the new angle θ at which $E(\theta)$ will be maximum in terms of λ_0 , θ_0 , and the actual wavelength λ . When the actual wavelength is 5 percent larger than λ_0 and the design steering angle is $\theta_0 =$
10°, what will be the actual steering angle (angle of the maximum of $E(\theta)$)? Repeat for $\theta_0 = 30^\circ$ and 70° .

- 19. Compute the integrated sidelobe ratios for the Barker codes in <u>Table 4.1</u>.
- 20. Determine the chip length and pulse length of a biphase-coded waveform for a pulsed radar to meet the following requirements:
 - a. Rayleigh range resolution = 0.3 meter.
 - b. Pulse compression gain > 15 dB.
 - c. Maximum allowable blind range within first range ambiguity = 50 meters.
 - d. Less than one-quarter cycle of Doppler phase rotation across the pulse for Doppler shifts up to 2000 Hz.

One or both parameters may have a range of allowable values. Give the full range if this is the case.

- 21. Consider Barker, MPS, and pseudorandom biphase codes. State whether each code type can meet the requirements of the previous problem. If not, state the reason; if so, state at least one specific length that will work.
- 22. Compute explicitly the $N = 4 = 2^2$ Frank code. What is the sequence of phases ϕ_n in the code (expressed as an angle in radians, e.g., 0, $\pi/3$, etc.)? Compute the autocorrelation function of the detected code sequence exp $(j\phi_n)$ explicitly by hand. Sketch the magnitude of the autocorrelation function. What is the peak sidelobe level, relative to the peak of the autocorrelation function, in dB?
- 23. Repeat the previous problem for an M = 4 P4 code. (Be sure to use the complex autocorrelation function.)
- 24. Equation (4.134) expressed the continuous autocorrelation function $s_x(t)$ of a phase-coded waveform for $t = k\tau_c + \eta$ as a linear interpolation between the possibly complex discrete autocorrelation values $s_A[k]$ and $s_A[k+1]$ of the code sequence $s_A[k]$. Show that this linear interpolation of the complex values also linearly interpolates the real and imaginary parts of $s_A[k]$ and $s_A[k+1]$, but that the magnitude of the interpolated value is not the linear interpolation of the magnitudes of $s_A[k]$ and $s_A[k+1]$.
- 25. Some waveform/matched filter pairs are more sensitive to Doppler mismatch ("less Doppler tolerant") than others. Consider three different waveforms, all using pulses of length τ seconds: a single simple pulse, a single LFM pulse with $\beta \tau = 1000$, and a pulse burst composed of 30 simple pulses of length τ with a PRI of 10τ seconds. Denote the time of the peak matched filter output when there is no Doppler shift as t_{max} . Suppose a target with a Doppler shift of $F_D = 1/\tau$ Hz is present. The matched filter does not compensate for this Doppler shift. For each waveform, what will

be the magnitude of the matched filter output waveform at $t = t_{max}$ compared to the value at t_{max} when there is no Doppler shift? Which of these waveforms is most Doppler tolerant in this case? Which is least Doppler tolerant?

⁴ A slightly more complicated version of this result holds when $T \le 2\tau$, see (Levanon and Mozeson, 2004).

⁶ The lower case w is used for the frequency-domain window function w(F) to emphasize that the multiplying function is the window function itself (e.g., a Hamming window) rather than its Fourier transform.

 7 Array antennas can also be steered using time delay units at each element or a combination of phase steering within a subarray and time delay steering across subarrays. Pure time-delay-steered arrays do not suffer antenna steering errors due to wideband waveforms.

 $\frac{8}{2}$ While "fine" is preferred in this text to "high" to describe small values of resolution, the term "high range resolution profile" is well-established in the literature.

⁹ This figure may appear to violate the earlier claim that the continuous autocorrelation function is a linear interpolation between the discrete autocorrelation values of the code sequence $\exp(j\phi_n)$. However, linear interpolation of the complex values does not result in linear interpolation of the magnitude; see Prob. 4.24.

¹ This phase shift term was absorbed into the effective reflectivity ρ' in <u>Chap. 2</u>.

² Note that a frequency component of $1/\tau$ hertz goes through exactly one full cycle during a pulse of duration τ seconds.

³ Some authors define the term "ambiguity function" as $|\hat{A}(t, F_D)|^2$ or as $\hat{A}(t, F_D)$ itself. Also, some authors define the ambiguity function as $|\int_{-\infty}^{\infty} x(s)\exp(j2\pi F_D s)x^*(s+t) ds|$ instead of $|\int_{-\infty}^{\infty} x(s)\exp(j2\pi F_D s)x^*(s-t) ds|$. The definition used here is consistent with that given in Rihaczek (1996).

⁵ Group delay in seconds is the negative of the derivative of the frequency-domain phase function $\Phi(\Omega)$ of $H(\Omega) = |H(\Omega)| \exp[j\Phi(\Omega)]$ with respect to Ω . It is a measure of the filter delay for inputs of a given frequency. See Oppenheim and Schafer (2010).

CHAPTER 5

Doppler Processing

Doppler processing is the term applied to filtering or spectral analysis of the signal received from a fixed range over a period of time corresponding to several pulses. The purpose is generally to suppress clutter returns and to enable the detection of targets in the presence of significant clutter.

Figure 5.1*a* shows a notional scenario where a down-looking stationary radar observes four moving targets in a ground clutter background. The gray dashed lines represent range bins. Receding targets are in bins 4, 11, and 18, while an approaching target shares bin 11. Figure 5.1*b* is a stylized representation of the range-Doppler power spectrum that might result from this scenario. The light gray background represents the receiver noise floor, which is spread uniformly throughout the range-Doppler map. The band of energy extending through all of the range bins represents the ground clutter echo. Because the radar is stationary, the clutter is centered at zero Doppler shift, and its power fades with range in accordance with the range equation. The four small ovals represent the returns from the moving targets. Their echo energy is located in the appropriate range bins. Their Doppler coordinates depend on the direction and speed of each target with respect to the radar.



FIGURE 5.1 Notional range-Doppler spectrum for a down-looking stationary radar. Noise, clutter, and target components are shown: (*a*) scenario, (*b*) range-Doppler echo power distribution, (*c*) Doppler spectrum of range bin 11.

Figure 5.1c highlights the notional Doppler spectrum of the slow-time signal from range bin 11, containing the two middle moving targets. Since the slow-time data is sampled at a rate equal to the pulse repetition frequency of the radar, this spectrum is periodic with a period equal to the PRF, so only the principal period from -PRF/2 to +PRF/2 is shown. The portion of the spectrum where clutter is the dominant interference is often termed the *clutter region*. The width β_c in hertz of the clutter region is determined by the actual clutter motion, the RF frequency, and the PRF. The portion where noise is the dominant interference is called the *clear region*; note that the clear region is defined at the transition between the clutter and clear regions; in the skirt region, both noise

and clutter are significant interference sources. The moving targets appear in the spectrum at Doppler shifts consistent with their radial velocities relative to the radar.

Doppler processing is most often of interest when the relative amplitudes of the clutter, target, and noise signals are as shown: the target returns are above the noise floor (signal-to-noise ratio $SNR \gg 1$), but weaker than the clutter (signal-to-clutter ratio $SCR \ll 1$). In this case, targets cannot be detected reliably based on amplitude in the slow-time domain alone because the presence or absence of the target makes little difference to the total signal power. Doppler processing is used to separate the target and clutter signals in the frequency domain. The clutter can be filtered out, leaving the target return as the strongest signal present, or the spectrum can be computed explicitly so that targets outside of the clutter region can be located by finding frequency components that significantly exceed the noise floor.

In this chapter the two major classes of Doppler processing, *moving target indication* (MTI) and *pulse Doppler* processing, are described. In the terminology used here MTI refers to the case where the slow-time signal is processed entirely in the time domain. Pulse Doppler processing refers to the case where the signal is processed in the frequency domain.¹ As will be seen, MTI processing produces limited information at very low computational cost; pulse Doppler processing requires more computation but produces more information and greater *signal-to-interference ratio* (SIR) improvement. Only coherent Doppler processing using digital implementations is considered since this is the approach taken in most modern radars. Good general references include Richards (2010) and Schleher (2010). Alternative systems using noncoherent Doppler processing and implementations based on analog technologies are described in Eaves and Reedy (1987), Nathanson (1991), and Schleher (2010).

5.1 Moving Platform Effects on the Doppler Spectrum

The notional Doppler spectrum of Fig. 5.1 represents a very simple case. While it is realistic for some scenarios, the Doppler spectrum for a given range bin can be greatly complicated by factors such as a moving radar platform, or range and Doppler ambiguities caused by aliasing of the target signatures.

The effect of the PRF in the Doppler dimension is straightforward. The PRF establishes the width of one period of the Doppler spectrum as discussed above. Clutter or target signals having Doppler shifts outside of the range \pm *PRF*/2 will alias into that interval. Figure 5.2 illustrates how the spectrum of Fig. 5.1c might look if the PRF were reduced by 40 percent. The clutter spectrum is unchanged but now represents a larger fraction of the total spectrum width. That is, the clear region is a smaller percentage of the spectrum width

while the clutter region is a larger percentage. The target originally at velocity v_1 is still unaliased and so still appears at Doppler shift $2v_1/\lambda$. The Doppler shift of the target originally at the (negative) velocity v_2 was outside the new spectrum range $\pm PRF'/2$ and so aliases to a new apparent Doppler shift of $2v_2/\lambda + PRF'$.



FIGURE 5.2 Effect of a 40 Percent PRF reduction on the Doppler spectrum of <u>Fig. 5.1*c*</u>.

Now consider the effect of platform motion on the range-Doppler map observed from an airborne or spaceborne platform. As described in Chap. 3, the Doppler shift observed for any target or clutter scatterer is increased by the platform-induced Doppler shift component of $F_D = 2v \cos \psi / \lambda$ Hz, where ψ is the cone angle between the platform velocity vector and the line of sight (LOS) to the scatterer. Figure 5.3a shows an aircraft in level flight with a forward but down-looking radar illuminating ground clutter. In this case the cone angle ψ is the depression angle ψ_{MLC} shown. While most of the echo energy will be in the radar mainbeam, it is common to have measurable (above the noise floor) clutter echo through both forward-and rear-looking sidelobes as well. Ground clutter echoes received from different directions will appear at both different Doppler shifts due to different cone angles, and at different ranges due to different distances to the ground. Target echoes will again appear at the appropriate Doppler shift and range bin, where the Doppler shift now is proportional to the total radial velocity due to both the platform motion and the target motion.



FIGURE 5.3 Notional Doppler spectrum of ground clutter observed from a moving radar platform: (*a*) sources of clutter return, (*b*) notional clutter spectrum in the boresight range bin, (*c*) notional spectrum in the altitude line range bin.

As shown in Fig. 5.3*a*, the mainlobe clutter (MLC) will occur at some positive Doppler shift and the appropriate range bins. Clutter echo received from below the aircraft will occur at a near range bin corresponding to the aircraft altitude and at zero Doppler shift. This clutter component is termed the *altitude line* (AL). Although it is transmitted and received through the radar sidelobes, the altitude line nonetheless tends to be relatively strong due to the relatively short vertical range and the high reflectivity of most clutter at normal

incidence. Other ranges may exhibit clutter seen through the antenna sidelobes, called *sidelobe clutter* (SLC). SLC is weaker than MLC because it is observed through the lower-gain sidelobes. It is also weaker than the AL due to longer ranges and non-vertical incidence. SLC can potentially extend from Doppler shifts approaching $+2\nu/\lambda$ (directly ahead of the radar) to $-2\nu/\lambda$ (directly behind the radar), though the amplitude will drop rapidly due to increasing range at those maximum Doppler shifts.

Figure 5.3*b* is a sketch of the notional MLC spectrum at the range bin corresponding to the radar boresight, assuming the PRF is high enough to avoid any aliasing. The MLC will be centered at $F_{MLC} = 2v\cos\psi_{MLC}/\lambda$, where ψ_{MLC} is the depression angle from the horizontal velocity vector to the radar boresight vector. The width β_{MLC} of the MLC will be approximately the sum of the intrinsic clutter spectrum width β_C and the mainlobe widening due to platform motion of $\beta_D = 2v\theta\sin\psi_{MLC}/\lambda$, where θ is the antenna azimuth beamwidth in radians. Figure 5.3*c* is a similar sketch for the altitude line, which would occur at a closer range bin. Range bins nearer than that of the AL can contain only noise.

Figure 5.4 is a more detailed simulated range-Doppler clutter spectrum for a down-looking airborne radar viewing flat homogeneous ground clutter. The aircraft is traveling level at an altitude of 10,000 feet (3048 m) and a velocity of 300 mph (134.1 m/s). The RF is 10 GHz. The radar antenna has a 2.5° beamwidth, and its boresight is scanned 45° clockwise (right) of the velocity vector in azimuth and 10° down in elevation. The boresight will then intersect the ground at a range of 17.3 km. The component of velocity along the boresight (i.e., $v\cos\psi_{MLC}$) is 93.4 m/s, giving a Doppler shift at the center of the mainlobe clutter of 6.23 kHz. The maximum clutter Doppler shift that could be observed is ±89.4 kHz, corresponding to relative velocities of ±134.1 m/s, the aircraft velocity.







FIGURE 5.4 Simulated clutter range-velocity spectrum for an airborne radar viewing ground clutter: (a) perspective view, (b) overhead view, (c) clutter power versus range, integrated over velocity, (d) clutter power versus velocity, integrated over range. See text for details.

The simulation used a constant- γ model for clutter reflectivity, a sinc² two-

way antenna power gain, and accounted for receiver noise, clutter cell size and power variation with range, and other factors. The resulting range-Doppler clutter spectrum, not including any aliasing in range or Doppler due to the PRF, is shown in Fig. 5.4. Part a of the figure gives a 3D view, while part b is shows a perhaps more useful 2D view. It is readily seen that the maximum clutter power occurs at the mainlobe peak, centered at 93.4 m/s and 17.3 km. The high antenna sidelobes (no weighting was used) create visible rings of clutter power around the mainlobe. The shortest-range clutter occurs at zero velocity and just over 3 km; this is the altitude line, and it is also relatively strong. It can be seen that at any longer range, which must be at some angle other than directly below the radar, the clutter must have a nonzero velocity, so the clutter energy spreads to the maximum relative velocities of ± 134.1 m/s (Doppler shifts of ± 89.4 kHz) as the range increases. In these figures, positive velocities (Doppler shifts) must come from clutter in front of the aircraft, while negative velocities come from clutter behind the aircraft. Figure 5.4c of the figure show the total clutter power versus range, integrated over velocity. This illustrates the high AL return, the fall-off of the clutter with range, and the large MLC return before the clutter falls off again. Figure 5.4d shows the power versus velocity, integrated over range, and illustrates the MLC, the relatively strong AL and the falloff at higher velocities, which imply greater cone angle, longer ranges, and shallower grazing angles with attendant lower clutter reflectivity.

These examples only hint at the complexity of the range-Doppler spectrum. The subject will be revisited in Sec. 5.3 as part of the discussion of pulse Doppler processing. Despite the potential complexity, the simple spectrum of Fig. 5.1*c* has all the features needed to introduce MTI.

5.2 Moving Target Indication

Figure 5.5 illustrates a two-dimensional data matrix formed from the coherently demodulated baseband returns from a series of M pulses comprising one coherent processing interval (CPI). This matrix corresponds to one two-dimensional horizontal plane from the radar datacube of Fig. 3.8. Thus, a similar matrix exists for each phase center in the antenna system. In a single-aperture system, or at the point in an array where the data from multiple phase centers have been combined, there is only a single two-dimensional data matrix as shown.²



FIGURE 5.5 Notional two-dimensional data matrix. Each cell is one complex number.



FIGURE 5.6 MTI filtering and detection process.

The samples in each column are successive samples of the returns from a single pulse, i.e., successive range bins. Each element of a column is one complex number representing the real and imaginary (I and Q) components of one range sample. Consequently, each row represents a series of measurements from the same range bin over successive pulses. The sampling rate in the fast time or range dimension (vertical column in Fig. 5.5) is at least equal to the transmitted pulse bandwidth and therefore is on the order of hundreds of kilohertz to tens or even hundreds of megahertz. The slow-time or pulse number dimension (horizontal row in Fig. 5.5) is sampled at the pulse repetition interval of the radar. Thus the sampling rate in this dimension is the PRF and is on the order of ones to tens, and sometimes hundreds of kilohertz. As indicated by the shading, Doppler filtering operates on rows of this matrix.

MTI processing applies a linear filter to the slow-time data sequence in

order to suppress the clutter component. Figure 5.6 illustrates the process. The type of filtering needed can be understood by considering Fig. 5.7. In this figure, it is again assumed that knowledge of the platform motion and scenario geometry has been used to center the clutter spectrum at zero Doppler frequency. Clearly, some form of highpass filter, as suggested by the notional frequency response $|H(F_D)|$ shown, is needed to attenuate the clutter without filtering out moving targets in the clear portions of the Doppler spectrum.



FIGURE 5.7 The concept of moving target indication filtering: (*a*) Doppler spectrum and MTI filter frequency response, (*b*) Doppler spectrum after filtering.

The output of the highpass MTI filter will be a new slow-time signal containing components due to noise and, possibly, one or more targets. This signal is passed to a detector typically consisting of a matched filter followed by a threshold test. If the peak matched filter output exceeds the threshold (i.e., its energy is too great to likely be the result of noise alone), a target will be declared. Note that in MTI processing, the presence or absence of a moving target in the range bin of interest is the only information obtained. The filtering process of Fig. 5.6 does not provide any estimate of the Doppler frequency at which the target energy causing the detection occurred, or even of its sign; thus, it "indicates" the presence of a moving target but does not determine whether the

target is approaching or receding, or at what radial velocity. Furthermore, it provides no indication of the number of moving targets present. If multiple moving targets are present in the slow-time signal from a particular range the result will still be only a "target present" decision from the detector. On the other hand, MTI processing is very simple and computationally undemanding. Despite its simplicity, a well-designed MTI can improve the SCR by several to sometimes 20 or more decibels in some clutter conditions.

5.2.1 Pulse Cancellers

The major MTI design decision is the choice of the particular MTI filter to be used. MTI filters are typically low order, simple designs. Indeed, some of the most common MTI filters are based on very simple heuristic design approaches. Suppose a fixed radar illuminates a moving target surrounded by perfectly stationary clutter. The clutter component of the echo signal from each pulse would be identical, while the phase of the moving target component would vary due to the changing range. Subtracting the echoes from successive pairs of pulses would cancel the clutter components completely. The target signal would not cancel in general due to the phase changes.

This observation motivates the two-pulse MTI canceller, also referred to as the single or first-order canceller. Figure 5.8*a* illustrates the flowgraph of a two-pulse canceller. The input data are a sequence of baseband complex (I and Q) data samples from the same range bin over successive pulses, forming a discrete-time sequence y[m] with an effective sampling interval *T* equal to the pulse repetition interval. The discrete time transfer function (also called the *system function*) of this linear finite impulse response (FIR, also called *tapped delay line* or *nonrecursive*) filter is simply $H(z) = 1 - z^{-1}$. The frequency response as a function of analog frequency *F* in hertz is obtained by setting $z = e^{j}$



FIGURE 5.8 Flowgraphs and transfer functions of basic MTI cancellers: (a) two-pulse canceller, (b) three-pulse canceller.

$$H(F) = (1 - z^{-1})|_{z=e^{j2\pi FT}}$$

= 1 - e^{-j2\pi FT}
= e^{-j\pi FT} (e^{+j\pi FT} - e^{-j\pi FT})
= 2 j e^{-j\pi FT} \sin(\pi FT)
(5.1)

It is common to work with normalized frequency f = FT cycles per sample or with the radian equivalent, $\omega = \Omega T = 2\pi FT$ radians per sample. In terms of normalized radian frequency the frequency response of the two-pulse canceller is

$$H(\omega) = 2je^{-j\omega/2}\sin(\omega/2)$$
 radians per sample (5.2)

Recall that as F ranges from -1/2T to +1/2T, f ranges from -0.5 to +0.5 and ω from $-\pi$ to $+\pi$.

Figure 5.9*a* plots the magnitude of the frequency response of the two-pulse canceller. The filter does indeed have a null at zero frequency to suppress the clutter energy. Spectral components representing moving targets may either be partially attenuated or amplified, depending on their location on the Doppler frequency axis. As with all discrete-time filters, the frequency response is periodic with a period of 1 in the normalized cyclical frequency variable *f*, corresponding to a period of 2π in the normalized frequency variable ω or a period of 1/T = PRF in actual frequency in hertz. The shaded area highlights the principal period from -PRF/2 to +PRF/2; this is all that is normally plotted. Considering only this frequency range, it is clear that the frequency response is highpass in nature. The implications of the periodicity will be considered in <u>Sec. 5.2.4</u>.



FIGURE 5.9 Frequency response of basic MTI cancellers: (a) two-pulse canceller, (b) three-pulse canceller.

The two-pulse canceller is a very simple filter. Its implementation requires no multiplications and only one subtraction per output sample. As Fig. 5.9*a* shows, however, it is a poor approximation to an ideal highpass filter for clutter suppression. The next traditional step up in MTI filtering is the three-pulse (second-order or double) canceller obtained by cascading two two-pulse cancellers. The flowgraph and frequency response are shown in Figs. 5.8*b* and 5.9*b*. The three-pulse canceller clearly improves the null breadth in the vicinity of zero Doppler, but it does not improve the consistency of response to moving targets at various Doppler shifts away from zero Doppler. It requires only two subtractions per output sample.

Despite their simplicity, the two-and three-pulse cancellers can be very effective against clutter with moderate-to-high pulse-to-pulse correlation. Figure 5.10 shows a simulated clutter sequence formed by passing a white noise sequence through a filter with a Gaussian power spectrum having a standard

deviation of $\sigma_f = 0.05$ on a normalized frequency scale; i.e., the standard deviation is 5 percent of the full width of the principal period of the spectrum. Also shown are the outputs from the two- and three-pulse cancellers. The power in the two-pulse canceller output has been reduced by 13.4 dB relative to the unfiltered clutter sequence. For the three-pulse canceller, the reduction is 21.9 dB.



FIGURE 5.10 Clutter cancellation using two-and three-pulse cancellers. See text for details.

The idea of cascading two-pulse canceller sections to obtain higher order filters can be extended to the *N*-pulse canceller obtained by cascading N - 1 two-pulse canceller sections. The transfer function of the *N*-pulse canceller is therefore

$$H_N(z) = (1 - z^{-1})^{N-1}$$
(5.3)

The corresponding impulse response coefficients of the filter are given by the binomial series

$$h_N[m] = (-1)^m \binom{N-1}{m} = (-1)^m \frac{(N-1)!}{m!(N-1-m)!}, \quad m = 0, \dots, N-1$$
(5.4)

Other types of digital highpass filters could also be designed for MTI

filtering. For example, an FIR highpass filter could be designed using standard digital filter design techniques such as the window method or the Parks-McClellan algorithm. To be suitable as an MTI filter, the FIR filter frequency response should have a zero at F = 0. In terms of the four recognized classes of FIR digital filters (see Oppenheim and Schafer, 2010, Sec. 5.7), the MTI filter can be either type I (even order with symmetric impulse response) or type IV (odd order with antisymmetric impulse response). The transfer functions of type IV filters always have a zero at z = 1 so that the frequency response is zero at f = 0, ideal for an MTI filter. The two-pulse canceller (which has an *order* of 1) is an example of a type IV filter. Type I filters do not necessarily have a zero at f = 0, but can be made to have one by requiring that the sum of the impulse response coefficients h[m] equal zero. The three-pulse canceller is an example of a type I filter that has been designed to be suitable as an MTI filter.

Type II and III filters are unsuitable because they always have a zero at z = -1, corresponding to a frequency response null at a normalized frequency of f = 0.5; this creates extra undesirable blind speeds (see Sec. 5.2.4). Alternatively, *infinite impulse response* (IIR) highpass filters could be designed. Many operational radar systems, however, use two- or three-pulse cancellers for the primary MTI filtering due to their computational simplicity.

5.2.2 Vector Formulation of the Matched Filter

The *N*-pulse cancellers described previously can be remarkably effective and have been widely used. Nonetheless, they are motivated by heuristic ideas. Can a more effective pulse canceller be designed? Since the goal of MTI filtering is to improve the signal-to-clutter ratio, it should be possible to apply the matched filter concept of <u>Chap. 4</u> to this problem. To do so for discrete-time signals, it is convenient to first restate the matched filter using vector notation. This will also aid in generalizing the matched filter somewhat.

Consider a complex signal column vector $\mathbf{y} = [y[m] \ y[m-1] \cdots y[m-N + 1]]^T$ and a filter weight vector $\mathbf{h} = [h \ [0] \cdots h[N-1]]^T$. The superscript *T* represents matrix transpose so that \mathbf{y} and \mathbf{h} are *N*-element column vectors. A single output sample *z* of the filter is given by $z = \mathbf{h}^T \mathbf{y}$. The power in the output sample is given by

Power in
$$z = |z|^2 = z^* z^T = \mathbf{h}^H \mathbf{y}^* \mathbf{y}^T \mathbf{h}$$
 (5.5)

where a superscript *H* represents the Hermitian (complex) transpose.

The matched filter is obtained by finding the filter coefficient vector **h** that maximizes the SIR of the filtered data. Denote the desired target signal vector by **t** and the interference vector by **w**, so that $\mathbf{y} = \mathbf{t} + \mathbf{w}$. The interference is a random process, but it is *not* assumed to be white or Gaussian, thus allowing modeling of both noise and clutter. The filtered signal and interference are,

respectively, $\mathbf{h}^T \mathbf{t}$ and $\mathbf{h}^T \mathbf{w}$. The power in the signal component is therefore $\mathbf{h}^H \mathbf{t}^* \mathbf{t}^T \mathbf{h}$, and in the interference component it is $\mathbf{h}^H \mathbf{w}^* \mathbf{w}^T \mathbf{h}$. Because the interference power is a random variable (RV), its expected value is used to get meaningful results. The expected value of $\mathbf{w}^* \mathbf{w}^T$ is the interference covariance matrix $\mathbf{S}_{\mathbf{I}}$

$$\mathbf{S}_{\mathbf{I}} \equiv \mathbf{E}\{\mathbf{w}^* \mathbf{w}^T\}$$
(5.6)

It follows also that $S_I^T = S_I^*$, $S_I = S_I^H$, and $(S_I^{-1})^T = (S_I^{-1})^*$. With this definition the SIR becomes

$$SIR = \frac{\mathbf{h}^{H} \mathbf{t}^{*} \mathbf{t}^{T} \mathbf{h}}{\mathbf{h}^{H} \mathbf{S}_{I} \mathbf{h}}$$
(5.7)

As in <u>Chap. 4</u>, the filter **h** that maximizes <u>Eq. (5.7)</u> is found using the Schwarz inequality, which in a form suitable for vector-matrix manipulations is

$$|\mathbf{p}^{H}\mathbf{q}|^{2} \le ||\mathbf{p}||^{2} ||\mathbf{q}||^{2}$$
(5.8)

where $\|\mathbf{p}\| = \sqrt{\mathbf{p}^H \mathbf{p}}$ and equality occurs if and only if $\mathbf{p} = k\mathbf{q}$ for some scalar constant k. To apply Eq. (5.8), first note that the matrix SI will be positive definite so that it can be factored into the form $S_I = \mathbf{A}^H \mathbf{A}$ for some matrix \mathbf{A} ; that is, \mathbf{A} is the "square root" of S_I in some sense. Define $\mathbf{p} = \mathbf{A}\mathbf{h}$ and $\mathbf{q} = (\mathbf{A}^H)^{-1}\mathbf{t}^*$. This choice is contrived so that $\mathbf{p}^H \mathbf{q} = \mathbf{h}^H \mathbf{t}^*$ and therefore $|\mathbf{p}^H \mathbf{q}|^2 = \mathbf{h}^H \mathbf{t}^* \mathbf{t}^T \mathbf{h}$, which is the numerator of Eq. (5.7). The Schwarz inequality then gives

$$\mathbf{h}^{H}\mathbf{t}^{*}\mathbf{t}^{T}\mathbf{h} \leq \|\mathbf{A}\mathbf{h}\|^{2} \|(\mathbf{A}^{H})^{-1}\mathbf{t}^{*}\|^{2}$$
$$= (\mathbf{h}^{H}\mathbf{S}_{\mathbf{I}}\mathbf{h})(\mathbf{t}^{T}\mathbf{S}_{\mathbf{I}}^{-1}\mathbf{t}^{*})$$
(5.9)

Rearranging Eq. (5.9) to isolate the SIR of Eq. (5.7) shows that

$$SIR \le \mathbf{t}^T \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^* \tag{5.10}$$

with equality only when $\mathbf{p} = k\mathbf{q}$. The optimal weight vector therefore satisfies $\mathbf{A}\mathbf{h}_{opt} = k(\mathbf{A}^H)^{-1}\mathbf{t}^*$, or, with k = 1

$$\mathbf{h}_{opt} = \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^*$$

(5.11)

Finally, the filtered data become, using $(\mathbf{S}_{\mathbf{I}}^{-1})^{\mathrm{T}} = (\mathbf{S}_{\mathbf{I}}^{-1})^{*}$

$$z = \mathbf{h}_{opt}^{T} \mathbf{y} = \mathbf{t}^{H} \left(\mathbf{S}_{\mathbf{I}}^{-1} \right)^{*} \mathbf{y}$$
(5.12)

Equation (5.11) is of fundamental importance and great versatility in signal processing generally and radar signal processing in particular. It will be used in this text not only for clutter filtering and pulse Doppler processing, but also for space-time adaptive processing, detection, and estimation.

5.2.3 Matched Filters for Clutter Suppression

The results of the previous section can now be applied to *N*-point (order N - 1) MTI filters that are more optimal than the *N*-pulse canceller. Equation (5.11) shows that the linear filter that optimizes detection performance in the presence of additive interference is the FIR matched filter, and that the coefficients of the filter are given by the matrix equation

$$\mathbf{h} = \begin{bmatrix} h[0] \\ \vdots \\ h[N-1] \end{bmatrix} = \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^*$$
(5.13)

where $\mathbf{h} = N \times 1$ column vector of filter coefficients

 $S_I = N \times N$ covariance matrix of the interference

 $\mathbf{t} = \frac{N \times 1}{\text{the filter is matched}}$ to which

and the asterisk denotes the complex conjugate.

To determine the optimal filter coefficients **h**, models are needed for the interference and target characteristics, S_I and **t**. For a simple example, consider the first-order (length N = 2) matched filter. Assume the interference w[m] consists of the sum of zero mean stationary white noise n[m] of power (variance) σ_n^2 and zero mean stationary colored clutter c[m] of power σ_c^2

$$w[m] = n[m] + c[m]$$
$$w = [w[m] \quad w[m-1]]^T$$
(5.14)

The clutter exhibits a correlation from one pulse to the next given by $E[c[m]c^*[m+1]] = E[c^*[m]c[m-1]] = s_c[1] = \sigma_c^2 \rho_c[1];$ the first normalized

autocorrelation coefficient ρ_c [1] is denoted as simply ρ for simplicity. The noise and clutter are uncorrelated with one another.

The definition of S_I is

$$\mathbf{S}_{\mathbf{I}} = \mathbf{E}\{\mathbf{w}^* \mathbf{w}^T\} \equiv \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$
(5.15)

Consider the s_{11} element. Using the fact that the noise and clutter are uncorrelated and that each is zero mean, it can be quickly concluded that

$$s_{11} = \mathbf{E}\{(n[m] + c[m])^* (n[m] + c[m])\}$$

= $\sigma_n^2 + \sigma_c^2$ (5.16)

Next consider the s_{12} element

$$s_{12} = \mathbf{E}\{(n[m] + c[m])^* (n[m-1] + c[m-1])\}$$

= $\mathbf{E}\{n^*[m]n[m-1]\} + \mathbf{E}\{n^*[m]c[m-1]\}$
+ $\mathbf{E}\{c^*[m]n[m-1]\} + \mathbf{E}\{c^*[m]c[m-1]\}$
(5.17)

Again, the noise and clutter are zero mean and uncorrelated and the noise is white, so that the expected values of the first three terms are zero. However, the last term becomes the pulse-to-pulse correlation of the clutter so that

$$s_{12} = 0 + 0 + 0 + \mathbb{E}\{c^*[m]c[m-1]\}\$$

= $\rho \sigma_c^2$ (5.18)

It is easy to see that $s_{22} = s_{11}$ and that $s_{21} = s_{12}^*$. Thus

$$\mathbf{S}_{\mathbf{I}} = \begin{bmatrix} \sigma_c^2 + \sigma_n^2 & \rho \sigma_c^2 \\ \rho^* \sigma_c^2 & \sigma_c^2 + \sigma_n^2 \end{bmatrix}$$
(5.19)

so that

$$S_{I}^{-1} = \frac{1}{\left(\sigma_{c}^{2} + \sigma_{n}^{2}\right)^{2} - |\rho|^{2} \sigma_{c}^{4}} \begin{bmatrix} \sigma_{c}^{2} + \sigma_{n}^{2} & -\rho\sigma_{c}^{2} \\ -\rho^{*}\sigma_{c}^{2} & \sigma_{c}^{2} + \sigma_{n}^{2} \end{bmatrix}$$
$$= k \begin{bmatrix} \sigma_{c}^{2} + \sigma_{n}^{2} & -\rho\sigma_{c}^{2} \\ -\rho^{*}\sigma_{c}^{2} & \sigma_{c}^{2} + \sigma_{n}^{2} \end{bmatrix}$$
(5.20)

where k absorbs constants resulting from the matrix inversion.

 S_I represents the available information on the interference. The elements on the main diagonal will always be identical and equal to the total interference power, which is the sum of the independent interference source powers. The off-diagonal elements represent the correlation properties of the interference over one PRI. Because the noise is white, it does not contribute to the offdiagonal elements, whereas the clutter does contribute provided $\rho = \rho_c[1] \neq 0$. More generally, an *N*th order filter will require the $N \times N$ covariance matrix and will involve correlation coefficients up to $\rho_c[N-1]$.

To finish computing **h**, a model is needed for the assumed target signal phase history **t**. For a target moving at a constant radial velocity, the expected target signal is just a discrete complex sinusoid at the appropriate Doppler frequency F_D . Following the discussion in Sec. 2.6.3, assume the waveform is a train of *M* simple pulses with PRI *T* and RF transmit frequency F_t . If the target is at a nominal range R_0 and is moving toward the radar at a radial velocity of *v* meters per second, the slow-time phase history will be of the form

$$y[m] = A \exp\left[-j\frac{4\pi}{\lambda}(R_0 - 2\beta_v R_s)\right] \exp\left[+j2\pi\left(\frac{2v}{\lambda}\right)mT\right]$$
$$= A \exp\left[+j2\pi F_D mT\right]$$
(5.21)

where $F_D = 2\nu/\lambda$ is the usual Doppler shift, $R_s \approx R_0$ is the range corresponding to the sampling time, and all constants are absorbed into *A* at each step.

Only *N* samples at a time of y[m] are of interest in analyzing an *N*-sample canceller. Assuming $N \le M$ and recalling the results of <u>App. B</u> on vector representation of linear filtering, the series of *N* samples ending at $m = m_0$, $\{y[m_0], y[m_0-1], ..., y[m_0-N+1]\}$, can be represented in vector form as

$$\mathbf{t} = A \begin{bmatrix} e^{j2\pi F_D m_0 T} & e^{j2\pi F_D (m_0 - 1)T} & \cdots & e^{j2\pi F_D (m_0 - N + 1)T} \end{bmatrix}^T$$

= $A \begin{bmatrix} 1 & e^{-j2\pi F_D T} & \cdots & e^{-j2\pi F_D (N - 1)T} \end{bmatrix}^T$
(5.22)

where the phase terms due to the delay to the first sample of interest, m_0 , have

been absorbed into A and the signal has been renamed t to emphasize that it is only the target component. For the specific case M = 2 this becomes

$$\mathbf{t} = A \begin{bmatrix} 1 & e^{-j2\pi F_D T} \end{bmatrix}^T$$
(5.23)

In practice, the target velocity and therefore Doppler shift are unknown; a target might be anywhere in the Doppler spectrum. The Doppler shift F_D is therefore modeled as a random variable with a uniform probability density function over [-PRF/2, +PRF/2) and the expected value of t is computed. The expected value of the constant 1 is, of course, 1. The expected value of the second component of t is

$$E\{e^{-j2\pi F_{D}T}\} = \frac{1}{PRF} \int_{-PRF/2}^{+PRF/2} e^{-j2\pi F_{D}T} dF_{D}$$
$$= \frac{1}{PRF} \int_{-PRF/2}^{+PRF/2} e^{-j2\pi F_{D}/PRF} dF_{D} = 0$$
(5.24)

The signal model then becomes simply

$$\mathbf{t} = A \begin{bmatrix} 1 & 0 \end{bmatrix}^T \tag{5.25}$$

Finally, combining Eqs. (5.20) and (5.25) in Eq. (5.13) gives the coefficients of the optimum two-pulse filter

$$\mathbf{h} = \begin{bmatrix} \sigma_c^2 + \sigma_n^2 & -\rho^* \sigma_c^2 \end{bmatrix}^T$$
(5.26)

In this equation the constant A has been dropped because it affects target, clutter, and noise equally and is therefore of no consequence.

To interpret this result, consider the case where the clutter is the dominant interference. Then σ_n^2 is negligible compared to σ_c^2 and $\mathbf{h} \approx [1 - \rho^*]^T$. Now suppose the clutter is highly correlated over one PRI so that ρ is close to one. Then $\mathbf{h} \approx [1 - 1]^T$, nearly the same as the two-pulse canceller. Despite its simplicity, the two-pulse canceller is therefore nearly a first-order matched filter for MTI processing when the clutter-to-noise (CNR) is high and the clutter is highly correlated over one PRI. In the limit of very high CNR and perfectly correlated clutter, the two-pulse canceller is exactly the first-order matched MTI filter.

The vector matched filter derivation of the optimum two-pulse MTI filter

given previously is easily extended to higher order MTI filters. As the order increases, the corresponding *N*-pulse canceller becomes a poorer approximation of the matched filter (Schleher, 2010).

It is interesting to consider the form of the optimum filter when the dominant interference is noise rather than clutter, that is, $\sigma_c^2 \ll \sigma_n^2$. In this case the optimum first-order MTI filter of Eq. (5.26) reduces to (ignoring overall scale factors again)

$$\mathbf{h} \approx \begin{bmatrix} 1 & 0 \end{bmatrix}^T \tag{5.27}$$

Equation (5.27) states that in the presence of completely *un*correlated interference and with no knowledge of the target velocity, the filter impulse response reduces to a single impulse, $h[m] = \delta[m]$. Since convolving any signal with $\delta[m]$ just returns the same signal, the filter does nothing. In the clutter-dominated case, the filter combined the two slow-time samples because, even though constructive interference of the target could not be guaranteed, the high correlation of the clutter did guarantee that the clutter signal would be suppressed. On average the overall effect was beneficial. In the noise-dominated case, there is still no guarantee that the target signal will be reinforced, and in addition there is now no guarantee that the noise at all.

The previous analysis assumes that the target Doppler shift is unknown and therefore considers all target Doppler frequencies equally likely. It is easy to modify the analysis to match the MTI filter to a specific Doppler shift or to the case where the target Doppler extends only over a portion of the Doppler spectrum. These alternative assumptions manifest themselves as alternate models for the desired signal vector \mathbf{t} . The second case is treated in Schleher (2010) and in Prob. 8; here the case of a known Doppler shift for the target and a two-pulse canceller is considered. The interference and signal models are exactly the same as given earlier except that now the target Doppler shift in \mathbf{t} is not a random variable but a specific, fixed value. Therefore, it is not necessary to take an expected value of \mathbf{t} . The filter coefficient vector is

$$\mathbf{h} = \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^{*} = \begin{bmatrix} \sigma_{c}^{2} + \sigma_{n}^{2} & -\rho\sigma_{c}^{2} \\ -\rho^{*}\sigma_{c}^{2} & \sigma_{c}^{2} + \sigma_{n}^{2} \end{bmatrix} \begin{bmatrix} 1 \\ e^{+j2\pi F_{D}T} \end{bmatrix}$$
$$= \begin{bmatrix} (\sigma_{c}^{2} + \sigma_{n}^{2}) - \rho\sigma_{c}^{2}e^{+j2\pi F_{D}T} \\ (\sigma_{c}^{2} + \sigma_{n}^{2})e^{+j2\pi F_{D}T} - \rho^{*}\sigma_{c}^{2} \end{bmatrix}$$
(5.28)

where again all overall constants have been dropped. While this result is easy

to implement if the interference statistics are known, it is difficult to interpret. However, in the noise-limited case $(\sigma_n^2 \gg \sigma_n^2)$, it reduces to

$$\mathbf{h} = \begin{bmatrix} 1 & e^{+j2\pi F_D T} \end{bmatrix}^T$$
(5.29)

Equation (5.29) shows that in this case the optimum filter adds the two target samples together with a phase correction to the second so that they add in phase. In other words, the filter performs a coherent integration of the two target samples.

5.2.4 Blind Speeds and Staggered PRFs

The frequency response of all discrete-time filters is periodic, repeating with a period of one in the normalized cyclical frequency, corresponding to a period of PRF = 1/T Hz of Doppler shift. Figure 5.9 illustrated this for the two- and three-pulse cancellers. Since MTI filters are designed to have a null at zero frequency, they will also have nulls at Doppler frequencies that are multiples of the pulse repetition frequency. Consequently, a target moving with a radial velocity that results in a Doppler shift equal to a multiple of the PRF will be suppressed by the MTI filter. Velocities that result in these unfortunate Doppler shifts are called *blind speeds* because the target return will be suppressed; the system is "blind" to such targets. From a digital signal processing point of view, blind speeds represent target velocities that will be aliased to zero velocity. Equivalently, they correspond to Doppler shifts that will be aliased to zero frequency.

For a given PRF, the unambiguous range is

$$R_{ua} = \frac{c}{2PRF}$$
(5.30)

The first blind speed is

$$v_b = \frac{\lambda PRF}{2} = \frac{c PRF}{2F_0}$$
(5.31)

The corresponding Doppler shift F_b simple equals the PRF. As the PRF is increased for a given RF, the unambiguous range decreases and the first blind speed increases. Figure 5.11 shows the unambiguous range-Doppler coverage regions that are possible. For example, each point on the line marked "1 GHz" represents a combination of R_{ua} and v_b corresponding to some PRF. The dotted lines mark one example, corresponding to 400 m/s for the first blind speed and a 56.25 km unambiguous range. Equation (5.31) can be used to see that these

values correspond to PRF = 2667 pulses per second.



FIGURE 5.11 Ambiguity-free range-Doppler coverage regions.

There is some confusion in the terminology regarding the meaning of "unambiguous velocity" or "unambiguous Doppler shift." The blind speed defined in Eq. (5.31) is the velocity interval between zeros of the MTI filter response. Generally, the range of velocities that are considered to be measurable without ambiguity is then $[-v_b/2,+v_b/2]$. The unambiguous velocity is then said to be $v_b/2$ because targets having a velocity magnitude greater than this will be aliased. In terms of Doppler shift, the blind Doppler shift is *PRF* Hz, allowing unambiguous measurements over the interval [-PRF/2, +PRF/2]. The unambiguous Doppler shift is considered to be *PRF*/2 Hz. However, in some sources the unambiguous velocity (Doppler shift) is taken to be v_b m/s (*PRF* Hz), so caution is needed in interpreting the literature. In this text, unambiguous velocity or Doppler shift will mean one-half of the first blind velocity or Doppler shift.

The unambiguous range could be thought of as the range *coverage* of the radar for a given PRF while the corresponding blind speed or Doppler could be considered the velocity or Doppler coverage. Equations (5.30) and (5.31) can be combined to show that for a given RF and single PRF the combined range-velocity coverage is a constant, independent of the PRF. The combined range-Doppler coverage is also independent of the RF:

$$R_{ua}v_b = \frac{c}{2PRF} \cdot \frac{\lambda PRF}{2} = \frac{\lambda c}{4}$$
$$R_{ua}F_b = \frac{c}{2PRF} \cdot PRF = \frac{c}{2}$$

(5.32)

Blind speeds could be avoided by choosing the PRF high enough so that the first unambiguous velocity exceeds any actual velocity likely to be observed for targets of interest. Unfortunately, higher PRFs also correspond to shorter unambiguous ranges. It is frequently not feasible to operate at a PRF that allows unambiguous coverage of both the range and velocity intervals of interest. For example, suppose a designer requires unambiguous range of 100 km and at least ± 112.5 m/s of unambiguous velocity coverage, corresponding to a 225 m/s blind speed. Figure 5.11 shows that the maximum RF at which this is possible is 1 GHz. If the radar is required to be at X band (10 GHz), the combination of 100 km unambiguous range coverage and 112.5 m/s unambiguous velocity coverage is not obtainable and some ambiguity must be accepted in range, Doppler, or both.

The use of *staggered PRFs* or *staggered PRIs* is an alternative approach that raises the first blind speed significantly with only a modest reduction in the unambiguous range (Levanon, 1988; Schleher, 2010). PRF staggering can be performed on either a pulse-to-pulse or CPI-to-CPI basis. The latter case is common in airborne pulse Doppler radars and is deferred to Sec. 5.3.8. Pulse-to-pulse stagger varies the PRI from one pulse to the next within a single CPI or dwell. One common approach is to cycle through a set of *P* preselected PRIs from one pulse to the next, repeating when all of the PRIs have been used. Figure 5.12 illustrates the pulse timing sequence for a case with P = 2. The resulting slow-time data for a given range bin is then passed through an MTI filter. As will be seen, this process has the advantage of achieving increased Doppler coverage with a single dwell.



FIGURE 5.12 Pulse sequence timing for two staggered PRIs.

One disadvantage of staggering is that the slow-time data are now a nonuniformly sampled sequence, making coherent Doppler filtering impractical and complicating analysis. Another is that ambiguous mainlobe clutter can cause large pulse-to-pulse amplitude changes as the PRI varies since the range of the second-time-around clutter that folds into each range cell will change as the PRI changes. Consequently, pulse-to-pulse PRI stagger is generally used only in low PRF modes where there are no range ambiguities.

Staggered PRI operation can be analyzed in terms of either the PRIs or the corresponding PRFs. The former is more direct and is used here. Consider a system using *P* staggered PRIs $\{T_0, T_1, ..., T_{P-1}\}$. The corresponding set of pulse repetition intervals is $\{PRF_p\} = \{1/T_p\}$. Assume that each of the PRIs is selected as an integer multiple of a base interval T_g ,

$$T_p = k_p T_g, \quad p = 0, \dots, P - 1$$
 (5.33)

with corresponding PRFs, $PRF_p = 1/T_p$ and $F_g = 1/T_g$. This is reasonable for many radar systems, where T_g may correspond to the fast-time sampling interval and the set of integers $\{k_p\}$ to the number of range bins in each PRI. The $\{k_p\}$ are called the *staggers*³ and the ratio k_m : k_p of any two of them is called a *stagger ratio*. In many practical cases the staggers are chosen to be relatively prime integers.

For a given PRI, any MTI filter will exhibit blind Doppler frequencies at all integer multiples of the corresponding PRF. Consequently, the first true blind Doppler frequency F_{bs} of a system using staggered PRIs will be the lowest frequency that is blind at all of the corresponding PRFs, i.e., the least common multiple (LCM) of the set

$$F_{bs} = \operatorname{lcm}(PRF_0, \dots, PRF_{P-1}) = \operatorname{lcm}(1/T_0, \dots, 1/T_{P-1})$$

= $F_g \operatorname{lcm}(1/k_0, \dots, 1/k_{P-1})$ (5.34)

A complete cycle through the set of PRIs takes a total period T_{tot} equal to the sum of each of the staggered PRIs

$$T_{\text{tot}} = \sum_{p=0}^{P-1} T_p = T_g \sum_{p=0}^{P-1} k_p$$
(5.35)

It is of interest to determine how much the blind Doppler of the staggered system is increased relative to a comparable unstaggered system. A reasonable choice for an unstaggered system to serve as a baseline reference is one whose PRI equals the average PRI of the staggered system. The time required to collect *N* pulses will then be approximately the same for the two systems. The average PRI is

$$T_{\text{avg}} = \frac{T_{\text{tot}}}{P} = \frac{T_g}{P} \sum_{p=0}^{P-1} k_p$$

(5.36)

The blind Doppler frequency that would be observed in an unstaggered waveform with this average PRI is

$$F_b = \frac{1}{T_{\text{avg}}}$$
(5.37)

Using Eqs. (5.34), (5.36) and (5.37) and noting that $F_gT_g = 1$ gives an expression for the first blind Doppler frequency of the staggered PRI system in terms of the staggers $\{k_p\}$ and the blind Doppler of the reference unstaggered system

$$\frac{F_{bs}}{F_b} = \frac{1}{P} \operatorname{lcm}(1/k_0, \dots, 1/k_{P-1}) \sum_{p=0}^{P-1} k_p$$
(5.38)

For example, a two-PRI system with a stagger ratio of 3:4 would have a first blind Doppler that is 3.5 times that of a system using a fixed PRI equal to the average of the two individual PRIs. If a third PRF is added to give the set of staggers $\{3, 4, 5\}$, the first blind Doppler will be four times that of the comparable unstaggered system.

These equations simplify if all of the staggers are indeed mutually prime. In that event, the LCM of the set of inverse staggers $\{1/k_p\}$ equals 1 (See Prob. 5.14). The blind Doppler shift [Eq. (5.34)] of the staggered system then equals F_g and the factor by which the blind Doppler frequency is expanded relative to the unstaggered case [Eq. (5.38)] is just the average of the staggers.

If a pure sinusoid $Aexp(j\Omega t)$ is input to a linear time-invariant (LTI) system, the output will be another pure sinusoid at the same frequency but with possibly different amplitude and phase, $Bexp(j\Omega t + \phi)$. However, if a pure sinusoid is sampled at nonuniform time intervals the resulting series of samples, if interpreted as a conventional discrete-time sequence, will *not* be equivalent to a uniformly sampled pure sinusoid at the appropriate frequency so that the sampled signal will contain multiple frequency components. Any subsequent processing, even though itself LTI, will still result in an output spectrum containing multiple frequency components. Thus, a system utilizing nonuniform time sampling is not LTI and the frequency response of a pulse-to-pulse staggered system does not exist in a conventional sense. Instead, an approach based on first principles can be used to explicitly compute the effect of a twopulse canceller on a complex sinusoid of arbitrary frequency and initial phase for the MTI filter structure of interest. Repeating for each possible sinusoid frequency, the effect of the combination of staggered sampling and MTI filtering can be determined for targets of different Doppler shifts (Roy and Lowenschuss,

1970; Levanon, 1988; Schleher, 2010).

A constant-PRI system with PRI *T* transmits pulse #m in an *N*-pulse sequence at time $t_m = mT$, m = 0, ..., N - 1. In a *P*-stagger system, the sequence of sampling times $\{t_m\}$ is

$$t_m = \begin{cases} 0, & m = 0\\ \sum_{p=0}^{m-1} T_{((p))p}, & m = 1, \dots, N-1 \end{cases}$$
(5.39)

The notation $((\cdot))_P$ denotes evaluation of the argument modulo *P*. Note that the $\{T_p\}$ are the sampling time *increments*, not the absolute sampling times, and that $t_m - t_{m-1} = T_{((m-1))P}$. The slow-time phase history of Eq. (5.21) for a constant PRI and constant-velocity target is easily generalized to the following form for a staggered PRI system:

$$y[m] = k \exp[+j 2\pi F_D t_m]$$
(5.40)

Now consider the two-pulse canceller network of Fig. 5.8*a*.⁴ Using Eqs. (5.39) and (5.40), the output z[m] = y[m] - y[m-1] can be written explicitly as

$$z[m] = k \exp[j2\pi F_0 t_m] - k \exp[j2\pi F_0 t_{m-1}]$$

= $ke^{j\pi F_0 (t_m + t_{m-1})} \{ \exp[+j\pi F_0 (t_m - t_{m-1})] - \exp[-j\pi F_0 (t_m - t_{m-1})] \}$
= $ke^{j\pi F_0 (t_m + t_{m-1})} \{ \exp[+j\pi F_0 T_{((m-1))p}] - \exp[-j\pi F_0 T_{((m-1))p}] \}$
= $2j ke^{j\pi F_0 (t_m + t_{m-1})} \sin(\pi F_0 T_{((m-1))p})$
(5.41)

The magnitude of the frequency response for the sampling + filter system can be defined as the square root of the ratio of the power of the filter output sequence to that of the input sequence. The power of each input sample is $|y[m]|^2$ $= |k|^2$. The power of the output samples $|z[m]|^2$ depends on the index *m* due to the varying PRIs. The average output power is therefore computed over one cycle of the staggered PRIs. The sum from m = 0 to P - 1 of $T_{((m-1))P}$ is the same as the sum of T_m so the expression for the squared magnitude of the average two-pulse canceller filter frequency response becomes

$$\left|H_{2,P}(F_0)\right|^2 = \frac{\frac{1}{P}\sum_{m=0}^{P-1} \left|z[m]\right|^2}{\left|x[m]\right|^2} = \frac{4k^2 \sum_{m=0}^{P-1} \sin^2(\pi F_0 T_m)}{Pk^2}$$

where the notation $|H_{N, P}(F)|^2$ indicates the power gain or attenuation of an *N*-pulse canceller using *P* staggers when the input is a sinusoid of frequency F_0 . Generalizing the specific F_0 to an arbitrary frequency *F* and renaming the summation index gives the squared magnitude of the frequency response of the two-pulse canceller with staggered PRIs

$$\left|H_{2,p}(F)\right|^{2} = \frac{4}{P} \sum_{p=0}^{P-1} \sin^{2}(\pi FT_{p}) = \frac{4}{P} \sum_{p=0}^{P-1} \sin^{2}(\pi F/PRF_{p})$$
(5.43)

The actual frequency in hertz rather than normalized frequency is used in Eq. (5.43) because the nonuniform sampling rate invalidates the usual definition of normalized frequency. The response of more general MTI filters can be obtained using a similar approach.

Figure 5.13*a* compares the frequency response of a two-pulse canceller using two (P = 2) PRIs versus the reference baseline single-PRF system. The staggered case uses PRIs of 4/3 ms and 1 ms (thus PRFs of 750 and 1000 pulses per second). $T_g = 1/3$ ms, giving $F_g = 3$ kHz. The set of staggers k_p is {3, 4}. The first blind Doppler shift F_{bs} occurs at the least common multiple of 750 and 1000 Hz, which is also 3000 Hz. The reference unstaggered PRF and blind Doppler F_b , which is the reciprocal of the average PRI $T_{avg} = 7/6$ ms, is 857.14 Hz. The baseline unstaggered response collected using $PRI = T_{avg}$ shows blind Doppler frequencies at integer multiples of 857.14 Hz. Thus, staggering the PRF has increased the blind Doppler frequency and thus the velocity and Doppler coverage by a factor of 3.5 (= 3000/857.14) consistent with Eq. (5.38).



FIGURE 5.13 Comparison of two-pulse canceller frequency response with unstaggered waveform and 3:4 staggered waveform: (*a*) Two PRIs in a 3:4 stagger ratio, (*b*) five PRIs in a 51:62:53:61:58 stagger ratio. The inset expands the portion of the graph near zero frequency.

It remains to determine the effect of the pulse stagger on the unambiguous range. The unambiguous range R_{uas} of the staggered-PRI system is simply the

shortest of the unambiguous ranges for each individual PRI

$$R_{uas} = \frac{c}{2} \min[\{T_p\}] = \frac{cT_g}{2} \min[\{k_p\}]$$
(5.44)

The unambiguous range for the reference unstaggered system is $R_{ua} = cT_{avg}/2$. The ratio is

$$\frac{R_{uas}}{R_{ua}} = \frac{(cT_g/2)\min[\{k_p\}]}{(cT_{avg}/2)} = \frac{P\min[\{k_p\}]}{\sum_{p=0}^{P-1} k_p}$$
(5.45)

which is 6/7 for this example, about a 14 percent reduction. The increase in the total range-Doppler coverage is the product of Eqs. (5.38) and (5.45),

$$\frac{R_{uas}F_{bs}}{R_{ua}F_{b}} = \min[\{k_{p}\}]lcm(\{1/k_{p}\})$$
(5.46)

For mutually prime staggers, $lcm(\{1/k_p\}) = 1$ and this reduces to simply min $[\{k_p\}]$, equal to a factor of three in the example above. Thus, the use of two PRIs with a 3:4 stagger ratio has reduced the unambiguous range by 14 percent, but has expanded the Doppler coverage by 350 percent and the combined range-Doppler coverage by 300 percent.

Figure 5.13*b* illustrates the frequency response obtained using $Tg = 100 \ \mu s$ and five mutually prime staggers in the ratio 51:62:53:61:58 with a two-pulse canceller.⁵ Only the staggered response is shown; the equivalent unstaggered response would have nulls every 175.4 Hz. The insert expands the power spectrum around zero Doppler shift to make the clutter null more visible. The first blind Doppler shift is at 10 kHz, a 57× increase as expected from Eq. (5.38). The unambiguous range reduction is 10.5 percent. The combined range-Doppler coverage is increased by 51×.

Additional examples, special cases suited to weather radar, and the use of both infinite impulse response (IIR) filters and time-varying filters for MTI are discussed in Shrader and Gregers-Hansen (2008). An alternative design approach based on randomized PRIs to extend the blind speed is discussed in Vergara-Domingues (1993).

5.2.5 MTI Figures of Merit

The goal of MTI filtering is to suppress clutter. In doing so, it also attenuates or amplifies the target return, depending on the particular target Doppler shift. The change in signal and clutter power then affects the probabilities of detection and false alarm achievable in the system in a manner dependent on the particular design of the detection system.

There are three principal MTI figures of merit in use. *Clutter attenuation* measures only the reduction in clutter power at the output of the MTI filter as compared to the input, but is simplest to compute. *Improvement factor* quantifies the increase in signal-to-clutter ratio due to MTI filtering; as such, it accounts for the effect of the filter on the target as well as on the clutter. *Subclutter visibility* is a more complex measure that also takes into account the detection and false alarm probabilities and the detector characteristic. Because of its complexity, it is less often used. In this chapter, attention is concentrated on clutter attenuation *CA* and improvement factor *I*.

There are several ways to approach the calculation of the improvement factor. These include frequency domain approaches using clutter power spectra and MTI filter transfer functions, autocorrelation functions of the input and output of the MTI filter, and the vector method. Each will be illustrated in turn, starting with the frequency domain approach, which is perhaps the most intuitive.

Clutter attenuation is simply the ratio of the clutter power at the input of the MTI filter to the clutter power at the output

$$CA = \frac{\sigma_{ci}^{2}}{\sigma_{co}^{2}} = \frac{\int_{-PRF/2}^{PRF/2} S_{c}(F) dF}{\int_{-PRF/2}^{PRF/2} S_{c}(F) |H(F)|^{2} dF}$$
(5.47)

where σ_{ci}^2 and σ_{co}^2 = clutter power at the filter input and output, respectively

 $S_c(F)$ = sampled clutter power spectrum

H(F) = discrete-time MTI filter frequency response

Since the MTI filter presumably reduces the clutter power, the clutter attenuation will be greater than one. In fact, it can be 13 dB or more in favorable conditions. However, it also depends on the clutter itself through $S_c(F)$. The shape of the clutter power spectrum and its spread in hertz are determined by the physical phenomenology and RF. The percentage of the digital spectrum width to which a given clutter spectrum is mapped depends on the PRF and therefore is determined by the system design. Consequently, a change in RF, PRF, or clutter power spectrum due to changing terrain or weather conditions will alter the achieved clutter cancellation.

Improvement factor *I* is defined formally as the signal-to-clutter ratio at the filter output divided by the signal-to-clutter ratio at the filter input, averaged over all target radial velocities of interest (IEEE, 2008). Considering for the

moment only a specific target Doppler shift, the improvement factor can be written in the form

$$I = \frac{(S/C)_{\text{out}}}{(S/C)_{\text{in}}} = \left(\frac{S_{\text{out}}}{S_{\text{in}}}\right) \left(\frac{C_{\text{in}}}{C_{\text{out}}}\right) = G \cdot CA$$
(5.48)

where *G* is the filter *gain*. Figure 5.9 makes clear that the effect of the MTI filter on the target signal is a strong function of the target Doppler shift. Thus, *G* is a function of target velocity, while clutter attenuation *CA* is not. The improvement factor is the signal processing gain G_{sp} in the radar range equation due to MTI filtering.

To reduce *I* to a single number instead of a function of target Doppler, the definition calls for averaging uniformly over all target Doppler shifts "of interest." If a target is known to be at a specific velocity, the improvement factor can be obtained by simply evaluating Eq. (5.48) at the known target Doppler. It is more common to assume the target velocity is unknown a priori and use the average target gain over all possible Doppler shifts, which is just

$$G = \frac{1}{PRF} \int_{-PRF/2}^{PRF/2} |H(F)|^2 dF$$
(5.49)

An alternative expression for the gain that is often easier to compute for simple MTI filters follows from converting Eq. (5.49) back to normalized frequency units and applying Parseval's theorem:

$$G = \frac{1}{PRF} \int_{-PRF/2}^{PRF/2} |H(F)|^2 dF = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega)|^2 d\omega = \sum_{m=-\infty}^{\infty} |h[m]|^2$$
(5.50)

For example, a two-pulse canceller has only two nonzero coefficients, +1 and – 1, giving immediately G = 2. Combining Eqs. (5.47) and (5.49) in gives the expression for the improvement factor

$$I = G \cdot CA = \frac{\frac{-PRF/2}{\int} S_c(F) dF}{\frac{PRF/2}{PRF/2}} \frac{|H(F)|^2 dF}{|H(F)|^2 dF}$$

$$\frac{PRF}{\frac{PRF/2}{PRF/2}} S_c(F) |H(F)|^2 dF$$
(5.51)

Equivalent expressions for improvement factor can be developed in terms

of the autocorrelation function of the clutter and the MTI filter impulse response (Levanon, 1988; Nathanson, 1991). For low-order filters such as two- or threepulse cancellers and clutter power spectra with either measured or analytically derivable autocorrelation functions, the resulting equations can be easier to evaluate than the frequency domain versions.

As an example of the autocorrelation approach, consider the output of the two-pulse canceller when the input is just clutter; this is c'[m] = c[m] - c[m - 1]. The expected value of the filter output power is

$$\mathbf{E}\{|c'[m]|^2\} = \mathbf{E}\{|c[m]|^2 - 2\operatorname{Re}\{c[m]c^*[m-1]\} + |c[m-1]|^2\}$$
(5.52)

where $\operatorname{Re}\{\cdot\}$ denotes the real part of the argument.⁶ Assuming that c[m] is stationary,

$$\mathbf{E}\{|c'[m]|^2\} = 2s_c[0] - 2\operatorname{Re}\{s_c[1]\}$$
(5.53)

where

$$s_c[k] \equiv \mathbf{E}\{c[m]c^*[m+k]\}$$

(5.54)

is the autocorrelation function of c[m]. Note that $s_c[k] = s_c^*[-k]$. Also define the normalized autocorrelation function

$$\rho_{c}[k] \equiv \frac{1}{s_{c}[0]} s_{c}[k]$$
(5.55)

The clutter attenuation component of the improvement factor in Eq. (5.48) can now be written as

$$CA = \frac{\mathbf{E}\{|c[m]|^2\}}{\mathbf{E}\{|c'[m]|^2\}} = \frac{s_c[0]}{2(s_c[0] - \operatorname{Re}\{s_c[1]\})}$$
$$= \frac{1}{2(1 - \operatorname{Re}\{\rho_c[1]\})}$$
(5.56)

As noted above, the gain for a two-pulse canceller is G = 2; thus the improvement factor for the two-pulse canceller is

$$I = CA \cdot G = \frac{1}{1 - \operatorname{Re}\{\rho_c[1]\}}$$
A similar analysis can be used to derive the improvement factor for a threepulse canceller; it is

$$I = CA \cdot G = \frac{1}{1 - \frac{4}{3} \operatorname{Re}\{\rho_c[1]\} + \frac{1}{3} \operatorname{Re}\{\rho_c[2]\}}$$
(5.58)

To see how these formulas are used, consider the case where the clutter spectrum is Gaussian with variance (in normalized radian frequency units) σ_{ω}^2 , that is, $S_c(\omega) = A \exp(-\omega^2/\sigma_{\omega}^2)$. Assuming $\sigma_{\omega} \ll \pi$ so that the continuous-time Fourier transform pair for Gaussian functions can be used to a good approximation, the normalized autocorrelation function for c [m] at lag *k* is (Richards, 2006)

$$\rho_c[k] \approx e^{-(\sigma_{\omega}k)^2/2} \tag{5.59}$$

Using Eq. (5.59) in Eqs. (5.57) and (5.58) gives the improvement factor for a Gaussian clutter spectrum with a two- or three-pulse canceller:

$$I = \begin{cases} \frac{1}{1 - e^{-\sigma_{\omega}^{2}/2}} & \text{(two-pulse canceller)} \\ \frac{1}{1 - \frac{4}{3}e^{-\sigma_{\omega}^{2}/2} + \frac{1}{3}e^{-2\sigma_{\omega}^{2}}} & \text{(three-pulse canceller)} \end{cases}$$
(5.60)

Table 5.1 shows the improvement factor predicted for two- and three-pulse cancellers for the case of a Gaussian clutter power spectrum of various spectral widths using Eq. (5.60). If the clutter spectrum is narrow compared to the PRF, the improvement factor can be 13 dB or more even for the simple two-pulse canceller. If the clutter spectrum is wide, much of the clutter power will be in the passband of the MTI highpass filter and the improvement factor will be small.

Standard Deviation of Clutter Power Spectrum, Hz	Improvement Factor, dB			
	Two-Pulse Canceller	Three-Pulse Canceller		
PRF/3	0.5	0.7		
PRF/10	7.5	12.5		
PRF/20	13.2	21.7		
PRF/100	24	51		

 TABLE 5.1
 Improvement Factor for Gaussian Clutter Power Spectrum

The third approach for computing the improvement factor uses the vector analysis techniques employed in determining the matched filter for MTI. For comparison with the autocorrelation analysis given previously, consider the case where $\sigma_n^2 = 0$ (clutter only) and $\mathbf{h} = [1 - 1]^T$ (two-pulse canceller). Improvement factor is the ratio of the signal-to-interference ratio at the filter output to the SIR at the filter input. While the optimum MTI filter was derived by averaging over possible target Doppler frequencies, in evaluating the improvement factor it is assumed that any specific target has a specific Doppler frequency. The improvement factor is calculated for that specific target Doppler frequencies. Since SIR at the input does not depend on Doppler frequency, it is sufficient to do the averaging on the output SIR.

Consider therefore the signal vector given by Eq. (5.23) and the clutter covariance matrix given by Eq. (5.19) (with $\sigma_n^2 = 0$). The input SIR is just $|A|^2/\sigma_c^2$. Equation (5.7) gave an explicit expression for the output SIR. The numerator of this expression is

$$\mathbf{h}^{H} \mathbf{t}^{*} \mathbf{t}^{T} \mathbf{h} = |A|^{2} [1 - 1] \begin{bmatrix} 1 & e^{-j2\pi F_{D}T} \\ e^{+j2\pi F_{D}T} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$= 2 |A|^{2} (1 - \operatorname{Re}\{e^{j2\pi F_{D}T}\})$$
(5.61)

This is the two-pulse canceller MTI filter output signal power for a target at Doppler shift F_D Hz. Averaging over all target Doppler shifts gives for the numerator

$$\mathbf{E}_{F_{D}}\{\mathbf{h}^{H}\mathbf{t}^{*}\mathbf{t}^{T}\mathbf{h}\} = 2|A|^{2}$$
(5.62)

The denominator of Eq. (5.7) is

$$\mathbf{h}^{H}\mathbf{S}_{\mathbf{I}}\mathbf{h} = \sigma_{c}^{2} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & \rho_{c} \\ \rho_{c}^{*} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$= 2\sigma_{c}^{2} (1 - \operatorname{Re}\{\rho_{c}\})$$
(5.63)

Dividing Eq. (5.62) by Eq. (5.63) gives the output SIR; further dividing that ratio by the input SIR gives the improvement factor for a two-pulse canceller operating against clutter only (no noise)

$$I = \frac{1}{1 - \operatorname{Re}\{\rho_c\}}$$
(5.64)

Since ρ_c in the matrix formulation is the same as $\rho_c[1]$, this is the same expression obtained using autocorrelation methods in Eq. (5.57). The same result could also have been obtained in the vector analysis by simply using the target model vector averaged over Doppler, $\mathbf{t} = [1 \ 0]^T$.

Additional MTI metrics can be defined. Improvement factor I is an average of the improvement in signal-to-clutter ratio over one Doppler period. At some Doppler shifts, the target is above the clutter energy, while at others it is below the clutter and therefore not detectable. I does not indicate over what percentage of the Doppler spectrum a target can be detected. The concept of *MTI visibility factor* or *target visibility V* has been proposed to quantify this effect (Kretschmer, 1986). V is the percentage of the Doppler spectrum over which the improvement factor for a target at a specific frequency is greater than or equal to the average improvement factor I. A related metric is the *usable Doppler space fraction* (UDSF), which in turn is determined by the *minimum detectable velocity* (MDV) or *minimum detectable Doppler* (MDD). These metrics are common in space-time adaptive processing, so their discussion is deferred to <u>Chap. 9</u>.

5.2.6 Limitations to MTI Performance

The basic idea of MTI processing is that repeated measurements of stationary clutter yield the same echo amplitude and phase; thus successive measurements, when subtracted from one another, should cancel. Any effect internal or external to the radar that causes the received echo from a stationary target to vary will cause imperfect cancellation, limiting the improvement factor.

Perhaps the simplest example is transmitter amplitude instability. Consider a two-pulse canceller and suppose that the amplitude of each pulse may differ in amplitude from the nominal amplitude by up to ± 5 percent (equivalent to 20 $\log_{10}(1.05/1) = 0.42$ dB). The signal resulting from subtracting two echoes from a perfectly stationary target can have an amplitude that is as large as 10 percent that of the nominal echo amplitude. Consequently, clutter attenuation may be as poor as $20 \log_{10}(1/0.1) = 20$ dB even though the clutter is perfectly stationary. For a two-pulse canceller with an average signal gain *G* of 2 (6 dB), the maximum achievable improvement factor is 26 dB.

A more realistic analysis of the limitations due to amplitude jitter can be obtained by modeling the amplitude of the *m*th transmitted pulse as A[m] = k(1 + a[m]), where a[m] is a zero mean, white random process with variance σ_a^2 that represents the percentage variation in transmitted amplitude, and *k* is a constant. The received signal will have a complex amplitude of the form $k'(1 + a[m]) \exp(j\phi)$, where ϕ is the phase of the received slow-time sample and the constant k' absorbs all the radar range equation factors. The average power of this signal, which is the input to the pulse canceller, is

$$\mathbf{E}\left\{|y[m]|^{2}\right\} = k'^{2}\mathbf{E}\left\{1 + 2a[m] + a^{2}[m]\right\} = k'^{2}\left(1 + \sigma_{a}^{2}\right)$$
(5.65)

The expected value of the two-pulse canceller output power will be

$$\mathbf{E} \{ |z[m]|^{2} \} = \mathbf{E} \{ |(y[m] - y[m-1])|^{2} \} = \mathbf{E} \{ |k'e^{j\phi}(a[m] - a[m-1])|^{2} \}$$

= $k'^{2} \mathbf{E} \{ a^{2}[m] \} - 2 \mathbf{E} \{ a[m]a[m-1] \} + k'^{2} \mathbf{E} \{ a^{2}[m-1] \}$
= $2k'^{2} \sigma_{a}^{2}$ (5.66)

The achievable clutter cancellation is thus

$$\frac{\text{input power}}{\text{output power}} = \frac{k^{\prime 2} \left(1 + \sigma_a^2\right)}{2k^{\prime 2} \sigma_a^2} = \frac{1 + \sigma_a^2}{2\sigma_a^2}$$
(5.67)

For example, an amplitude variance of 1 percent ($\sigma_a^2 = 0.01$) limits two-pulse clutter cancellation to a factor of 50.5, or 17 dB. Because the average target gain *G* of the two-pulse canceller is G = 2 (3 dB), the limit to the improvement factor *I* is 50.5 × 2 = 101, or 17 + 3 = 20 dB.

Another example is phase drift in either the transmitter or receiver. This can occur due to instability in coherent local oscillators used either as part of the waveform generator on the transmit side or in the demodulation chains on the receiver side. Consider the weighted coherent integration of M data samples y[m] with a zero-mean stationary white phase error $\phi[m]$

$$Z = \sum_{m=0}^{M-1} a_m y[m] e^{j\phi[m]}$$
(5.68)

where the $\{a_m\}$ are the integration weights. Assume that each data sample y[m] is a (possibly complex) constant *A*. Then the power of the weighted coherent sum in the absence of phase error $\phi[m]$ is

$$|Z|^{2} = |A|^{2} \left| \sum_{m=0}^{M-1} a_{m} \right|^{2}$$
(5.69)

It can be shown that the integrated power when a white Gaussian phase error is present is (Richards, 2003)

$$|Z|^{2} = |A|^{2} \left(\sum_{m=0}^{M-1} |a_{m}|^{2} + \exp\left(-\sigma_{\phi}^{2}\right) \sum_{\substack{m=0\\m\neq k}}^{M-1} \sum_{\substack{k=0\\m\neq k}}^{M-1} a_{m} a_{k}^{*} \right)$$
(5.70)

where σ_{\bullet}^2 is the variance of the phase noise in radians. This can be applied to the two-pulse canceller by letting M = 2 and $a_0 = 1$, $a_1 = -1$. Equation (5.70) then gives the power at the canceller output as $2|A|^2 [1 - \exp(-\sigma_{\bullet}^2)]$. Since the power of a clutter sample before the canceller is $|A|^2$, the limitation on two-pulse cancellation due to the phase noise becomes $(1/2)[1 - \exp(-\sigma_{\bullet}^2)]$.

A Gaussian PDF for phase error is reasonable when the error is small, but not when it becomes large because the PDF is not confined to the interval $[-\pi, \pi]$. An extension to this analysis uses instead the Tikhonov distribution for phase:

$$p_{\phi}(\phi) = \frac{1}{2\pi I_0(\alpha)} e^{\alpha \cos\phi}$$
(5.71)

As the parameter α varies from zero to infinity, the PDF varies from uniform random to a fixed value of zero radians, as shown in Fig. 5.14. Using this PDF produces very similar results to the Gaussian analysis, differing only in replacing the quantity $\exp(-\sigma_{\phi}^2)$ in Eq. (5.70) and the two-pulse canceller improvement factor limit by the quantity $[I_1(\alpha)/I_0(\alpha)]^2$, where $I_1(\alpha)$ and $I_0(\alpha)$ are the modified Bessel functions of the first kind and orders one and zero, respectively (Richards, 2011).



FIGURE 5.14 The Tikhonov PDF for phase.

Other sources of limitation due to radar system instabilities include instability in transmitter or oscillator frequencies, transmitter phase drift, coherent oscillator locking errors, PRI jitter, pulse width jitter, and quantization noise. Simple formulas can be developed to bound the achievable clutter attenuation due to each of these error sources, as well as some others not mentioned. Still another source is PRI stagger. When used in a range-ambiguous scenario, PRI stagger results in clutter from different distant ranges aliasing to the same near ranges on different PRIs, so that the clutter becomes less stationary and the clutter attenuation is degraded. Additional information and analysis on these and other issues is available in Shrader and Gregers-Hansen (2008), Nathanson (1991), and Schleher (2010).

External to the radar, the chief factor limiting MTI improvement factor is simply the width of the clutter spectrum itself. Wider spectra put more clutter energy outside of the MTI filter null, so that less of the clutter energy is filtered out. This effect is evident in Eq. (5.47) and was illustrated numerically in Table 5.1. The effective clutter spectrum width can be increased by radar system instabilities or by measurement geometry and dynamics. For instance, a scanning antenna adds some amplitude modulation due to antenna pattern weighting to the clutter return. The power spectrum of the measured clutter is then the convolution in the frequency domain of the actual clutter power spectrum and the squared magnitude of the Fourier transform of the amplitude modulation caused by the antenna scanning. This convolution increases the observed spectral width somewhat. In some cases, the clutter power spectrum may not be centered on zero Doppler shift. A good example is rain clutter. Moving weather systems will have a nonzero average Doppler representing the rate at which the rain cell is approaching or receding from the radar system. Unless this average

motion is detected and compensated, the MTI filter null will not be centered on the clutter spectrum and cancellation will be poor.

The largest source of clutter offset and spreading is radar platform motion. Recall from <u>Chap. 3</u>, <u>Eq. (3.5)</u>, that the motion-induced clutter bandwidth is

$$\beta_D \approx \frac{2\upsilon\theta_3}{\lambda}\sin\psi \tag{5.72}$$

The offset in center frequency of the clutter spectrum can be as much as a few kilohertz for fast aircraft when forward looking, while the motion-induced spectral spread can be tens to a few hundreds of hertz for a sidelooking configuration. This clutter spreading adds to the intrinsic spread of the clutter spectrum due to internal motion and can often be the dominant effect determining the observed clutter spectral width and therefore determining the MTI performance limits.

5.3 Pulse Doppler Processing

Pulse Doppler processing is the second major class of Doppler processing. Recall that in MTI processing the fast time-slow-time data matrix is highpass filtered in the slow-time dimension, yielding a new fast-time/slow-time data sequence in which the clutter components have been attenuated. Pulse Doppler processing differs in that filtering in the slow-time domain is replaced by explicit spectral analysis of the slow-time data for each range bin. Target detection is then performed directly on the range-Doppler matrix of data. Because the range-Doppler matrix is the fundamental data quantity in pulse-Doppler processing, the first step is to form it from the fast-time/slow-time CPI matrix by computing the one-dimensional spectrum of the slow-time signal in each range bin. The spectral analysis is most commonly by far performed using the fast Fourier transform (FFT) algorithm to compute the discrete Fourier transform (DFT) as shown in Fig. 5.15, but other techniques can also be used. The computed DFT is a frequency-sampled version of the DTFT of the slowtime signal. The figure emphasizes the fact that the number of frequency samples K does not have to equal the number of slow-time samples M; often K > M as shown.



FIGURE 5.15 Conversion of the fast-time/slow-time data matrix to a range-Doppler matrix.

Consider the notional pulse Doppler spectrum for one range bin shown in Fig. 5.16. The DTFT of the data is shown in a fashion similar to earlier figures. The white dots represent the samples of the DTFT computed by the DFT; these are the only actually available data. Assuming that the clutter has been centered at zero Doppler, spectral samples at or near zero frequency will be dominated by the strong clutter signal even though noise is also present. Spectral samples in the clear region have only thermal noise to interfere with signal detection. Each clear-region spectrum sample can be individually compared to a noisebased threshold to determine whether the signal at that range bin and Doppler frequency appears to be noise only or noise plus a target. If the sample crosses the threshold it not only indicates the presence of a target in that range bin but also its approximate velocity, since the Doppler frequency bin is known. The samples that are clutter-dominated are often simply discarded on the grounds that the SIR will be too low for successful detection. However, other systems use a technique called clutter mapping, discussed in Sec. 5.6.1, to attempt detection of strong targets in the clutter region using a clutter-based threshold.



FIGURE 5.16 The computed pulse Doppler spectrum is a frequency-sampled version of an underlying discrete-time Fourier transform.

The advantages of pulse Doppler processing are that it provides at least a coarse estimate of the radial velocity component of a moving target and that it provides a way to detect multiple targets, provided they are separated enough in Doppler to be resolved. The chief disadvantages are greater computational complexity of pulse Doppler processing as compared to MTI filtering and longer required dwell times due to the use of more pulses for the Doppler measurements. Thorough discussions of pulse Doppler processing are contained in Morris and Harkness (1996) and Stimson (1998). An excellent new addition to the literature is Alabaster (2012).

5.3.1 The Discrete-Time Fourier Transform of a Moving Target

To understand the issues in pulse Doppler processing it is useful to understand the appearance of noise, clutter, and target signals in the range-Doppler map. Begin by again considering the Fourier spectrum of an ideal constant radial velocity moving point target and the effects of a sampled Doppler spectrum. The issues are the same as those considered when discussing the sampling of the Doppler spectrum in <u>Chap. 3</u>. Consider a radar illuminating a moving target over a CPI of M pulses, and suppose a moving target is present in a particular range bin. If the target's velocity is such that the Doppler shift is F_D Hz, the slow-time received signal after quadrature demodulation is

$$y[m] = Ae^{j2\pi F_{D}mT} \quad m = 0, ..., M-1$$
(5.73)

where *T* is the radar's pulse repetition interval and is the effective sampling interval in slow time. The signal of Eq. (5.73) is the same signal considered in Chap. 3 [Eq. (3.19)], except for the change from normalized frequency ω_D in radians to analog frequency F_D in hertz; they are related according to ω_D =

 $2\pi F_D T$. Equation (3.20) gave the *M*-pulse discrete-time Fourier transform of this signal; converting to analog frequency gives

$$Y(F) = \sum_{m=-\infty}^{\infty} y[m] \exp(-j2\pi FTm)$$

= $A \frac{\sin[\pi(F - F_D)MT]}{\sin[\pi(F - F_D)T]} \exp[-j\pi(M - 1)(F - F_D)T], \quad F \in [-PRF/2, +PRF/2)$
(5.74)

Y(F) is an asinc² function with its peak at $F = F_D$ as expected and a peak magnitude of *MA*. Its magnitude is illustrated in Fig. 5.17*a* for the case where F_D = *PRF*/4, *M* = 20 samples, and *A* = 1. So long as $M \ge 4$, the Rayleigh (peak-tonull) mainlobe bandwidth is 1/*MT* Hz; this is also the 4-dB bandwidth. The width of the mainlobe at the -3 dB points is 0.89/*MT* Hz. The first sidelobe is 13.2 dB below the response peak. These mainlobe width measures determine the Doppler resolution of the radar system. They are all inversely proportional to *MT*, which is the total elapsed time of the set of pulses used to make the spectral measurement. Thus, Doppler resolution is determined by the observation time of the measurement as first discussed in <u>Chap. 1</u>. Longer observation allows finer Doppler resolution.



FIGURE 5.17 Magnitude of the discrete-time Fourier transform of an ideal moving target slow-time data sequence with $F_D = PRF/4$ and M = 20 samples: (*a*) no window, (*b*) Hamming window.

Because of the high sidelobes, it is common to use a data window to weight the slow-time data samples y[m] prior to computing the DTFT or DFT. To analyze this case, replace y[m] by $\omega[m]y[m]$ in the computation of the DTFT and again use Eq. (5.73) for the particular form of y[m]. Y(F) becomes

$$Y_w(F) = A \sum_{m=0}^{M-1} w[m] e^{-j2\pi(F-F_D)mT} = W(F-F_D)$$
(5.75)

where the notation $Y_w(F)$ is used to emphasize that the spectrum is computed with a window applied to the data. This is simply the Fourier transform of the window function itself, shifted to be centered on the target Doppler frequency F_D rather than at zero. Figure 5.17b illustrates the effect of the window on the DTFT for the same data used in part a of the figure. In fact, the asinc function of Fig. 5.17a is also just the Fourier transform of the rectangular window function (equivalent to no window). An extensive description of common window functions and their characteristics is in Harris (1978). In general, nonrectangular windows cause an increase in mainlobe width, a decrease in peak amplitude, and a decrease in SNR in exchange for large reductions in peak sidelobe level.

It is straightforward to compute the reduction in peak amplitude and the SNR loss given the window function w[n]. Consider the peak gain first. From Eq. (5.74), the peak value of $|Y(F)|^2$ when no window is used is A^2M^2 . Evaluating Eq. (5.75) at $F = F_D$ gives the peak power when a window is used:

$$|Y(F_D)|^2 = \left| A \sum_{m=0}^{M-1} w[m] e^{-j2\pi(0)mT} \right|^2 = A^2 \left| \sum_{m=0}^{M-1} w[m] \right|^2$$
(5.76)

The ratio $|Y(F)|^2/|Y(F_D)|^2$, called the loss in processing gain (LPG), is

$$LPG = \frac{M^2}{\left|\sum_{m=0}^{M-1} w[m]\right|^2}$$
(5.77)

With this definition $LPG \ge 1$, so the loss in dB is a positive number. Using Eq. (5.77), the LPG can be computed for any window. Values of 5 to 8 dB are common. Details depend on the specific window function, but the LPG is typically a weak function of the window length M, higher for small M and rapidly approaching an asymptotic value for large M (on the order of 100 or more).

Although the window reduces the peak amplitude of the DTFT substantially, it also reduces noise power. *Processing loss* (PL) is the reduction in SNR at the peak of the DTFT for a pure sinusoidal input. Denoting the SNR with and without the window as χ and χ_w , respectively, it is possible to separate the effects of the window on the target and noise components of the signal

$$\frac{\chi}{\chi_w} = \frac{(S/N)}{(S_w/N_w)} = \left(\frac{S}{S_w}\right) \left(\frac{N_w}{N}\right) = LPG\left(\frac{N_w}{N}\right)$$
(5.78)

To determine the window's effect on the noise power, suppose y[m] is a zero mean stationary white noise with variance σ^2 . Then the windowed noise power is

$$\sigma_{w}^{2} = \mathbf{E} \left\{ \left(\sum_{m=0}^{M-1} w[m] y[m] \right) \left(\sum_{l=0}^{M-1} w^{*}[l] y^{*}[l] \right) \right\}$$
$$= \mathbf{E} \left\{ \left(\sum_{m=0}^{M-1} |w[m] y[m]|^{2} \right) + \text{cross terms} \right\}$$
$$= \sigma^{2} \sum_{m=0}^{M-1} |w[m]|^{2} = N_{w}$$
(5.79)

The unwindowed noise power *N* can be obtained from Eq. (5.79) by setting w[m] = 1 for all *m*, giving $N = M\sigma^2$. Using Eq. (5.77) and these values for *N* and N_w in Eq. (5.79) gives the processing loss

$$PL = \frac{M \sum_{m=0}^{M-1} |w[m]|^2}{\left|\sum_{m=0}^{M-1} w[m]\right|^2}$$
(5.80)

Like the loss in peak gain, the processing loss is a weak function of M that is higher for small M but quickly approaches an asymptotic value. As an example, for the Hamming window the loss in SNR is 1.75 dB for a very short (M = 8) window, decreasing asymptotically to about 1.35 dB for long windows. <u>Table 5.2</u> summarizes the four key properties of 3-dB resolution, peak sidelobe level, loss in processing gain, and processing loss for several common windows. Also shown is worst-case straddle loss, to be discussed next. All of these characteristics of windows are the same as discussed in <u>Chap. 4</u> in the context of range side-lobe control. A much more extensive table, including both more metrics and many more types of windows, is given in Harris (1978).⁸

Window	Window Length M	3-dB Mainlobe Width (Relative to Rectangular Window)	LPG (dB Relative to Rectangular Window)	Peak Sidelobe, dB	PL, dB	Worst- Case Straddle Loss, dB
Rectangular	20	1.0	0.0	13.2	0.0	3.9
	100	1.0	0.0	13.3	0.0	3.9
Hann	20	1.71	6.5	31.5	2.0	1.3
	100	1.65	6.1	31.5	1.8	1.4
Hamming	20	1.52	5.7	40.4	1.5	1.6
	100	1.49	5.4	42.6	1.4	1.7
Kaiser, $\beta = 3$	20	1.26	3.6	25.3	0.7	2.4
	100	1.25	3.4	24.0	0.6	2.5
Dolph- Chebyshev 50-dB equiripple	20	1.55	5.8	50	1.6	1.6
	100	1.53	5.7	50	1.5	1.7

 TABLE 5.2
 Properties of Some Common Data Windows

5.3.2 Sampling the DTFT: The Discrete Fourier Transform

In practice, the DTFT is not computed because its frequency variable is continuous. Instead, the discrete Fourier transform is computed

$$Y[k] = \sum_{m=0}^{M-1} y[m] e^{-j2\pi mk/K} \quad k = 0, \dots, K-1$$
(5.81)

As discussed in <u>Chap. 3</u>, for a finite length data sequence Y[k] is just Y(F) evaluated at F = k/KT = k (*PRF/K*) Hz. These samples are called *Doppler bins*. Thus, the DFT computes *K* samples of the DTFT evenly spaced across one period of the DTFT. The DFT is almost invariably computed using an FFT algorithm. The appearance of a plot of the Doppler spectrum computed using a DFT can be a strong function of the relation between the actual DTFT shape and the number and location of the DFT frequency samples.

In some situations, the number of data samples available can be greater than the desired DFT size, that is, M > K. This occurs when there is a need to reduce the DFT size for computational reasons, or when the radar timeline permits the collection of more pulses than the number of Doppler bins required and it is desirable to use the extra pulses to improve the SNR of the Doppler spectrum measurement. The data turning procedure described in <u>Chap. 3</u> and <u>Fig. 3.10</u> allows the use of a *K*-point DFT while taking advantage of all of the data.

Some caution is needed in applying a data window when the data are modified by zero padding or turning. In either case a length M window should be applied to the data *before* it is either zero padded or turned. Applying a K-point window to the full length of a zero-padded sequence has the effect of

multiplying the data by a truncated, asymmetric window (the portion of the actual window that overlaps the M nonzero data points), resulting in greatly increased sidelobes. Applying a shortened K-point window to a turned data sequence results in DFT samples that do not equal samples of the DTFT of the windowed M-point original data sequence.

Because the DFT is a sampled version of the DTFT, the peak value of the DFT obtained for a pure sinusoidal signal is greatest when the Doppler frequency coincides exactly with one of the DFT sample frequencies, and decreases when the target signal is between DFT frequencies. This reduction in amplitude is called a Doppler *straddle loss*. The amount of loss depends on the particular window used and the ratio K/M. For a given signal length M, the straddle loss is always greatest for signal frequencies exactly halfway between DFT sample frequencies. The calculation can be simplified by assuming that the signal frequency is $F_D = 0$ so that y[m] = w[m] and then evaluating Eq. (5.81) with k = 0 (DFT sample on the "sinusoid" peak) and k = 1/2 (1/2 bin away from the sinusoid peak). To be explicit, consider the rectangular window case; then

$$|Y[k]| = \left| \sum_{m=0}^{M-1} e^{-j2\pi mk/K} \right| = \left| \frac{\sin(\pi Mk/K)}{\sin(\pi k/K)} \right|$$
(5.82)

Assuming $K \ge M$ and evaluating at k = 1/2 gives

$$\left|Y\left[\frac{1}{2}\right]\right| = \frac{\sin(\pi M/2K)}{\sin(\pi/2K)} \approx \frac{2K}{\pi} \sin(\pi M/2K)$$
(5.83)

The last step was obtained by assuming that *K* is large enough to allow a small angle approximation to the sine function in the denominator. *Y*[0] is obtained either by applying L'HÔpital's rule to the second form of Eq. (5.82) or computing it explicitly from the first form; the result is Y[0] = M. The maximum straddle loss for the DFT filterbank with no windowing is

Maximum straddle loss
$$\approx \frac{\pi M}{2K\sin(\pi M/2K)} = \frac{1}{\sin(M/2K)}$$
(5.84)

This equation verifies that the loss depends on the ratio K/M. The worst case occurs when the sinc term is minimized; this happens when K = M. In decibels, this worst-case loss for a rectangular window is sinc $(1/2) = \sin(\pi/2)/(\pi/2) = 2/\pi$, equivalent to -3.92 dB.

The straddle loss for shaped windows varies somewhat with M as can be seen in <u>Table 5.2</u>. A calculation for the Hamming window similar to that

leading to Eq. (5.84) results in a smaller maximum straddle loss of 1.75 dB for very long windows (M > 300 or so), dropping to 1.5 dB for very short (M = 8) windows. Thus, while any nonrectangular window causes a reduction in peak gain, typical windows have the desirable property of having less variability in gain as the Doppler shift of the target varies. This effect is illustrated in Fig. 5.18, which shows the maximum DFT output amplitude as a function of the target Doppler shift for rectangular and Hamming windows for the case of M = K = 16. Observe the general reduction in peak amplitude for the Hamming-windowed data compared to the unwindowed data. On the other hand, the *variation* in amplitude is significantly less for the windowed data (1.6 dB versus 3.9 dB); i.e., the amplitude response is more consistent, an underappreciated benefit of windowing.



FIGURE 5.18 Variation of DFT output with complex sinusoid input frequency for two different data analysis windows.

5.3.3 The DFT of Noise

The target signals discussed in the previous sections compete with interference, primarily noise and clutter, so the characteristics of the DTFT and DFT of the interference are of interest. First consider noise. Assume the usual model of additive complex WGN with variance σ_{π}^2 in the *M*-sample slow-time signal y[m]. The value of the DTFT at any particular frequency or of any DFT sample is a weighted sum of the same *M* i.i.d. (independent identically distributed) complex Gaussian time samples and so is also a complex Gaussian RV. If no window is used the weights are the DTFT or DFT kernel values $\exp(-j\omega m)$ or $\exp(-j2\pi km/K)$. Because these weights have unit magnitude, the resulting

spectrum (DTFT or DFT) sample will have variance $M\sigma_w^2$. If the slow-time data is windowed, the variance becomes $E_w \sigma_w^2$, where $E_w = \sum_w |w[m]|^2$ is the energy of the window sequence. Because the DTFT and DFT samples are complex Gaussian, it also follows that the magnitude of a spectrum sample has a Rayleigh PDF, the magnitude-squared has an exponential PDF, and the phase has a uniform PDF over $[-\pi, \pi]$. The PDF and its parameters are important in setting detection thresholds as will be seen in <u>Chap. 6</u>. These and additional properties of the DTFT and DFT of noise are derived in Richards (2007).

DTFT and DFT values at different frequencies or indices for random inputs are, in general, correlated. In the unwindowed case the normalized autocorrelation function in frequency is an asinc function of peak amplitude 1 and a zero spacing of $2\pi/M$ radians per sample. For the windowed case, the normalized autocorrelation function follows the shape of the DTFT of the window function. In the case of an unwindowed DFT of size K = M the DFT sample spacing matches the zero spacing of the asinc function so that the DFT samples will be uncorrelated. Because the complex samples are Gaussian, they are also independent in this case.

5.3.4 Pulse Doppler Processing Gain

The DTFT and DFT represent forms of coherent integration in the slow-time domain and result in a processing gain. Equation (5.74) shows that each sample of y[m] is phase-adjusted by multiplication with the DTFT kernel and then integrated. If no window is used, then at the value of F that matches the signal frequency, $F = F_D$, the kernel values exactly compensate the phase history of the data such that all of the signal samples add in phase. At this frequency $Y(F_D) = A \cdot M$. The value of the noise component at the same frequency value will be $M\sigma_w^2$. The SNR is therefore $(A \cdot M)^2 / M\sigma_w^2 = M \cdot (A^2 / \sigma_w^2)$ a processing gain of a factor of $G_{sp} = M$ compared to the SNR in the slow-time signal before the DTFT. The same result applies for the DFT. Note that in the DFT case, the processing gain is determined by M, the number of slow-time samples and not by K, the DFT size. Computing a DFT larger than the number of time domain samples, K > M, does not increase the processing gain.

Various issues can reduce G_{sp} . When a window is used with either a DTFT or DFT, G_{sp} is decreased by the processing loss *PL* described earlier [Eq. (5.80)]. In this case, the range equation signal processing gain due to Doppler processing becomes $G_{sp} = M/PL$. When a DFT is used, G_{sp} is also reduced by any straddle loss that may occur.

5.3.5 Matched Filter and Filterbank Interpretations of Pulse Doppler Processing with the DFT

Equation (5.13) defined the coefficients of the matched Doppler filter. In MTI filtering, it is assumed that the target Doppler shift is unknown. The resulting

signal model of Eq. (5.25) leads to the pulse canceller as a near-optimum MTI filter for small order N. In contrast, DFT-based pulse Doppler processing attempts to separate target signals based on their particular Doppler shift. Assume that the signal is a pure complex sinusoid (ideal moving target) at a Doppler shift of F_D Hz. Based on Eq. (5.21), the model of the signal vector is then

$$\mathbf{t} = [y[m] \quad y[m+1] \quad \cdots \quad y[m+M-1]]^T$$

= $A[1 \quad e^{j2\pi F_D T} \quad \cdots \quad e^{j2\pi F_D (M-1)T}]^T$
(5.85)

where the complex scalar *A* absorbs all constants. If the interference consists only of white noise (no correlated clutter), S_I reduces to $\sigma_n^2 I$. It follows that for an arbitrary data vector **y** the output $\mathbf{h}^T \mathbf{y}$ of the matched filter becomes

$$\mathbf{h}^{T}\mathbf{y} = A \sum_{m=0}^{M-1} y[m] e^{-j2\pi F_{D}mT}$$
(5.86)

This is the DTFT of y[m] to within the scale factor *A*. When $F_D = k/KT = kPRF/K$ for some integer *k*, Eq. (5.86) is simply the *K*-point DFT of the data sequence y[m] to within the scale factor *A*. Consequently, the DFT is a matched filter to ideal constant radial velocity moving target signals, provided that the Doppler shift equals one of the DFT sample frequencies and the interference is white. This result is very closely related to the two-pulse canceller for a specific target Doppler and noise interference only considered in Sec. 5.2.3.

Since the *K*-point DFT computes *K* different outputs from each input vector, it effectively implements a bank of *K* matched filters at once, each tuned to a different Doppler frequency. The frequency response shape of each matched filter is just the asinc function. To see this, denote the impulse response vector in Eq. (5.86) when $F_D = k/MT$ as \mathbf{h}_k . To within a scale factor

$$\mathbf{h}_{k} = \begin{bmatrix} 1 & e^{-j2\pi k/K} & e^{-j4\pi k/K} & \cdots & e^{-j2\pi (K-1)k/K} \end{bmatrix}^{T}$$
(5.87)

The corresponding discrete time frequency response $H_k(\omega)$ is

$$H_{k}(\omega) = \sum_{m=0}^{K-1} h_{k}[m]e^{-j\omega m}$$
$$= \sum_{m=0}^{K-1} e^{-j(\omega + 2\pi k/K)m}$$
(5.88)

This summation evaluates to the asinc function of Eq. (5.74) shifted to a center frequency of $\omega = -2\pi k/K$, which is equivalent to $\omega = 2\pi (K - k)/K$. The *k*th DFT sample therefore corresponds to filtering the data with a bandpass filter having a frequency response with an asinc function shape centered at the frequency of the (K - k)th DFT sample.

If the data are windowed before processing with a window function w[m], Eq. (5.86) becomes (again for $F_D = k/KT$)

$$\mathbf{h}_{k}^{T}\mathbf{y} = A \sum_{m=0}^{M-1} w[m] y[m] e^{-j2\pi mk/K}$$
(5.89)

The impulse response vector and frequency response are then

$$\mathbf{h}_{k} = [w[0] \quad w[1]e^{-j2\pi k/K} \quad w[2]e^{-j4\pi k/K} \quad \cdots \quad w[M-1]e^{-j2\pi (M-1)k/K}]^{T}$$

$$H_{k}(\omega) = \sum_{m=0}^{M-1} w[m]e^{-j(\omega + 2\pi k/K)m}$$

$$= W\left(\omega + \frac{2\pi k}{K}\right)$$
(5.90)

$$= W\left(\omega + \frac{2\pi k}{K}\right) \tag{5.91}$$

The DFT still implements a bandpass filter centered at each DFT frequency but the filter frequency response shape now becomes that of the window function.

The relation between the DFT and a bank of filters can be made more explicit. Consider a slow-time signal y[m] obtained with a long series of pulses and an *M*-point window function w[m]. The window function can be slid along the data sequence to select a portion of the data for spectral analysis as shown in Fig. 5.19. The DTFT of the resulting sequence w[m - p]y[p] is, in terms of analog frequency *F*,



FIGURE 5.19 Relationship between data sequence x[m] and *M*-point sliding analysis window w[m].

$$Y_{m}(F) = \sum_{p=-\infty}^{\infty} w[m-p]y[p]e^{-j2\pi FpT}$$

= $e^{-j2\pi FmT} \sum_{p=-\infty}^{\infty} w[m-p]e^{-j2\pi F(p-m)T}y[p]$
= $e^{-j2\pi FmT} \{(w[p]e^{+j2\pi FpT}) * y[p]\}_{p=m}$
(5.92)

Equation (5.92) shows that, aside from a phase factor, the DTFT at a particular frequency is equivalent to the convolution of the input sequence and a modulated window function, evaluated at time m. Furthermore, if W(F) is the discrete-time Fourier transform of w[m] (converted to an analog frequency scale), the DTFT of $w[p]\exp(+j2\pi F_DpT)$ is $W(F + F_D)$, which is simply the Fourier transform of the window shifted so that it is centered at Doppler frequency $-F_D$ Hz. This means that measuring the DTFT at a frequency F_D is equivalent to passing the signal through a bandpass filter centered at $-F_D$ and having a passband shape equal to the Fourier transform of the window function. Since the DFT evaluates the DTFT at K distinct frequencies at once, it follows that pulse Doppler spectral analysis using the DFT is equivalent to passing the data through a bank of bandpass filters.

Of course, it is possible to build a literal bank of bandpass filters, each one perhaps individually designed, and some systems are constructed in this way. For example, the zero-Doppler filter in the filterbank can be optimized to match the expected clutter spectrum or even made adaptive to account for changing clutter conditions. Most commonly, however, the DFT is used for Doppler spectrum analysis. This places several restrictions on the effective filterbank design. There will be *K* filters in the bank, where *K* is the DFT size; the filter

center frequencies will be equally spaced, equal to the DFT sample frequencies; and all the passband filter frequency response shapes will be identical, determined by the window used and differing only in center frequency. The advantages to this approach are simplicity and speed with reasonable flexibility. The FFT provides a computationally efficient implementation of the filterbank: the number of filters can be changed by changing the DFT size; the filter shape can be changed by choosing a different window; and the filter optimizes the output SNR for targets coinciding with a DFT filter center frequency in a noise-limited interference environment.

5.3.6 Fine Doppler Estimation

Peaks in the DFT output that are sufficiently above the noise level to cross an appropriate detection threshold are interpreted as responses to moving targets, i.e., as samples of the peak of an asinc component of the form of Eq. (5.74). As has been emphasized, there is no guarantee that a DFT sample will fall exactly on the asinc function peak. Consequently, the amplitude and frequency of the DFT sample giving rise to a detection are only approximations to the actual amplitude and frequency of the asinc peak. In particular, the estimated Doppler frequency of the peak can be off by as much as one-half Doppler bin, equal to PRF/2K Hz.

If the DFT size *K* is significantly larger than the slow-time data sequence length *M*, several DFT samples will be taken on the asinc mainlobe and the largest may well be a good estimate of the amplitude and frequency of the asinc peak. Frequently, however, K = M and sometimes, with the use of data turning, it is even true that K < M. In these cases, the Doppler samples are far apart and a half-bin error may be intolerable. One way to improve the estimate of the true Doppler frequency F_D is to interpolate the DFT in the vicinity of the detected peak.

The most obvious way to interpolate the DFT is to zero pad the data and compute a larger DFT. In the absence of noise the interpolated values are exact. However, this approach is computationally expensive and interpolates all of the spectrum. If finer sampling is needed only over a small portion of the spectrum around a detected peak the zero padding approach is inefficient.

Computing a larger DFT is tantamount to interpolation using an asinc interpolation kernel. To see this, consider evaluating the DTFT at an arbitrary value of ω using only the available DFT samples. This can be done by computing the inverse DFT to recover the original time-domain data and then computing the DTFT from those samples:

$$Y(\omega) = \sum_{m=0}^{M-1} y[m] e^{-j\omega m}$$

= $\sum_{m=0}^{M-1} \left(\frac{1}{K} \sum_{k=0}^{K-1} Y[k] e^{+j2\pi mk/K} \right) e^{-j\omega m}$
= $\frac{1}{K} \sum_{k=0}^{K-1} Y[k] \left\{ \sum_{m=0}^{M-1} \exp\left[-jm\left(\omega - \frac{2\pi k}{K}\right) \right] \right\}$
(5.93)

The term in braces is the interpolating kernel. It can be expressed in closed form as

$$\sum_{m=0}^{M-1} \exp\left[-jm\left(\omega - \frac{2\pi k}{K}\right)\right]$$

= $\exp\left[-j\left(\omega - \frac{2\pi k}{K}\right)(m-1)/2\right]\frac{\sin\left[\left(\omega - \frac{2\pi k}{K}\right)M/2\right]}{\sin\left[\left(\omega - \frac{2\pi k}{K}\right)/2\right]}$
= $Q_{M,K}(\omega, k)$ (5.94)

Combining these gives

$$Y(\omega) = \frac{1}{K} \sum_{k=0}^{K-1} Y[k] Q_{M,K}(\omega, k)$$
(5.95)

Equations (5.94) and (5.95) can be used to compute the DTFT at any single value of ω from the DFT samples. It can be applied to interpolate the values of the DFT over localized regions with any desired sample spacing and the result is exact in the absence of noise. However, it remains relatively computationally expensive.

A simpler but very serviceable technique for interpolating local peaks is illustrated in Fig. 5.20. For each detected peak in the magnitude of the DFT output, a second-order polynomial is fit through that peak and the two adjacent magnitude data samples. Once the polynomial coefficients are known, the amplitude and frequency of its peak are easily found by differentiating the formula for the polynomial and setting the result to zero. It can be shown that the second-order polynomial passing through the three samples can be expressed as



FIGURE 5.20 Refining the estimated target amplitude and Doppler shift by quadratic interpolation around the DFT peak.

$$|Y[k_0 + \Delta k]| = \frac{1}{2} \{ (\Delta k - 1) \Delta k | Y[k_0 - 1] | - 2(\Delta k - 1)(\Delta k + 1) | Y[k_0] | + (\Delta k + 1) \Delta k | Y[k_0 + 1] | \}$$
(5.96)

where Δk is the location of the interpolated peak relative to the index k_0 of the apparent peak so that the estimated peak location becomes $k' = k_0 + \Delta k$ (see Fig. 5.20). Differentiating this equation with respect to Δk , setting the result to zero, and solving for Δk gives the estimated location of the polynomial peak relative to k_0 as

$$\Delta k = \frac{-(1/2)\{|Y[k_0 + 1]| - |Y[k_0 - 1]|\}}{|Y[k_0 - 1]| - 2|Y[k_0]| + |Y[k_0 + 1]|}$$
(5.97)

The amplitude of the estimated peak $A' = |Y|[k_0 + \Delta k]|$ is found by computing Δk and then using that result in Eq. (5.96). Note that the formula for Δk behaves in intuitively satisfying ways. If the first and third DFT magnitude samples are equal, $\Delta k = 0$; the middle sample is the estimated peak. If the second and third samples are equal, $\Delta k = 1/2$, indicating the estimated peak is halfway between the two samples; a similar result applies if the first and second DFT magnitude samples are equal.

This interpolation technique is less effective when the width of the presumed mainlobe response that is to be interpolated is so narrow that the apparent peak and its two neighbors are not on the same lobe of the response.

This occurs when the spectrum is sampled at the Nyquist rate in Doppler, K = M; no window is applied to the data; and the data frequency does not happen to fall on a DFT frequency sample (the very situation in which interpolation is most needed). Figure 5.21*a* illustrates this case. The data are M = 20 samples of a 1.35-kHz complex sinusoid sampled at 5 kHz. The spectrum was computed using a K = 20 point DFT. Without interpolation, the largest DFT magnitude sample has a value of 15.15. This apparent peak amplitude is in error by 24.3 percent from the true DTFT peak value of 20. The DFT samples occur every 5000/20 = 250 Hz, so the apparent peak occurs at a frequency of 1500 Hz, an error of 150 Hz.



FIGURE 5.21 Ideal Doppler spectrum due to a moving target, sampled at the Nyquist rate: (*a*) no window, (*b*) Hamming window.

If the interpolation procedure is applied to these data, poor results will be achieved because the assumption that the three points are on a quadratic curve segment is not valid. For these particular data, the interpolation technique will estimate the true frequency and amplitude of the spectral peak as 1295.4 Hz and 15.41, respectively. The amplitude estimate is improved only slightly, to a 23 percent error. The frequency error is reduced significantly, to 54.6 Hz, but is

still large.

This problem can be avoided by ensuring that the sample set is dense enough to guarantee that the three samples are all on the mainlobe. One way to do this is to oversample in Doppler, i.e., choose K > M. Another is to window the data. For most common windows, the expansion of the mainlobe that results is sufficient to guarantee that the apparent peak samples and its two neighbors fall on the same lobe, so that the basic assumption of a quadratic segment is more valid. Figure 5.21b illustrates this effect by applying a Hamming window to the same data used for Fig. 5.21a and again applying a 20-point DFT. Note that the peak DTFT amplitude is now 10.34 due to the effect of the Hamming window. Applying quadratic interpolation to this spectrum gives an estimated spectral peak frequency and amplitude of 1336.6 Hz and 9.676, respectively, errors of 6.4 percent in amplitude and 13.4 Hz in frequency. A hybrid technique can be defined that combines attributes of the quadratic interpolation and the more exact asinc interpolation. The parabolic method is used to identify the frequency of the peak, and then Eq. (5.95) is used to estimate the amplitude. This approach improves amplitude accuracy while avoiding the need to compute Eq. (5.95) more than once.

Figure 5.22 illustrates the frequency estimation performance of the quadratic interpolator both with and without Hamming windowing on a sinusoidal data sequence of length M = 30. Part *a* of the figure illustrates the minimally sampled case K = M. The interpolated frequency estimates are best when the actual frequency is either very close to a sample frequency or exactly half way between two sample frequencies. If no window is used, the worst case error of 0.23 bins occurs when the actual frequency is 0.35 bins away from a sample frequency. A Hamming window reduces this maximum error to 0.067 bins at an offset of 0.31 bins. Figure 5.22*b* shows the performance when the spectrum sampling density is slightly more than doubled to K = 64. The maximum frequency estimation error is only 0.022 bins without the window and 0.014 bins with the Hamming window.



FIGURE 5.22 Noise-free frequency estimation performance of quadratic interpolator M = 30. (*a*) Minimally sampled case (K = 30), (*b*) oversampled case (K = 64).

Figure 5.23 shows the amplitude estimation performance for the same cases. Figure 5.23a shows that the interpolator reduces the worst-case straddle loss from 3.92 to 3.22 dB when no window is used, and from 1.68 to 0.82 dB

when a Hamming window is used. The worst-case amplitude error with the oversampled spectrum using K = 64 is reduced to 0.17 dB without a window and only 0.05 dB with the Hamming window.



FIGURE 5.23 Noise-free amplitude estimation performance of quadratic interpolator M = 30. (*a*) Minimally sampled case (K = 30), (*b*) oversampled case (K = 64).

Many other interpolation-based estimators have been proposed for the DFT; several are described in Jacobsen and Kootsookos (2007) and MacLeod (1998). They include versions of Eqs. (5.96) and (5.97) that use the complex DFT data Y[k] instead of its magnitude and achieve significantly better frequency estimation accuracy in noise-free data, as well as adjustments to the weighting coefficients in either the complex or magnitude version which improve accuracy when a window is used on the data. Another family of interpolators uses the apparent peak and the larger of its two neighbors to compute Δk as

$$\Delta k = \frac{|Y[k_0]|}{|Y[k_0]| + |Y[k_{\pm}]|}$$
(5.98)

where $Y[k_{\pm}]$ is the larger of $Y[k_0 - 1]$ and $Y[k_0 + 1]$. While this estimator uses only two DFT values explicitly, it implicitly uses three since two must be examined to determine which is $Y[k_{\pm}]$. Complex and magnitude-only versions of this estimator also exist. Its frequency estimation performance is surprisingly good in the absence of noise.

These interpolation techniques are not limited to Doppler frequency estimation only. The same issues of sampling density and straddle loss arise at the output of the matched filter in fast time, for instance, and the same interpolation techniques can be applied to improve the range estimates. The application to time delay (range) estimation, along with simulation results for this and alternative techniques, is discussed in <u>Chap. 7</u>.

The above results are all for noise-free data, an unrealistic assumption. Estimation accuracy and the effect of interpolation algorithms will be revisited more formally in <u>Chap. 7</u>. It will be seen there the two-point interpolator is of little value in the presence of noise, but that the quadratic interpolator is still effective at high SNR. Other effects unrelated to interpolation dominate the estimation precision at mid to low SNRs.

5.3.7 Modern Spectral Estimation in Pulse Doppler Processing

So far, the DFT has been used exclusively to compute the spectral estimates needed for pulse Doppler processing. Other spectral estimators can be used. One that has been applied to radar is the *autoregressive* (AR) model, which models the actual spectrum $Y(\omega)$ of the slow-time signal with a spectrum of the form

$$\hat{Y}(\omega) = \frac{\alpha}{1 + \sum_{p=1}^{p} a_p e^{-j\omega p}}$$
(5.99)

The algorithm finds the set of model coefficients $\{a_p\}$ that optimally fits $\hat{Y}(\omega)$ to $Y(\omega)$ for a given model order *P*. These coefficients are found by solving a set of *normal equations* (Hayes, 1996) derived from the autocorrelation of the slow-time data y[m]; the actual spectrum $Y(\omega)$ is not needed. The $\{ap\}$ are then used to compute an estimated spectrum according to Eq. (5.99), which can be analyzed for target detection, pulse pair processing, or other functions.

Modeling the spectrum as shown in Eq. (5.99) is equivalent to modeling the slow-time signal y[m] as the impulse response of an IIR filter with frequency response $(1+\sum_{p=1}^{p} a_p e^{-j\omega p})^{-1}$. The inverse filter is an FIR filter with impulse response coefficients $h[m] = a_m$ and $a_0 = 1$. If y[m] is passed through this filter the output spectrum will be approximately constant provided that the actual signal spectrum is accurately modeled by Eq. (5.99). It follows that if the $\{a_p\}$ are chosen such that $|\hat{Y}(\omega)|^2$ is a good model of the power spectrum of random process data such as noise and clutter, then passing that data through the inverse filter will produce a new random process with an approximately flat power spectrum. Thus, the FIR filter designed from the model coefficients *whitens* the signal, removing any correlated signal components such as clutter.

Figure 5.24 illustrates the application of AR spectral estimation to design a clutter filter to enhance detection of windshear from an airborne radar (Keel, 1989). Part *a* shows the Fourier spectrum of the slow-time data from one range bin. Two peaks are evident above the noise floor. The one at zero velocity is ground clutter. The smaller peak at approximately 8 m/s is due to windblown rain. The middle plot shows the frequency response of an optimal clutter filter implemented from the $\{a_p\}$. The third plot shows the Fourier spectrum of the slow-time data after processing with the clutter filter. The ground clutter has been significantly suppressed and the weather echo is now the dominant spectral feature.



FIGURE 5.24 Clutter suppression and windshear detection using an autoregressive Doppler spectrum estimate: (*a*) Fourier spectrum of raw data, (*b*) frequency response of clutter suppression filter derived from the AR coefficients, (*c*) Fourier spectrum of filtered data. (Figure courtesy of Dr. Byron M. Keel, GTRI.)

5.3.8 CPI-to-CPI Stagger and Blind Zone Maps

Pulse Doppler processing sometimes is combined with pulse cancellers. In this case, the applicability of the concept of blind speeds is clear. If a pulse canceller is not used, there is no highpass filter and therefore a target whose Doppler shift equals an integer multiple of the PRF will not be filtered out as in MTI processing. However, the target energy will be indistinguishable from clutter energy since it will alias to the DC portion of the spectrum and combine with the clutter energy. Targets having Doppler shifts equal to a multiple of the PRF will still go undetected, so the corresponding target velocities are still blind speeds.

In CPI-to-CPI PRF stagger, a coherent processing interval of M pulses is transmitted at a fixed PRF. A second CPI is then transmitted at a different fixed PRF. Because the blind speeds are different for each PRF used, a target that falls in a blind speed of one PRF may be visible in the others. This concept is illustrated in Fig. 5.25, which shows a notional Doppler spectrum for two different PRFs. The plots are shown on the same frequency scale. First consider the upper spectrum plot, which corresponds to data collected at PRF_1 . A target whose Doppler equals PRF_1 will be aliased to zero Doppler shift, where it will be undetectable if clutter is present. If the same target scenario is measured with a lower pulse repetition frequency PRF_2 as shown in the lower half of the figure, the Doppler shift of the target no longer matches the PRF. The target energy aliases to a nonzero Doppler where it does not compete with the clutter and is still detectable.



FIGURE 5.25 The use of two PRFs to avoid blind speeds in pulse Doppler radar. The target Doppler shift equals the PRF in the upper sketch, but not in the lower sketch.

In some systems, as many as eight PRFs may be used. The first velocity that is blind at all of the PRFs is the LCM of the individual blind speeds, which will usually be much higher than any one of them alone. Target detections are typically accepted and passed to subsequent processing only if they occur in some minimum fraction of the PRFs used, for example two out of two or three out of eight PRFs.

The advantages of a CPI-to-CPI stagger system are that range-aliased clutter can be canceled within each CPI using coherent MTI and that the radar system stability, particularly in the transmitter, is not as critical as with a pulse-to-pulse stagger system (Schleher, 1991). The disadvantages are that the overall velocity response may not be very good and that the transmission of multiple CPIs consumes large amounts of the radar timeline and energy.

This discussion shows that for a given PRF there are periodic *blind zones* in Doppler spaced by the PRF. A target whose Doppler shift is such that it either is in the clutter region or aliases to the clutter region is unlikely to be detected. Equivalently, if the target falls into a clutter region caused by the periodically repeated clutter spectrum will likely go undetected. The idea is illustrated in Fig. 5.26*a*, which indicates clutter at zero Doppler and its first two periodic repetitions at $\pm PRF$ Hz. The target shown just above *PRF* Hz but still within the clutter region is considered to be in the blind zone centered at $F_D = PRF$ Hz.



FIGURE 5.26 Blind zones in Doppler and range: (*a*) blind zones in Doppler due to the periodic repetition of the mainlobe clutter spectrum, (*b*) blind zones in range due to pulse transmission and near-in clutter in a monostatic radar. Partial eclipsing of two of the three target echoes is also shown.

Recall that there are also blind zones in range or fast time in a monostatic radar as discussed in Sec. 3.1.3. The range blind zones due to pulse eclipsing are deliberately extended in some systems by keeping the receiver disconnected from the antenna for an additional period of time beyond the pulse length. The purpose is to avoid saturating the receiver with very strong near-in clutter returns, such as those from the altitude line. Figure 5.26*b* illustrates range blind zones and eclipsing. Pulses of duration τ are transmitted every *T* seconds as shown in the upper line. The middle line shows three target echoes after each pulse and strong near-in clutter returns. The bottom line shows that the receiver will be off during the time of transmission of the pulse and of reception of the

near-in clutter echoes. These "off" periods represent the blind zones in range. In this example, the first target is close enough to the radar that its echo is 50 percent eclipsed because it arrives during the near-in clutter interval. The third echo is also 50 percent eclipsed because the target is far enough away that a portion of the echo is received during transmission of the next pulse. The middle echo is not eclipsed.

Blind zones in range and Doppler collectively combine to form a twodimensional *blind zone map* as shown in Fig. 5.27. Targets that fall within a range blind zone are undetectable no matter what their Doppler shift. Targets that fall within a Doppler blind zone are undetectable at any range. The spacing of the blind zones in both dimensions is determined by the PRF of the radar. If the PRF is increased, the Doppler blind zones spread further apart but the range blind zones contract to become closer together. This observation leads to the idea of using CPI-to-CPI stagger in combination with an "M out of N" detection logic to maximize the combinations of range and Doppler shift at which targets can be detected.



FIGURE 5.27 Blind zone map for one PRF formed by combined blind zones in Doppler and range.

Figure 5.28 shows blind zone maps for a 10-GHz radar having a 10- μ s pulse length and using two PRIs, 100 μ s and 120 μ s, on successive CPIs. The range blind zones are not lengthened for near-in clutter. The clutter velocity spread is assumed to be ± 20 m/s. The velocity axis is shown only for positive velocity; the negative portion of the axis would be its mirror image. Both maps are shown to the same scale in range and velocity. Notice that the first range blind zone and the zero-velocity blind zone are the same for any PRF and so always overlap in all PRFs used. This is not true for their repetitions in range and Doppler due to the change in periodicity when the PRF is changed. For example, a target at a velocity of 300 m/s and range of 30 km is blind in both range and velocity at the 100- μ s PRI but is not in a blind zone at the 120- μ s PRI.



FIGURE 5.28 Blind zone maps for two different PRIs. The maps are drawn to the same velocity and range scales. Left: $PRI = 100 \ \mu s$. Right: $PRI = 120 \ \mu s$.

The left image of Fig. 5.29 results from overlaying the two blind zone maps in Fig. 5.28 on common intervals in range and velocity. Range-velocity coordinates that are in the blind zone at both PRFs are shown as black; those in only one of the two blind zones are shown in gray; and those in the clear on both PRFs are shown in white. For example, the target at 300 m/s and 30 km mentioned above falls in the gray area, suggesting that a radar collecting a CPI of data at each of these two PRIs would detect the target on only one of them (120 μ s). The image on the right is the blind zone map that would result from the use of these two PRIs and a "1 of 2" detection logic. This means that a detection at particular range-velocity coordinates is accepted as being a true target if it is detected on at least one of the CPIs; detection on both is not required. With this logic, a target at any range-velocity pair in the white or gray area of the left figure would be expected to be detectable (assuming adequate SIR). Only the coordinate pairs in the black area of the left image would be considered blind, leading to the blind zone map shown on the right.


FIGURE 5.29 Blind zones using two PRIs and a "1 of 2" detection rule. Left: Overlay of the two blind zone maps of Fig. 5.28. Right: Resulting blind zone map using "1 of 2" detection logic.

A problem with a "1 of N" rule is that a single false alarm would be accepted as a target. A more conservative "2 of 2" rule would mean a target was detectable only if it was in the clear on both PRIs. The effective blind zone map would be black in all of the black and gray regions of the left image in Fig. 5.29, with the result that very few combinations of range and velocity would correspond to detectable targets. Thus, there is a tradeoff between the rangevelocity coverage area in which targets are detectable and the immunity to false alarms and other error sources. Using values of N larger than two offers more compromise options and better results. Figure 5.30 shows an exceptionally good blind zone map for a system using eight PRFs with a "3 of 8" detection rule. The radar operates at 10 GHz with a 1-µs pulse length, blanking of 10 near-in clutter cells, a ± 17 m/s mainlobe clutter spread, and a 10-ms CPI. An evolutionary algorithm was used to select PRIs that maximized the detectable region within the range and velocity limits shown; the result was the set {51, 53, 60, 63, 67, 84, 89, 93} µs (Davis and Hughes, 2002). Nearly all of the range-velocity space shown is in the clear except for the first range and Doppler blind zones. "M of N' detection logic is discussed further in <u>Chap. 6</u>.



FIGURE 5.30 "3 of 8" blind zone map obtained using evolutionary algorithms to select the PRIs. See text for details.

5.4 Pulse Pair Processing

Pulse pair processing (PPP) is a form of Doppler processing common in meteorological radar. Unlike the MTI and pulse Doppler techniques discussed so far in this chapter, the goal of pulse pair processing is not clutter suppression to enable the detection of moving targets. In PPP it is assumed that the spectrum of the slow-time data consists of noise and a single Doppler peak, generally not located at zero Doppler (though it could be), that is due to echo from moving weather events such as wind-blown rain or other particulates. The goal of PPP is to estimate the parameters of that spectral peak. These are then used as the input to higher-level algorithms for estimating precipitation types and rates, forecasting severe weather, and so forth in both ground-based and airborne weather radars. In airborne radars, it is also one of the techniques used for windshear detection.

Pulse pair processing assumes the radar is looking generally upward if it is ground-based, or forward if it is airborne. Consequently, it is assumed that ground clutter competing with the weather signatures is small or negligible, or has been removed by MTI filtering. The notional Doppler spectrum $S_y(F)$ assumed by PPP is shown in Fig. 5.31. It consists only of white noise and a single spectral peak due to backscatter from weather-related phenomena



FIGURE 5.31 Notional slow-time power spectrum assumed in pulse pair processing.

$$S_y(F) = S_w(F) + S_n(F)$$
 (5.100)

The weather peak $S_w(F)$ is often assumed to be approximately Gaussian shaped and is characterized by its amplitude, mean, and standard deviation. The total area of the $S_w(F)$ power spectrum component equals the power of the meteorological echo. PPP is used to estimate the power, mean Doppler shift F_0 , and variance σ_F^2 (commonly called the *spectral width*) of the weather component. Each of these can be estimated using either time- or frequencydomain algorithms, all of which are included under the PPP rubric.

Consider time-domain measurements first. The autocorrelation and power spectrum of the slow-time data sequence y[m], m = 0, ..., M - 1 obtained from M pulses sampled at a particular range bin are

$$s_{y}[k] \equiv \sum_{m=0}^{M-k-1} y[m] y^{*}[m+k]$$
(5.101)

$$S_y(\omega) = \mathbf{F}\{s_y[k]\} = |Y(\omega)|^2$$
(5.102)

The power in the slow-time signal can be estimated in the time domain from the peak of the autocorrelation function

$$\hat{P}_{y} = \frac{1}{M} s_{y}[0] = \frac{1}{M} \sum_{m=0}^{M-1} y[m] y^{*}[m] = \frac{1}{M} \sum_{m=0}^{M-1} |y[m]|^{2}$$
(5.103)

To see how to estimate the mean frequency, ignore the noise for the moment and assume that the signal component is a pure sinusoid; the power spectrum of a finite segment would then be an asinc function squared. Now compute the first autocorrelation lag

$$y[m] = Ae^{j2\pi F_0 Tm} \quad m = 0, ..., M - 1$$

$$s_y[1] = \sum_{m=0}^{M-2} y[m]y^*[m+1]$$

$$= \sum_{m=0}^{M-2} Ae^{+j2\pi F_0 Tm} A^* e^{-j2\pi F_0 T(m+1)} = |A|^2 \sum_{m=0}^{M-2} e^{-j2\pi F_0 T}$$

$$= |A|^2 e^{-j2\pi F_0 T} \sum_{m=0}^{M-2} (1) = |A|^2 (M-1)e^{-j2\pi F_0 T}$$
(5.104)

The argument of the exponential, $-2\pi F_0 T$, is simply the negative of the amount of phase rotation in one sample period for a sampled sinusoid of original analog frequency F_0 Hz. The frequency can be estimated from Eq. (5.104) as

$$\hat{F}_0 = \frac{-1}{2\pi T} \arg\{s_y[1]\} \text{ Hz}$$
(5.105)

Multiplying \hat{F}_0 by $\lambda/2$ converts the result into units of velocity. Although derived for the pure sinusoid, this time-domain PPP frequency estimator works well for more general signals provided there is a single dominant frequency component with adequate SNR. The frequency estimate will be aliased if the Doppler frequency is outside the interval $\pm PRF/2$.

Since the complex exponential inside the summation in Eq. (5.104) does not depend on *m* and so could be brought out of the sum, it is not necessary to compute the full autocorrelation $\log s_y[1]$; it would suffice to simply compute $y[m]y^*[m + 1]$ using only two slow-time samples. In reality, noise is present in all of the samples and using all *M* available samples in the full summation averages the noise and improves the estimate quality.

To obtain a time-domain estimate of σ_F^2 , assume that the Doppler power spectrum exhibits a Gaussian shape with standard deviation σ_F . The estimate of F_0 can be used to remove the mean Doppler component, giving a modified sequence y'[m] with its Doppler spectrum centered at zero frequency. It is convenient to start with the continuous-time equivalent y'(t). $S_{y'}(F)$ will be a zero-mean Gaussian function

$$S_{y'}(F) = \frac{|A|^2}{\sqrt{2\pi\sigma_F}} e^{-F^2/2\sigma_F^2}$$
(5.106)

It follows that the continuous-time autocorrelation function is also Gaussian

$$s_{y'}(z) = |A|^2 e^{-2\pi^2 \sigma_F^2 z^2}$$
(5.107)

where the variable *z* represents the autocorrelation lag. If the sampling interval *T* is chosen sufficiently small to guarantee that $S_{y'}(1/2T) \approx 0$, the discrete time spectrum and autocorrelation will also form a Gaussian pair to a very good approximation (Richards, 2006). The sampled autocorrelation function will be

$$s_{y'}[k] = s_{y'}(z)|_{z=kT} = |A|^2 e^{-2\pi^2 \sigma_F^2 k^2 T^2}$$
(5.108)

and the corresponding DTFT, still in units of hertz, is

$$S_{y'}(F) = \frac{|A|^2}{\sqrt{2\pi}\sigma_F T} e^{-F^2/2\sigma_F^2}$$
(5.109)

Because $s_{v'}[0] = |A|^2$, the first autocorrelation lag can be written

$$s_{y'}[1] = |A|^2 \exp\left(-2\pi^2 \sigma_F^2 T^2\right) = s_{y'}[0] \exp\left(-2\pi^2 \sigma_F^2 T^2\right)$$
(5.110)

Equation (5.110) is easily solved to give an estimate for the spectrum standard deviation in terms of only $s_{y'}[0]$ and $s_{y'}[1]$

$$\hat{\sigma}_F^2 = -\frac{1}{2\pi^2 T^2} \ln\left\{\frac{s_{y'}[1]}{s_{y'}[0]}\right\} \quad \text{Hz}^2$$
(5.111)

Equations (5.103), (5.105), and (5.111) are the time-domain pulse pair processing estimators. They can be computed from only two autocorrelation lags of the slow-time data in each range bin.

Equation (5.111) is sometimes simplified to avoid the natural logarithm calculation. Consider the following series expansion and approximation of ln(x)

$$\ln x = \frac{x-1}{x} + \frac{1}{2} \left(\frac{x-1}{x}\right)^2 + \frac{1}{3} \left(\frac{x-1}{x}\right)^3 + \cdots$$
$$\approx \frac{x-1}{x} = 1 - \frac{1}{x}$$
(5.112)

Applying Eq. (5.112) to Eq. (5.111) gives the simplified spectral width estimator

$$\hat{\sigma}_F^2 = -\frac{1}{2\pi^2 T^2} \left\{ 1 - \frac{s_{y'}[0]}{s_{y'}[1]} \right\} \quad \text{Hz}^2$$
(5.113)

The basic PPP measurements of signal power, frequency, and spectral width can also be performed in the frequency domain. The power is obtained by applying Parseval's theorem to Eq. (5.103)

$$\hat{P}_{y} = \frac{1}{2\pi M} \int_{-\pi}^{+\pi} S_{y}(\omega) d\omega = \frac{1}{2\pi M} \int_{-\pi}^{+\pi} |Y(\omega)|^{2} d\omega$$
(5.114)

A practical calculation uses the DFT version of Parseval's theorem with the DFT Y[k] of y[m]

$$\hat{P}_{y} = \frac{1}{M^{2}} \sum_{k=0}^{M-1} |Y[k]|^{2}$$
(5.115)

There are two frequency-based methods for estimating the mean frequency of the signal. The first is a direct analog to Eq. (5.105)

$$\hat{F}_{0} = \frac{-1}{2\pi T} \arg\{s_{y}[1]\} \\ = \frac{-1}{2\pi T} \arg\{\frac{1}{2\pi} \int_{-\pi}^{+\pi} |Y(\omega)|^{2} e^{j\omega} d\omega\} \quad \text{Hz}$$
(5.116)

The integrand in Eq. (5.116) is real except for the $e^{j\omega}$ term. Noting that for a complex number *z*, $\arg\{z\} = \operatorname{atan}\{\operatorname{Im}(z)/\operatorname{Re}(z)\}, \operatorname{Eq.}(5.116)$ gives the estimator

$$\hat{F}_{0} = \frac{-1}{2\pi T} \operatorname{atan} \left\{ \frac{\int_{-\pi}^{+\pi} |Y(\omega)|^{2} \sin \omega \, d\omega}{\int_{-\pi}^{+\pi} |Y(\omega)|^{2} \cos \omega \, d\omega} \right\} \quad \text{Hz}$$
(5.117)

In practice, the DFT version is used:

$$\hat{F}_{0} = \frac{-1}{2\pi T} \operatorname{atan} \left\{ \frac{\sum_{k=0}^{K-1} |Y[k]|^{2} \sin(2\pi k/K)}{\sum_{k=0}^{K-1} |Y[k]|^{2} \cos(2\pi k/K)} \right\}$$
Hz (5.118)

The other frequency-domain mean frequency and spectral width estimators result from viewing the signal spectrum of Fig. 5.31 as a probability density function. A valid PDF must be real and nonnegative, a condition met by the power spectrum. However, a PDF must also have unit area, so the power spectrum must be normalized to ensure this is the case. By Parseval's theorem, the integral of $|Y(\omega)|^2 = 2\pi E_y$, where E_y is the energy in y[m]; this is the required normalization factor. For any arbitrary PDF $p_z(z)$ the mean and variance are given by

$$\overline{z} = \int_{-\infty}^{+\infty} z p_z(z) dz$$

$$\sigma_z^2 = \int_{-\infty}^{+\infty} (z - \overline{z})^2 p_z(z) dz$$
(5.119)

Applying the first of these definitions to the power spectrum gives an alternative mean frequency estimator

$$\hat{F}_{0} = \frac{1}{2\pi T} \int_{-\pi}^{+\pi} \omega (|Y(\omega)|^{2} / 2\pi E_{y}) d\omega = \frac{1}{4\pi^{2} T E_{y}} \int_{-\pi}^{+\pi} \omega |Y(\omega)|^{2} d\omega \quad \text{Hz}$$
(5.120)

Similarly, an estimator of the spectral width is

$$\hat{\sigma}_{F}^{2} = \frac{1}{(2\pi T)^{2}} \int_{-\pi}^{+\pi} (\omega - \omega_{0})^{2} (|Y(\omega)|^{2} / 2\pi E_{y}) d\omega$$
$$= \frac{1}{8\pi^{3} T^{2} E_{y}} \int_{-\pi}^{+\pi} (\omega - \omega_{0})^{2} |Y(\omega)|^{2} d\omega \quad \text{Hz}^{2}$$
(5.121)

Generally, the time-domain estimators are preferred if the SNR is low or the spectral width is very narrow (Doviak and Zrnić, 1993). In the latter case, the signal is closer to the pure sinusoid assumption that motivated the timedomain estimator. In addition, the time-domain methods are more computationally efficient because no Fourier transform calculations are required. Conversely, the frequency-domain estimators tend to provide better estimators at high SNR and large spectral widths. The frequency domain estimator also allows reduction of the noise before the estimates are calculated, reducing the estimate bias. This process, called *spectral subtraction*, is depicted in Fig. 5.32. The noise power spectrum $N(\omega)$ is estimated from a presumed clear region of the spectrum and then simply subtracted off to form a reduced-noise power spectrum



FIGURE 5.32 Spectral subtraction.

$$S'_{y}(\omega) = S_{y}(\omega) - N(\omega)$$
(5.122)

Because of the statistical variations of any given data set, it is possible that $S'_{y}(\omega)$ may have some negative values; these are usually set to zero.

Figure 5.33 shows two images from the KFFC WSR-88D NEXRAD weather radar located in Peachtree City, Georgia, just south of Atlanta and operated by the U.S. National Weather Service. The images were collected on March 19, 1996 and show a heavy storm in the area. While these images are much more easily viewed in color than in grayscale, some features are visible. The image on the left is a map of the power estimate, proportional to the volume reflectivity η . Lighter grays represent areas of heavier rainfall. The image on the right is a map of the radial velocity measured by the radar and thus of the wind speed. The large area on the left and top-left of the image with the circular inner boundary represents a range-aliased region where reliable velocity estimates are not available. The radar itself is at the center of the circle having this boundary. Inside that radius, the black area at the top represents a high velocity toward the radar, while the dark gray area inside the circular contour and to the left and bottom of the image represents high wind speeds away from the radar. Thus, the wind is blowing from the top to the bottom in this image. The various square and round markers in the power and velocity images are created by the analysis software and flag various features in the storm.



FIGURE 5.33 Sample weather radar images obtained with pulse pair processing: (*a*) power image, (*b*) velocity image. (Images courtesy of National Severe Storms Laboratory.)

5.5 Additional Doppler Processing Issues

5.5.1 Combined MTI and Pulse Doppler Processing

It is not unusual to have both MTI filtering for gross clutter removal and pulse Doppler spectrum analysis for detailed examination of the Doppler spectrum. Since both operations are linear, the order in which they are applied would appear to make no difference to the final Doppler spectrum used for detection. However, differences in signal dynamic range can make their order significant in considering hardware effects when finite-wordlength hardware is used.

Clutter is usually the strongest component of the signal; it can be several tens of decibels above the target signals of interest. If the DFT of the slow-time signal is computed prior to MTI filtering, the sidelobes of the response from the clutter around DC may swamp potential target responses at near-in velocities, masking these targets from possible detection. If the processor dynamic range is limited as well, the effect of the strong clutter signal on processor automatic gain control may drive the target signal amplitude below the minimum detectable signal of the processor, effectively filtering out the target.

For these reasons, the MTI filter is generally placed first if both processes are used. The MTI filter will attenuate the clutter component selectively so that the target signals become the dominant components. Subsequent finite wordlength processing will adapt the dynamic range to the targets rather than the now-suppressed clutter. In the increasingly common floating point processors, dynamic range is less of an issue.

5.5.2 Transient Effects

All of the discussion in this chapter has assumed a steady-state scenario in the sense that the clutter spectrum is stationary and that filter transient effects have been ignored. In range-ambiguous medium and high PRF modes the received signal sample in each range bin contains contributions from multiple ranges because of the multiple contributing pulses. Whenever the radar PRF changes several pulses, known as *clutter fill pulses*, they must be transmitted before a steady-state situation is achieved. Here "steady state" means that the physical clutter intervals contributing to a given range bin are the same for each pulse so that the clutter statistics can be expected to be stationary from one pulse to the next. For example, suppose that in steady state each range gate contains significant contributions from L = 4 pulses (four range ambiguities). Then the clutter signature in each range bin will reach a steady-state condition only for the fourth and subsequent pulses in a CPI. This issue was first discussed in <u>Chap. 4</u>.

Steady-state operation of the digital filters used for MTI processing occurs when the output value depends only on actual data input values rather than any initial (typically zero-valued) samples used to initialize the processing. For FIR filters of length N, the first and last N - 1 samples of the complete convolution are transient in the sense that the filter impulse response does not fully overlap the finite data sequence. These transient output samples are often discarded.

These two effects are independent. To see how many pulses are needed in total to obtain an *M*-point non-transient, steady-state sequence $y_{ss}[m]$, consider Fig. 5.34. Assume P total pulses are transmitted. This sketch assumes L = 4range ambiguities and a three pulse canceller (N = 3) MTI filter, but is labeled for general L and N. The notional data sequence v[m] is shown as ramping up in amplitude over the first L samples. While actual data would vary unpredictably depending on the clutter scenario, this represents the increasing number of range ambiguities present in each sample, stabilizing at four when m = 3 (the fourth sample). Recall that the convolution of h and y is given by $y_{ss}[n] = \sum y[m]h[n - 1]$ *m*]. The three-sample sequence in the figure represents the three-pulse canceller filter coefficients h[n - m]. (The actual coefficient values would be $\{+1, -2, -2\}$ +1}). It can be seen from the figure that in general the first value of *n* for which the filter coefficients will overlap only with steady-state measured data occurs when L - 1 = n - N + 1, i.e. n = L + N - 2 (n = 5 in this example). The last value for which this is true occurs when n = P - 1. The number of output samples in this interval is M = (P-1) - (L+N-2) + 1. Therefore, P = M + L + N - 2pulses are needed to obtain M valid outputs for further processing.



FIGURE 5.34 Determination of the relationship between number of range ambiguities L, MTI filter length N, and number of steady-state slow-time samples P needed to obtain M non-transient steady-state samples.

For example, suppose 20 valid stationary pulses are needed for the pulse Doppler DFT. Also suppose a three-pulse canceller (N = 3) is used and the unambiguous range and radar sensitivity are such that L = 4 range ambiguities are present in the measured data. Then the CPI should collect 25 pulses of data, discard the first five outputs of the MTI filter (three for the range ambiguity buildup and two for the filter transient) and pass only the last 20 outputs to the pulse Doppler DFT or other processing. Additional pulses may be used to set the automatic gain control of the receiver and are also not used for Doppler processing.

5.5.3 PRF Regimes

As was seen in <u>Chap. 4</u>, measurements made with a pulse burst waveform can be ambiguous in range, Doppler, or both. Pulse Doppler radars in particular frequently operate in scenarios that are ambiguous in one or both of the range and Doppler dimensions. Modern airborne pulse Doppler radars operate in a dizzying variety of modes having various range and Doppler span and resolution requirements. Pulse burst waveforms using a variety of constituent pulses, including simple pulses, LFM, and Barker phase codes at a minimum, are common. To meet the various mode requirements, PRFs ranging from several hundred hertz to 100 kHz or more are used.

Pulse Doppler radar operation is commonly divided into three *PRF* regimes according to their ambiguity characteristics. Many radars operate in all three regimes, depending on the requirements of the moment. The dividing lines are not absolute, but depend on the mission requirements. Given an unambiguous range R_{ua} and unambiguous velocity v_{ua} of interest, where v_{ua} is one-half the blind speed v_b , the radar is considered to be in a *low PRF mode* if the PRF is sufficiently low to be unambiguous in range over the interval of interest, but is ambiguous in velocity, meaning targets of interest have velocities outside the range $\pm v_{ua}$. The *high PRF mode* is the opposite: the system is ambiguous in soft. This tradeoff is summarized in Fig. 5.35. The line plots the achievable combinations of R_{ua} and v_{ua} at 10 GHz. Suppose the desired range and velocity

coverage are 60 km and 300 m/s (± 150 m/s), respectively. The range limit might be set by the expected maximum detection range of the radar, while the velocity limit might be set by the maximum expected relative velocity of targets of interest. The shaded area indicates a range of PRFs (in this case, 2.5 to 20 kHz) that will result in ambiguities in both dimensions. PRFs below 2.5 kHz will be ambiguous in Doppler but not in range; those from 2.5 to 20 kHz will be ambiguous in both; and those above 20 kHz will be ambiguous in range but not in Doppler.



FIGURE 5.35 Low, medium, and high PRF regimes for a notional X-band radar.

The choice of PRF regime has major effects on the range-Doppler target and clutter spectrum. Consider again the radar and clutter scenario associated with Fig. 5.4 and suppose the maximum range and velocity intervals of interest are 75 km and ± 150 m/s. Now consider viewing the same scenario with low, medium, and high PRFs of 2, 10, and 30 kHz, respectively. Table 5.3 lists the unambiguous range, Doppler, and velocity intervals for each of these PRFs for reference.

PRF	PRF Regime	Unambiguous, Range km	Unambiguous Doppler Shift Interval, kHz	Unambiguous Velocity Interval, m/s
2	Low	75	±1	±15
10	Medium	15	±5	±75

30	High	5	±15	±225	

TABLE 5.3 Unambiguous Range, Doppler, and Velocity Intervals for SelectedPRFs and a 10-GHz RF

Figure 5.36 compares the unaliased range-Doppler clutter spectrum of Fig. 5.4b to the aliased versions at each of these PRFs. In all cases the clutter echo out to a range of 75 km has been included in the computation. Figure 5.36a repeats the unaliased spectrum for convenient comparison. The MLC is centered at 17.3 km and 93.4 m/s as noted earlier. The SLC spreads over ± 134.1 m/s but is mostly confined to ranges of 15 km or less.





FIGURE 5.36 Effect of low, medium, and high PRFs of <u>Table 5.3</u> on the ground clutter range-Doppler spectrum: (*a*) unaliased spectrum, (*b*) viewed with a 2-kHz PRF, (*c*) viewed with a 10-kHz PRF, (*d*) viewed with a 30-kHz PRF. See text for details.

Part b of the figure is the low PRF spectrum. The mainlobe clutter is

centered at its unambiguous range of 17.3 km, but the velocity has aliased from its actual value of 93.4 m/s to an ambiguous velocity of $3.4 \text{ m/s} (93.4 - 3 \times 30)$. The SLC is heavily aliased in velocity but fades with range so that most targets do not compete with it. If MTI filtering is applied to suppress the MLC there will be blind speeds every 30 m/s in the affected range bins.

Part *c* of the figure is the medium PRF case. The MLC is ambiguous in both range and Doppler, having aliased to a range of 2.3 km (17.3 - 15) and a velocity of -56.6 m/s (93.4 - 150). Notice that the MLC and its sidelobes now wrap around in the range dimension and that the SLC wraps around in Doppler. SLC is now present at essentially all ranges and Dopplers, though in varying amounts and patterns in different range cells. The SLC also wraps in range but this is less evident.

Part *d* of the figure is the high PRF case. The MLC again wraps to the ambiguous range of 2.3 km $(17.3 - 3 \times 5)$ but is spread fairly uniformly throughout the short 5 km unambiguous range. It is located at its unambiguous velocity of 93.4 m/s. The narrow spread of the MLC in velocity allows it to be filtered out with a relatively narrow MTI or other notch filter with little risk of filtering out moving targets. The clutter out to 75 km has folded over 15 times to "fit" into the 5 km unambiguous range at this PRF. There is now significant and relatively constant SLC at all ranges, though the AL and other near-in clutter is still visible beginning at just over 2 km. On the other hand, the radar is now unambiguous in Doppler and the full SLC spread in velocity of ± 134.1 m/s can be seen. In addition, there is now a clear region in the Doppler spectrum for velocities having a magnitude between 134.1 and 225 m/s that was not present in the other figures, enabling noise-limited detection of targets at these high relative velocities.

Table 5.4 summarizes the major strengths and weaknesses of low, medium, and high PRF operation, especially from the viewpoint of an airborne radar. Broadly speaking, low PRF modes are very effective for ranging, mapping, and imaging modes, but poor at detection of moving targets due to the lack of a sizable clear region. High PRF operation is complementary to low PRF operation in both its strengths and weaknesses. High PRF modes are good for detection of high-Doppler shift targets (e.g., rapidly closing aircraft or missiles) in high clutter due to the large clear region, but poor at detection of low-Doppler targets (slow-moving closing targets or opening targets) due to high sidelobe clutter and little or no range gating capability. Medium PRF operation is a compromise that retains most of the strengths of each without the inherited weaknesses becoming too severe. In-depth discussion of the properties and processing for all three regimes is given in Alabaster (2012), Morris and Harkness (1996), and Stimson (1998).

PRF Regime	Advantages	Disadvantages
Low	 No range ambiguities Precise range measurement Fine range resolution possible Sidelobe clutter rejection via range gating Simple processing 	 Existence of blind speeds Highly ambiguous Doppler, esp. at higher RFs Poor detection performance in look- down modes High peak power or pulse compression required
Medium	 Good detection over wide range of target Dopplers Good rejection of both mainlobe and sidelobe clutter Accurate ranging Reduced eclipsing compared to high PRF operation 	 Sidelobe clutter at all velocities Large number of PRF and pulse width combinations Complex range and Doppler ambiguity resolution processing Poor performance for large targets in sidelobes
High	 High average power Unambiguous Doppler No blind speeds Mainlobe clutter rejection without rejecting targets 	 Highly ambiguous range Increased eclipsing of targets Complex, reduced-accuracy ranging Reduced sensitivity to low-Doppler targets due to sidelobe clutter

TABLE 5.4 Partial List of Advantages and Disadvantages of PRF Regimes forAirborne Radar

For a given PRF regime, the problem of picking the particular PRFs to be used remains. Two commonly cited approaches are the "major-minor method" and the "*M-of-N* method." The major-minor method begins by picking at least two and more likely three "major" PRFs in the regime of interest. The PRFs are separated by at least the MLC Doppler width β_{MLC} to guarantee that the first repetitions of the mainlobe clutter portion of the Doppler spectrum at each PRF do not overlap, thus making detection presumably possible on at least one of the PRFs for targets outside of the zero-Doppler clutter region. For each major PRF, two "minor" PRFs are then selected for use in range ambiguity resolution. This technique results in sets of six or nine PRFs in total.

The *M*-of-*N* method decides how many total PRFs will be used (often eight in airborne radar) and on how many a detection will be required (usually three). Any three PRFs on which a detection is observed can be used for range ambiguity resolution. A variety of techniques have been used to search for sets that result in good blind zone maps within bounds on the maximum and minimum PRF; Fig. 5.30 was a particularly good example. A typical ad hoc approach begins by choosing PRFs that are stepped by at least $\beta_{MLC}/(N - M)$ Hz. This choice ensures that the first Doppler blind zones for each PRF do not overlap for more than N - M PRFs, meaning that any target Doppler shift in that vicinity will be in the clear on at least M PRFs.

There are some constraints that must be satisfied by the PRFs in either method. The minimum desired percentage of the Doppler spectrum that is outside the MLC region sets a minimum PRF. The maximum allowable duty cycle of the radar transmitter and the pulse length (derived from the required range resolution when a simple pulse is used) establish a maximum PRF. Candidate PRFs are often adjusted so that there are an integer number of range bins within the unambiguous range for each PRI; this eases ambiguity resolution. In addition to these practical considerations, the set of PRFs used for disambiguating range and velocity have to satisfy the "decodability constraints."

$$LCM(PRI_{1},...,PRI_{M}) \ge \frac{2R_{ua}}{c} \quad (\text{range decodability constraint})$$
$$LCM(PRF_{1},...,PRF_{M}) \ge F_{Db} \quad (\text{Doppler decodability constraint}) \quad (5.123)$$

where R_{ua} and F_{Db} are the maximum range and Doppler shift coverage of interest. These constraints ensure that the ambiguity resolution algorithms discussed next can provide a unique solution within that coverage area. Additional constraints for robust operation include decodability and blindness "margins" in range and Doppler and minimum transmit time. Generally, the best PRF sets are obtained using advanced constrained search techniques rather than the simpler major-minor or *M*-of-*N* methods.

Extensive discussion of PRF selection considerations and examples with an emphasis on airborne multimode radars is given in Alabaster (2012). PRF selection may be less complex in radars with less varied missions and less complicated environments. For example, weather radars typically map weather conditions in their surrounding region using fewer and less complex PRF sets. In weather radar, volume clutter is the target of interest. Ground clutter is minimal because the system is not generally down-looking (although airborne weather radars may be). Weather radars typically require relatively long unambiguous ranges and therefore low PRFs to provide adequate area coverage. The WSR-88D radar used by the U.S. National Weather Service for long-range weather observation uses PRFs of 322 Hz to 1282 Hz, giving unambiguous ranges of 466 to 117 km. The RF is approximately 3 GHz, so the corresponding unambiguous velocity intervals are ± 8 m/s (about ± 18 mph) and ± 32 m/s (about ± 72 mph). They are well short of the approximately ± 100 mph velocity interval considered adequate by meteorologists. Consequently, ambiguous windspeed measurements are a common problem in weather radar. Because weather radars are interested in continuous reflectivity fields (wind flows, storm cells, etc.) distributed in three dimensions instead of discrete targets, they can also take advantage of special ambiguity resolution methods that rely on continuity of the measured velocities and other special features of the measured data; some of these techniques are described in Doviak and Zrnić (1993). Another approach combines pulse pair processing with two staggered PRFs to compute two autocorrelation values that can be combined to extend the unambiguous range.

Another novel technique applies a pulse-to-pulse phase code similar to the LFM-like P3 and P4 codes discussed in <u>Chap 4</u>. Correlating with specific elements of the code on receive can emphasize a particular range foldover region at the expense of the others. These two techniques are discussed in Zrnić (2008).

5.5.4 Ambiguity Resolution

Several techniques exist to resolve range and Doppler ambiguities when multiple-PRF data is available. Consider range ambiguity resolution first. Once a PRF is selected, it establishes an unambiguous range $R_{ua} = c/2PRF$. A target at an actual range $R_t > R_{ua}$ will be detected at an apparent range R_a that satisfies

$$R_t = R_a + kR_{ua} \tag{5.124}$$

for some integer k. Equivalently

$$R_a = ((R_t))_{R_{ua}}$$
(5.125)

where the notation ((·)) _{*x*} denotes modulo *x*. Normalize the range measurements to the range bin spacing ΔR , for example, $n_a = R_a / \Delta R$; then

$$n_t = n_a + kN \quad \Rightarrow \quad n_a = ((n_t))_N \tag{5.126}$$

The basic approach to resolving range ambiguities relies on multiple PRFs. Suppose that there are N_i range bins in the unambiguous range interval on PRF *i*; thus, $R_{uai} = N_i \Delta R$. Note that the unambiguous range is different for each PRF. For simplicity, assume that the range bin spacing is the same in each PRF. Then

$$n_t = n_{a_0} + k_0 N_0 = n_{a_1} + k_1 N_1 = \dots \implies n_{a_i} = ((n_t))_{N_i}$$
(5.127)

The set of equations in Eq. (5.127) is called a set of *congruences*.

The set of congruences can be solved using the *Chinese remainder* theorem (CRT) (Trunk and Brockett, 1993). The CRT states that given a set of r relatively prime integers N_0 , N_1, \ldots, N_{r-1} and the set of congruences in Eq. (5.127), there exists a unique solution (modulo $N = N_0 N_1 \dots N_{r-1}$) for n_t given by the equations

$$\begin{split} n_t &= k_0 \beta_0 n_{a_0} + k_1 \beta_1 n_{a_1} + \dots + k_{r-1} \beta_{r-1} n_{a_{r-1}}, \\ k_i &= N/N_i = \prod_{j=0, \, j \neq i}^{r-1} N_j, \quad \beta_i = \left(\left(k_i^{-1} \right) \right)_{N_i} \implies ((\beta_i k_i))_{N_i} = 1 \end{split}$$

To make the procedure clearer, consider the case of r = 3 PRFs. Then n_t satisfies

$$n_t = ((\alpha_0 n_{a_0} + \alpha_1 n_{a_1} + \alpha_2 n_{a_2}))_{N_0 N_1 N_2}$$
(5.129)

where

$$\alpha_i = \beta_i k_i = \beta_i \prod_{\substack{j=0\\j\neq i}}^2 N_j$$
(5.130)

(for example, $\alpha_1 = \beta_1 N_0 N_2$) and the β_i are the smallest integers such that

$$((\beta_0 N_1 N_2))_{N_0} = 1, ((\beta_1 N_0 N_2))_{N_1} = 1, ((\beta_2 N_0 N_1))_{N_2} = 1$$
(5.131)

To illustrate the procedure, suppose that the true range of a target, normalized to the range bin size, is $n_t = 19$. Further suppose that the three PRFs are chosen such that the number of range bins in the unambiguous range for each PRF are $N_0 = 11$, $N_1 = 12$, and $N_2 = 13$. On the first PRF the target will be detected in the apparent range bin $n_{a_0} = ((19))_{11} = 8$. Similarly, $n_{a_1} = ((19))_{12} = 7$ and $n_{a_2} = ((19))_{13} = 6$. From Eq. (5.131), β_0 is the smallest integer that satisfies $((\beta_0 \times 12 \times 13))_{11} = 1$; that is, β_0 satisfies $\beta_0 \times 12 \times 13 = 156\beta_0 = 11k + 1$ for some integer k. The solution is $\beta_0 = 6$. In the same manner it is found that $\beta_1 = 11$ and $\beta_2 = 7$. Equation (5.130) then gives $\alpha_0 = 6 \times 12 \times 13 = 936$, $\alpha_1 = 1573$, and $\alpha 2 = 924$. Finally, Eq. (5.129) gives the estimate of the true range bin as

$$\hat{n}_t = ((\alpha_0 n_{a_0} + \alpha_1 n_{a_1} + \alpha_2 n_{a_2}))_{N_0 N_2 N_3} = 19$$
(5.132)

which is the correct result.

A serious problem with the CRT is its extreme sensitivity to errors induced by noise and range quantization. There is no guarantee that the actual range R_t will be an integer multiple of the range bin spacing ΔR as assumed previously; the target may in fact straddle range bins. In addition, noise in the measurements may cause the target to be located in an incorrect range bin. To illustrate the effect of such errors, repeat the previous example but suppose that n_{d_2} is for some reason measured to be 7 instead of the correct value of 6. Carrying out the previous calculations will give $\hat{n}_t = 943$ instead of 19, a huge error.

A "robust CRT" that controls the maximum error is given in Li et al.

(2010). However, it is more useful to introduce the *coincidence algorithm* for determining n_t . This technique is essentially a graphical implementation of the CRT (Hovanessian, 1976; Morris and Harkness, 1996). The method is best illustrated with an example. Again presume that r = 3 PRFs are used. Suppose that there are two targets with true ranges corresponding to range bins $n_a = 6$ and $n_b = 11$. Further suppose that PRFs are such that the number of range bins in each unambiguous range interval are $N_0 = 7$, $N_1 = 8$, and $N_2 = 9$. This means that the first target is actually unambiguous at each PRF, while the second is ambiguous at each PRF. The measured data will be

$$n_{a_0} = n_{a_1} = n_{a_2} = 6$$

 $n_{b_0} = 4, \quad n_{b_1} = 3, \quad n_{b_2} = 2$
(5.133)

This measurement scenario is illustrated in Fig. 5.37.



FIGURE 5.37 Notional measured data for illustrating coincidence algorithm for range ambiguity resolution.

The graphical technique proceeds by taking the pattern of detections at each PRF and replicating it as shown in Fig. 5.38. In essence, the replication implements Eq. (5.126), placing a detection at each value of $n_a + kN_0$ and $n_b + kN_0$ within the maximum detection range of the radar. These detections represent the plausible ranges for each target at each PRF. The algorithm then searches for a range bin that exhibits a detection at all three PRFs, indicating that that range bin is consistent with the measurements at all three PRFs. As shown in Fig. 5.38, this process correctly detects the true range bins $n_a = 6$ and $n_b = 11$ in this example.



FIGURE 5.38 Coincidence detection of target ranges in replicated range data.

The graphical interpretation suggests various methods to reduce the sensitivity of the CRT to measurement errors. In one approach, exact coincidence is not required to declare a target. Instead, a tolerance N_T is established and a detection is declared if a detection occurs in all three PRFs at some range bin $n_t \pm N_T$. Depending on the range bin size and SNR, N_T will typically be only 1 or 2 range bins. A more sophisticated version of this basic idea is described in Trunk and Kim (1994). Their method combines a systematic approach to clustering plausible ranges from each PRF into candidate target ranges with a maximum likelihood calculation to recognize multiple target situations.

In the last example three PRFs proved sufficient to resolve two different range-ambiguous targets. In general, N PRFs are required to successfully disambiguate N - 1 targets. If the number of targets exceeds N - 1, ghosts can appear (Morris and Harkness, 1996). Ghosts are false targets resulting from false coincidences of range-ambiguous data from different targets. The problem is illustrated in Fig. 5.39, which repeats the example of Fig. 5.38 using only two of the previous three PRFs. While targets will still be detected at the correct bins $n_a = 6$ and $n_b = 11$, a third coincidence occurs between detections from targets 1 and 2 at range bin $n_c = 20$, representing an apparent third target. Unless additional data such as tracking information is available, the signal processor has no way of recognizing that the last coincidence is among detections from different targets. Thus, the processor will declare the presence of three targets in this example, the two correct targets and one "ghost." Use of a third PRF as in Fig. 5.38 eliminates this ghost. More extensive discussion of ghosting in range and Doppler and of ghosting due to false alarms is given in Alabaster (2012).



FIGURE 5.39 Formation of ghosts in range ambiguity resolution.

In a medium or high PRF mode the radar may also suffer velocity ambiguities. This problem is identical to that of range ambiguities: given an apparent Doppler shift F_t , the actual Doppler shift must be of the form $F_t + k \cdot PRF$ for some integer k. Use of the DFT for spectral estimation results in quantization of the Doppler spectrum into Doppler bins (equivalently, velocity bins), analogous to range bins in the range dimension. The same techniques used for range disambiguation can therefore be used to resolve velocity ambiguities as well.

The coincidence algorithm can be readily extended to simultaneously disambiguate detections in both range and Doppler. Other extensions to deal with radar systems that do not use the same range resolution in each PRF (typically so as to maintain constant duty cycle) or the same velocity resolution, as well as references to other disambiguation algorithms are given in Alabaster (2012). A newer approach using the emerging technique of sparse reconstruction has been proposed in Shaban and Richards (2013). It also works in range and Doppler simultaneously and appears tolerant of reasonable numbers of missed detections due to blind zones or other causes, false alarms, and other data inconsistencies.

5.6 Clutter Mapping and the Moving Target Detector

5.6.1 Clutter Mapping

All of the MTI and pulse Doppler processing discussed so far has been focused on reducing the clutter power that interferes with the signature of a moving target so as to improve the SIR and ultimately the probability of detection. These techniques are not effective for targets with little or no Doppler shift, and that therefore are not separable from the clutter based on Doppler shift. *Clutter mapping* is a technique for detection of moving targets with zero or very low Doppler shift. It is intended for maintaining detection of targets on crossing paths, that is, passing orthogonal to the radar line of sight so that the radial velocity is zero; such targets are discarded by MTI and pulse Doppler processing. Clutter mapping can be effective if the target RCS is relatively large and the competing clutter is relatively weak, a situation depicted in Fig. 5.40 that can arise for instance in a ground-based air surveillance radar. In that situation the antenna is tilted upward so that mainlobe ground clutter is not present to compete with the target echo (though weather clutter may be); the clutter is primarily from the sidelobes.



FIGURE 5.40 Pulse Doppler spectrum for a large RCS crossing target in weak clutter.

The concept of clutter mapping is shown in Fig. 5.41, which presumes that conventional pulse Doppler processing is applied to targets having Doppler shifts sufficient to separate them from the ground clutter, i.e., those in the clear region of the Doppler spectrum. The output of the zero-Doppler bin and others in the clutter region is used to create a stored map of recent clutter echo power for each range-azimuth cell in the radar's search area. This map is updated continuously to allow for clutter variations due to weather and other environmental changes. On each scan, the received power in the clear region Doppler bins is applied to a conventional threshold detector using a threshold based on the noise that dominates the interference in those bins. Instead of being discarded, the current received power in the clutter region Doppler bins for each range-azimuth cell is applied to a separate detector using a threshold based on the stored clutter power level for that cell. The clutter map procedure is a form of constant false alarm rate (CFAR) detection but with the interference power estimated by averaging in time instead of in space. The details of threshold detection and CFAR are discussed in Chap. 6



FIGURE 5.41 The concept of clutter mapping for detection of strong targets in clutter.

Instead of using the zero-Doppler output of a pulse Doppler processor (typically an FFT of the slow-time data), many clutter maps systems pass the I/Q slow-time data through a separate "zero velocity filter" as shown in Fig. 5.42. The zero velocity filter serves the opposite purpose of an MTI filter. It is a lowpass design, the output of which consists only of ground clutter and crossing target returns. The design of the zero-Doppler filter can be optimized for the clutter environment at a specific radar site and can also be made adaptive to clutter changes, for instance due to weather in the area.



FIGURE 5.42 Zero-Doppler filter used to isolate low-Doppler targets and ground clutter.

5.6.2 The Moving Target Detector

The *moving target detector* (MTD) is a term applied to the Doppler processing system used in many airport surveillance radars. The MTD combines most of the techniques discussed previously and others to achieve good overall moving

target detection performance. A block diagram of the original MTD is shown in Fig. 5.43 (Nathanson, 1991). The upper channel begins with a standard threepulse canceller. The clutter-cancelled output is then applied to an 8-point FFT for pulse Doppler analysis. Two PRFs are used in a CPI-to-CPI stagger to extend the unambiguous velocity region. The "frequency domain weighting" is an implementation of time-domain windowing of the data for Doppler sidelobe control. Certain windows, including for example the Hamming, can be efficiently implemented as a convolution in the frequency domain with a 3-point kernel. The individual FFT samples are applied to a 16-range-bin cell-averaging CFAR threshold detector (to be discussed in Chap. 6) with thresholds selected separately for each frequency bin.



FIGURE 5.43 Block diagram of a complete "moving target detector" system.

To provide some detection capability for crossing targets, the lower channel uses a site-specific zero-velocity filter to isolate the echo from clutter and low-Doppler targets. The output is again applied to a clutter map threshold detector. The original MTD updated the clutter map using an 8-scan moving average, corresponding to 32 seconds of data history (Skolnik, 2001).

The MTD design has progressed through several generations since the implementation described here. Versions used in the ASR-9 and ASR-12 airport surveillance radars are described respectively in Taylor and Brunins (1985) and Cole et al (1998). Additional discussion of the design and performance of the Doppler filterbank and zero-velocity filters is given in Shrader and Gregers-Hansen (2008).

5.7 MTI for Moving Platforms: Adaptive Displaced Phase Center Antenna Processing

5.7.1 The DPCA Concept

MTI filtering and pulse Doppler processing provide an effective way to detect moving targets whose Doppler shift is in the clear region of the spectrum on at least one PRF. Airborne targets can generally be detected in this manner. However, slow-moving ground targets having actual Doppler shifts only slightly higher than the ground clutter or less will appear within the clutter spectrum or on its skirts at all PRFs and are therefore very difficult to detect. Recall that platform motion can substantially spread the ground clutter spectrum as described in Eq. (5.72). This spread of mainlobe clutter exacerbates the problem, raising the minimum velocity at which slow-moving ground targets can be detected. This phenomenon is illustrated in Fig. 5.44, which shows the spreading of the mainlobe clutter by the platform motion after the change in Doppler center frequency has been removed. Because of this spreading, clutter energy may compete directly with relatively slow-moving targets ("slow movers," typically surface targets such as vehicles on land and ships on the sea), making MTI processing less effective and detection difficult. Processing techniques intended to detect such "slow movers" from moving platforms are referred to as ground moving target indication (GMTI) or surface moving *target indication* (SMTI).



FIGURE 5.44 Illustration of the effect of a moving radar platform on the Doppler spectrum and the detection of "slow movers." The change in Doppler center frequency has been removed.

Displaced phase center antenna (DPCA) processing is a technique for countering the platform-induced clutter spectral spreading. By minimizing the clutter spectral width, DPCA improves the probability of detection for slowmoving targets. It is a special case of the more general *space-time adaptive processing* (STAP) introduced in <u>Chap. 9</u>. Heuristically, the basic concept is to create the effect of a stationary antenna even though the platform is moving forward by electronically moving the receive aperture backward during operation, thus avoiding the clutter spreading. More specifically, DPCA processing attempts to compensate for aircraft motion by using multiple receive subapertures to create carefully controlled multiple virtual phase centers such that data received on one subaperture have the same virtual phase center as the data received on a *different* subaperture some time later. By properly combining samples staggered by this delay time from the slow-time data streams from the same range bin but different receive subapertures, effective MTI cancellation can be achieved. References for basic DPCA are Skolnik (2001), Shaw and McAulay (1983), Staudaher (1990), and Lightstone et al. (1991).

Figure 5.45 illustrates the concept using an electronic antenna that has two subapertures. The entire antenna is used on transmission for maximum gain so the phase center for transmission is the point **T** in the middle of the antenna. Each half of the antenna has its own receiver so there are in effect two receive apertures having respective phase centers **R1** and **R2** which are each d_{pc} meters from the transmit phase center. If the transmit phase center is located at position x_0 on the first pulse transmitted, the forward receive phase center is at $x_0 + d_{pc}$ and the aft receive phase center is at $x_0 - d_{pc}$, a separation of $2d_{pc}$. The effective phase center for a complete transmit-receive path is approximately halfway between the transmit and receive phase centers. Thus, for the common full-array transmit apertures and the two receive apertures the effective two-way phase centers are at $x_0 + d_{pc}/2$ and $x_0 - d_{pc}/2$, a separation of d_{pc} meters.



FIGURE 5.45 Relationship of transmit and receive aperture phase centers in DPCA processing.

Now consider the motion of the platform over M_s pulses. If the pulse repetition interval is T and the platform velocity is v, the effective transmitreceive phase centers move forward by vM_sT meters in M_s PRIs. The idea of DPCA is to achieve effective MTI cancellation by combining pulses measured from the same phase center location in space in a pulse canceller. Specifically, if the **T-R1** phase center is at position $x_0 + d_{pc}/2$ on the first pulse, then M_s pulses later the **T-R2** transmit phase center will be at position $x_0 - d_{pc}/2 + vM_sT$. Equating these two positions gives

$$M_{\rm s} = \frac{d_{\rm pc}}{vT} \tag{5.134}$$

 M_s is the "time slip" in pulses.

The significance of the time slip given by Eq. (5.134) is that the clutter component of the slow-time signal received on the aft receive aperture is highly correlated with the slow-time signal in the same range bin received on the forward receive aperture M_s pulses earlier because it represents the same ground clutter measured from the same point in space. Consequently, two-pulse cancellation can be implemented by taking each sample from the **R1** data stream and subtracting the sample from the **R2** data stream taken M_s pulses later as illustrated in Fig. 5.46. Even though these data samples were collected on different receive apertures and one or more pulses apart in time, their effective transmit-receive phase centers are the same so they appear equivalent to successive pulses from a *stationary* antenna. The effective stationarity of the antenna then implies that the clutter spectral width is not spread by the platform motion, therefore improving the clutter cancellation and the detection of slowmoving ground targets.



FIGURE 5.46 Illustration of two-pulse cancellation across two received data streams in DPCA for a time slip of approximately $M_s = 3$ PRIs.

In general, M_s will not be integer. For example, if $d_{pc} = 3$ meters, v = 200

m/s and T = 2 ms, then $M_s = 3.75$ pulses. A typical DPCA implementation will round M_s to the nearest integer for coarse alignment of the two data streams and then use adaptive processing as described next to achieve good clutter cancellation.

5.7.2 Adaptive DPCA

While conventional bandlimited interpolation could be used to implement fractional-PRI timing adjustments, in practice there will also be mismatches between channels that will make it impossible to achieve high cancellation ratios even if the time alignment is perfect. Adaptive processing can be combined with the basic DPCA cancellation to minimize the clutter residue at the processor output and therefore maximize the improvement factor. The following discussion of adaptive DPCA is modeled after the "suboptimum matched filter algorithm" in Shaw and McAulay (1983). This algorithm assumes that an integer PRI delay of one channel with respect to the other is used to achieve coarse time alignment of the two-phase center channels to be combined. Each received signal channel is then divided into Doppler bins using a DFT. MTI cancellation is performed independently in each subband, allowing the adaptive cancellation weight to be optimized separately for each Doppler bin and improving overall performance.

The vector analysis approach will be used to model the signals and develop the adaptive filtering. Transmit a CPI of $M + M_s$ pulses and collect L range bins of data for each pulse and each of the N = 2 phase centers, resulting in an $L \times (M + M_s) \times 2$ datacube y[l, m, n]. Advance the aft channel slow-time data (n = 1) to coarse-align it with the fore channel (n = 0) in each range bin and retain only the overlapped slow-time samples to obtain the $L \times M \times 2$ datacube

$$\tilde{\mathbf{y}}[l,m,n] = \left[\frac{\mathbf{y}[l,m,0]}{\mathbf{y}[l,m+M_{s},1]}\right] \equiv \left[\frac{y_{f}[l,m]}{y_{a}[l,m+M_{s}]}\right]$$
(5.135)

where $y_f[l, m]$ is the "fore" channel data plane, $y_a[l, m]$ is the "aft" channel data plane, and the dotted horizontal line represents vertical concatenation of datacube planes. The pulse number (slow time) index *m* is now in the range $0 \le m \le M - 1$. The separation of the datacube into phase center planes is done for convenience because the DPCA filter will apply weighting in that dimension only. Now take the *K*-point DFT of the data in each range bin to get the $L \times K \times 2$ range-Doppler datacube

$$\tilde{Y}[l,k,n] = \begin{bmatrix} -\frac{Y_f[l,k]}{e^{-j2\pi M_s k/K} Y_a[l,k]} \end{bmatrix}$$

(5.136)

Equation (5.136) points out that the time slip adjustment adds a linear phase term to the aft channel Doppler data. This effect applies to both target and interference signal components. For a given range-Doppler bin [l, k], $\tilde{Y}[l, k, n]$ is a 2 × 1 column vector.

The signals at each subaperture consist of clutter, noise, and (if present) target components. Because the weighting will be in the phase center dimension, a model of the covariance matrix $S_I = \mathbf{E}\{\tilde{\mathbf{Y}}^* \; \tilde{\mathbf{Y}}^T\}$ of the phase center data for each range-Doppler bin similar to that of Eq. (5.19) is needed. Begin with the clutter. It is not white in slow time, so its power spectrum is not flat and the clutter covariance $\sigma_c^2[l,k]$ is a function of the range index *l* and Doppler index *k*. The clutter at the same range-Doppler bin is correlated across the spatial phase center channels to some degree determined by the platform motion and time slip correction. That correlation is denoted by the normalized phase-center dimension correlation function $\rho[k]$ and also varies with the Doppler bin. The thermal noise is assumed uncorrelated between channels and is white. Therefore, for a fixed range-Doppler bin \mathbf{S}_I will take the 2 × 2 form

$$\mathbf{S}_{\mathbf{I}}[l,k] = \mathbf{S}_{\mathbf{I}}[k] = \begin{bmatrix} \sigma_{c}^{2}[k] + \sigma_{n}^{2} & \rho[k]\sigma_{c}^{2}[k] \\ \rho^{*}[k]\sigma_{c}^{2}[k] & \beta[k](\sigma_{c}^{2}[k] + \sigma_{n}^{2}) \end{bmatrix}$$
(5.137)

The coefficient $\beta[k]$ accounts for any mismatch in the gain and frequency response or subaperture antenna patterns of the two channels.

Next, a model similar to Eq. (5.23) or (5.25) for the target data in the Doppler domain is needed. A CPI of fast-time/slow-time data for a moving point target in range bin l_t is modeled as

$$\mathbf{t}[l,m] = A_t \,\,\delta[l-l_t]e^{j2\pi \,F_{D}mT} \begin{bmatrix} \gamma_f\\ \gamma_a \end{bmatrix}$$
(5.138)

complex scalar constants representing the unknown target where γ_f and γ_a = phases and (possibly unequal) amplitudes in the fore and aft receive channels

 A_t = target amplitude

 δ [·] = discrete impulse function

The target phases represented by γ_f and γ_a are determined by the absolute range and the angle of arrival as well as the electrical lengths of the receive paths.

It is useful to note the relationship between the angle of arrival and the

phase difference between the signals observed at the same moment on the fore and aft antenna subapertures. Figure 5.47 shows a wavefront impinging on two receivers representing the two antenna phase centers separated by a distance d_{pc} . The angle of arrival of the wavefront, measured from the normal to the line connecting the two receivers, is θ_a . As shown in the figure, the additional distance the wavefront must travel after arriving at the first receiver before it reaches the second receiver is $d_{pc} \sin \theta_a$. Consequently, the phase difference in the signals received at the two sites becomes



FIGURE 5.47 Geometry for relating angle of arrival to phase difference at two subaperture receive phase centers. The angle of arrival shown is considered negative; positive angles of arrival are clockwise from the normal to the line connecting the phase centers.

$$\Delta \phi = \frac{2\pi}{\lambda_k} d_{\mu c} \sin \theta_a \tag{5.139}$$

The subscript k on the wavelength has been added to emphasize that the appropriate wavelength should be used for each Doppler bin to which Eq. (5.139) will be applied.

Using Eq. (5.139), assuming $|\gamma_f| = |\gamma_a| = 1$, and letting *R* be the range from the scatterer to the fore aperture (rightmost in Fig. 5.47), γ_f and γ_a of Eq. (5.138) can be expanded as follows

$$\gamma_{f} = \exp\left[j\left(\frac{4\pi}{\lambda_{k}}R + \psi_{f}\right)\right]$$
$$\gamma_{a} = \exp\left[j\left(\frac{4\pi}{\lambda_{k}}R + \psi_{a} - \Delta\phi\right)\right]$$

The terms ψ_f and ψ_a represent the receiver phase shifts, which are different for each channel in general.

After the Doppler DFT, the target data model of Eqs. (5.138) and (5.140) results in the following target range-Doppler domain signal model

$$\mathbf{t}[l,k] = MA_t \,\delta[l-l_t] \,\delta[k-k_t] \begin{bmatrix} \gamma_f \\ e^{-j2\pi \,M_s k/K} \gamma_a \end{bmatrix}$$
(5.141)

where $k_t = (F_D KT/2\pi)$ is the target Doppler shift converted to an equivalent DFT bin number. The two impulse functions serve to confine the response to range bin l_t and Doppler bin k_t . If K > M there will be multiple DFT samples on the target DTFT mainlobe and the assumption that the target response is essentially confined to one Doppler bin is less valid.

As before, the SIR can be maximized with a matched filter that computes the scalar quantity for each range-Doppler bin

$$\mathbf{z}[l,k] = \left(\mathbf{S}_{\mathbf{I}}^{-1}[k]\mathbf{t}^*\right)^T \mathbf{Y}[l,k] = \mathbf{t}^H[l,k] \left(\mathbf{S}_{\mathbf{I}}^{-1}[k]\right)^* \mathbf{Y}[l,k]$$
(5.142)

Since the target signal AOA is not known a priori the target signal vector is again averaged over all values of AOA in $[-\pi/2, \pi/2]$. Absorbing out all common constants into the complex amplitude, the new target vector becomes simply

$$\mathbf{t}[l,k] = \mathbf{t} = M\hat{A}_t \begin{bmatrix} 1\\0 \end{bmatrix}$$
(5.143)

The target location in range-Doppler space is not assumed known so the same target model is used in each range and Doppler bin.

An alternative to assuming an unknown AOA is to assume specific values for θ_a . For example, a series of values in an interval equal to the antenna mainlobe width and centered on the nominal steering angle used on transmit might be generated and used. However, the loss in SIR from using the much simpler Eq. (5.143) is very small (Shaw and McAulay, 1983).

The exact clutter and noise statistics are also not known a priori. Consequently, S_I cannot be known exactly, but it can be estimated from the data. Since the clutter covariance is expected to have essentially the same form in every range bin, one way to estimate S_I would be to compute a sample average

$$\hat{\mathbf{S}}_{\mathbf{I}}[k] = \frac{1}{L_2 - L_1 + 1} \sum_{l=L_1}^{L_2} \mathbf{Y}^*[l, k] \mathbf{Y}'[l, k]$$
(5.144)

This estimate of $\hat{\mathbf{S}}_{\mathbf{I}}$ then replaces the actual $\mathbf{S}_{\mathbf{I}}$ in Eq. (5.142). Since the coefficients used to combine the fore and aft data streams are computed from the data itself, this is now an *adaptive* DPCA processor. This method for estimating $\mathbf{S}_{\mathbf{I}}$ is analogous to cell-averaging CFAR interference estimation to be discussed in <u>Chap. 6</u> and revisited in <u>Chap. 9</u>.

Equation (5.144) implicitly assumes that the covariance matrices in the range bins adjoining bin k are all the same as the covariance matrix in bin k itself. Even if the physical clutter is the same over the averaging interval, this assumption also requires that a preprocessing gain control step compensate for the expected R^3 variation⁹ in clutter power with range. Noise power does not vary with range or Doppler.

Combining Eqs. (5.136), (5.137), (5.142), and (5.143) gives the output of the DPCA system. Assuming that $\hat{\mathbf{S}}_{\mathbf{i}}$ is a good approximation to $\mathbf{S}_{\mathbf{i}}$ and absorbing all constants into a single constant α gives

$$z[l,k] = \alpha \left\{ \beta[k] \left(1 + \frac{\sigma_n^2}{\sigma_c^2} \right) Y_f[l,k] - \rho^*[k] e^{+j2\pi M_s k/K} Y_a[l,k] \right\}$$
(5.145)

While complicated, the structure of a two-pulse canceller is clearly present in the subtraction of $\mathbf{Y}_a[l, k]$ from $\mathbf{Y}_f[l, k]$. If the interference is clutter-limited so that $\sigma_c^2 \gg \sigma_n^2$ and also highly correlated across phase centers so that $\rho[k] \rightarrow 1$ (implying that the coarse alignment was very successful), the output simplifies to

$$z[l, k] = \alpha \left\{ \beta[k] \mathbf{Y}_{f}[l, k] - e^{+j2\pi M_{s}k/K} \mathbf{Y}_{a}[l, k] \right\}$$
(5.146)

The two-pulse canceller structure is clearer here. The complex exponential in the Doppler index k is equivalent to a time-domain shift of M_s samples, in accordance with the DPCA condition discussed earlier. The factor $\beta[k]$ provides a Doppler-dependent weighting factor that can be optimized to maximize cancellation in each Doppler channel.

The matched filter design assumes that $\hat{\mathbf{S}}_{\mathbf{I}}$ is an estimate of the covariance of the interference only, i.e., it should not contain any target signal components. A practical system must take steps to ensure this is the case, perhaps by skipping

range bins containing targets already in track, averaging over enough range bins to minimize any unknown target influences, prescreening the data for large amplitudes that might indicate a target, or other means. Many of the required techniques are similar to those used in constant false alarm rate detection, discussed in <u>Chap. 6</u>.

DPCA is a clutter suppression technique, and as such it makes sense to apply it only on the Doppler bins in the clutter region. It is easy to see that if the interference is noise-limited the adaptive DPCA operation reduces to performing no operation (see Prob. 36). Consequently, DPCA should not be applied in the clear region of the Doppler spectrum.

References

- Alabaster, C., *Pulse Doppler Radar: Principles, Technology, Applications.* SciTech Publishing, Edison, NJ, 2012.
- Cole, E. L. et al., "ASR-12: A Next Generation Solid State Air Traffic Control Radar," *Proceedings of the 1988 IEEE Radar Conference*, Dallas, TX, pp. 9–13, May 1998.
- Davis, P. G. and E. J. Hughes, "Medium PRF Set Selection Using Evolutionary Algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 933–939, Jul. 2002.
- Doviak, D. S., and R. J. Zrnić, *Doppler Radar and Weather Observations*, 2d ed. Academic Press, San Diego, CA, 1993.
- Eaves, J. L., and E. K. Reedy (eds.), *Principles of Modern Radar*. Van Nostrand Reinhold, New York, 1987.
- Harris, F J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proceedings of the IEEE*, vol. 68, no. 1, pp. 51–83, Jan. 1978.
- Hayes, M. H., *Statistical Digital Signal Processing and Modeling*. Wiley, New York, 1996.
- Hovanessian, S. A., "An Algorithm for Calculation of Range in a Multiple PRF Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 12, no. 2, pp. 287–290, Mar. 1976.
- *IEEE Standard Radar Definitions*, IEEE Standard 686-2008. Institute of Electrical and Electronics Engineers, New York.
- Jacobsen, E., and P. Kootsookos, "Fast, Accurate Frequency Estimators," *IEEE Signal Processing Magazine*, pp. 123–125, May 2007.
- Keel, B. M., "Adaptive Clutter Rejection Filters for Airborne Doppler Weather Radar," M.S. Thesis, Clemson University, Clemson, AL, 1989.
- Kretschmer, F. F., Jr., "MTI Visibility Factor," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 2, pp. 216–218, Mar. 1986.

Levanon, N., Radar Principles. Wiley, New York, 1988.

- Lightstone, L., D. Faubert, and G. Rempel, "Multiple Phase Center DPCA for Airborne Radar," *Proceedings of the 1991 IEEE National Radar Conference*, pp. 36–40, Mar. 1991.
- MacLeod, M. D., "Fast Nearly ML Estimation of the Parameters of Real of Complex Single Tones or Resolved Multiple Tones," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 141–148, Jan. 1998.
- Morris, G. V., and L. Harkness (eds.), *Airborne Pulse Doppler Radar*, 2d ed. Artech House, Boston, MA, 1996.
- Nathanson, F. E., *Radar Design Principles*, 2d ed. McGraw-Hill, New York, 1991.
- Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing*, 3d ed. Prentice Hall, Englewood Cliffs, NJ, 2010.
- Richards, M. A., "Coherent Integration Loss due to White Gaussian Phase Noise," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 208–210, Jul. 2003.
- Richards, M. A., "Discrete-Time Gaussian Fourier Transform Pair, and Generating a Random Process with Gaussian PDF and Power Spectrum," unpublished technical note, Oct. 3, 2006. Available at <u>http://www.radarsp.com</u>.
- Richards, M. A., "The Discrete-Time Fourier Transform and Discrete Fourier Transform of Windowed Stationary White Noise," unpublished technical note, Oct. 27, 2007. Available at <u>http://www.radarsp.com</u>.
- Richards, M. A., "Doppler Processing," Chap. 17 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.
- Richards, M. A., "A Slight Extension of 'Coherent Integration Loss due to White Gaussian Phase Noise," unpublished technical note, Mar. 3, 2011. Available at <u>http://www.radarsp.com</u>.
- Roy, R., and O. Lowenschuss, "Design of MTI Detection Filters with Nonuniform Interpulse Periods," *IEEE Transactions on Circuit Theory*, vol. CT-17, no. 4, pp. 604–612, Nov. 1970.
- Schleher, D. C, *MTI and Pulse Doppler Radar*, 2d ed. Artech House, Boston, MA, 2010.
- Shaban, F., and M. A. Richards, "Application of L¹ Reconstruction of Sparse Signals to Ambiguity Resolution in Radar," *Proceedings IEEE 2013 Radar Conference*, Ottawa, May 2013.
- Shaw, G. A., and R. J. McAulay "The Application of Multichannel Signal Processing to Clutter Suppression for a Moving Platform Radar," *IEEE Acoustics, Speech, and Signal Processing (ASSP) Spectrum Estimation Workshop II*, Nov. 10–11, Tampa, FL, 1983.
- Shrader, W. W., and V. Gregers-Hansen, "MTI Radar," Chap. 2 in M. I. Skolnik (ed.), *Radar Handbook*, 3d ed. McGraw-Hill, New York, 2008.
- Skolnik, M. I., *Introduction to Radar Systems*, 3d ed. McGraw-Hill, New York, 2001.
- Staudaher, F. M., "Airborne MTI," Chap. 16 in M. I. Skolnik (ed.), *Radar Handbook*, 2d ed. McGraw-Hill, New York, 1990.
- Stimson, G. W., *Introduction to Airborne Radar*, 2d ed. SciTech Publishing, Mendham, NJ, 1998.
- Taylor, J. W., Jr., and G. Brunins, "Design of a New Airport Surveillance Radar (ASR-9)," *Proceedings of the IEEE*, vol. 73, no. 2, pp. 284–289, Feb. 1985.
- Trunk, G., and S. Brockett, "Range and Velocity Ambiguity Resolution," *Record* of the *1993 IEEE National Radar Conference*, pp. 146–149, Apr. 20–22, 1993.
- Trunk, G., and M. W. Kim, "Ambiguity Resolution of Multiple Targets Using Pulse-Doppler Wave forms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 4, pp. 1130–1137, Oct. 1994.
- Vergara-Dominguez, L., "Analysis of the Digital MTI Filter with Random PRI," *IEEE Proceedings*, *Part F*, vol. 140, no. 2, pp. 129–137, Apr. 1993.
- Zrnić, D., "Weather Radar—Recent Developments and Trends," Proceedings Microwaves, Radar and Remote Sensing Symposium, 2008 (MRRS 2008), pp. 174–178, Kiev, Ukraine, Sept. 2008.

Problems

- 1. An aircraft has a 4° azimuth 3-dB beamwidth. The RF is 10 GHz and the antenna is steered to a squint angle ψ of 30°. If the aircraft flies at 100 m/s, what is the Doppler spread of the clutter echoes induced by the aircraft motion?
- 2. Suppose the aircraft in the previous problem has a PRF of 10 kHz. Sketch a Doppler spectrum similar to that of Fig. 5.1*c*, but with noise and mainlobe clutter components only. What range of Doppler shifts lie in the mainlobe clutter region of the spectrum? What range of Doppler shifts lie in the clear region of the spectrum? What percentage of the total spectrum width from -PRF/2 to +PRF/2 is in the clear region?
- 3. If the PRF in the previous problem is changed to 1 kHz, what percentage of the total spectrum width will lie in the clear region?
- 4. Consider a 10-GHz airborne radar traveling straight, level, and forward at 200 mph at an altitude of 30,000 feet. The antenna is pointed at an azimuth angle of 0° and an elevation angle of -20° , similar to Fig. 5.3*a*. What will be the radial velocity in meters per second of echoes from stationary scatterers in the following locations: (a) directly ahead of the aircraft, (b) at the point where the antenna boresight intercepts the ground, (c) directly

below the aircraft, (d) on the ground directly to the left or right of the aircraft (azimuth angle of $\pm 90^{\circ}$), and (e) directly behind the aircraft?

- 5. Verify all of the calculations regarding Fig. 5.4 given in the text. These include boresight range to the ground, component of velocity along the boresight, Doppler shift at the middle of the mainlobe clutter, maximum Doppler shift of clutter, and maximum radial velocity of clutter.
- 6. For the aircraft in Prob. 4, sketch the approximate unaliased slant rangevelocity distribution of the ground clutter (mainlobe + sidelobe) in a "bird's-eye" format similar to that of Fig. 5.36*a*. The slant range axis of the sketch should cover 0 to 100 km and the velocity axis should cover $\pm v_{max}$, where v_{max} is the maximum possible radial velocity in meter per second that could be observed from scatterers in front of the radar. It is not necessary to represent antenna gain effects; concentrate on indicating where the mainlobe clutter will be centered and the intervals in range and velocity where clutter energy will be seen.
- 7. Suppose the radar in the previous problem has an operating frequency of 10 GHz and a PRF of 3 kHz. What are the unambiguous slant range R_{ua} and blind velocity interval v_b ? Sketch the approximate aliased range-velocity distribution of the ground clutter (mainlobe + sidelobe) in a format similar to that of "bird's-eye" format similar to that of Fig. 5.36c. The slant range axis of the sketch should cover 0 to R_{ua} and the velocity axis should cover $\pm v_{ua} = \pm v_{ub}/2$. Concentrate on indicating where the mainlobe clutter will be centered and the intervals in range and velocity where clutter energy will be seen.
- 8. Use the vector form of the matched filter to find the coefficients of an optimum two-pulse (N = 2) MTI filter under the assumptions that (a) the interference is white noise only, and (b) only approaching targets are of interest, that is, those with positive Doppler shifts; however, targets can be approaching with any positive velocity between 0 and $\lambda \cdot PRF/4$ (corresponding to $F_D = PRF/2$) being equally likely. Repeat for receding targets. Interpret the result: that is, explain how the particular form of the filter coefficients will maximize the SIR at the filter output for this target and interference model.
- 9. Suppose an MTI radar is placed on a moving platform such that the clutter spectrum, rather than being centered at normalized radian frequency $\omega = 0$, is instead centered on $\omega = \pi/2$. If the clutter observed by a stationary platform was c[m], the new clutter process can be modeled crudely as $c'[m] = c[m]\exp(j\pi m/2)$. (This is an oversimplified model because it does not account for the broadening of the clutter spectrum caused by platform motion, but that effect is ignored in this problem for simplicity.) Assume the target can be at any velocity with equal likelihood. Ignore the noise,

i.e., assume $\sigma_n^2 = 0$. Assume the stationary clutter c[m] has power σ_c^2 and first autocorrelation lag $\rho \sigma_c^2$. Use the vector approach to find the coefficients of the optimum two-pulse (N = 2) matched filter for this case. Compare to the N = 2 vector matched filter for the stationary radar. Explain how the shift in the clutter power spectrum center frequency changes the coefficients. Interpret the resulting coefficients as in the previous problem.

10. Use the vector form of the matched filter to find the coefficients of an optimum two-pulse MTI filter under the assumptions that (a) the interference is white noise plus a pure complex sinusoid at some specific normalized radian frequency ω_J , and (b) target velocity is uniform random over the entire spectrum. Specifically, the interference signal becomes

w[m] = n[m] + q[m] $q[m] = A_J \exp(j\omega_J m)$

where n[m] is a stationary white noise process with power σ_n^2 . This could be a simple model for a jammer at frequency ω_J . What is the frequency response $H(\omega)$ of the filter having the resulting coefficients? Explain how this frequency response will affect the jammer and target signals.

- 11. Consider a pulse-to-pulse staggered PRF system using a series of P = 3 PRIs, namely $\{66.\overline{6} \ \mu s, 83.\overline{3} \ \mu s, 100 \ \mu s\}$.
 - a. What are the base interval T_g and the set of staggers $\{k_p\}$?
 - b. What is the average PRI, T_{avg} ?
 - c. What is the first blind Doppler frequency, F_{bs} , of the staggered system?
 - d. What is the ratio F_{bs}/F_b of the first blind Doppler of the staggered system to the first blind speed of a constant-PRF system having the average PRI T_{avg} ?
- 12. Now consider the effect of the staggered PRIs of the previous problem on range coverage.
 - a. What is the unambiguous range, R_{us} , corresponding to the average PRI from Prob. 11? (This would be the unambiguous range of a constant-PRI system that used the same amount of time to collect N pulses as the staggered-PRF system.)
 - b. For the three PRFs in Prob. 11, what is the maximum unambiguous range R_{uas} ?
 - c. What is the factor by which the range coverage (unambiguous range) is reduced in the staggered PRF system of Prob. 11 relative to the unstaggered system?
- 13. Using the numerical results from the previous two problems, determine the

percentage by which the total range-velocity coverage, defined as the product of the unambiguous range and first blind Doppler, increased in the staggered case as compared to the unstaggered case? In other words, compute $R_{uas}F_{bs}/R_{ua}F_b$. Show that the numerical results agree with the result predicted by Eq. (5.46).

- 14. Assume a set of integer staggers $\{k_p\}$ is mutually prime. Show that $lcm\{1/k_p\} = 1$. *Hint* #1: For any set of numbers β , $\alpha \cdot lcm(\beta_1, ..., \beta_p) = lcm(\alpha\beta_1, ..., \alpha\beta_p)$. *Hint* #2: Use hint #1 with $a = \beta_1 \cdot \beta_2 \cdot ... \cdot \beta_p$ and the prime factorization method for finding the least common multiple.
- 15. Consider a three-pulse canceller; thus the filter coefficients are $h[m] = \{1, -2, 1\}$. Suppose the clutter c[m] at the canceller input has an autocorrelation function given by

$$s_c[k] = \sigma_c^2(\delta[k] + 0.5\delta[|k| - 1] + 0.25\delta[|k| - 2])$$

- a. Find the clutter power spectrum on a normalized radian frequency scale, $S_c(\omega)$.
- b. The improvement factor *I* was expressed as the product of "gain" and "clutter attenuation," $I = G \cdot CA$, in Eq. (5.48). Find the gain *G* for the three-pulse canceller using Eq. (5.49).
- c. Use the autocorrelation method to find the improvement factor *I* for the three-pulse canceller with this clutter process.
- 16. Repeat Prob. 15(c) using the frequency domain approach. The value of G from Prob. 15(b) can be reused.
- 17. Repeat Prob. 15(c) using the vector matched filter approach. Recompute both G and CA using the vector method.
- 18. Compute the clutter attenuation CA when using a two-pulse canceller, assuming the clutter Doppler power spectrum is of the form

$$S_{c}(F_{D}) = \begin{cases} \sigma_{c}^{2}, |F_{D}| \leq F_{co} \\ 0, F_{co} < |F_{D}| \leq PRF/2 \end{cases}$$

Assume $F_{co} \leq PRF/2$ and express the answer in terms of F_{co} .

19. Consider an M = 32 pulse sequence of slow-time data collected with *PRF* = 10 kHz. A radix 2 FFT algorithm is used to compute the Doppler spectrum of the data. If the Doppler frequency samples are to have a spacing of 100 Hz or less, what is the minimum FFT size *K* that should be used? What is the resulting spacing of the Doppler frequency samples in hertz?

- 20. Consider a C band (5 GHz) radar using a pulse repetition frequency of *PRF* = 3500 pulses per second. The radar collects 30 pulses of data. For a given range, the slow-time data sequence is zero-padded and input to a 64-point DFT to compute the Doppler spectrum. What is the spacing of the DFT samples in normalized radian frequency (i.e, on the $-\pi$ to $+\pi$ scale)? What is the spacing in hertz? In meters per second? What is the Rayleigh resolution (peak-to-first null width) in Doppler, in hertz? In meters per second?
- 21. An X-band (10 GHz) pulse-Doppler radar collects a fast-time/slow-time matrix of 30 pulses by 200 range bins per pulse. This is converted to a range-Doppler matrix by applying a Hamming window and then a 64-point fast Fourier transform to each slow-time row. Suppose that there is a target with a constant radial velocity of 30 m/s approaching the radar at a range corresponding to range bin #100. The PRF is 6000 samples per second. There is no ground clutter, and noise can be ignored as well. For which FFT sample index k_0 is $|Y[k_0]|$ the largest? (Remember that the DC sample is k = 0.) What velocity in meters per second does this sample correspond to? What is the error between the apparent velocity based on the largest FFT sample and the actual velocity?
- 22. Continuing Prob. 21: in terms of the window function w[m], give an expression for the peak value of the DTFT (not DFT) of the windowed data in range bin #100, assuming that each slow-time sample has an amplitude of 1 before windowing. What is the numerical value of this peak? (MATLAB® or a similar computational tool can be used to compute this value). Now suppose the peak value of the magnitude of the FFT of the data $|Y[k_0]| = 15.45$. What is the straddle loss in dB?
- 23. Continuing with Probs. 21 and 22, suppose also that $|Y[k_0 1]| = 11.61$ and $|Y[k_0 + 1]| = 14.61$. Use the amplitude-based quadratic interpolation technique of Eqs. (5.96) and (5.97) to estimate the velocity of the target and the peak amplitude of the DTFT. Compute the new values of velocity error and straddle loss and compare to those found in Probs. 21 and 22.
- 24. Consider two radars. The first is a 3-GHz weather radar having a desired unambiguous range of $R_{ua} = 300$ km and unambiguous velocity $v_{ua} = v_b/2$ of 50 m/s (about 112 mph). The second is a 10-GHz airborne radar having a desired unambiguous range of $R_{ua} = 100$ km and unambiguous velocity $v_{ua} = v_b/2$ of 250 m/s (about 560 mph). For each radar, is a 1 kHz PRF considered to be a low, medium, or high PRF?
- 25. What is the lowest PRF that would be considered "high" for each of the two radars in the previous problem?
- 26. Consider range ambiguity resolution using three PRFs. Suppose the three

PRFs correspond to $N_0 = 4$, N1 = 5, $N_2 = 7$ range cells. A single target is detected in the first range bin on the first PRF, the fourth range bin on PRF #2, and the second range bin on PRF #3; i.e., $n_{a \ 0} = 1$, $n_{a \ 1} = 4$, $n_{a \ 2} = 2$. Assume the radar sensitivity is such that it could possibly detect targets out to the 15th range bin. Use the coincidence algorithm to determine the true range bin number for this target.

- 27. Repeat the previous problem using the Chinese Remainder Theorem approach.
- 28. Again consider range ambiguity resolution using three PRFs. Suppose the three PRFs correspond to $N_0 = 3$, $N_1 = 4$, and $N_2 = 5$ range cells. Two targets are detected in each PRF: in range bins 1 and 2 of the first PRF, bins 1 and 4 of the second PRF, and bins 2 and 4 of the third PRF. Assume the radar sensitivity is such that it could possibly detect targets out to the 25th range bin. How many targets will be reported by the coincidence algorithm, and in which range bins?
- 29. Now assume that in the previous problem, one false alarm occurs on the third PRF in range bin #3. How many targets will be reported by the coincidence algorithm, and in which range bins?
- 30. Now suppose that in Prob. 28, the target in range bin #4 on the third PRF is not detected (a missed detection). How many targets will be reported by the coincidence algorithm, and in which range bins?
- 31. Finally, suppose only the second and third PRFs in problem 28 are used. How many targets will be reported by the coincidence algorithm, and in which range bins?
- 32. Suppose a radar has a pulse length of $t = 10 \ \mu s$ and a PRF of 10 kHz. Assume that the clutter observed by the radar has a two-sided spectral width of 1 kHz (i.e., the clutter spectrum occupies the range from -500 to +500 Hz). Sketch the blind zone map for these operating conditions. For the vertical axis, use time in seconds from 0 to 400 μs ; for the horizontal axis, use Doppler frequency in hertz from -10,000 to +10,000 Hz.
- 33. A weather radar has a PRF of 2 kHz. Using a series of 50 samples of data from a particular range bin and look direction, the following values of the autocorrelation function are computed: $s_y[0] = 50$, $s_y[1] = 30 \exp(jp/3)$. Use the pulse-pair processing (PPP) time domain method to compute the estimated mean frequency of the echo in hertz.
- 34. Continuing with the same weather radar, the mean Doppler shift is now removed from the data to obtain a new sequence y'[m] with autocorrelation lags $s_{y'}[0] = 50$, $s_{y'}[1] = 30$. Use the PPP time domain method to compute the estimated spectral width of the echo in hertz. Use the version of the spectral width estimator that contains the ln(\cdot) function. Remember that

spectral width is the variance of the spectrum, not the standard deviation.

- 35. Repeat the spectral width calculation using the version based on the series approximation to ln(x) given in Eq. (5.113). What is the percentage error in this estimate compared to the estimate in the previous problem?
- 36. Assume the simplified target model of Eq. (5.143) and suppose the adaptive DPCA system is noise limited at a particular Doppler bin, $\sigma_n^2[l,k] \gg \sigma_c^2[l,k]$. Show that the output z[l, k] of Eq. (5.145) reduces to simply Yf [l, k] to within a scale factor. That is, the processor does not combine the fore and aft channels in noise-limited Doppler bins.

 $\frac{7}{2}$ "Aliased sinc," also called the Dirichlet function. It is the discrete-variable equivalent of the usual continuous-variable sinc function.

¹ Skolnik (2001) distinguishes MTI and pulse Doppler by defining pulse Doppler as a system that uses a PRF high enough to avoid blind speeds (see Sec. 5.2.4). In this text the two terms are instead differentiated based on the type of processing used and the information obtained.

 $[\]frac{2}{2}$ Not all digital processors necessarily form a data matrix similar to Fig. 5.5 explicitly. MTI processors in particular can be implemented more simply. However, the data matrix is used explicitly in many other processors and is useful for illustration of Doppler filtering concepts.

 $[\]frac{3}{2}$ Some authors work in terms of the PRFs instead of the PRIs and use the term "staggers" to refer to ratios of PRF_n .

 $[\]frac{4}{1}$ It is important to realize that the number of staggers used is independent of the canceller order selected. A two-pulse canceller can be implemented with two staggers or 10 staggers.

 $[\]frac{5}{2}$ The order of the PRIs makes no difference to the power spectrum. Alternating shorter and longer PRIs keeps the transmitter duty cycle more nearly constant.

⁶ Since the power spectrum is real-valued, the autocorrelation function must be Hermitian symmetric. It can therefore be complex valued. However, clutter is usually modeled by a power spectrum that is also an even function of frequency (i.e., symmetric about F = 0), for example a Gaussian clutter spectrum with a zero mean. With this extra constraint, the autocorrelation function must be real-valued also, so the Re{} operator can be dropped in Eq. (5.52) and other equations in this section.

⁸ Harris uses a slightly different definition of the windows than is now conventional in data analysis and simulation packages for reasons having to do with DFT symmetry properties. Effectively, his version of an M-point-shaped window (e.g., Hamming) is the first M points of the (M + 1)-point symmetric window commonly used. The difference is of minor consequence, especially as M gets large.

 $[\]frac{9}{2}$ This correction factor assumes pulse-limited ground clutter as described in <u>Chap. 2</u>. Other appropriate factors are used if this is not the case.

CHAPTER 6

Detection Fundamentals

As was noted in <u>Chap. 1</u>, the primary functions to be carried out by a radar signal processor are detection, tracking, and imaging. In this chapter, the concern is detection. In radar, this means deciding whether a given radar measurement is the result of an echo from a target or simply represents the effects of interference. If it is decided that the measurement indicates the presence of a target further processing is usually undertaken. This additional processing might, for instance, take the form of tracking via precise range, angle, or Doppler measurements.

Detection decisions can be applied to signals present at various stages of the radar signal processing, from raw echoes to heavily preprocessed data such as Doppler spectra or even synthetic aperture radar images. In the simplest case, each range bin (fast-time sample) for each pulse can be individually tested to decide if a target is present at the range corresponding to the range bin, and the spatial angles corresponding to the antenna pointing direction for that pulse. Since the number of range bins can be in the hundreds or even thousands and pulse repetition frequencies can range from a few kilohertz to tens or hundreds of kilohertz, the radar can be making many thousands to millions of detection decisions per second.

It was seen in <u>Chap. 2</u> that both the interference and the echoes from complex targets are best described by statistical signal models. Consequently, the process of deciding whether or not a measurement represents the influence of a target or only interference is a problem in statistical hypothesis testing. In this chapter, it will be shown how this basic decision strategy leads to the concept of threshold testing as the most common detection logic in radar. Performance curves will be derived for the most basic signal and interference models.

Clutter (echoes from the ground) is sometimes interference and sometimes the target. If one is trying to detect a moving vehicle, ground clutter, along with noise and possibly jamming, is the interference; but if one is trying to image a region of the earth, this same terrain becomes the desired target and only noise and jamming are the interference.

An excellent concise reference for modern detection theory is given in <u>Chap. 5</u> of Johnson and Dudgeon (1993). When greater depth is needed, another excellent modern reference with a digital signal processing point of view is Kay (1998). An important classical textbook in detection theory is Van Trees (1968), while Meyer and Mayer (1973) provide a classical in-depth analysis and many detection curves specifically for radar applications.

6.1 Radar Detection as Hypothesis Testing

For any radar measurement that is to be tested for the presence of a target, one of two hypotheses can be assumed to be true:

- 1. The measurement is the result of interference only.
- 2. The measurement is the combined result of interference and echoes from a target.

The first hypothesis is denoted as the *null hypothesis* H_0 and the second as the non-null hypothesis H_1 . The detection logic therefore must examine each radar measurement to be tested and select one of the hypotheses as best accounting for that measurement. If H_0 best accounts for the data, the system declares that a target was not present at the range, angle, or Doppler coordinates of that measurement; if H_1 best accounts for the data, the system declares that a target was present.¹

Because the signals are described statistically, the decision between the two hypotheses is an exercise in statistical decision theory. A general approach to this problem is described in many texts (e.g., Kay, 1998). The analysis starts with a probability density function (PDF) that describes the measurement to be tested under each of the two hypotheses. If the sample to be tested is denoted as y, the following two PDFs are required:

 $p_y(y|H_0) = PDF$ of y given that a target was *not* present $p_y(y|H_1) = PDF$ of y given that a target was present

Thus, part of the detection problem is to develop models for these two PDFs. In fact, analysis of radar performance is dependent on estimating these PDFs for the system and scenario at hand. Furthermore, a good deal of the radar system design problem is aimed at manipulating these two PDFs in order to obtain the most favorable detection performance.

More generally, detection will be based on N samples of data y_n forming a column vector **y**

$$\mathbf{y} \equiv [y_0 \dots y_{N-1}]^T \tag{6.1}$$

The *N*-dimensional joint PDFs $p_{\mathbf{y}}(\mathbf{y}|H_0)$ and $p_{\mathbf{y}}(\mathbf{y}|H_1)$ are then used.

Assuming the two PDFs are successfully modeled, the following probabilities of interest can be defined:

Probability of	The probability that a target is declared (i.e., H_1 is
Detection, P_D :	chosen) when a target is in fact present.

Probability of The probability that a target is declared (i.e., H_1 is False Alarm P_{FA} : The probability that a target is right as HOt declared (i.e., H_0 is Miss, P_M : chosen) when a target is in fact present.

Note that $P_M = 1 - P_D$. Thus, P_D and P_{FA} suffice to specify all of the probabilities of interest. As the latter two definitions imply, it is important to realize that, because the problem is statistical, there will be a finite probability that the decisions will be wrong.²

6.1.1 The Neyman-Pearson Detection Rule

The next step in making a decision is to decide what the rule will be for deciding what constitutes an optimal choice between the two hypotheses. This is a rich field. The Bayes optimization criterion assigns a cost or risk to each of the four possible combinations of actual state (target present or not) and decision (select H_0 or H_1). In radar, it is more common to use a special case of the Bayes criterion called the Neyman-Pearson criterion. Under this criterion, the decision process is designed to maximize the probability of detection P_{D} under the constraint that the probability of false alarm P_{FA} does not exceed a set value. The achievable combinations of P_D and P_{FA} are affected by the quality of the radar system and signal processor design. However, it will be seen that for a fixed system design, increasing P_D implies increasing P_{FA} as well. The radar system designer will generally decide what rate of false alarms can be tolerated based on the implications of acting on a false alarm, which may include using radar resources to start a track on a nonexistent target, or in extreme cases even firing a weapon! Recalling that the radar may make tens or hundreds of thousands, even millions of detection decisions per second, values of P_{F4} must generally be quite low. Values in the range of 10^{-4} to 10^{-8} are common, and yet may still lead to false alarms every few seconds. Higher-level logic implemented in downstream data processing, beyond the scope of this book, is often used to reduce the number or impact of false alarms.

Each vector of measured data values y can be considered to be a point in *N*-dimensional space. To have a complete decision rule, each point in that space (each possible combination of *N* measured data values) must be assigned to one of the two allowed decisions, H_0 or H_1 . Then, when the radar measures a particular set of data values ("observation") y, the system declares either "target absent" or "target present" based on the preexisting assignment of y to either H_0 or H_1 . Denote the set of all observations y for which H_1 will be chosen as the region 1. Note that 1 is not necessarily a connected region. General expressions can now be written for the probabilities of detection and false alarm as integrals of the joint PDFs over the region 1 in a *N*-dimensional space:

$$P_{D} = \int_{\Re_{1}} p_{\mathbf{y}}(\mathbf{y}|H_{1}) d\mathbf{y}$$
$$P_{FA} = \int_{\Re_{1}} p_{\mathbf{y}}(\mathbf{y}|H_{0}) d\mathbf{y}$$
(6.2)

Because probability density functions are nonnegative, Eq. (6.2) proves a claim made earlier, namely that P_D and P_{FA} must rise or fall together. As the region 1 grows to include more of the possible observations y, either integral encompasses more of the *N*-dimensional space and therefore integrates more of the nonnegative PDF. The opposite is true if 1 shrinks. That is, as 1 grows or shrinks, *both* P_D and P_{FA} must decrease or increase.³ In order to increase detection probability, the false alarm probability must be allowed to increase as well. Loosely speaking, to achieve a good balance of performance, the points that contribute more probability mass to P_D than to P_{FA} are assigned to 1. If the system can be designed so that $p_y(y|H_0)$ and $p_y(y|H_1)$ are as disjoint as possible, this task becomes easier and more effective. This point will be revisited later.

6.1.2 The Likelihood Ratio Test

The Neyman-Pearson criterion is motivated by the goal of obtaining the best possible detection performance while guaranteeing that the false alarm probability does not exceed some tolerable value. Thus, the Neyman-Pearson decision rule is to

choose
$$\Re_1$$
 such that P_D is maximized, subject to $P_{FA} \le \alpha$
(6.3)

where α is the maximum allowable false alarm probability.⁴ This optimization problem is solved by the method of Lagrange multipliers. Construct the function

$$F \equiv P_D + \lambda (P_{FA} - \alpha) \tag{6.4}$$

To find the optimum solution, maximize F and then choose λ to satisfy the constraint criterion $P_{FA} = \alpha$. Substituting Eq. (6.2) into Eq. (6.4)

$$F = \int_{\Re_1} p_{\mathbf{y}}(\mathbf{y}|H_1) d\mathbf{y} + \lambda \left(\int_{\Re_1} p_{\mathbf{y}}(\mathbf{y}|H_0) d\mathbf{y} - \alpha \right)$$

= $-\lambda \alpha + \int_{\Re_1} \{ p_{\mathbf{y}}(\mathbf{y}|H_1) + \lambda p_{\mathbf{y}}(\mathbf{y}|H_0) \} d\mathbf{y}$
(6.5)

Remember that the design variable here is the choice of the region $_1$. The first term in the second line of Eq. (6.5) does not depend on $_1$, so *F* is maximized by maximizing the value of the integral over $_1$. Since λ could be negative, the

integrand can be either positive or negative, depending on the values of λ and the relative values of $p_y(\mathbf{y}|H_0)$ and $p_y(\mathbf{y}|H_1)$. The integral is therefore maximized by including in _1 all the points, and only the points, in the *N*-dimensional space for which $p_y(\mathbf{y}|H_1) + \lambda p_y(\mathbf{y}|H_0) > 0$, that is, _1 is all points \mathbf{y} for which $p_y(\mathbf{y}|H_1) > -\lambda p_y(\mathbf{y}|H_0)$. This leads directly to the decision rule:

$$\frac{p_{\mathbf{y}}(\mathbf{y} \mid H_1)}{p_{\mathbf{y}}(\mathbf{y} \mid H_0)} \stackrel{H_1}{\underset{H_0}{>}} - \lambda$$
(6.6)

Equation (6.6) is known as the *likelihood ratio test* (LRT). Although derived from the point of view of determining what values of y should be assigned to decision region 1, it in fact allows one to skip over explicit determination of 1 and gives a rule for optimally guessing, under the Neyman-Pearson criterion, whether a target is present or not based directly on the observed data y and a threshold $-\lambda$ (which must still be computed). This equation states that the ratio of the two PDFs, each evaluated for the particular observed data v, should be compared to a threshold. If that "likelihood ratio" exceeds the threshold, choose hypothesis H_1 , i.e., declare a target to be present. If it does not exceed the threshold, choose H_0 and declare that a target is not present. Under the Neyman-Pearson optimization criterion, the probability of a false alarm cannot exceed the original design value P_{FA} . Note again that models of $p_{\mathbf{v}}(\mathbf{y}|H_0)$ and $p_{\mathbf{v}}(\mathbf{y}|H_1)$ are required in order to carry out the LRT. Finally, realize that in computing the LRT, the data processing operations to be carried out on the observed data y are being specified. What exactly the required operations are depends on the particular PDFs.

The LRT is as ubiquitous in detection theory and statistical hypothesis testing as is the Fourier transform in signal filtering and analysis. It arises as the solution to the hypothesis testing problem under several different decision criteria, such as the Bayes minimum cost criterion, or maximization of the probability of a correct decision. Substantial additional detail is provided in Johnson and Dudgeon (1993) and Kay (1998). As a convenient and common shorthand, it is convenient to express the LRT in the following notation:

$$\Lambda(\mathbf{y}) \underset{H_0}{\overset{H_1}{\leq}} \eta \tag{6.7}$$

From Eq. (6.6), $\Lambda(\mathbf{y}) = p_{\mathbf{y}}(\mathbf{y}|H_1)/p_{\mathbf{y}}(\mathbf{y}|H_0)$ and $\eta = -\lambda$.

Because the decision depends only on whether the LRT exceeds the threshold or not, any monotone increasing⁵ operation can be performed on both sides of Eq. (6.7) without affecting the values of observed data y that cause the

threshold to be exceeded, and therefore without affecting the performance (P_D and P_{FA}). A well-chosen transformation can sometimes greatly simplify the computations required to actually carry out the LRT. Most common is to take the natural logarithm of both sides of Eq. (6.7) to obtain the *log likelihood ratio test*:

$$\ln \Lambda(\mathbf{y}) \underset{H_0}{\overset{>}{\leq}} \ln \eta \tag{6.8}$$

To make these procedures more concrete, consider what is perhaps the simplest example, detection of the presence or absence of a constant in zeromean Gaussian noise of variance σ_w^2 . Let **w** be a vector of independent identically distributed (i.i.d.) zero mean Gaussian random variables. When the constant is absent (hypothesis H_0) the data vector $\mathbf{y} = \mathbf{w}$ follows an *N*-dimensional normal distribution with a scaled identity covariance matrix. When the constant is present (hypothesis H_1), $\mathbf{y} = \mathbf{m} + \mathbf{w} = m\mathbf{1}_N + \mathbf{w}$ and the distribution is simply shifted to a nonzero positive mean⁶:

$$H_0: \mathbf{y} \sim \mathbf{N}(\mathbf{0}_N, \sigma_w^2 \mathbf{I}_N)$$

$$H_1: \mathbf{y} \sim \mathbf{N}(m\mathbf{1}_N, \sigma_w^2 \mathbf{I}_N)$$

(6.9)

where m > 0 and $\mathbf{0}_N$, $\mathbf{1}_N$, and \mathbf{I}_N are, respectively, a vector of *N* zeros, a vector of *N* ones, and the identity matrix of order *N*. The model of the required PDFs is therefore

$$p(\mathbf{y}|H_0) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{1}{2} \left(\frac{y_n}{\sigma_w}\right)^2\right\}$$
$$p(\mathbf{y}|H_1) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left\{-\frac{1}{2} \left(\frac{y_n - m}{\sigma_w}\right)^2\right\}$$
(6.10)

The likelihood ratio $\Lambda(\mathbf{y})$ and the log-likelihood ratio can be directly computed from Eq. (6.10):

$$\Lambda(\mathbf{y}) = \frac{\prod_{n=0}^{N-1} \exp\left\{-\frac{1}{2} \left(\frac{y_n - m}{\sigma_w}\right)^2\right\}}{\prod_{n=0}^{N-1} \exp\left\{-\frac{1}{2} \left(\frac{y_n}{\sigma_w}\right)^2\right\}}$$
(6.11)

$$\ln \Lambda(\mathbf{y}) = \sum_{n=0}^{N-1} \left\{ -\frac{1}{2} \left(\frac{y_n - m}{\sigma_w} \right)^2 + \frac{1}{2} \left(\frac{y_n}{\sigma_w} \right)^2 \right\}$$
$$= \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} m y_n - \frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} m^2$$
(6.12)

Because of its simpler form, the log-likelihood ratio will be used. Substituting Eq. (6.12) into Eq. (6.8) and rearranging gives the decision rule:

$$\sum_{n=0}^{N-1} y_n \stackrel{>}{\underset{<}{\sim}} \frac{\sigma_w^2}{m} \ln(-\lambda) + \frac{Nm}{2} \equiv T$$
(6.13)

where all the right-hand-side constants have been combined into a single constant *T*. Note that the right-hand side of the equation consists only of constants, though not all are yet known. Equation (6.13) thus specifies that the available data samples y_n be integrated (summed) and the integrated data compared to a threshold. This integration is an example of how the LRT specifies the data processing to be performed on the measurements. Note also that Eq. (6.13) does *not* require specifically evaluating the PDFs, let alone determining what exactly is the region in *N*-space comprising $_1$ or whether the observation **y** is in it.

In many cases of interest, the specific form of the log-likelihood ratio can be further rearranged to isolate on the left-hand side of the equation only those terms explicitly including the data samples y_n , moving all other constants to the right-hand side. Equation (6.13) is such a rearrangement of Eqs. (6.8) and (6.12). The term Σy_n is called a *sufficient statistic* for this problem, and is denoted by $\Upsilon(\mathbf{y})$. The sufficient statistic, if it exists, is a function of the data \mathbf{y} that has the property that the likelihood ratio (or log-likelihood ratio) can be written as a function of $\Upsilon(\mathbf{y})$, i.e., the data appear in the likelihood ratio *only* through $\Upsilon(\mathbf{y})$ (Van Trees, 1968). This means that in making a decision that is optimal under the Neyman-Pearson criterion, knowing the sufficient statistic $\Upsilon(\mathbf{y})$ is as good as knowing the actual data \mathbf{y} . In particular, the decision criterion in Eq. (6.8) can be expressed as

$$\Upsilon(\mathbf{y}) \underset{H_0}{\overset{H_1}{\underset{H_0}{>}}} T$$

(6.14)

The idea of a sufficient statistic is quite rich. For example, it can be interpreted as a geometric coordinate transformation chosen to place all of the useful information in the first coordinate (Van Trees, 1968). Procedures for verifying that a statistic (a function of the data y) is sufficient are given in Kay (1993), as is the Neyman-Fisher factorization theorem for identifying sufficient statistics. Detailed consideration of the properties of sufficient statistics is beyond the scope of this text; the reader is referred to the references for greater depth.

The specific value of the threshold $\eta = -\lambda$ that will ensure that $P_{FA} = \alpha$ as desired has not yet been found. The original expression for P_{FA} was given in Eq. (6.2), but this is not very useful since its evaluation requires the *N*-dimensional joint PDF of **y** and an explicit definition of the region ₁, which have been defined only implicitly as the points in *N*-space for which the LRT exceeds the still-unknown threshold. Since they are functions of the random data **y**, Λ and Υ are also random variables and thus have their own probability density functions. An alternate approach to computing the LRT threshold is thus to express P_{FA} in terms of Λ or Υ and then solve those expressions for η or equivalently for *T*. The required expressions are

$$P_{FA} = \int_{\eta = -\lambda}^{+\infty} p_{\Lambda}(\Lambda | H_0) d\Lambda = \alpha$$
(6.15)

or

$$P_{FA} = \int_{T}^{+\infty} p_{\Upsilon}(\Upsilon | H_0) d\Upsilon = \alpha$$
(6.16)

Because false alarms are the result of interference and do not involve targets, the result depends only on the PDF of the likelihood ratio [if using Eq. (6.15)] or the sufficient statistic [if using Eq. (6.16)] when a target is not present. Given a specific model of that PDF, a specific value can be computed for η (equivalently, λ or *T*.

To illustrate these results, continue the "constant in Gaussian noise" example by finding the threshold and then evaluating the performance, working with the sufficient statistic $\Upsilon(\mathbf{y})$. In this case $\Upsilon(\mathbf{y})$ is the sum of the individual data samples y_n . Under hypothesis H_0 (no target), the samples are i.i.d. It follows that $\Upsilon \sim N(0, N\sigma_w^2)$. A false alarm occurs anytime $\Upsilon > T$, so

$$\begin{aligned} \alpha &= P_{FA} = \int_{T}^{+\infty} p_{\Upsilon}(\Upsilon | H_0) d\Upsilon \\ &= \int_{T}^{+\infty} \frac{1}{\sqrt{2\pi N \sigma_w^2}} \exp\left[\frac{-\Upsilon^2}{2N\sigma_w^2}\right] d\Upsilon \end{aligned}$$

Equation (6.17) is the integral of a Gaussian PDF, so the *error function* erf(x) will appear in the solution. The standard definition is (Abramowitz and Stegun, 1972)²

$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$
 (6.18)

Also define the *complementary error function* erfc(x) corresponding to erf(x)

$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_{x}^{+\infty} e^{-t^{2}} dt = 1 - \operatorname{erf}(x)$$

(6.19)

One will generally be interested in finding the value of x that results in a certain value of erf(x) or erfc(x); thus the inverse error and complementary error functions, denoted by erf⁻¹(z) and erfc⁻¹(z), respectively, are of interest. It follows from Eq. (6.19) that the two are related by erfc⁻¹(z) = erf⁻¹(1 - z).⁸

With the change of variables $t = \Upsilon / \sqrt{2N\beta^2}$, Eq. (6.17) can be written as

$$\alpha = P_{FA} = \frac{1}{\sqrt{\pi}} \int_{T/\sqrt{2N\beta^2}}^{+\infty} e^{-t^2} dt$$
$$= \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{T}{\sqrt{2N\beta^2}}\right) \right]$$
(6.20)

Finally, Eq. (6.20) can be solved to obtain the threshold *T* in terms of the tabulated inverse error function:

$$T = \sqrt{2N\beta^2} \operatorname{erf}^{-1}(1 - 2P_{FA})$$
(6.21)

Equations (6.20) and (6.21) show how to compute P_{FA} given T and vice versa.

All of the information needed to carry out the LRT in its sufficient statistic form of Eq. (6.14) is now available. $\Upsilon(\mathbf{y})$ is just the sum of the data samples, while the threshold *T* can be computed from the number *N* of samples, the variance σ_w^2 of the noise, which is assumed known, and the desired false alarm probability P_{F4} .

The performance of this detector is evaluated by constructing a *receiver operating characteristic* (ROC) curve. There are four interrelated variables of

interest: P_D , P_{FA} , the noise power σ_w^2 , and the constant *m* whose presence or absence is to be decided. The latter two are characteristics of the given signals, while P_{FA} is generally fixed as part of the system specifications at whatever level is deemed tolerable. Thus it is necessary only to determine P_D . The approach is identical to that used for determining P_{FA} : determine the probability density function of the sufficient statistic Υ under the hypothesis H_0 and integrate the area under it from the threshold to $+\infty$.

Continuing the example, note that the only change under hypothesis H_1 is that the individual data samples y_n now each have mean m, so their sum Υ has mean Nm. Thus, $\Upsilon(\mathbf{y}) \sim N(Nm, N\sigma_w^2)$ and

$$P_{D} = \int_{T}^{+\infty} p_{\Upsilon}(\Upsilon | H_{1}) d\Upsilon$$
$$= \int_{T}^{+\infty} \frac{1}{\sqrt{2\pi N \sigma_{w}^{2}}} \exp\left[\frac{-(\Upsilon - N_{m})^{2}}{2N \sigma_{w}^{2}}\right] d\Upsilon$$
(6.22)

Again applying the definition of the error function in Eq. (6.18) leads to

$$P_{D} = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{T - Nm}{\sqrt{2N\sigma_{w}^{2}}} \right) \right]$$
(6.23)

Since the primary concern is the relationship between the performance metrics P_D and P_{FA} , Eq. (6.21) can be used in Eq. (6.23) to eliminate the threshold *T* and arrive at

$$P_{D} = \frac{1}{2} \left[1 - \operatorname{erf} \left\{ \operatorname{erf}^{-1} (1 - 2P_{FA}) - \frac{\sqrt{N}m}{\sqrt{2\sigma_{w}^{2}}} \right\} \right]$$
$$= \frac{1}{2} \operatorname{erfc} \left\{ \operatorname{erfc}^{-1} (2P_{FA}) - \frac{\sqrt{N}m}{\sqrt{2\sigma_{w}^{2}}} \right\}$$
(6.24)

Nm is considered to be the signal voltage of interest in the sufficient statistic $\Upsilon(\mathbf{y})$. The corresponding signal power is $(Nm)^2$. The power of the noise component of $\Upsilon(\mathbf{y})$ is $N\sigma_w^2$. Thus, the term $m\sqrt{N}/\sigma_w$ is the square root of the signal-to-noise ratio χ for this problem, and Eq. (6.24) can be rewritten as

$$P_{D} = \frac{1}{2} \left[1 - \operatorname{erf} \left\{ \operatorname{erf}^{-1} (1 - 2P_{FA}) - \sqrt{\chi/2} \right\} \right]$$

= $\frac{1}{2} \operatorname{erfc} \left[\operatorname{erfc}^{-1} (2P_{FA}) - \sqrt{\chi/2} \right]$
(6.25)

Figure 6.1 illustrates how the detection and false alarm probabilities follow from the PDFs under the two hypotheses and the threshold, and how their relative values depend on the relation between the two PDFs. Two Gaussian PDFs with variance equal to one are shown. The leftmost has a zero mean, while the rightmost has a mean of 1. With $N\sigma_w^2 = 1$ and m = 1/N these fit the model of Eq. (6.9) and the subsequent analysis. P_D and P_{FA} are the areas under the right and left PDFs, respectively, from the threshold (shown as a vertical line at about $\Upsilon = 1.5$) to $+\infty$. The receiver design thus consists of adjusting the position of the threshold until the black area equals the acceptable false alarm probability. The detection probability is then the gray area (which includes the black area). This figure again makes it clear that P_D and P_{FA} must increase or decrease together as the threshold moves lower or higher. The achievable combinations of P_D and P_{FA} are determined by the degree to which the two distributions overlap.



FIGURE 6.1 Gaussian probability density functions of the sufficient statistic when $N\sigma_w^2 = 1$ under hypothesis H_0 (left) and H_1 (right).

Figure 6.2 illustrates the ROC for this problem with the SNR χ as a parameter. Part *a* of the figure plots the ROC using linear scales for both P_{FA}

and P_D . Several features are worth noting. First, $P_{FA} = P_D$ when $\chi = 0$ (implying m = 0). This is to be expected since in that case, the PDF of $\Upsilon(\mathbf{y})$ is the same under either hypothesis. For a given P_{FA} and $\chi > 0$, P_D increases as the SNR increases, a result which should also be intuitively satisfying. Finally, note how abrupt the transition between near-zero and near-unity detection probabilities becomes as the SNR increases. This is a little misleading, since radars normally operate with very low values of P_{FA} ; depending on the type of system, P_{FA} is typically no higher than 10^{-3} and very often is in the range of 10^{-6} to 10^{-8} or even lower. Figure 6.2*b* plots the same data on a logarithmic scale for P_{FA} , which better reveals the characteristics of the ROC for false alarm probabilities of interest in radar signal processing.



FIGURE 6.2 Receiver operating characteristic for the Gaussian example: (*a*) displayed on a linear P_{FA} scale, (*b*) displayed on a logarithmic P_{FA} scale.

If the achievable combinations of P_D and P_{FA} do not meet the performance specifications, what can be done? Consideration of Fig. 6.1 suggests two answers. First, for a given P_{FA} , P_D can be increased by causing the two PDFs to move further apart when a target is present. That is, the presence of a target must cause a larger shift in the mean *m* of the distribution of the sufficient statistic. It was shown that the signal-to-noise ratio equals m^2/σ_w^2 . Thus, one way to improve the detection/false alarm tradeoff is to increase the SNR. This conclusion is borne out in Fig. 6.2. Figure 6.3*a* illustrates the effect of increased SNR on the PDFs and performance probabilities under the two hypotheses using the same threshold as Fig. 6.1.



FIGURE 6.3 Two ways to modify the PDFs of Fig. 6.1 to improve the tradeoff between detection and false alarms: (a) increasing the signal power, (b) reducing the noise power.

The second way to improve the performance tradeoff is to reduce the overlap of the PDFs by reducing their variance. Reducing the noise power σ_w^2 will reduce the variance of both PDFs, leading to the situation shown in Fig. 6.3b (where the area corresponding to P_{FA} is too small to be seen) and again improving performance. As with the first technique of increasing *m*, reducing σ_w^2 again constitutes increasing the SNR. Thus, consistent with Eq. (6.25), improving the tradeoff between P_D and P_{FA} requires increasing the SNR χ . This is a fundamental result that will arise repeatedly.

Radar systems are designed to achieve specified values of P_D and P_{FA} subject to various conditions, such as specified ranges, target types, interference environments, and so forth. The designer can work with antenna design, transmitter power, waveform design, and signal processing techniques, all within cost and form factor constraints. The job of the designer is therefore to develop a radar system design which ultimately results in a pair of "target absent" and "target present" PDFs at the point of detection with a small enough overlap to allow the desired P_D and P_{FA} to be achieved. If the design does not do this, the designer must redesign one or more of these elements to reduce the variance of the PDFs, shift them further apart, or both until the desired performance is obtained. Thus, a significant goal of radar system design is controlling the overlap of the two PDFs analogous to those in Fig. 6.1, or equivalently, maximizing the SNR.

6.2 Threshold Detection in Coherent Systems

The Gaussian problem considered so far is useful to introduce and explain the major elements of Neyman-Pearson detection such as the likelihood ratio test, probabilities of detection and false alarm, receiver operating characteristics, and the major design tradeoffs that follow. The problem seems "radar-like": under one hypothesis, only Gaussian noise is observed; under the other, a constant was added to the noise, which could be interpreted as the echoes from a steady target. Figure 6.4 summarizes the design and analysis strategy that was used. Beginning with models of the PDF of the data under hypotheses H_0 and H_1 , the LRT or log-LRT is written down and manipulated to isolate the terms involving the measured data. If necessary or useful, a simplified detector law is substituted (see Section 6.2.3). The sufficient statistic is then identified and its PDF under each hypothesis determined. The PDF under H_0 is integrated to get a relationship between the threshold T and the P_{FA} which is solved analytically or

numerically to determine the threshold that gives the desired P_{FA} . Finally, that threshold value is used with the PDF of Υ under H_1 to get P_D . This same analysis strategy will be used repeatedly to develop the detector design and evaluate its performance for various data models. The only difference will be in the raw data PDFs $p_y(y|H_0)$ and $p_y(y|H_1)$ assumed in the first step. However, changes to these PDFs ripple through the entire process, changing the likelihood ratio, then the sufficient statistic and thus the detector design, the threshold setting, and finally the detection performance.



FIGURE 6.4 Strategy for radar detector design, analysis, and performance evaluation.

The real constant-in-white Gaussian noise (WGN) example is not a good

model for any radar detection problem due to at least three major limitations. First, coherent radar systems that produce complex-valued measurements are of most interest. The approach must therefore be extended to the complex case. Second, there are unknown parameters. The analysis so far has assumed that such signal parameters as the noise variance and target amplitude are known, when in fact these are not known a priori but must be estimated if needed. To complicate matters further, some parameters are linked. Specifically in radar, the (unknown) echo amplitude varies with the (unknown) echo arrival time according to the appropriate version of the radar range equation. Thus, the LRT must be generalized to develop a technique that can work when some signal parameters are unknown. Finally, as seen in <u>Chap. 2</u>, there are a number of established models for radar signal phenomenology that must be incorporated. In particular, it is necessary to account for fluctuating targets, i.e., statistical variations in the amplitude of the target components of the measured data when a target is present. Furthermore, while the Gaussian PDF remains a good model for noise, in many problems the dominant interference is clutter which may have one of the distinctly non-Gaussian PDFs discussed in Chap. 2. The next subsections begin addressing these shortcomings by extending the LRT to coherent systems.

6.2.1 The Gaussian Case for Coherent Receivers

An appropriate model for noise at the output of a coherent receiver was developed in <u>Chap. 2</u>. It was shown there that if the noise in the system prior to quadrature signal generation is a zero mean, white Gaussian process with power $\sigma_w^2 = KT$,² the I and Q channels will each contain independent, identically distributed zero-mean white Gaussian processes with power $kT/2 = \sigma_w^2/2$ That is, the noise power splits evenly but independently between the two channels. A complex noise process for which the real and imaginary parts are i.i.d. is called a *circular random process*. The expression for the joint PDF of N complex samples of the circular Gaussian random process is

$$p_{y}(y) = \frac{1}{\det\{\pi S_{y}\}} \exp\{-(y-m)^{H} S_{y}^{-1}(y-m)\}$$
(6.26)

where **m** is the $N \times 1$ vector mean of the $N \times 1$ vector signal $\mathbf{y} = \mathbf{m} + \mathbf{w}$, \mathbf{S}_y is the $N \times N$ covariance matrix of \mathbf{y}

$$\mathbf{S}_{\mathbf{y}} = \mathbf{E}\{\mathbf{y}\mathbf{y}^H\} - \mathbf{m}\mathbf{m}^H$$
(6.27)

and *H* is the Hermitian (conjugate transpose) operator. In most cases the noise samples are i.i.d. so that $S_y = \sigma_w^2 I_N$, which in turn means that $\det[\pi S_y] = \pi^N \sigma_w^{2N}$.

Treatment of the case where the noise samples are not equal-variance, and the colored noise case where S_y is not even diagonal, is beyond the scope of this text. The reader is referred to Dudgeon and Johnson (1993) and Kay (1998) for these more complex situations.

Equation (6.26) now reduces to

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{\pi^{N} \sigma_{w}^{2N}} \exp\left\{-\frac{1}{\sigma_{w}^{2}} (\mathbf{y} - \mathbf{m})^{H} (\mathbf{y} - \mathbf{m})\right\}$$
(6.28)

Further simplifications occur when all of the means under H_1 are identical so that $\mathbf{m} = m\mathbf{1}_N$, where *m* can now be complex-valued. In this case Eq. (6.28) reduces slightly further to

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{1}{\pi^{N} \sigma_{w}^{2N}} \exp\left\{-\frac{1}{\sigma_{w}^{2}} (\mathbf{y} - m\mathbf{1}_{N})^{H} (\mathbf{y} - m\mathbf{1}_{N})\right\}$$
(6.29)

The LRT for the coherent version of the previous Gaussian example can be obtained by repeating the steps in the example of Eqs. (6.10) through (6.25) using the PDF of Eq. (6.28), with $\mathbf{m} = \mathbf{0}_N$ under hypothesis H_0 and $\mathbf{m} \neq \mathbf{0}_N$ under H_1 . The log-likelihood ratio is

$$\ln \Lambda = \frac{1}{\sigma_w^2} \{2 \operatorname{Re}[\mathbf{m}^H \mathbf{y}] - \mathbf{m}^H \mathbf{m}\}$$

= $\frac{2}{\sigma_w^2} \operatorname{Re}\left\{\sum_{n=0}^{N-1} m^* y_n\right\} - \frac{1}{\sigma_w^2} N |m|^2$
(6.30)

where the second line of Eq. (6.30) applies only to the case where the means are identical ($\mathbf{m} = m\mathbf{1}_N$).

Some interpretation of Eq. (6.30) is in order. The term $\mathbf{m}^H \mathbf{y}$ is the dot product of the complex vectors \mathbf{m} and \mathbf{y} . As seen in <u>App. B</u>, this dot product represents an FIR filtering operation evaluated at the particular instant when the equivalent impulse response \mathbf{m}^H and the data vector \mathbf{y} completely overlap. Furthermore, since the impulse response of the filter is identical to the signal whose presence is to be detected under hypothesis H_1 , namely $m\mathbf{1}_N$, it is a matched filter. The same reasoning applies if the elements of \mathbf{m} are the samples of a modulated waveform or any other function of interest.

The second term in $\ln \Lambda$, which is the complex dot product of **m** with itself, expands to $\mathbf{m}^{H}\mathbf{m} = \sum_{n=0}^{N-1} |m_n|^2$. This is just the energy *E* in **m**. In the equal means case $E = N|m|^2$.

Finally, note the Re{·} operator applied to the matched filter output $\mathbf{m}^{H}\mathbf{y}$. Because \mathbf{m} and \mathbf{y} are complex, one might be concerned that the dot product could be purely imaginary or nearly so, such that Re{ $\mathbf{m}^{H}\mathbf{y}$ } ≈ 0 . The measured data \mathbf{y} would then have little or no effect on the threshold test. For this example, $\mathbf{m} = \mathbf{0}_{N}$ under hypothesis H_{0} and the Re{·} operator is inconsequential. Under hypothesis H_{1} each element of \mathbf{m} is a complex number $m_{n}e^{j\theta_{n}}$. If the target is actually present the elements of the measured data vector $\mathbf{y} = \mathbf{m} + \mathbf{w}$ will be of the form $m_{n}e^{j\theta_{n}} + w_{n}$ where w_{n} is a zero mean complex Gaussian noise sample. It follows that

$$\mathbf{m}^{H}\mathbf{y} = \mathbf{m}^{H}\mathbf{m} + \mathbf{m}^{H}\mathbf{w}$$

= $\sum_{n=0}^{N-1} |m_{n}|^{2} + \sum_{n=0}^{N-1} w_{n} m_{n} e^{-j\theta_{n}}$
(6.31)

The first term is again the energy *E* in the signal **m**; this is real-valued and therefore unaffected by the $\text{Re}\{\cdot\}$ operator. The second term is simply weighted and integrated noise samples. The phase of this noise component and therefore the effect of the $\text{Re}\{\cdot\}$ operator is random. Its effect on the phase of the sum will be large when the SNR is low, but minimal when the SNR is high.

It is evident by inspection of Eq. (6.30) that the sufficient statistic is now Re{ $\mathbf{m}^{H}\mathbf{y}$ }. Expressing the LRT in its sufficient statistic form for the complex case gives

$$\Upsilon = \operatorname{Re}\{\mathbf{m}^{H}\mathbf{y}\}_{\substack{<\\ H_{0} \\ }}^{H_{1}} \frac{\sigma_{w}^{2}}{2} \ln(-\lambda) + \frac{E}{2} = T$$
(6.32)

Note that if $\mathbf{m} = m\mathbf{1}_N$, the term $\operatorname{Re}\{\mathbf{m}^H\mathbf{y}\} = m \Sigma y_n$ and Eq. (6.32) is very similar to Eq. (6.13).

To complete consideration of the complex Gaussian case, its performance, i.e., P_D and P_{FA} , must be determined. The sufficient statistic $\Upsilon = \operatorname{Re}\{\mathbf{m}^H\mathbf{y}\} = \operatorname{Re}\{\Sigma m_n^* y_n\}$ is just a sum of Gaussian random variables and so will also be Gaussian. To determine the performance of the coherent detector, the PDF of Υ must be determined under each hypothesis. To do so it is useful to first consider the quantity $\mathbf{z} = \mathbf{m}^H \mathbf{y}$, which will be a complex Gaussian. First suppose hypothesis H_0 is true. In this case the $\{y_n\}$ are zero mean and therefore so is z. Because the $\{y_n\}$ are independent the variance of z is just the sum of the variances of the individual weighted samples:

$$\operatorname{var}(z) = \sum_{n=0}^{N-1} \operatorname{var}(m_n^* y_n)$$

= $\sum_{n=0}^{N-1} |m_n|^2 \sigma_w^2 = E \sigma_w^2$
(6.33)

Thus, under hypothesis H_0 , $z \sim N(E, E\sigma_w^2/2)$. Similarly, under hypothesis H_1 **y** = **m** + **w** and $z \sim N(E, E\sigma_w^2/2)$. Note that the mean of *z* is real in both cases. The power of the complex Gaussian noise splits evenly between the real and imaginary parts of *z*. Since $\Upsilon = \text{Re}\{\mathbf{m}^H\mathbf{y}\}$, it follows that $\Upsilon \sim N(0, E\sigma_w^2/2)$ under H_0 and $\Upsilon \sim N(E, E\sigma_w^2/2)$ under H_1 . Following the procedure used in Sec. 6.1.2, it can be seen that

$$P_{FA} = \frac{1}{2} \left[1 - \operatorname{erf}\left(\frac{T}{\sqrt{E\sigma_w^2}}\right) \right]$$
(6.34)

Repeating the development of Eqs. (6.22) to (6.24) gives the probability of detection

$$P_{D} = \frac{1}{2} \left[1 - \operatorname{erf} \left\{ \operatorname{erf}^{-1} (1 - 2P_{FA}) - \sqrt{\frac{E}{\sigma_{w}^{2}}} \right\} \right]$$
$$= \frac{1}{2} \operatorname{erfc} \left\{ \operatorname{erfc}^{-1} (2P_{FA}) - \sqrt{\frac{E}{\sigma_{w}^{2}}} \right\}$$
(6.35)

Note again that the last term in Eq. (6.35) is the square root of the energy in the signal **m**, divided this time by the noise power σ_w^2 , i.e., the signal-to-noise ratio. Thus Eq. (6.35) can be written as

$$P_{D} = \frac{1}{2} [1 - \operatorname{erf} \{\operatorname{erf}^{-1}(1 - 2P_{FA}) - \sqrt{\chi}\}]$$

= $\frac{1}{2} \operatorname{erfc} \{\operatorname{erfc}^{-1}(2P_{FA}) - \sqrt{\chi}\}$
(6.36)

Finally, in the equal means case when $\mathbf{m} = m\mathbf{1}_N$, Eq. (6.35) is similar (but not identical) to Eq. (6.24). The coherent case includes the term $\sqrt{Nm^2/\sigma_w^2}$ instead of $\sqrt{Nm^2/2\sigma_w^2}$ because all of the signal energy competes with only half of the noise power.

Figure 6.5 shows the receiver operating characteristic for this example. It

is identical in general form to the real-valued case of Fig. 6.2, however, for a given signal-to-noise ratio the performance is better because in the coherent receiver the signal competes with only half the noise power. For example, in this coherent case a signal-to-noise ratio of 13 dB produces $P_D = 0.94$ at a $P_{FA} = 10^{-6}$. Figure 6.2 shows that in the real case the same χ and P_{FA} produce a P_D of just under 0.39.



FIGURE 6.5 Performance of the coherent receiver on the complex Gaussian example.

6.2.2 Unknown Parameters and Threshold Detection

In general, perfect knowledge of each of the parameters of the PDFs $p_{\Upsilon}(\Upsilon | H_0)$ and $p_{\Upsilon}(\Upsilon | H_1)$ is required to carry out the LRT, which usually means having perfect knowledge of $p_y(\mathbf{y} | H_0)$ and $p_y(\mathbf{y} | H_1)$. In the Gaussian example, for instance, it was assumed that the expected signal \mathbf{y} is known under the various hypotheses, as well as the noise sample variance σ_w^2 . This is not the case in the real world, where the PDFs that form the likelihood ratio may depend on one or more parameters ξ that are unknown. Depending on the available information three cases arise;

- 1. ξ is a random variable with a known probability density function.
- 2. ξ is a random variable with an unknown probability density function.

3. ξ is deterministic but unknown.

Different techniques are used to handle each of these cases. The first is the most important because it has the greatest effect on the structure of the optimal Neyman-Pearson detector.

To illustrate the approach for handling a random parameter with a known PDF, consider yet again the complex Gaussian case. The optimal detector implemented a matched filter operation $\mathbf{m}^H \mathbf{y}$ followed by the Re{·} operator. The success of the matched filter structure depended on knowing exactly the constant component of $\mathbf{y} = \mathbf{m} + \mathbf{w}$ under hypothesis H_1 , so that the filter coefficients could be set equal to \mathbf{m} and the filter output would be real-valued. Recall that, when applied to radar, \mathbf{y} under H_0 is considered to consist only of samples \mathbf{w} of receiver noise, and under H_1 to consist of noisy samples $\mathbf{m} + \mathbf{w}$ of the echoes from a radar target over multiple pulses, or alternately successive fast-time samples of the waveform of one pulse echo from a target.

Claiming perfect knowledge of **m** implies knowing the range to the target very precisely, since a variation in one-way range of only $\lambda/4$ causes the received echo phase to change by 180°. A quarter-wavelength is only 30 cm at L band and 3.16 mm at 95 GHz. Because this precision is usually unrealistic, it is more reasonable to assume **m** is known only to within a phase factor $\exp(j\theta)$, where the phase angle θ is considered to be a random variable distributed uniformly over $(0, 2\pi]$ and independent of the random variables $\{m_n\}$. In other words, $\mathbf{m} = \tilde{\mathbf{m}} \exp(j\theta)$, where $\tilde{\mathbf{m}}$ is known exactly but θ is a random phase. Note that the energy in $\tilde{\mathbf{m}}$ is the same as that in **m**, that is, $\mathbf{m}^H \mathbf{m} = \tilde{\mathbf{m}}^H \tilde{\mathbf{m}}$. This "unknown phase" assumption cannot usually be avoided in radar. What is its effect on the optimal detector and its performance?

The goal remains to carry out the LRT, so it is necessary to return to its basic definition of Eq. (6.6) and determine $p_y(\mathbf{y}|H_0)$ and $p_y(\mathbf{y}|H_1)$, both of which now presumably depend on θ , and use the technique known as the *Bayesian approach* for random parameters with known PDFs (Kay, 1998).¹⁰ Specifically, compute the PDF under H_i by averaging the conditional PDFs $p_y(\mathbf{y}|H_i, \theta)$ over θ

$$p_{\mathbf{y}}(\mathbf{y} \mid H_i) = \int p_{\mathbf{y}}(\mathbf{y} \mid H_i, \theta) \, p_{\theta}(\theta) \, d\theta, \quad i = 0, 1$$
(6.37)

The unconditional PDFs $p_y(\mathbf{y}|H_i)$ are then used to define the likelihood ratio in the usual way.

As an example of the Bayesian approach for random parameters, consider again the complex Gaussian case, but now with an unknown phase in the data, $m = \tilde{m} \exp(i\theta)$. The conditional PDF of the observations y becomes, under each of the two hypotheses,

$$p_{\mathbf{y}}(\mathbf{y}|H_{0},\theta) = \frac{1}{\pi^{N}\sigma_{w}^{2N}} \exp\left[-\frac{1}{\sigma_{w}^{2}}\mathbf{y}^{H}\mathbf{y}\right]$$
$$p_{\mathbf{y}}(\mathbf{y}|H_{1},\theta) = \frac{1}{\pi^{N}\sigma_{w}^{2N}} \exp\left[-\frac{1}{\sigma_{w}^{2}}(\mathbf{y}-\tilde{\mathbf{m}}e^{j\theta})^{H}(\mathbf{y}-\tilde{\mathbf{m}}e^{j\theta})\right]$$
(6.38)

Expanding the exponent in Eq. (6.38) gives

$$p_{\mathbf{y}}(\mathbf{y}|H_{1},\theta) = \frac{1}{\pi^{N}\sigma_{w}^{2N}} \exp\left[-\frac{1}{\sigma_{w}^{2}}(\mathbf{y}^{H}\mathbf{y} - 2\operatorname{Re}\left\{\tilde{\mathbf{m}}^{H}\mathbf{y}e^{-j\theta}\right\} + E)\right]$$
$$= \frac{1}{\pi^{N}\sigma_{w}^{2N}} \exp\left[-\frac{1}{\sigma_{w}^{2}}(\mathbf{y}^{H}\mathbf{y} - 2|\tilde{\mathbf{m}}^{H}\mathbf{y}|\cos(\phi - \theta) + E)\right]$$
(6.39)

where ϕ is the unknown but fixed phase of the inner product $\tilde{\mathbf{m}}^H \mathbf{y}$.

Notice that $p_y(\mathbf{y}|H_0,\theta)$ does not depend on θ after all (not surprising since there is no target present in this case to present an unknown phase), so it is not necessary to apply Eq. (6.37). However, in $p_y(\mathbf{y}|H1)$ the dependence on θ is explicit. Assuming a uniform random phase, defining $\theta' = \phi - \theta$, and applying Eq. (6.37) under H_1 gives

$$p_{\mathbf{y}}(\mathbf{y}|H_{1}) = \frac{1}{\pi^{N} \sigma_{w}^{2N}} e^{-(\mathbf{y}^{H}\mathbf{y}+\mathbf{E})/\sigma_{w}^{2}} \frac{1}{2\pi} \int_{0}^{2\pi} \exp\left[\frac{2}{\sigma_{w}^{2}} |\tilde{\mathbf{m}}^{H}\mathbf{y}| \cos\theta'\right] d\theta'$$
(6.40)

Equation (6.40) is a standard integral. Specifically, integral 9.6.16 in Olver et al. (2010) is

$$\frac{1}{\pi} \int_0^{\pi} e^{\pm z \cos\theta} \, d\theta = I_0(z) \tag{6.41}$$

where $I_0(z)$ is the modified Bessel function of the first kind. Using this result and properties of the cosine function, Eq. (6.40) becomes

$$p_{\mathbf{y}}(\mathbf{y}|H_1) = \frac{1}{\pi^N \sigma_w^{2N}} \exp\left[-(\mathbf{y}^H \mathbf{y} + \mathbf{E})/\sigma_w^2\right] I_0\left(\frac{2|\tilde{\mathbf{m}}^H \mathbf{y}|}{\sigma_w^2}\right)$$
(6.42)

The log-LRT now becomes

$$\ln \Lambda = \ln \left[I_0 \left(\frac{2 |\tilde{\mathbf{m}}^H \mathbf{y}|}{\sigma_w^2} \right) \right] - \frac{E}{\sigma_w^2} \mathop{\geq}_{\substack{K \\ H_0}}^{H_1} \ln(-\lambda)$$
(6.43)

or in sufficient statistic form

$$\Upsilon = \ln \left[l_0 \left(\frac{2 \left| \tilde{\mathbf{m}}^H \mathbf{y} \right|}{\sigma_w^2} \right) \right] \stackrel{H_1}{\underset{H_0}{>}} \ln(-\lambda) + \frac{E}{\sigma_w^2} = T$$
(6.44)

Equation (6.44) defines the signal processing required for optimum detection in the presence of an unknown phase. It calls for taking the magnitude of the matched filter output $\tilde{\mathbf{m}}^H \mathbf{y}$, passing it through the memoryless nonlinearity $\ln[I_0(\cdot)]$, and comparing the result to a threshold. This result is appealing in that the matched filter is still applied to utilize the *internal* phase structure of the known signal and maximize the integration gain, but then a magnitude operation is applied because the absolute phase of the result cannot be known. Also, note that the argument of the Bessel function is the energy in the matched filter output divided by half the noise power; again, a signal-to-noise ratio. Only half of the noise power appears because the total noise power in the complex case is split between the real and imaginary channels.

As a practical matter, it is desirable to avoid having to compute the natural logarithm and Bessel function for every threshold test, since these might occur millions of times per second in some systems. Because the function $\ln[I_0(\cdot)]$ is monotonically increasing, the same detection results can be obtained by simply comparing its argument $2|\tilde{\mathbf{m}}^H\mathbf{y}|/\sigma_w^2$ to a modified threshold. Equation (6.44) then becomes simply

$$\left\|\tilde{\mathbf{m}}^{H}\mathbf{y}\right\|_{\substack{\leq\\H_{0}\\H_{0}}}^{H_{1}}T'$$
(6.45)

Figure 6.6 illustrates the optimal detector for the coherent detector with an unknown phase.



FIGURE 6.6 Structure of optimal detector when the absolute signal phase is unknown.

The performance of this detector will now be established. Let $z = |\tilde{\mathbf{m}}^H \mathbf{y}|$. The detection test becomes simply z < T'; thus the distribution of z under each of the two hypotheses is needed. As in the known phase case, under hypothesis H_0 (target absent) $\tilde{\mathbf{m}}^H \mathbf{y} - N(0, E\sigma_w^2)$; thus the real and imaginary parts of $\tilde{\mathbf{m}}^H \mathbf{y}$ are independent of one another and each distributed as $N(0, E\sigma_w^2/2)$. It follows (see <u>Chap. 2</u> or <u>App. A</u>) that z is Rayleigh distributed

$$p_{z}(z \mid H_{0}) = \begin{cases} \frac{2z}{E\sigma_{w}^{2}} \exp\left(-\frac{z^{2}}{E\sigma_{w}^{2}}\right), & z \ge 0\\ 0, & z < 0 \end{cases}$$

$$(6.46)$$

The probability of false alarm is

$$P_{FA} = \int_{T'}^{+\infty} P_z(z|H_0) dz = \exp\left(\frac{-T'^2}{E\sigma_w^2}\right)$$
(6.47)

It is convenient to invert this equation to obtain the threshold setting in terms of P_{FA} :

$$T' = \sqrt{-E\sigma_w^2 \ln P_{FA}} \tag{6.48}$$

Now consider hypothesis H_1 , i.e., target present. In this case $E\sigma_w^2 \sim N(E, E\beta^2)$. Since *E* is real-valued, the real part $\tilde{\mathbf{m}}^H \mathbf{y}$ is distributed as $N(E, E\beta^2/2)$ while the imaginary part is distributed as N(0, $E\beta^2/2$). It again follows that the PDF of *z* is

$$p_{z}(z \mid H_{1}) = \begin{cases} \frac{2z}{E\sigma_{w}^{2}} \exp\left[-\frac{1}{E\sigma_{w}^{2}}(z^{2} + E^{2})\right] I_{0}\left(\frac{2z}{\sigma_{w}^{2}}\right), & z \ge 0\\ 0, & z < 0 \end{cases}$$
(6.49)

where $I_0(z)$ is again the modified Bessel function of the first kind. Equation (6.49) is the Rician PDF. The probability of detection is obtained by integrating it from *T*' to $+\infty$.

In normalized form, the required integral is

$$Q_M(\alpha,\gamma) = \int_{\gamma}^{+\infty} t \exp\left[-\frac{1}{2}(t^2 + \alpha^2)\right] I_0(\alpha t) dt$$
(6.50)

The expression $Q_M(\alpha, \gamma)$ is known as *Marcum's Q function*. It arises frequently in radar detection calculations. A closed form for this integral is not known. Algorithms for evaluating $Q_M(\alpha, \gamma)$ are compared in Cantrell and Ojha (1987). The Communications ToolboxTM and Signal Processing ToolboxTM optional packages of MATLAB® includes a marcumq function to evaluate $Q_M(\alpha, \gamma)$; another MATLAB® algorithm is given by Kay (1998).

By defining a change of variables the integral of Eq. (6.49) can be put into the form of Eq. (6.50). Specifically, choose $t = z/\sqrt{E\sigma_w^2/2}$ and $\alpha = \sqrt{2E}/\sigma_w$. Substituting into Eq. (6.49) and doing the integration gives

$$P_D = Q_M \left(\sqrt{\frac{2E}{\sigma_w^2}}, \sqrt{\frac{2T'^2}{E\sigma_w^2}} \right)$$
(6.51)

Finally, noting that E/σ_w^2 is the signal-to-noise ratio χ and expressing the threshold in terms of the false alarm probability using Eq. (6.48) gives

$$P_D = Q_M \left(\sqrt{2\chi} \,, \sqrt{-2\ln P_{FA}} \right) \tag{6.52}$$

It is usually the case that the energy E in **m** or $\tilde{\mathbf{m}}$ is not known. Fortunately, Eq. (6.52) does not depend on E (or the noise power σ_w^2 explicitly, but only on their ratio χ , so that it is possible to generate the ROC without this information. However, actually implementing the detector requires a specific value of the threshold T' as given in Eq. (6.48), and this does require knowledge of both Eand σ_w^2 . One way to avoid this problem is to replace the matched filter coefficients $\tilde{\mathbf{m}}$ with a normalized coefficient vector $\hat{\mathbf{m}} = \tilde{\mathbf{m}}/E$. This choice simply normalizes the gain of the matched filter to 1. The energy in this modified sequence is $\hat{E} = 1$, leading to a modified threshold

$$\hat{T} = \sqrt{-\sigma_w^2 \ln P_{FA}} \tag{6.53}$$

The modified matched filter gain and threshold result in no change to the ROC so that Eq. (6.52) remains valid. Setting of the threshold \hat{T} still requires knowledge of the noise power σ_w^2 ; removal of this restriction is the subject of Sec. 6.5. The handling of unknown amplitude parameters is discussed in

somewhat more detail in Sec. 6.2.4.

The performance of the envelope detector in this example is given in Fig. 6.7. The general behavior is very similar to the known phase coherent detector case of Fig. 6.5. Closer inspection, however, shows that for a given P_{FA} , the coherent detector obtains a higher P_D . To make this point clearer, Fig. 6.8 compares the detection curves for the coherent and envelope detectors (known and unknown phase, respectively), plotted two different ways. Part (*a*) of the figure simply repeats the 10-dB curves from the two earlier figures. At $P_{FA} = 10^{-4}$ and $\chi = 10$ dB, for example, P_D is about 0.74 for the coherent detector. This figure drops to 0.6 when the envelope detector is used.



FIGURE 6.7 Performance of the linear envelope detector for the Gaussian example with unknown phase.



FIGURE 6.8 Performance difference between coherent and envelope detectors for the complex Gaussian example: (*a*) difference in P_D for $\chi = 10$ dB, (*b*) difference in P_D for $P_{FA} = 10^{-6}$.

Figure 6.8*b* plots the detection performance as a function of χ with P_{FA} fixed (at 10⁻⁶ in this example). This figure shows that, to achieve the same

probability of detection, the envelope detector requires about 0.6 dB higher SNR than the coherent detector at $P_D = 0.9$, and about 0.7 dB more at $P_D = 0.5$. The extra signal-to-noise ratio required to maintain the detection performance of the envelope detector compared to the coherent case is called an *SNR loss*. SNR losses can result from many factors; this particular one is often called the *detector loss*. It represents extra SNR that must be obtained in some way if the performance of the envelope detector is to match that of the ideal coherent detector. Increasing the SNR in turn implies one or more of many radar system changes, such as greater transmitter power, a larger antenna gain, reduced range coverage, and so forth.

The phenomenon of detector loss illustrates a very important point in detection theory: the less that is known about the signal to be detected, the higher must be the SNR to achieve a given combination of P_D and P_{FA} . In this case, not knowing the absolute phase of the signal has cost about 0.6 dB. Inconvenient though it may be, this result is intuitively satisfying: the worse the knowledge of the signal details, the worse the performance of the detector will be.

6.2.3 Linear and Square-Law Detectors

Equation (6.44) defines the optimal Neyman-Pearson detector for the Gaussian example with an unknown phase in the data. It was shown that the $\ln[I_0(x)]$ function could be replaced by its argument x without altering the performance. In Sec. 6.3.2 a simpler detector characteristic than $\ln[I_0(\cdot)]$ will again be desirable for noncoherent integration, but it will not be possible to simply substitute any monotonic increasing function. It is therefore useful to see what approximations can be made to the $\ln[I_0(\cdot)]$ function.

A standard series expansion for the Bessel function holds that

$$I_0(x) = 1 + \frac{x^2}{4} + \frac{x^4}{64} + \dots$$
(6.54)

Thus for small x, $I_0(x) \approx 1 + x^2/4$. Furthermore, one series expansion of the natural logarithm has $\ln(1 + z) = z - z^2/2 + z^3/3 + \dots$ Combining these gives

$$\ln[I_0(x)] \approx \frac{x^2}{4} \qquad x \ll 1$$
(6.55)

Equation (6.55) shows that if x is small, the optimal detector is well approximated by a matched filter followed by a so-called *square law detector*, i.e., a magnitude squaring operation. The factor of four can be incorporated into the threshold in Eq. (6.44).
For large values of x, $I_0(x) \approx e^x / \sqrt{2\pi x}$; then

$$\ln[I_0(x)] \approx x - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(x)$$
(6.56)

The constant term on the right of Eq. (6.56) can be incorporated into the threshold in Eq. (6.44), while the linear term in x quickly dominates the logarithmic term for $x \gg 1$. This leads to the *linear detector* approximation for large x

$$\ln[I_0(x)] \approx x \qquad x \gg 1 \tag{6.57}$$

Figure 6.9 illustrates the fit between the square law and linear approximations and the exact $\ln[I_0(\cdot)]$ functions. The square law detector is an excellent fit for x < 3 dB, while the linear detector fits the $\ln[I_0(x)]$ very well for x > 10 dB.



FIGURE 6.9 Approximation of the $\ln[I_0(\cdot)]$ detector characteristic by the square law detector when its argument is small, and the linear detector when its argument is large.

Finally, note that it is easy enough to compute the squared magnitude of a complex-valued test sample as simply the sum of the squares of the real and

imaginary parts. The linear magnitude requires a square root and is less computationally convenient.

6.2.4 Other Unknown Parameters

The preceding sections have shown the effect of unknown phase of the received signal on the optimal detector. However, other parameters of the received signal are also unknown in practice. The amplitude of the echo depends on all of the factors in the radar range equation, including especially the unknown target radar cross section and, at least until it is successfully detected, its range. In addition, the target may be moving relative to the radar, so that the echo is modified by a Doppler shift.

The derivation of the magnitude-based detectors of Secs. 6.2.2 and 6.2.3 included an assumption that the received signal amplitude was known. Specifically, it was assumed that the received signal sample vector $\tilde{\mathbf{m}}$ was known, except for its absolute phase. However, as noted the absolute amplitude is also unknown in general. To determine the effect of an unknown amplitude, assume that the received signal is $A\tilde{\mathbf{m}}$, where A is an unknown but deterministic scale factor.¹¹ The analysis of Sec. 6.2.2 can be repeated under this assumption. The detector output under hypothesis H_0 is unchanged as would be expected, since the target echo with its unknown amplitude is not present in this case. Under hypothesis H_1 , the detector output is now $\operatorname{Re}\{\tilde{\mathbf{m}}^H\mathbf{y}\} - N(A^2E,E\sigma_w^2/2)$. Note that the detector still considers the quantity $\tilde{\mathbf{m}}^H\mathbf{y}$ rather than $A\tilde{\mathbf{m}}^H\mathbf{y}$ because the $\tilde{\mathbf{m}}$ arises from the matched filter applied to the data and thus does not include the unknown amplitude factor A of the signal echo. Also, the quantity $E = \tilde{\mathbf{m}}^H\mathbf{m}$ is now the energy of the matched filter reference signal, while the actual signal energy becomes A^2E .

The equivalent of Eq. (6.51) is now

$$P_D = Q_M \left(\sqrt{\frac{2A^2 E}{\sigma_w^2}}, \sqrt{\frac{2T'^2}{E\sigma_w^2}} \right)$$
(6.58)

As before, the second argument of Eq. (6.58) can be written in terms of the probability of false alarm. Furthermore, because the actual signal energy is now A^2E , the first argument is still $\sqrt{2\chi}$. Thus, the detection performance is still given by Eq. (6.52). The unknown echo amplitude neither requires any change in the detector structure nor changes its performance.

Despite the unknown amplitude, the sufficient statistic was not changed. Furthermore, the probability of false alarm could be computed without knowledge of the amplitude. When both these conditions hold, the detection test is called a *uniformly most powerful* (UMP) test (Dudgeon and Johnson, 1993).

A UMP does not exist for the case where the signal delay (range) is

unknown, which again is the only realistic assumption that can be made in radar. It is therefore necessary to resort to a *generalized likelihood ratio test* (GLRT), in which the likelihood ratio is written as a function of the unknown signal delay Δ , and then the value of D that maximizes the likelihood ratio is found. Details are given in Dudgeon and Johnson (1993). The problem of estimating the time delay or range that maximizes the likelihood ratio is a major topic in <u>Chap. 7</u>. The result simply requires evaluation of the matched filter output to identify the range that produces the maximum output. In practice, each matched filter output sample is compared to a threshold. If the threshold is crossed, a detection is declared and the value of Δ at which the threshold crossing occurs is taken as an estimate of the target delay.

If the target is moving, an unknown Doppler shift will be imposed on the incident signal. The received echo will then be proportional not to $\tilde{\mathbf{m}}$, but to a modified signal $\tilde{\mathbf{m}}'$ where the samples of the reference signal $\tilde{\mathbf{m}}$ have been multiplied by the complex exponential sequence $\exp(j\omega_D n)$, where ω_D is the normalized Doppler shift. The required matched filter impulse response is now $\tilde{\mathbf{m}}'$; if $\tilde{\mathbf{m}}$ is replaced by $\tilde{\mathbf{m}}'$ in the derivations of Sec. 6.2.2, the same performance results as before will be obtained. Because ω_D is unknown, however, it is necessary to test for different possible Doppler shifts by conducting the detection test for multiple possible values of ω_D , similar to the procedure used to test for unknown range. If a set of K potential Doppler frequencies uniformly spaced from -PRF/2 to +PRF/2 is to be tested, the matched filter can be implemented for all K frequencies at once using the pulse Doppler processing techniques described in Chap. 5.

6.3 Threshold Detection of Radar Signals

The results of the preceding sections can now be applied to some reasonably realistic scenarios for detecting radar targets in noise. These scenarios will almost always include unknown parameters of the signal to be detected (the target), specifically, its amplitude, absolute phase, time of arrival, and Doppler shift. Both detection using a single sample of the target signal and, when available, multiple samples are of interest. In the latter case, as discussed in Chap. 2, the target signal is often modeled as a random process, rather than a simple constant; the discussion in this chapter will be limited to the four Swerling models to illustrate both the approach and the classical, and still very useful, results obtained in these cases. Furthermore, it will be seen that the idea of pulse integration is needed in the case of multiple samples. Finally, a square-law detector will be assumed, though one important approximation that applies to linear detectors will also be introduced. Figure 6.10 represents one possible taxonomy of the most common variations on the radar detection problem.



FIGURE 6.10 Taxonomy of detection problems considered. (Adapted from Levanon, 1988.)

Each of these will be discussed in turn in this section, with the exception of the Swerling 3 and 4 cases, for which the strategy is shown but the details are not carried out; the partially correlated case, which is not considered; and adaptive threshold-setting techniques (CFAR), which are the subject of <u>Sec. 6.5</u>.

6.3.1 Coherent, Noncoherent, and Binary Integration

The ability to detect targets is inhibited by the presence of noise and clutter. Both are modeled as random processes; the noise as uncorrelated from sample to sample, the clutter as partially correlated (including possibly uncorrelated) from sample to sample. The target is modeled as either nonfluctuating (i.e., a constant) or a random process that can be either completely correlated or completely uncorrelated from sample to sample (the Swerling models), or partially correlated from sample to sample. The signal-to-interference ratio (SIR) and thus the detection performance are often improved by *integrating* (adding) multiple samples of the target and interference, motivated by the idea that the interference can be "averaged out" by adding multiple samples. This idea was first discussed in Chap. 1. Thus, in general detection will be based on N samples of the target + interference. Note that care must be taken to integrate

samples that represent the same range and Doppler resolution cells.

Integration may be applied to the data at three different stages in the processing chain

- 1. After coherent demodulation, to the baseband complex-valued (I and Q, or magnitude and phase) data. Combining complex data samples is referred to as *coherent integration*.
- 2. After envelope detection, to the magnitude (or squared or log magnitude) data. Combining magnitude samples after the phase information is discarded is referred to as *noncoherent integration*.
- 3. After threshold detection, to the target present/target absent decisions. This technique is called *binary integration*.

A system could elect to use none, one, or any combination of these techniques. Many systems use at least one integration technique, and a combination of either coherent or noncoherent with post-detection binary integration is also common. The major cost of integration is the time and energy required to obtain multiple samples of the same range, Doppler, and/or angle cell (or multiple threshold detection decisions for that cell); this is time that cannot be spent searching for targets elsewhere, or tracking already-known targets, or imaging other regions of interest. Integration also increases the signal processing computational load. Modern systems vary as to whether this is an issue: the required operations are simple, but must be performed at a very high rate in many systems.

In coherent integration, complex data samples y_n are combined to form a new complex variable *y*:

$$y = \sum_{n=0}^{N-1} y_n \tag{6.59}$$

As shown in <u>Chap. 1</u>, if the SNR of a single sample y_n is χ_1 the integrated data sample y has an SNR $\chi_N = N\chi_1$ provided that all of the samples add in phase. That is, coherent integration attains an *integration gain* of a factor of N. This is the signal processing gain G_{sp} in the radar range equation due to coherent integration. Detection calculations are then based on the result for a single sample of target + noise having the improved SNR χ_N .

In noncoherent integration, phase information is discarded. Instead, the magnitudes or squared magnitudes of the data samples are integrated. (Sometimes another function of the magnitude, such as the log-magnitude, is used.) Most classical detection results have been developed for the square law detector, which bases detection on the quantity

$$z = \sum_{n=0}^{N-1} |y_n|^2 \tag{6.60}$$

Consideration will be largely restricted to the square law detector in this section.

When coherent integration is used, detection results are obtained by using single-sample (N = 1) results with χ_1 replaced by the integrated χ_N . The situation for noncoherent integration is more complicated. As shown in <u>Chap. 1</u>, the integrated signal *z* cannot be expressed as the sum of a target-only part and a noise-only part, so that an integrated SNR cannot be defined directly. It will prove necessary to determine the actual probability density function of the integrated random variable *z* to compute detection results; this is done in the next subsection.

Binary integration takes place after an initial detection decision has taken place. That initial decision may be based on a single sample or on data that have already been coherently or noncoherently integrated. Whatever the processing before the threshold detection, after it the result is a choice between hypothesis H_0 , "target absent," and H_1 , "target present." Because there are only two possible outputs of the detector each time a threshold test is made, the output is said to be binary. Multiple binary decisions can be combined in an "*M* out of *N*" decision logic in an attempt to further improve the performance. This type of integration is discussed in <u>Sec. 6.4</u>.

6.3.2 Nonfluctuating Targets

Now consider detection based on noncoherent integration of *N* samples of a nonfluctuating target (sometimes called the "Swerling 0" or "Swerling 5" case) in white Gaussian noise. The amplitude and absolute phase of the target component are unknown. Thus, an individual data sample y_n is the sum of a complex constant $m = \tilde{m} \exp(j\theta)$ for some real amplitude \tilde{m} and phase θ , and a complex white Gaussian noise sample w_n of power $\sigma_w^2/2$ in each of the I and Q channels (total noise power σ_w^2)

$$y_n = m + w_n \tag{6.61}$$

Under hypothesis H_0 , the target is absent and $y_n = w_n$. The PDF of $z_n = |y_n|$ is Rayleigh

$$p_{z_n}(z_n | H_0) = \begin{cases} \frac{2z_n}{\sigma_w^2} e^{-z_n^2/\sigma_w^2}, & z_n \ge 0\\ 0, & z_n < 0 \end{cases}$$

(6.62)

Under hypothesis H_1 , z_n is a Rician voltage density

$$p_{z_n}(z_n | H_1) = \begin{cases} \frac{2z_n}{\sigma_w^2} e^{-(z_n^2 + \tilde{m}^2)/\beta^2} I_0\left(\frac{2\tilde{m}z_n}{\sigma_w^2}\right), & z_n \ge 0\\ 0, & z_n < 0 \end{cases}$$
(6.63)

For a vector **z** of *N* such samples the joint PDFs are, for each $z_n \ge 0$

$$p_{z}(z|H_{0}) = \prod_{n=0}^{N-1} \frac{2z_{n}}{\sigma_{w}^{2}} \exp\left[-z_{n}^{2}/\sigma_{w}^{2}\right]$$
(6.64)

$$p_{z}(z|H_{1}) = \prod_{n=0}^{N-1} \frac{2z_{n}}{\sigma_{w}^{2}} \exp\left[-(z_{n}^{2} + \tilde{m}^{2})/\sigma_{w}^{2}\right] I_{0}\left(\frac{2\tilde{m}z_{n}}{\sigma_{w}^{2}}\right)$$
(6.65)

The LRT and log-LRT become

$$\Lambda = \prod_{n=0}^{N-1} \exp\left(-\tilde{m}^2 / \sigma_w^2\right) I_0\left(\frac{2\tilde{m}z_n}{\sigma_w^2}\right) = e^{-N\tilde{m}^2 / \sigma_w^2} \prod_{n=0}^{N-1} I_0\left(\frac{2\tilde{m}z_n}{\sigma_w^2}\right) \stackrel{H_1}{\underset{H_0}{>}} -\lambda$$
(6.66)

$$\ln \Lambda = -\frac{N\tilde{m}^2}{\sigma_w^2} + \sum_{n=0}^{N-1} \ln \left[I_0 \left(\frac{2\tilde{m}z_n}{\sigma_w^2} \right) \right] \stackrel{H_1}{\underset{H_0}{>}} \ln(-\lambda)$$
(6.67)

Incorporating the term involving the ratio of signal power and noise power on the left-hand side into the threshold gives

$$\sum_{n=0}^{N-1} \ln \left[I_0 \left(\frac{2\tilde{m}z_n}{\sigma_w^2} \right) \right] \stackrel{H_1}{\underset{H_0}{\leq}} \ln(-\lambda) + \frac{N\tilde{m}^2}{\sigma_w^2} \equiv T$$
(6.68)

Equation (6.68) shows that, given N noncoherent samples of a nonfluctuating target in white noise, the optimal Neyman-Pearson detection test scales each sample by the quantity $2\tilde{m}/\sigma_w^2$, passes it through the monotonic nonlinearity $\ln [I_0(\cdot)]$, and then integrates the processed samples and performs a

threshold test. There are two practical problems with this equation. First, it is desirable to avoid computing the function $\ln[I_0(\cdot)]$ possibly millions of times per second. Second, both the target amplitude $\tilde{\mathbf{m}}$ and the noise power σ_w^2 must be known to perform the required scaling. The test can be simplified by using the results of Sec. 6.2.3. Applying the square law detector approximation of Eq. (6.55) to Eq. (6.68) gives the test:

$$\sum_{n=0}^{N-1} \left(\frac{\tilde{m}^2 z_n^2}{\sigma_w^4} \right) \quad \stackrel{H_1}{\underset{H_0}{>}} \quad T \tag{6.69}$$

Combining all constants into the threshold gives us the final detection rule:

$$z = \sum_{n=0}^{N-1} z_n^2 \qquad \stackrel{H_1}{\underset{H_0}{<}} \qquad \frac{\sigma_w^4 T}{\tilde{m}^2} \equiv T'$$
(6.70)

Equation (6.70) states that the squared magnitudes of the data samples are simply integrated and the integrated sum compared to a threshold to decide whether a target is present or not. The integrated variable z is the sufficient statistic Υ for this problem.

The performance of the detector must now be determined. It is convenient to scale the z_n , replacing them with the new variables $z'_n = z_n/\sigma_w$ and thus replacing z with $z' = \sum (z'_n)^2 = z/\sigma_w^2$; such a scaling does not change the performance, but merely alters the threshold value that corresponds to a particular P_D or P_{FA} . The PDF of z'_n is still either Rayleigh or Rician as in Eqs. (6.62) and (6.63), but now with unit noise variance:

$$p_{z'_n}(z'_n \mid H_0) = \begin{cases} 2z'_n e^{-z'_n^2}, & z'_n \ge 0\\ 0, & z'_n < 0 \end{cases}$$
(6.71)

$$p_{z'_{n}}(z'_{n} \mid H_{1}) = \begin{cases} 2z'_{n}e^{-(z'_{n}^{2}+\chi)}I_{0}(2z'_{n}\sqrt{\chi}), & z'_{n} \ge 0\\ 0, & z'_{n} < 0 \end{cases}$$
(6.72)

where $\chi = \tilde{m}^2 / \sigma_w^2$ is the SNR. Since a square law detector is being used, define $r_n = (z'_n)^2$; then $z' = \Sigma r_n$. The PDF of r_n is exponential under H_0 and a generalized noncentral chi-squared density under H_1 :

$$p_{r_n}(r_n | H_0) = \begin{cases} e^{-r_n}, & r_n \ge 0\\ 0, & r_n < 0 \end{cases}$$
(6.73)

$$p_{r_n}(r_n|H_1) = \begin{cases} e^{-(r_n + \chi)} I_0(2\sqrt{\chi r_n}), & r_n \ge 0\\ 0, & r_n < 0 \end{cases}$$
(6.74)

Since z' is the sum of N scaled random variables r_n , the PDF of z' is the Nfold convolution of the PDF given in Eq. (6.73) or (6.74). This is most easily found using *characteristic functions* (CFs; see App. A). If $C_z(q)$ is the CF corresponding to a PDF $p_z(z)$, the CF of the N-fold convolution of the PDFs is the product of their individual characteristic functions, i.e. $C_z^N(q)$.

Under hypothesis H_0 the CF of r_n can be readily shown to be

$$C_{r_n}(q) = \frac{1}{1 - jq}$$
(6.75)

The characteristic function of z' is therefore

$$C_{z'}(q) = [C_{r_n}(q)]^N = \left(\frac{1}{1 - jq}\right)^N$$
(6.76)

The PDF of z' is obtained by inverting its characteristic function using the Fourier-like inverse CF transform, giving

$$p_{z'}(z'|H_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C_{z'}(q) e^{-jqz'} dq$$
(6.77)

Using Eq. (6.76) in Eq. (6.77) and referring to any good Fourier transform table (with allowance for the reversed sign of the Fourier kernel in the definition of the characteristic function), the Erlang density is obtained

$$p'_{z}(z'|H_{0}) = \begin{cases} \frac{(z')^{N-1}}{(N-1)!}e^{-z'}, & z' \ge 0\\ 0, & z' < 0 \end{cases}$$

(6.78)

This reduces to the exponential PDF when N = 1 as would be expected since in that case z' is the magnitude squared of a single sample of complex Gaussian noise.

The probability of false alarm is obtained by integrating Eq. (6.78) from the threshold value to $+\infty$. The result is (Olver et al., 2010)

$$P_{FA} = \int_{T}^{\infty} \frac{(z')^{N-1}}{(N-1)!} e^{-z'} dz' = 1 - I\left(\frac{T}{\sqrt{N}}, N-1\right)$$
(6.79)

where

$$I(u, M) = \int_{0}^{u\sqrt{M+1}} \frac{e^{-\tau} \tau^{M}}{M!} d\tau$$
(6.80)

is Pearson's form of the incomplete gamma function.¹² For a single sample (N = 1), Eq. (6.79) reduces to the especially simple result

$$P_{FA} = e^{-T} \implies T = -\ln P_{FA} \quad \text{(normalized data)}$$
(6.81)

This value of the threshold applies to the normalized statistic $z' = z/\sigma_w^2$. The corresponding result for the unnormalized statistic z is

$$P_{FA} = e^{-T/\sigma_w^2} \implies T = -\sigma_w^2 \ln P_{FA}$$
 (unnormalized data) (6.82)

Equation (6.79) can be used to determine the probability of false alarm P_{FA} for a given threshold *T* or, more likely, the required value of *T* for a desired P_{FA} .

Now the probability of detection P_D corresponding to the same threshold must be determined. Start by finding the PDF of the normalized, integrated, and square-law-detected samples under hypothesis H_1 . Each individual data sample r_n has a generalized noncentral chi-squared PDF [Eq. (6.74)]; the corresponding characteristic function is

$$C_{r_n}(q) = \frac{1}{1 - jq} \exp\left[\frac{j\chi q}{1 - jq}\right]$$
(6.83)

The CF of the sum z' of N such samples is

$$C_{z'}(q) = \left(\frac{1}{1 - jq}\right)^N \exp\left[\frac{jN\chi q}{1 - jq}\right]$$
(6.84)

and the PDF of z' is $\frac{13}{13}$

$$p_{z'}(z'|H_1) = \left(\frac{z'}{N_{\chi}}\right)^{\frac{N-1}{2}} \exp(-z' - N\chi) I_{N-1}(2\sqrt{N\chi z'})$$
(6.85)

 P_D is found by integrating Eq. (6.85). One version of the result is (Meyer and Mayer, 1973)

$$P_{D} = \int_{T}^{\infty} \left(\frac{z'}{N_{\chi}}\right)^{\frac{N-1}{2}} \exp\left(-z' - N\chi\right) I_{N-1}(2\sqrt{N\chi z'}) dz'$$
$$= Q_{M}\left(\sqrt{2N\chi}, \sqrt{2T}\right) + e^{-(T+N\chi)} \sum_{r=2}^{N} \left(\frac{T}{N\chi}\right)^{\frac{r-1}{2}} I_{r-1}\left(2\sqrt{N\chi T}\right)$$
(6.86)

Note that the summation term in the second line of Eq. (6.86) only contributes when $N \ge 2$. Equations (6.79) and (6.86) define the performance achievable for a nonfluctuating target with noncoherent integration using a square law detector (Meyer and Mayer, 1973; DiFranco and Rubin, 1980).

Figure 6.11 shows the effect of the number of samples noncoherently integrated, N, on the receiver operating characteristic when $P_{FA} = 10^{-8}$. This figure shows that noncoherent integration reduces the required single-sample SNR required to achieve a given P_D and P_{FA} , but not by the factor N achieved with coherent integration. For example, consider the single-sample SNR required to achieve $P_D = 0.9$. For N = 1, this is 14.2 dB; for N = 10, it drops to 6.1 dB, a reduction of 8.1 dB, but less than the 10 dB that corresponds to the factor of 10 increase in the number of pulses integrated. This reduction in required single-sample SNR is called the *noncoherent integration gain*.



FIGURE 6.11 Effect of noncoherent integration on detection performance for a nonfluctuating target in complex Gaussian noise as a function of single-sample SNR.

6.3.3 Albersheim's Equation

The performance results for the case of a nonfluctuating target in complex Gaussian noise are given by Eqs. (6.79) and (6.86). While relatively easy to implement in a modern software analysis system such as MATLAB[®], these equations do not lend themselves to manual calculation. Fortunately, there does exist a simple closed-form expression relating P_D , P_{FA} , and SNR *c* that can be computed with simple scientific calculators. This expression is known as *Albersheim's equation* (Albersheim, 1981; Tufts and Cann, 1983).

Albersheim's equation is an empirical approximation to the results in (Robertson, 1967) for computing the single-sample SNR χ_1 required to achieve a given P_D and P_{FA} . It applies under the following conditions:

- Nonfluctuating target in Gaussian (i.i.d. in I and Q) noise
- Linear (not square-law) detector
- Noncoherent integration of *N* samples

The estimate is given by the series of calculations

$$A = \ln\left(\frac{0.62}{P_{FA}}\right), \quad B = \ln\left(\frac{P_D}{1 - P_D}\right)$$
$$\chi_1 = -5\log_{10}N + \left[6.2 + \left(\frac{5.54}{\sqrt{N + 0.44}}\right)\right] \cdot \log_{10}(A + 0.12AB + 1.7B) \, dB$$
(6.87)

Note that χ_1 is in decibels. The error in the estimate of χ_1 is less than 0.2 dB for $10^{-7} \leq P_{FA} \leq 10^{-3}$, $0.1 \leq P_D \leq 0.9$, and $1 \leq N \leq 8096$, a useful range of parameters. For the special case of N = 1, Eq. (6.87) reduces to

$$A = \ln\left(\frac{0.62}{P_{FA}}\right), \quad B = \ln\left(\frac{P_D}{1 - P_D}\right)$$

$$\chi_1 = 10\log_{10}(A + 0.12AB + 1.7B) \, \mathrm{dB}$$
(6.88)

On a linear (not decibel) scale, the last line of Eq. (6.88) is just $\chi_1 = A + 0.12AB + 1.7B$.

To illustrate, suppose $P_D = 0.9$ and $P_{FA} = 10^{-6}$ are required for a nonfluctuating target in a system using a linear detector. If detection is to be based on a single sample, what is the required SNR of that sample? This is a direct application of Albersheim's equation. Compute $A = \ln(0.62 \times 10^6) = 13.34$ and $B = \ln(9) = 2.197$. Equation (6.88) then gives $\chi_1 = 13.14$ dB; on a linear scale, this is 20.59.

If N = 100 samples are noncoherently integrated, it should be possible to obtain the same P_D and P_{FA} with a lower single-sample SNR. To confirm this, use Eq. (6.87). The intermediate parameters A and B are unchanged. χ_1 is now reduced to -1.26 dB, a reduction of 14.4 dB corresponding to a factor of 27.54 on a linear scale. This value closely matches that obtained using the exact expressions. It is much better than the \sqrt{N} rule of thumb sometimes given for noncoherent integration, which would give a gain factor of only 10 for N = 100 samples integrated. Rather, the gain is approximately $N^{0.7}$ in this example. Albersheim's equation will be used shortly to develop an expression for estimating the noncoherent integration gain.

Albersheim's equation is useful because it requires no function more exotic than the natural logarithm and square root for its evaluation so it can be used with virtually any scientific calculator. If a somewhat larger error can be tolerated, it can also be used for square-law detector results for the nonfluctuating target, Gaussian noise case. Specifically, square law detector results are within 0.2 dB of linear detector results (Robertson, 1967; Tufts and Cann, 1983). Thus, the same equation can be used for rough calculations over the range of parameters given previously with errors not exceeding 0.4 dB.

Equations (6.87) and (6.88) provide for calculation of χ_1 given P_D , P_{FA} ,

and *N*. It is possible to solve Eq. (6.87) for either P_D or P_{FA} in terms of the other and χ_1 and *N*, further extending its usefulness. For instance, the following calculations show how to estimate P_D given the other factors (χ_1 is in dB)

$$A = \ln\left(\frac{0.62}{P_{FA}}\right), \quad Z = \frac{\chi_1 + 5\log_{10}N}{6.2 + (4.54/\sqrt{N} + 0.44)}, \quad B = \frac{10^Z - A}{1.7 + 0.12A}$$
$$\Rightarrow \quad P_D = \frac{1}{1 + e^{-B}} \tag{6.89}$$

In Eq. (6.89) A and B are the same values as in Eq. (6.87), though B cannot be computed in terms of P_D since P_D is now the unknown. A result similar to Eq. (6.89) can be derived for computing P_{FA} (see Prob. 11).

Albersheim's equation can also be used to write a relatively compact formula for estimating the noncoherent integration gain G_{nc} for nonfluctuating targets. G_{nc} is the reduction in single-sample SNR required to achieve a specified P_D and P_{FA} when N samples are combined and is the range equation signal processing gain G_{sp} for noncoherent integration. In dB, it is given by

$$G_{nc}(N) (dB) = \chi_{\parallel 1 \text{ pulse}} - \chi_{\parallel N \text{ pulse}}$$

= $5 \log_{10} N - \left[6.2 + \left(\frac{4.54}{\sqrt{N + 0.44}} \right) \right] \cdot \log_{10}(A + 0.12AB + 1.7B)$
+ $10 \log_{10}(A + 0.12AB + 1.7B) dB$
= $5 \log_{10} N - \left[\left(\frac{4.54}{\sqrt{N + 0.44}} \right) - 3.8 \right] \cdot \log_{10}(A + 0.12AB + 1.7B) dB$
(6.90)

On a linear scale this becomes

$$G_{nc}(N) = \frac{\sqrt{N}}{k^{f(N)}}$$
(6.91)

where

$$k = A + 0.12AB + 1.7B$$

$$f(N) = \left(\frac{0.454}{\sqrt{N+0.44}}\right) - 0.38\tag{6.92}$$

The constant k depends only on P_D and P_{FA} , while the term f(N) is a slowly

declining function only of N.

Figure 6.12 plots this estimate of G_{nc} in decibels for Albersheim's nonfluctuating, linear detector case as a function of N for $P_D = 0.9$ and $P_{FA} = 10^{-6}$. Also shown are curves corresponding to $N^{0.7}$ and $N^{0.8}$. The noncoherent gain is slightly better than $N^{0.8}$ for very few samples integrated (N = 2 or 3), with the effective exponent on N declining slowly as N increases. G_{nc} is bracketed by $N^{0.7}$ and $N^{0.8}$ to in excess of N = 100 samples integrated; the gain eventually slows asymptotically to become proportional to \sqrt{N} for very large N. This can be seen from Eqs. (6.91) and (6.92), which show that as $N \rightarrow \infty$, $f(N) \rightarrow -0.38$ and G_{nc} becomes proportional to \sqrt{N} .¹⁴ Large N to achieve a given P_D and P_{FA} implies a very poor single-sample SNR so that a large amount of integration is needed, while small N implies a relatively large single-sample SNR. Another conclusion is then that noncoherent integration is more efficient when the single-sample SNR is higher to begin with. In any event, the simplicity and robustness of noncoherent integration, requiring no knowledge of the phase, means it is widely used to improve the SNR before the threshold detector.



FIGURE 6.12 Noncoherent integrations gain G_{nc} for a nonfluctuating target, estimated using Albersheim's equation.

6.3.4 Fluctuating Targets

The analysis in the preceding section considered only nonfluctuating targets, sometimes called the "Swerling 0" or "Swerling 5" case. A more realistic model allows for target fluctuations. If one of the Swerling fluctuation models is used, the target RCS is drawn from either the exponential or chi-squared PDF

and the RCS of a group of *N* samples follows either the uncorrelated or fully correlated model as described in <u>Chap. 2</u>. Note that using a fluctuating target model has no effect on the probability of false alarm. P_{FA} is determined only by the PDF when no target is present; thus Eq. (6.79) still applies.

The strategy for determining the probability of detection depends on the target fluctuation model used. Figure 6.13 illustrates the approach. In all cases, the PDF of the magnitude-squared of a single sample of signal + noise is a generalized noncentral chi-squared so that the CF of a single square-law detected sample is still given by Eq. (6.83). However, the SNR in that expression is now a random variable because the target RCS is a random variable.



FIGURE 6.13 Strategy for computing P_D for fluctuating target models.

In a fully correlated case (e.g., Swerling 1 or 3) the target RCS is a fixed value for all N pulses integrated to form z'. Thus, the CF of z' is the N-fold product of Eq. (6.83) with itself

$$C_{z'}(q;\chi,N) = \left(\frac{1}{1-jq}\right)^N \exp\left[\frac{jN\chi q}{1-jq}\right]$$
(6.93)

This is the same expression as Eq. (6.84) except that now $C_{z'}$ is written explicitly as a function of all of q, χ , and N. Next take the expected value of the CF over the SNR

$$\overline{C}_{z'}(q;\overline{\chi},N) = \int_0^\infty p_{\chi}(\chi) C_{z'}(q;\chi,N) d\chi$$
(6.94)

where $p_{\chi}(\chi)$ is the PDF of the SNR.

For the Swerling 1 model the PDF of the SNR is exponential:

$$P_{\chi}(\chi) = \frac{1}{\overline{\chi}} e^{-\chi/\overline{\chi}}$$
(6.95)

where \overline{x} is the mean value of the SNR. Using Eqs. (6.93) and (6.95) in Eq. (6.94) gives the characteristic function averaged over the signal fluctuations:

$$\overline{C}_{z'}(q;\overline{\chi},N) = \frac{1}{(1-jq)^{N-1}[1-j(1+N\,\overline{\chi})q]}$$
(6.96)

The N = 1 and N > 1 cases are best handled separately. For N = 1 the CF simplifies to just $\overline{C}_{z'}(q; \overline{\chi}, 1) = (1 - j(1 + \overline{\chi})q)^{-1}$. The inverse is the PDF of z' under hypothesis H_1 for Swerling case 1 and N = 1,

$$p_{z'}(z'|H_1) = \frac{1}{1+\overline{\chi}} \exp\left[\frac{-z'}{1+\overline{\chi}}\right] \quad (N=1)$$
(6.97)

The inverse of the more general CF for N > 1 of Eq. (6.96) is the PDF

$$p_{z'}(z'|H_1) = \frac{1}{N\overline{\chi}} \left(1 + \frac{1}{N\overline{\chi}} \right)^{N-2} I\left[\frac{z'}{(1+1/N\overline{\chi})\sqrt{N-1}}, N-2 \right] \exp\left[\frac{-z'}{1+N\overline{\chi}} \right] \quad (N > 1)$$

$$(6.98)$$

Integrating the PDF of Eqs. (6.97) and (6.98) from the threshold T to $+\infty$ leads to the following expression for the probability of detection in the Swerling 1 case (Meyer and Mayer, 1973):

$$N = 1: P_{D} = e^{-T/(1+\bar{\chi})}$$

$$N > 1: P_{D} = 1 - I \left[\frac{T}{\sqrt{N-1}}, N-2 \right]$$

$$+ \left(1 + \frac{1}{N\bar{\chi}} \right)^{N-1} I \left[\frac{T}{\left(1 + \frac{1}{N\bar{\chi}}\right)\sqrt{N-1}}, N-2 \right] \exp\left[-T/(1+N\bar{\chi}) \right]$$
(6.99)

The expression for N > 1 can be simplified when $P_{\text{FA}} \ll 1 \text{ and } N \bar{\chi} > 1$; both conditions will be true in any scenario where target detection is likely to be successful. The result is

$$P_D \approx \left(1 + \frac{1}{N\,\overline{\chi}}\right)^{N-1} \exp\left[-T/(1 + N\,\overline{\chi})\right] \quad (P_{FA} \ll 1, N\,\overline{\chi} > 1)$$
(6.100)

Furthermore, Eq. (6.100) is exact when N = 1; in this case it reduces to the N = 1 case in Eq. (6.99). For the N = 1 case, Eq. (6.81) can then be used in Eq. (6.99) to write a direct relationship between P_D and P_{FA} for a single sample of a Swerling 1 target

$$P_D = (P_{FA})^{1/(1+\bar{\chi})}$$
(6.101)

In an uncorrelated fluctuation model such as the Swerling 2 or 4 cases, each of the N samples noncoherently integrated has a different value of SNR. Consequently, it is appropriate to average over the SNR in the single-sample CF first

$$\overline{C}_{r_n}(q;\overline{\chi}) = \left(\frac{1}{1-jq}\right) \int_0^\infty p_{\chi}(\chi) \exp\left(\frac{jq\chi}{1-jq}\right) d\chi$$
(6.102)

and then perform the N-fold multiplication of the averaged single-sample CF to get the CF of the integrated data

$$\overline{C}_{z'}(q;\overline{\chi},N) = [\overline{C}_{r}(q;\overline{\chi})]^{N}$$
(6.103)

For the Swerling 2 model specifically the exponential PDF can again be used for the SNR [Eq. (6.95)], applying it this time in Eq. (6.102) to arrive at

$$\overline{C}_{r_n}(q;\overline{\chi}) = \frac{1}{[1 - jq(1 + \overline{\chi})]}$$
(6.104)

and thus

$$\overline{C}_{z'}(q;\overline{\chi},N) = \frac{1}{\left[1 - jq(1+\overline{\chi})\right]^N}$$
(6.105)

Inverse transforming Eq. (6.105) gives the PDF of z' under hypothesis H_1 for Swerling case 2:

$$p_{z'}(z'|H_1) = \frac{z'^{N-1}e^{-z'/(1+\overline{\chi})}}{(1+\overline{\chi})^N(N-1)!}$$
(6.106)

Integrating Eq. (6.106) gives the probability of detection, which can be shown to be (Meyer and Mayer, 1973)

$$P_D = 1 - I \left[\frac{T}{(1 + \overline{\chi})\sqrt{N}}, N - 1 \right]$$
 (6.107)

When N = 1 correlation models are irrelevant. Because they are based on the same PDF for RCS the Swerling 1 and 2 cases therefore produce the same outcome. This is most easily seen by observing that the CFs of the detection statistic z' for each case given in Eqs. (6.96) and (6.105) and the corresponding PDFs of Eqs. (6.97) and (6.106) are identical for N = 1. Consequently, the simple results of Eqs. (6.99) and (6.101) can be used for both models.

Results for Swerling 3 and 4 targets can be obtained by repeating the previous analyses for the Swerling 1 and 2 cases, but with a chi-squared instead of exponential density function for the SNR:

$$p_{\chi}(\chi) = \frac{4\chi}{\overline{\chi}^2} e^{-2\chi/\overline{\chi}}$$
(6.108)

Derivations of the resulting expressions for P_D can be found in Meyer and Mayer (1973), DiFranco and Rubin (1980), and many other radar detection texts. Table 6.1 summarizes one form of the resulting expressions. For N = 1 the Swerling 3 and 4 results are in fact identical.

Case	P _D	Comments	
0 or 5	$Q_{M}\left(\sqrt{2N\overline{\chi}},\sqrt{2T}\right) + e^{-(T+N\overline{\chi})} \sum_{r=2}^{N} \left(\frac{T}{N\overline{\chi}}\right)^{\frac{r-1}{2}} I_{r-1}\left(2\sqrt{NT\overline{\chi}}\right)$	Nonfluctuating case	
1	$\left(1\!+\!\frac{1}{N\overline{\chi}}\right)^{\!N\!-\!1} \exp\!\left[\frac{-T}{1\!+\!N\overline{\chi}}\right]$	Approximate for $P_{FA} \ll 1$ and $N\overline{\chi} > 1$; exact for $N = 1$	
2	$1 - I\left[\frac{T}{(1+\overline{\chi})\sqrt{N}}, N-1\right]$		
3	$\left(1+\frac{2}{N\overline{\chi}}\right)^{N-2} \left[1+\frac{T}{1+(N\overline{\chi}/2)}-\frac{2(N-2)}{N\overline{\chi}}\right] \exp\left[\frac{-T}{1+N\overline{\chi}/2}\right]$	Approximate for $P_{FA} \ll 1$ and $N\overline{\chi}/2 > 1$; exact for $N = 1$ or 2	
4	$\begin{split} c^N \sum_{k=0}^N \frac{N!}{k!(N-k)!} & \left(\frac{1-c}{c}\right)^{N-k} \left\{\sum_{l=0}^{2N-1-k} \frac{e^{-cT}(cT)^l}{l!}\right\} & T > N(2-c) \\ & 1-c^N \sum_{k=0}^N \frac{N!}{k!(N-k)!} & \left(\frac{1-c}{c}\right)^{N-k} \sum_{l=2N-k}^\infty \frac{e^{-cT}(cT)^l}{l!} & T < N(2-c) \end{split}$	$c \equiv \frac{1}{1 + (\overline{\chi}/2)}$	
$P_{FA} = 1 -$	$I\left(\frac{T}{\sqrt{N}}, N-1\right)$ in all cases	1	
$I(\cdot,\cdot)$ is function	Pearson's form of the incomplete Gamma function; $I_k(\cdot)$ is the of the first kind and order k.	modified Bessel	

TABLE 6.1 Probability of Detection for Swerling Model Fluctuating Targets

 with a Square-Law Detector

Figure 6.14 compares the detection performance of the four Swerling model fluctuating targets and the nonfluctuating target for N = 10 samples as a function of the mean single-sample SNR for a fixed $P_{FA} = 10^{-8}$. Assuming that the primary interest is in relatively high (> 0.5) values of P_D , the upper half of the figure is of greatest interest. In this case, the nonfluctuating target is the most favorable in the sense that it achieves a given probability of detection at the lowest SNR. The worst case (highest required SNR for a given P_D) is the Swerling case 1, which corresponds to fully correlated samples and an exponential PDF of the target RCS. For instance, $P_D = 0.9$ requires $\chi \approx 6.1$ dB for the nonfluctuating case, but $\chi \approx 14.5$ dB for the Swerling 1 case, a difference of 8.4 dB.



FIGURE 6.14 Comparison of detection performance for nonfluctuating and Swerling fluctuating target models using noncoherent integration of 10 pulses (N = 10) and a fixed probability of false alarm $P_{FA} = 10^{-8}$ with a square law detector.

At least two general conclusions can be drawn from Fig. 6.14. First, for values of $N\bar{x}$ in the mid-teens of decibels (certainly the case if detection is to be very likely), P_D is greatest for nonfluctuating targets. Evidently target fluctuations make detection more difficult, i.e., require a higher SNR to achieve a given P_D and P_{FA} . Second, given that a target exhibits a fluctuating RCS, uncorrelated fluctuations (e.g., Swerling 2 and 4) aid target detectability compared to correlated fluctuations (e.g., Swerling 1 and 3). The last observation suggests that it is desirable to be able to force the data collected from a complex target and subsequently combined noncoherently to be fully decorrelated. Many radars use *frequency agility* to accomplish this. As discussed in Chap. 2, stepping the radar RF from one CPI or pulse to the next, as appropriate, will decorrelate successive target RCS measurements provided the frequency step size $\Delta F \ge c/2L_d$, where L_d is the target depth as viewed from the radar.

6.3.5 Shnidman's Equation

The analytic results in <u>Table 6.1</u> are too complex for "back-of-the-envelope" calculations or even for calculation on programmable calculators. Albersheim's equation provided a simple approximation for the nonfluctuating target case, but it is not applicable to fluctuating targets in general or the Swerling models in particular. This is a serious limitation since the nonfluctuating case provides overly optimistic results for most parameter ranges of interest.

Fortunately, empirical approximations have also been developed for the Swerling cases. One example is *Shnidman's equation* (Shnidman, 2002). Similar to Albersheim's equation, this series of equations gives the single-pulse SNR χ_1 required to achieve a specified P_D and P_{FA} with noncoherent integration of *N* samples. Unlike Albersheim's equation, the results are for a square law detector. However, as noted previously the differences in the required SNR for linear and square-law detectors are typically no more than 0.2 dB.

Shnidman's equation is given by the following series of calculations:

$$K = \begin{cases} \infty, & \text{nonfluctuating target ("Swerling 0/5")} \\ 1, & \text{Swerling 1} \\ N, & \text{Swerling 2} \\ 2, & \text{Swerling 3} \\ 2N & \text{Swerling 4} \\ \alpha = \begin{cases} 0 & N < 40 \\ 1/4 & N \ge 40 \end{cases}$$
(6.109)

$$\eta = \sqrt{-0.8 \ln(4P_{FA}(1-P_{FA}))} + \text{sign} (P_D - 0.5) \sqrt{-0.8 \ln(4P_D(1-P_D))}$$

$$X_{\infty} = \eta \left(\eta + 2\sqrt{\frac{N}{2}} + \left(\alpha - \frac{1}{4} \right) \right)$$

$$C_1 = \{ [(17.7006P_D - 18.4496)P_D + 14.5339]P_D - 3.525 \} / K$$

$$C_2 = \frac{1}{K} \left\{ \exp(27.31P_D - 25.14) + (P_D - 0.8) \left[0.7 \ln\left(\frac{10^{-5}}{P_{FA}}\right) + \frac{(2N - 20)}{80} \right] \right\}$$
(6.110)

$$C_{\rm dB} = \begin{cases} C_1, & 0.1 \le P_D \le 0.872 \\ C_1 + C_2, & 0.872 < P_D \le 0.99 \end{cases}$$
$$C = 10^{C_{\rm dB}/10} \tag{6.111}$$

$$\chi_{1} = \frac{C \cdot X_{\infty}}{N}$$

$$\chi_{1} (dB) = 10 \log_{10}(\chi_{1})$$
(6.112)

The function sign(x) is +1 if x > 0 and -1 if x < 0. Note that several of the equations simplify in the nonfluctuating case, $K = \infty$. Specifically, in this case $C_1 = C_2 = 0$ so that in turn $C_{dB} = 0$ and C = 1.

The accuracy bounds on Shnidman's equation are somewhat looser than those specified for Albersheim's equation. Except at the extreme values of P_D for the Swerling 1 case, the error in the estimate of χ_1 is less than 0.5 dB for 0.1 $\leq P_D \leq 0.99$, $10^{-9} \leq P_{FA} \leq 10^{-3}$, and $1 \leq N \leq 100$. This is a much wider range for P_D than used in Albersheim's equation. The range of N is much smaller but still large enough for almost all problems of interest. Figure 6.15 illustrates the error in the estimate of the single-pulse SNR χ_1 using Shnidman's equation for the case $P_{FA} = 10^{-6}$, N = 5, and P_D over the specified range of 0.1 to 0.99.



FIGURE 6.15 Example of error in estimating χ_1 via Shnidman's equation for $P_{FA} = 10^{-6}$ and N = 5.

A still more accurate approximation for the nonfluctuating and Swerling 1 cases is described in Hmam (2003). However, it is not applicable to all of the Swerling cases and the computations, while easy, are more extensive.

6.4 Binary Integration

Any coherent or noncoherent integration is followed finally by comparing the integrated data to a threshold. The result is a choice between two hypotheses, "target present" or "target absent," so the output is binary in the sense that it takes one of only two possible outcomes. If the entire detection process is repeated *N* times for a given range, Doppler, or angle cell, *N* binary decisions will be available. Each decision of "target present" will have some probability P_D of being correct and a probability P_{FA} of being incorrect. To improve the reliability of the detection decision, the decision rule can require that a target be detected on some number *M* of the *N* decisions before it is finally accepted as a valid target detection. This process is called *binary integration*, "*M* of *N*" *detection*, or *coincidence detection* (Levanon, 1988; Skolnik, 2001).

To analyze binary integration, begin by assuming a nonfluctuating target so that the probability of detection P_D is the same for each of N threshold tests. Then the probability of *not* detecting an actual target (i.e., the probability of a miss) on one trial is $1 - P_D$. If there are N independent trials, the probability of missing the target on all N trials is $(1 - P_D)^N$. Thus, the probability of detecting the target on at least one of N trials, denoted the *binary integrated probability* P_{BD} , is

$$P_{BD} = 1 - (1 - P_D)^N \tag{6.113}$$

Table 6.2 shows the single-trial probability of detection required to achieve $P_{BD} = 0.99$ as a function of *N*. Clearly, a "1 of *N*" decision rule achieves a high binary integrated probability of detection with relatively low single-trial probabilities of detection. In other words, the "1 of *N*" rule increases the effective probability of detection. This has the effect of reducing the SNR required to achieve the final target value of P_D .

Ν	1	2	4	10	20	100
P_D	0.99	0.90	0.68	0.37	0.2	0.045

TABLE 6.2 Single-Trial P_D Needed to Achieve $P_{BD} = 0.99$

The trouble with the "1 of *N*" rule is that it "works" for the probability of false alarm also. The probability of at least one false alarm in *N* trials is the binary integrated probability of false alarm, P_{BFA} :

$$P_{BFA} = 1 - (1 - P_{FA})^N \tag{6.114}$$

Assuming that $P_{FA} \ll 1$, Eq. (6.114) can be approximated as

$$P_{BFA} = 1 - (1 - P_{FA})^{N}$$

= $1 - \left(1 - N \cdot P_{FA} + \frac{N(N-1)}{2}P_{FA}^{2} - \cdots\right)$
 $\approx 1 - (1 - N \cdot P_{FA}) + N \cdot P_{FA}$ (6.115)

where the binomial series expansion was used to obtain the second line. Equation (6.115) shows that the "1 of N" rule increases P_{FA} by a factor of approximately N, an undesirable result. What is needed is a binary integration rule that increases P_{BD} compared to P_D , while leaving P_{BFA} equal to or less than P_{FA} . An "M of N" strategy provides better results.

Consider the binary integrated probability P_B of M successes in N trials when the probability of success on a single trial is p; it is

$$P_{B} = \sum_{r=M}^{N} {\binom{N}{r}} p^{r} (1-p)^{N-r}$$
(6.116)

where

$$\binom{N}{r} \equiv \frac{N!}{(N-r)!r!}$$
(6.117)

Equation (6.116) can be applied to the probability of false alarm by letting $p = P_{FA}$, and to the probability of detection by letting $p = P_D$. In the former case, a "success" is a false alarm, i.e., the event that has a probability of p; in the latter case, a "success" is a correct detection. Consider the specific example of a "2 of 4" rule, that is, N = 4 and M = 2. Using these parameters in Eq. (6.116) gives

$$P_{B} = \sum_{r=2}^{4} \frac{24}{(4-r)!r!} p^{r} (1-p)^{4-r}$$

= $6p^{2}(1-p)^{2} + 4p^{3}(1-p) + p^{4}$
(6.118)

To determine the effect of this rule on the probability of false alarm, let $p = P_{FA}$. Assuming that $P_{FA} \ll 1$, Eq. (6.118) can be approximated by its first term and simplified to obtain

$$P_{BFA} \approx 6P_{FA}^2$$

Thus, the "2 of 4" rule will result in a binary integrated false alarm probability that is less than the single-trial P_{FA} as desired. Because single-trial values of P_D are not necessarily very close to 1, Eq. (6.118) cannot easily be approximated in a simple form similar to Eq. (6.119). Table 6.3 shows the binary integrated probability obtained using a "2 or 4" rule for various values of the single-trial probability p. The three cases above the dotted line are appropriate for considering the effect on example single-trial probabilities of detection, while the two cases below the line are examples of the effect on example single-trial probabilities of false alarm. This table shows that the "2 of 4" rule not only reduces the probability of false alarms, it also increase the probability of detection so long as the single-trial P_D is reasonably high.

р	P _c
0.5	0.688
0.8	0.973
0.9	0.996
10 ⁻³	$5.992 imes 10^{-6}$
10 ⁻⁶	$6.0 imes 10^{-10}$

TABLE 6.3 Binary Integrated Probability Using a "2 of 4" Rule

This example illustrates the characteristics required of an "M of N" rule. For small values of p, P_B should be less than or equal to p so that the rule reduces false alarm probabilities. For larger values of p, P_B should be greater than or equal to p so that detection probabilities are increased by binary integration. To show the effect of the "M of N" rule on large and small singletrial probabilities, Fig. 6.16 plots the ratio of P_B to p for N = 4 and all four possible choices of M. A ratio greater than 1 means P_B is greater than p; this should be the case for values of p appropriate to single-trial detection probabilities. Conversely, for small values of p appropriate to false alarm probabilities, the ratio should be less than 1. Figure 6.16 shows that the ratio is greater than 1 for all values of p for the "1 of 4" rule, consistent with the earlier discussion. Similarly, the "4 of 4" rule results in a ratio that is always less than 1, good for false alarm reduction but bad for improving detection. The "2 of 4" and "3 of 4" rules both provide good false alarm reduction for small values of p and detection improvement for large values of p. However, the "3 of 4" rule increases probabilities only for p equal to approximately 0.75 or higher, a relatively narrow range, and the increase is very slight. The "2 of 4" rule

improves detection for values of *p* down to approximately 0.23, still well above any likely single-trial false P_{FA} . Thus the "2 of 4" rule appears to be the best choice when N = 4.



FIGURE 6.16 Ratio of the binary integrated probability to the single-trial probability for an "M of 4" binary integration rule.

Recall that a nonfluctuating target has been implicitly assumed since the single-trial P_D was taken to be the same on each trial. The results can be extended to fluctuating targets (Weiner, 1991; Shnidman, 1998). One result of these analyses is that for a given Swerling model, P_{FA} specification, SNR, and number of trials N, there is a value M_{opt} of M that maximizes P_{BD} for a given P_{BFA} , N, and SNR. M_{opt} can be estimated as

$$M_{\rm opt} = 10^b N^a$$
 (6.120)

where the parameters *a* and *b* are given in <u>Table 6.4</u> for the various Swerling models with $P_D = 0.9$ and $10^{-8} \le P_{FA} \le 10^{-4}$ ("Swerling 0" is the nonfluctuating case) (Shnidman, 1998). M_{opt} must be rounded to the nearest integer.

Swerling Model	a	b	Range of N	
0	0.8	-0.02	5–700	
1	0.8	-0.02	6–500	
2	0.91	-0.38	9–700	
3	0.8	-0.02	6–700	
4	0.873	-0.27	10-700	

TABLE 6.4 Parameters for Estimating M_{opt}

Another term sometimes used for the binary integrated probability is *cumulative probability*. However, that term is more commonly restricted to describe the probability of detecting a target at least once in *N* tries, for example on *N* successive scans of a surveillance radar (IEEE, 2008). If the individual scan P_D is the same on each scan, the cumulative probability P_{CD} is the binary integrated probability for the "1 of *N*" case given in Eq. (6.113). If the target range changes significantly during the *N* scans, the individual scan SNR and thus P_D would be expected to change so that a more general formula would be required.

6.5 Constant False Alarm Rate Detection

Standard radar threshold detection, as discussed in the preceding sections, assumes that the interference level is known and constant. This in turn allows accurate setting of a threshold that guarantees a specified P_{FA} . In practice, interference levels are often variable. *Constant false alarm rate* (CFAR) detection, also frequently referred to as "adaptive threshold detection" or "automatic detection," is a set of techniques designed to provide predictable detection and false alarm behavior in more realistic interference scenarios.

6.5.1 The Effect of Unknown Interference Power on False Alarm Probability

In the preceding sections the detection and false alarm performance of a square law detector were considered for a target in complex white Gaussian interference as a function of the target fluctuation model and number of measurements noncoherently integrated. It was shown in Eq. (6.82) that for a single unnormalized data sample (N = 1) of a nonfluctuating target the false alarm probability and threshold were (repeating them here for convenience)

$$P_{FA} = e^{-T/\sigma_w^2} \quad \Rightarrow \quad T = -\sigma_w^2 \ln P_{FA} \tag{6.121}$$

The threshold *T* is seen to be proportional to the interference power, $T = \alpha \sigma_w^2$, with the multiplier α a function of the desired false alarm probability.

To tune the square law detector for a particular radar system, an acceptable value of P_{FA} must be chosen. The threshold is then computed according to Eq. (6.121). The probability of detection that will be achieved is determined by the target SNR.

Accurate setting of the threshold requires accurate knowledge of the interference power σ_w^2 . In some systems this is known, but in many it is not. When the interference is principally receiver noise it is possible to measure σ_w^2 and calibrate the detector. In day-to-day operation, however, the receiver noise will vary over time due to factors such as temperature changes and component aging. Temperature compensation and periodic recalibration, if possible, can combat this problem. If the total interference power is significantly affected by external sources, the variability can be much more severe. In very low noise radar systems, a significant part of the noise power is cosmic noise. The total receiver interference then varies with the look direction and the time of day. In conventional radars, the total interference power can be affected by in-band *electromagnetic interference* (EMI). For example, UHF radars can be affected by television transmissions, while certain wireless communication services can compete with higher frequency radars, especially in urban areas. If the dominant interference is ground clutter, its power will vary radically with the type of terrain being illuminated and even the weather and seasons. For instance, open desert has a relatively low reflectivity, while refrozen snow can have a very high reflectivity. Finally, the dominant interference can be hostile electromagnetic emissions deliberately directed at the radar system (jamming). In this case, the interference power can be extremely high.

In any of these cases the observed P_{FA} will vary from the intended value. To see how significant this variation might be, let P_{FA0} be the intended probability of false alarm when the actual interference power is the expected value of σ_{w0}^2 ; thus $T = -\sigma_{w0}^2 \ln P_{FA0}$. Now suppose the actual interference power is σ_w^2 . The actual P_{FA} , using Eq. (6.121) with the threshold designed assuming an interference power of σ_{w0}^2 , will be

$$P_{FA} = \exp\left(\frac{\sigma_{w0}^{2} \ln P_{FA0}}{\sigma_{w}^{2}}\right) = \exp\left[\ln P_{FA0}^{(\sigma_{w0}^{2}/\sigma_{w}^{2})}\right]$$
$$= P_{FA0}^{(\sigma_{w0}^{2}/\sigma_{w}^{2})}$$
(6.122)

and the increase in false alarm probability will be a factor of

$$\frac{P_{FA}}{P_{FA0}} = (P_{FA0})^{\left[(\sigma_{w0}^2/\sigma_w^2) - 1\right]}$$
(6.123)

Figure 6.17 plots Eq. (6.123) for three different values of the design false alarm probability. This figure shows that even modest increases of 2 dB can cause an unintended increase in P_{FA} of 1.5 to 3 orders of magnitude, with the largest changes occurring when the desired P_{FA} is lowest. When the increase is 3 dB (a factor of 2 ×), $P_{FA} = \sqrt{P_{FA0}}$. Clearly, such sensitivity to small changes in interference power or, equivalently, small errors in setting the threshold will have major impacts on radar performance.



FIGURE 6.17 Increase in probability of false alarm for fixed threshold due to increase in noise power.

The reason for the large increase in P_{FA} observed in Fig. 6.17 is that the threshold *T* was based on an incorrect value for the interference power σ_w^2 . More generally, as the interference power at the output of the radar receiver varies, the actual P_{FA} will vary widely. From a system point of view, this is highly undesirable. When the interference power rises, the number of false alarms will also rise, possibly by orders of magnitude. It might seem that the difference between a probability of false alarm of 10^{-8} and a rate of 10^{-6} may be insignificant, but consider the example of a simple radar system with a PRF of 10 kHz and 200 range bins. If each range bin is tested, this system makes

(10,000)(200) = two million detection decisions per second. With $P_{FA} = 10^{-8}$, false alarms occur on average only once every 50 seconds. If P_{FA} rises to 10^{-6} , the system is confronted with an average of two false alarms every second. How much of a concern this increase is depends on the impact of a false alarm in the overall radar system. This could include increased demand on radar or signal processor resources to confirm or reject the false alarm or to start unneeded tracks, increased cluttering of an operator display, or reduced time for search and tracking of other targets.

If the interference power drops below that assumed when calculating the threshold, the false alarm probability will drop. This may seem inconsequential or even desirable, but a reduced P_{FA} represents a threshold that is higher than necessary to achieve the system design goals. Since P_{FA} and P_D always increase or decrease together as discussed earlier, this means that the probability of detection is less than could be achieved with a correctly set threshold.

6.5.2 Cell-Averaging CFAR

In order to obtain predictable and consistent performance the radar system designer would usually prefer a constant false alarm rate. To achieve this the actual interference power must be estimated from the data in real time so that the detector threshold can be adjusted to maintain the desired P_{FA} . A detection processor that can maintain a constant P_{FA} is called a CFAR processor.

Figure 6.18 shows a generic radar detection processor. The detector shown is for a system using range-Doppler processing, but other systems might consider only a one-dimensional vector of range cells in making a decision. Still other systems might perform the detection process on a radar image, so that the individual cells are pixels in a two-dimensional image. Whatever the form of the data, the detector will test each available data sample for the presence or absence of a target. The current *cell under test* (CUT), denoted by x_i in Fig. 6.18, is compared against a threshold determined by the interference power. If the value of the data in the test cell exceeds the threshold, the processor declares a target present at the range and velocity (or range, or image location, as appropriate) corresponding to the CUT. The next cell is then tested and so forth until a target present/target absent decision has been made for all cells of interest.



FIGURE 6.18 Generic detection processor.

To set the threshold for testing cell x_i , the interference power *in the same cell* must be known. Since it may be variable, it must be estimated from the data. The approach used in CFAR processing is based on two major assumptions:

- The neighboring cells contain interference with the same statistics as the CUT (called *homogeneous* interference), so that they are representative of the interference that is competing with the potential target.
- The neighboring cells do *not* contain any targets; they contain interference only.

Under these conditions, the interference statistics in the CUT can be estimated from the measured samples in the adjoining cells.

The statistics that must be estimated are determined by the statistics needed to implement the threshold test. For Gaussian interference and linear or square law detectors, the interference will be Rayleigh or exponential distributed, respectively. In either case, the interference PDF has only one free parameter, the mean interference power. Thus, the CFAR processor must estimate the mean interference power in the CUT by using the measured data in the adjoining cells.

For a more specific example, consider the square law case. The PDF of a cell x_i , assuming the interference is i.i.d. WGN in the I and Q signals with power $\sigma_w^2/2$ in each (total power of σ_w^2), is

$$p_{x_i}(x_i) = \frac{1}{\sigma_w^2} \exp\left(-\frac{x_i}{\sigma_w^2}\right)$$
(6.124)

As seen in Eq. (6.121), knowledge of σ_w^2 is needed to set the threshold. When exact knowledge of σ_w^2 is not available, it must be estimated.

Assume that *N* cells in the vicinity of the cell under test are used to estimate σ_w^2 , and that the interference in each is i.i.d. WGN. The joint PDF of a vector *x* of *N* such samples is

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{\sigma_{w}^{2N}} \prod_{i=1}^{N} \exp\left(-x_{i}/\sigma_{w}^{2}\right)$$
$$= \frac{1}{\sigma_{w}^{2N}} \exp\left[-\left(\sum_{i=1}^{N} x_{i}\right)/\sigma_{w}^{2}\right] \equiv \ell(\mathbf{x})$$
(6.125)

Equation (6.125) is the *likelihood function* ℓ (see Chap. 7) for the observed data vector x. The maximum likelihood estimate of σ_w^2 is obtained by maximizing Eq. (6.125) with respect to σ_w^2 while Σx_i is held constant (Kay, 1993). It is equivalent and more convenient to maximize the log-likelihood function

$$\ln \ell = -N \ln(\sigma_w^2) - \frac{1}{\sigma_w^2} \left(\sum_{i=1}^N x_i\right)$$
(6.126)

Setting the derivative of Eq. (6.126) with respect to σ_w^2 equal to zero gives

$$\frac{d(\ln \ell)}{d(\sigma_w^2)} = 0 = -N\left(\frac{1}{\sigma_w^2}\right) - \left(-\frac{1}{\left(\sigma_w^2\right)^2}\right) \left(\sum_{i=1}^N x_i\right)$$
(6.127)

Solving Eq. (6.127) for σ_w^2 gives the unsurprising result that the maximum likelihood estimate is just the average of the available data samples:

$$\widehat{\sigma_w^2} = \frac{1}{N} \sum_{i=1}^N x_i \tag{6.128}$$

The required threshold is then estimated as a multiple of the estimated interference power:

$$\hat{T} = \alpha \widehat{\sigma_w^2} \tag{6.129}$$

Because the interference power and thus the threshold are estimated from an average of the power in the cells adjoining the test cell, this CFAR approach is referred to as *cell-averaging CFAR* (CA CFAR). Because the interference power is estimated rather than known exactly, the scale factor α will not have the same value as in Eq. (6.121); it will be derived in the next subsection.

Equation (6.128) states that the parameter of the exponential PDF describing the square-law detected data should be estimated from an average of N adjoining data samples. Figure 6.19 shows two examples of how the samples to be averaged are selected. Figure 6.19*a* shows a one-dimensional data vector of range cells with the CUT, x_i , in the middle. The data in the grey cells to either side, representing data from ranges nearer and farther from the radar than the CUT, are averaged to estimate the noise parameter. These cells are called the

reference cells. The cross-hatched cells immediately adjacent to the CUT, called guard cells, are excluded from the average. The reason is that a target, if present, might straddle range cells. In that case, the energy in the cell adjacent to x_i would contain both interference and target energy and would therefore not be representative of the interference alone. The extra energy from the target would tend to raise the estimate of the interference parameter. For instance, in the square law detector case, the estimate of σ_w^2 would be too high, resulting in a threshold that was too high and a lower P_{FA} and P_D than intended. If the system range resolution is such that anticipated targets could extend over multiple range cells, more than one guard cell would be skipped on each side of the CUT. The combined reference cells, guard cells, and cell under test are referred to as the *CFAR window*.



FIGURE 6.19 CFAR windows: (*a*) one-dimensional window for range-only processor, (*b*) two-dimensional window for range-Doppler processor.

Figure 6.19*b* shows a typical two-dimensional equivalent to the onedimensional case, in this case applied to a range-Doppler matrix after detection. Both the guard region and the reference window are now two dimensional. A two-dimensional CFAR window could also be applied to synthetic aperture radar imagery, in which case the two dimensions would be simply the range and cross-range spatial dimensions. In the range-Doppler case, the cell-averaging CFAR might be applied only over certain range and Doppler cells because ground clutter renders the interference nonhomogeneous in Doppler and possibly in range.

6.5.3 Analysis of Cell-Averaging CFAR

The intent of the adaptive calculation of the threshold is to provide a constant false alarm rate despite varying interference power levels. In this subsection, the detector performance is analyzed for the case of a square law detector to see if this goal has been achieved. The threshold computed according to Eq. (6.129) will be a random variable, and therefore so will be the probability of false alarm. The detector will be considered to be CFAR if the expected value of P_{FA} does not depend on the actual value of σ_w^2 .

Combining Eqs. (6.128) and (6.129) gives an expression for the estimated threshold

$$\hat{T} = \frac{\alpha}{N} \sum_{i=1}^{N} x_i$$
(6.130)

Define $z_i = (\alpha/N)x_i$; thus $\hat{T} = \sum_{i=1}^N z_i$ Standard results from probability theory and Eq. (6.124) give the PDF of z_i

$$p_{z_i}(z_i) = \frac{N}{\alpha \sigma_w^2} e^{-N z_i / \alpha \sigma_w^2}$$
(6.131)

The PDF of \hat{T} is the Erlang density

$$p_{\hat{T}}(\hat{T}) = \begin{cases} \left(\frac{N}{\alpha \sigma_w^2}\right)^N \frac{\hat{T}^{N-1}}{(N-1)!} e^{-N\hat{T}/\alpha \sigma_w^2}, & \hat{T} > 0\\ 0, & \hat{T} < 0 \end{cases}$$
(6.132)

The P_{FA} observed with the estimated threshold will be $\exp(-\hat{T}/\sigma_w^2)$. This is now also a random variable; its expected value is

$$\overline{P}_{FA} = \int_0^\infty \exp\left(-\hat{T}/\sigma_w^2\right) p_{\hat{T}}(\hat{T}) d\hat{T}$$

$$= \left(\frac{N}{\alpha \sigma_w^2}\right)^N \frac{1}{(N-1)!} \int_0^\infty \hat{T}^{N-1} \exp\left\{-\left[(N/\alpha) + 1\right]\hat{T}/\sigma_w^2\right\} d\hat{T}$$
(6.133)

Completing this standard integral and performing some algebraic manipulations

gives the final result:

$$\overline{P}_{FA} = \left(1 + \frac{\alpha}{N}\right)^{-N} \tag{6.134}$$

For a given desired \overline{P}_{FA} , the required threshold multiplier is obtained by solving Eq. (6.134) to obtain

$$\alpha = N \left(\overline{P}_{FA}^{-1/N} - 1 \right) \tag{6.135}$$

Note that \overline{P}_{FA} does not depend on the actual interference power σ_w^2 but only on the number N of neighboring cells averaged and the multiplier. Thus, the cell-averaging technique exhibits CFAR behavior.

Now that a rule for selecting the CA CFAR threshold has been determined, the detection performance can be determined. Equation (6.99) shows that for a single sample of a Swerling 1 or 2 target with the threshold $\hat{T}, P_D = \exp[-\hat{T}/(1+\bar{\chi})]$, where $\bar{\chi}$ is the mean signal-to-noise ratio. The expected value of P_D is obtained by averaging over the threshold

$$\overline{P}_{D} = \int_{0}^{\infty} \exp\left[-\hat{T}/(1+\overline{\chi})\right] p_{\hat{T}}(\hat{T}) d\hat{T}$$
(6.136)

This integral is the same general form as Eq. (6.133); the result is

$$\overline{P}_D = \left(1 + \frac{\alpha}{N(1 + \overline{\chi})}\right)^{-N}$$
(6.137)

Note that this also does not depend on the interference power. However, this result is specific to the assumptions of complex WGN, a square law detector, Swerling 1 or 2 target, and a single test sample.

Figure 6.20 illustrates the operation of cell-averaging CFAR. The simulated data correspond to additive complex WGN with power $10\log_{10}(\sigma_w^2) = 20$ dB. A single nonfluctuating target with a power of 35 dB is present in range bin 50; the SNR is thus $10\log_{10}(\chi) = 15$ dB.


FIGURE 6.20 Example of cell-averaging CFAR threshold behavior.

If the desired $P_{FA} = 10^{-3}$, Eq. (6.121) gives the ideal T = 691, equal to 28.4 dB. This threshold level is indicated on the plot. Note that the ideal threshold is a multiple of $-\ln(P_{FA}) = 6.91$ times the true interference power; equivalently, the threshold is 8.4 dB above the interference power level.

Now consider a CA CFAR with leading and lagging windows of 10 cells each after skipping a three-cell "guard region" to each side of the CUT. Thus N= 20 cells are averaged to estimate the interference power.¹⁵ From Eq. (6.135), the multiplier α will be 8.25, placing the threshold about 9.2 dB above the *estimated* mean power. The line labeled "CFAR threshold" shows the computed threshold as the reference window slides across the data. Except in the vicinity of the target, the estimated threshold tracks the ideal threshold well, staying within 2 dB across most of the data. Note that the data exceed the CFAR threshold only at range bin 50. In this example the CFAR detector works very well: a detection would correctly be declared when the CFAR test cell is located at range bin 50, but there are no false alarms (threshold crossings) at any other range bins.

The increase in the threshold to either side of the target location is characteristic of cell-averaging CFAR. For the particular CFAR window configuration used here the cell containing the target will be in the leading reference window and will be included in the estimate of the interference power when the test cell is between range bins 37 and 46. The estimated power σ_w^2 and, in turn, the computed threshold \hat{T} will be significantly raised. This phenomenon repeats when the test cell is between the test cell is between bins 54 and 63 so that the

target is in the lagging reference window. When a target is in the reference window, the assumption that all of the reference cells share the same interference statistics as the test cell is violated and the estimate of the interference power is unreliable. However, when the test cell is located at the cell containing the target, the reference windows contain only noise samples and the threshold falls to an appropriate level, allowing target detection. The extent of the elevated threshold regions to either side of a target equals the extent of the leading and lagging windows. The extent of the region of normal threshold level between the two elevated regions equals the total number of guard cells plus one (for the test cell).

As the number of reference cells *N* becomes large, the estimate σ_w^2 should converge to the true value σ_w^2 and the average probabilities of detection and false alarm should also converge to the values obtained in <u>Sec. 6.3.4</u>. To see this, it is easier to work with $\ln \overline{P}_{FA}$ than with \overline{P}_{FA} itself:

$$\ln \overline{P}_{FA} = \ln \left\{ \left(1 + \frac{\alpha}{N} \right)^{-N} \right\}$$
$$= -N \ln \left(1 + \frac{\alpha}{N} \right) = -N \left\{ \frac{\alpha}{N} - \frac{1}{2} \left(\frac{\alpha}{N} \right)^2 + \cdots \right\}$$
(6.138)

Taking the limit as $N \rightarrow \infty$

$$\lim_{N \to \infty} \left\{ \ln \overline{P}_{FA} \right\} = -N \left\{ \frac{\alpha}{N} \right\} \quad \Rightarrow \quad \lim_{N \to \infty} \overline{P}_{FA} = e^{-\alpha}$$
(6.139)

Similarly

$$\lim_{N \to \infty} \left\{ \overline{P}_D \right\} = e^{-\alpha/(1+\overline{\chi})}$$
(6.140)

Combining Eqs. (6.139) and (6.140) gives the relation:

$$\overline{P}_D = \left(\overline{P}_{FA}\right)^{1/(1+\overline{\chi})} \tag{6.141}$$

which is identical to Eq. (6.101) for a Swerling 1 target, no noncoherent integration, and a known interference power.

All of the previous discussion has been for a square-law detector. Similar analyses can be carried out for a linear detector, but the results are more difficult to obtain in closed form. Suppose that the measurements $\{w_i\}$ are the

output of a linear detector. The threshold will be set according to

$$\hat{T} = \kappa \left(\frac{1}{N} \sum_{i=1}^{N} w_i\right)$$
(6.142)

A formula can be found relating \overline{P}_{FA} and $\hat{\theta}$ that must be solved iteratively and is numerically difficult (Raghavan, 1992). The exact results show excellent agreement with the approximation (Di Vito and Moretti, 1989)

$$\kappa = \sqrt{N\left(\overline{P}_{FA}^{-1/N} - 1\right)\left[c - (c - 1)e^{1-N}\right]}$$
(6.143)

where $c = 4/\pi$. Notice the similarity to Eq. (6.135). The square root is a consequence of using a linear rather than square-law detector. For N > 4 the $(c - 1) \exp(1 - N)$ term is negligible and since $c \approx 1.27$, $\kappa \approx 1.13\sqrt{\alpha}$. Furthermore, although the square law detector CA CFAR performs marginally better than the linear detector for some parameter choices, its performance is virtually identical for parameter values of practical interest.

In the example of Fig. 6.20 it was noted that for the parameters given the ideal threshold would be 8.4 dB above the mean power if the interference power were known exactly, but if the power had to be estimated from the data with N = 20 the threshold would be 9.2 dB above the estimated power level. This higher threshold compared to the interference power level is necessary to compensate for the imperfectly known interference power and guarantee the desired \overline{P}_{FA} . Because the threshold multiplier is increased in CFAR, the average probability of detection for a target of a given SNR will be decreased relative to the known-interference case. Alternately, to achieve a specified \overline{P}_{D} for a given \overline{P}_{FA} , a higher SNR will be required than would be were the interference power known exactly. This increase in SNR required to achieve specified detection statistics when using CFAR techniques is called the *CFAR loss*.

To quantify the CFAR loss in the case of a CA CFAR, combine Eqs. (6.134) and (6.137) to eliminate the multiplier α and solve for the value of SNR required to achieve a specified combination of \overline{P}_{FA} and \overline{P}_{D} . The result is a function of the number of samples averaged and is denoted by $\overline{\chi}_{N}$:

$$\overline{\chi}_{N} = \frac{\left(\overline{P}_{D}/\overline{P}_{FA}\right)^{1/N} - 1}{1 - \overline{P}_{D}^{1/N}}$$
(6.144)

As $N \to \infty$, the estimate of interference power converges to the true value and so \overline{P}_{FA} and \overline{P}_{D} will converge to the values given by Eqs. (6.139) and (6.140).

Similarly combining these two equations gives the value of SNR, denoted by \bar{x}_{\cdot} , required to achieve the specified probabilities when the interference estimate is perfect:

$$\overline{\chi}_{\infty} = \frac{\ln(\overline{P}_{FA}/\overline{P}_D)}{\ln(\overline{P}_D)}$$
(6.145)

The CFAR loss is then simply the ratio (Levanon, 1988; Hansen and Sawyers, 1980)

CFAR loss =
$$\frac{\overline{\chi}_N}{\overline{\chi}_{\infty}}$$
 (6.146)

Figure 6.21 plots Eq. (6.146) for a \overline{P}_D of 0.9 and three values of \overline{P}_{FA} . The loss is greatest for lower values of \overline{P}_{FA} and decreases as expected when the number of reference cells increases. For small (N < 20) reference windows, the CFAR loss can be several dB. High losses make values of N less than 10 unacceptable in most cases. Although not shown here, the CFAR loss also increases with increasing \overline{P}_D for a given \overline{P}_{FA} and N. Also, although these results were derived for a Swerling 1 or 2 target, the literature shows that the CFAR loss is roughly the same for all of the Swerling target fluctuation models and the nonfluctuating case (Nathanson, 1991).



FIGURE 6.21 Cell-averaging CFAR loss for Swerling 1/2 target in complex WGN with $\overline{P}_D = 0.9$.

6.5.4 CA CFAR Limitations

The cell-averaging CFAR concept relies on two major assumptions:

- 1. Targets are isolated; specifically, targets are separated by at least the reference window size, so that no two are ever in the reference window at the same time.
- 2. All of the reference window interference samples are independent and identically distributed, and that distribution is the same as that of the interference component in the cell containing the target; in other words, the interference is homogeneous.

While useful in many situations, either or both of these conditions are frequently violated in real-world scenarios. The second assumption is particularly likely to be untrue when the dominant interference is clutter, i.e., echo from terrain, rather than thermal noise. In this section the effect on cell-averaging CFAR of violating these assumptions is discussed, and then some modifications that combat these effects are described.

Target *masking* occurs when two or more targets are present such that, when one target is in the test cell, one or more targets are located among the reference cells. Assuming that the power of the target in the reference cell exceeds that of the surrounding interference, its presence will raise the estimate of the interference power and thus of the CFAR threshold. The target(s) in the reference window can "mask" the target in the test cell because the increased threshold will cause a reduction in the probability of detection, i.e., the detection is more likely to be missed. Equivalently, a higher SNR will be required to achieve a specified \overline{P}_{D} .

Figure 6.22 is an example of target masking. As before, the interference level is 20 dB, the target in range bin 50 has an SNR of 15 dB, and the threshold is computed using 20 reference cells and a desired \bar{P}_{FA} of 10⁻³. However, a second target with an SNR of 20 dB in range bin 58 elevates the estimated interference power when the first target is in the test cell. This increase in threshold is sufficient to prevent detection of the first target in this case. On the other hand, the 15 dB target does not affect the threshold enough to prevent detection of the second, stronger target.



FIGURE 6.22 Illustration of target masking. See text for details.

Precise analysis of the effect of a target in the reference cells is conceptually simple but somewhat complicated in practice. However, a relatively simple approximation that illustrates the effect of an interfering target can be derived. Consider a single interfering target with power γ_i that contaminates only one of the *N* CFAR reference cells. The mean SNR of this interferer is $\bar{\chi}_i = \gamma_i / \sigma_w^2$. The expected value of the new threshold will be

$$E\{\hat{T}'\} = E\left\{\frac{\alpha}{N}\left(\gamma_{i} + \sum_{i=0}^{N-1} x_{i}\right)\right\}$$
$$= \frac{\alpha\gamma_{i}}{N} + \alpha\sigma_{w}^{2} = \alpha\left(1 + \frac{\overline{\chi}_{i}}{N}\right)\sigma_{w}^{2}$$
(6.147)

Thus, $\mathbf{E}\{\hat{T}'\}$ is again a multiple of the interference power σ_w^2 as in Eq. (6.129) but with a multiplier α' given by

$$\alpha' \equiv \alpha \left(1 + \frac{\chi_i}{N} \right) \tag{6.148}$$

The elevated threshold will decrease both the P_D and P_{FA} . Using Eq. (6.148) and Eq. (6.135) in Eq. (6.137) gives an expression for the new value of \overline{P}_D in terms of the *original* design value of \overline{P}_{FA}

$$\overline{P}_D' = \left[1 + \left(\overline{P}_{FA}^{-1/N} - 1 \right) \left(\frac{1 + \overline{\chi}_i / N}{1 + \overline{\chi}} \right) \right]^{-N}$$
(6.149)

Note that if $\bar{\chi}_i \to 0$ (no interfering target) or $N \to \infty$ (target influence becomes negligible), then $\bar{P}'_D \to \bar{P}_D$. Figure 6.23*a* illustrates this behavior for one example, where $\bar{P}_{FA} = 10^{-3}$ and *N* is either 20 or 50 cells. The probability of detection without the interferer for these two cases is about 0.78 and 0.8, respectively.



FIGURE 6.23 Approximate effect of interfering target on cell-averaging CFAR. Threshold set for $P_{FA} = 10^{-3}$: (*a*) reduction in P_D , (*b*) equivalent masking loss.

Another way to characterize the effect of an interfering target is by the increase in SNR required to maintain the original value of \overline{P}_D . Let $\overline{\chi}'$ be the value of SNR required to attain the original \overline{P}_D using the elevated threshold \hat{T}' . Equation (6.137) expressed \overline{P}_D in terms of the original value of $\overline{\chi}$ and threshold

multiplier α . Approximately the same relationship will determine the detection probability \overline{P}'_D attained with the new threshold multiplier α' and SNR $\overline{\chi}'.\overline{P}'_D$ will equal \overline{P}_D if

$$\frac{\alpha}{N(1+\overline{\chi})} = \frac{\alpha'}{N(1+\overline{\chi}')}$$
(6.150)

Using <u>Eq. (6.148)</u> in <u>Eq. (6.150)</u> leads to

$$\overline{\chi}' = \left(1 + \frac{\overline{\chi}_i}{N}\right)(1 + \overline{\chi}) - 1 \tag{6.151}$$

<u>Figure 6.23*b*</u> plots the approximate "target masking loss" $\overline{x}'/\overline{x}$ in decibels for the same conditions as in <u>Fig. 6.23*a*</u>.

The results given in Eqs. (6.149) and (6.151) are only approximations. A more careful analysis would mimic the basic CA CFAR analysis of Sec. 6.5.3 by finding the PDF of the threshold in the presence of an interfering target, then using that PDF to find the expected values of P_D and P_{FA} . This approach is complicated by the fact that the interfering target changes the PDF of the cell containing it. For instance, if the interferer is nonfluctuating the PDF of the power in its cell will be a generalized noncentral chi-squared, while all of the remaining cells will still be exponentially distributed. The PDF of the threshold will be a mixture of the noncentral chi-squared and exponential PDFs. To avoid calculating this PDF, the expected value of the threshold was used in the expressions for the case of no interfering target. This gives a simple approximation that behaves correctly in the limits of large and small mean interferer SNR \bar{x}_i .

Figure 6.24 illustrates a related phenomenon, self-masking by a distributed target. The interference, detector, and target characteristics are the same as in Fig. 6.20, with the exception that the physical extent of the target is now greater than a range bin so that the target signature is spread over three consecutive cells. Figure 6.24*a* shows the effect when no guard cells are used. When one of the three target cells is the test cell, the other two contaminate the interference estimate, raising the threshold just enough to prevent detection. This effect is the reason for using guard cells. Their impact is illustrated in Fig. 6.24*b*, where three guard cells are used to each side of the test cell. This lengthens the total CFAR window slightly, but assures that when the test cell is centered on the target, the adjoining target cells do not contaminate the interference estimate and the target is now detected.



FIGURE 6.24 Self-masking and guard cells in CA CFAR: (*a*) threshold using no guard cells, (*b*) threshold using three guard cells to each side of the test cell. See text for details.

If the dominant interference is clutter rather than thermal noise or jamming, the interference will often be highly heterogeneous. The radar may collect data over a terrain region that is, for instance, part open field and part forested, or part land and part water. When the test cell is at or near the boundary between two clutter regions having different reflectivities, the statistics in the leading and lagging window will not be the same. Such *clutter edges* can cause both false alarms at the edge and masking of targets near the edge and in the lowerreflectivity region.

Figure 6.25 shows the false alarm effect of a clutter edge. The first 100 bins have a mean interference power of 20 dB. The clutter power rises suddenly by 10 dB to a mean of 30 dB in the last 100 bins, simulating a change in terrain type, perhaps from open field to a wooded area. One target is present at bin 50. Two ideal thresholds corresponding to $P_{FA} = 10^{-3}$ are shown, one for each of the two clutter regions. The threshold estimated by the cell-averaging CFAR tracks each region, but with a transition region in the range of bins 87 through 113. In this example, the clutter happens to include a high-amplitude fluctuation near the clutter edge. Because the CFAR threshold does not rise to the correct level for the new clutter level until several cells after the transition, the clutter spike crosses the CFAR threshold and a false alarm occurs. However, the target at bin 50 is detected normally.



FIGURE 6.25 False alarms at a clutter edge.

Figure 6.26 illustrates the second effect of clutter edges. The clutter regions are the same as in Fig. 6.25, but the particular clutter sequence is different. A target with an SNR of 15 dB is located at bin 95, five bins away from the clutter edge. The CFAR window uses 10 cells in both the leading and lagging windows, as well as three guard cells on either side of the test cell, for

a total CFAR window length of 27 cells. Thus, when the test cell is centered over the target, the leading reference window is mostly filled by clutter from the high-reflectivity region, elevating the threshold above the target and causing a missed detection. (Note that this example does not also suffer a clutter-edge false alarm.)



FIGURE 6.26 Target masking at a clutter edge. Clutter parameters are the same as in <u>Fig. 6.25</u>.

6.5.5 Extensions to Cell-Averaging CFAR

The performance limitations caused by nonhomogeneous clutter and interfering targets have led to the development of numerous extensions to the cell-averaging CFAR concept, each designed to combat one or more of the deleterious effects. These techniques are often heuristically motivated and can be difficult to analyze exhaustively due to the many variations in clutter non-homogeneity, target and interfering target SNR, CFAR window size, and CFAR detection logic. Additional information on many of the techniques described here is available in Keel (2010).

One common CFAR extension is the *smallest-of cell-averaging CFAR* (SOCA CFAR); the method is also known as the *least-of cell-averaging CFAR*. This technique is intended to combat the masking effect caused by an interfering target among the CFAR reference cells seen in Fig. 6.22. In an *N*-cell SOCA approach, the lead and lag windows are averaged separately to create two independent estimates σ_{w1}^2 and σ_{w2}^2 of the interference mean, each based on *N*/2 reference cells. The threshold is then computed from the smaller of the two

estimates in a manner similar to Eq. (6.129)

$$\hat{T} = \alpha_{\rm SO} \min\left(\sigma_{w1}^2, \sigma_{w2}^2\right) \tag{6.152}$$

If an interfering target is present in one of the two windows it will raise the interference power estimate in that window. Thus, the lesser of the two estimates is more likely to be representative of the true interference level and should be used to set the threshold.

Because the interference power is estimated from N/2 cells instead of N cells, the threshold multiplier α required for a given design value of \overline{P}_{FA} will be increased. It is tempting to conclude that the threshold multiplier α_{SO} for SOCA CFAR could be calculated using Eq. (6.135) with N replaced by N/2. A more careful analysis shows that the required multiplier is the solution of the equation (Weiss, 1982)

$$\overline{P}_{FA}/2 = \left(2 + \frac{\alpha_{SO}}{(N/2)}\right)^{-N/2} \left\{\sum_{k=0}^{N-1} \left(\frac{N}{2} - 1 + k \atop k\right) \left(2 + \frac{\alpha_{SO}}{N/2}\right)^{-k}\right\}$$
(6.153)

This equation must be solved iteratively. As an example, for $\overline{P}_{FA} = 10^{-3}$ and N = 20, $\alpha_{SO} = 11.276$. In contrast, the CA CFAR multiplier is $\alpha = 8.25$ for the same conditions.

Figure 6.27*a* compares the behavior of conventional CA CFAR and SOCA CFAR on simulated data containing two closely spaced targets of mean SNR 15 and 20 dB, a 10-dB clutter edge, and a third 15-dB target near the clutter edge. As before, the lead and lag windows are both 10 samples (thus N = 20) and there are three guard cells to each side of the test cell. The ideal threshold shown is based on $\overline{P}_{FA} = 10^{-3}$. The threshold multipliers are $\alpha = 8.25$ for CA CFAR and $\alpha_{SO} = 11.276$ for SOCA CFAR as discussed previously.



FIGURE 6.27 Comparison of conventional, "smallest-of," and "greatest of" cell-averaging CFAR with multiple targets and a clutter edge: (*a*) CA and SOCA CFAR, (*b*) CA and GOCA CFAR.

The CA CFAR masks the weaker of the two closely spaced targets and also fails to detect the target near the clutter edge, but does not exhibit a false alarm at the clutter edge in this instance. The SOCA threshold, in contrast, easily allows detection of both targets; the half of the CFAR window contaminated by the other target is simply ignored by the SOCA logic. Similarly, the SOCA CFAR detects the target near the clutter edge, again because the half of the window containing the higher-power clutter is ignored.

Figure 6.27*a* also shows the principal failing of the SOCA method. Although the CA CFAR did not exhibit a false alarm at the clutter edge, the SOCA CFAR does. This is a natural consequence of the SOCA logic. As the CFAR window crosses a clutter edge, there will be a region in which the test cell is in the higher interference power region, while one of the lead or lag windows is filled mostly or entirely with samples of the lower power interference. The SOCA logic ensures that the threshold is then based on the lower interference power, significantly raising the probability that the clutter in the test cell will cross the threshold.

For systems and environments in which closely spaced targets are unlikely but the clutter is highly nonhomogeneous, clutter-edge false alarms may be of much more concern than target masking. In this case, the observations above suggest that a *greater-of cell-averaging CFAR* (GOCA CFAR) logic be used. As with the SOCA technique, the lead and lag windows are averaged separately, but now the threshold is based on the larger of the two averages:

$$\hat{T} = \alpha_{\rm GO} \max\left(\sigma_{w1}^2, \sigma_{w2}^2\right) \tag{6.154}$$

Similar to the SOCA case, the GOCA threshold multiplier is the solution of the equation (Weiss, 1982)

$$\overline{P}_{FA}/2 = \left(1 + \frac{\alpha_{GO}}{(N/2)}\right)^{-N/2} - \left(2 + \frac{\alpha_{GO}}{(N/2)}\right)^{-N/2} \times \left\{\sum_{k=0}^{(N/2)-1} \left(\frac{N}{2} - 1 + k \atop k\right) \left(2 + \frac{\alpha_{GO}}{(N/2)}\right)^{-k}\right\}$$
(6.155)

For N = 20 and $\overline{P}_{FA} = 10^{-3}$, $\alpha_{GO} = 7.24$.

Figure 6.27*b* illustrates the performance of the GOCA CFAR logic on the same example used in Fig. 6.27*a*. The GOCA threshold is now equal to or higher than the CA CFAR threshold. Not surprisingly, the GOCA logic successfully avoids the false alarm at the clutter edge. However, the strong target masks the weaker target. Furthermore, the weaker target even makes detection of the stronger target more marginal, although in this case the detection is successful. The GOCA CFAR also misses the target near the clutter edge due to the masking effect of the elevated clutter. Additional analysis of the GOCA

CFAR for the case of a linear detector is provided in Pace and Taylor (1994).

The SOCA and GOCA CFARs estimate the threshold using only half of the reference cells. Consequently, they exhibit a higher CFAR loss above than the conventional CA CFAR. This additional loss is less than 0.3 dB for the GOCA CFAR over a wide range of parameters (Hansen and Sawyers, 1980). The additional loss for the SOCA CFAR is greater, especially for small values of *N*. It is necessary to use N > 32, approximately, to ensure that the additional loss is less than 1 dB over a wide range of \overline{P}_{FA} values (Weiss, 1982). A mitigating influence for either approach is that the use of a split window may allow a larger value of *N* than would normally be used in a conventional cell-averaging CFAR. The reason is that the window size in CA CFAR is often limited by concern over nonhomogeneous clutter. Since it is known that only half the window will actually be used, a larger value of *N* can be tolerated.

Still another way to combat the target masking problem is *censored* or trimmed mean CFAR (Ritcey, 1986). In these techniques, the M reference cells (M < N) having the highest power are discarded and the interference power is estimated from the remaining N - M cells. In some versions of trimmed mean CFAR both the highest and lowest power reference cells are discarded. Consider an example where M = 2. If an interfering target is present, but is confined to only one or two cells (or if two interferers are present, each confined to one cell), the censoring process will completely eliminate their elevating effect on the estimate of interference power. There will, however, be a small additional CFAR loss due to the use of only N - M cells instead of N cells (Ritcey and Hines, 1989). Proper selection of M requires some knowledge of the maximum number of interferers to be expected, as well as whether they will be confined to one cell or will be distributed over multiple cells. Typically, one-quarter to one-half of the reference window cells are discarded (Nathanson, 1991). In addition, implementation of the technique requires logic to rank order the reference cell data, sometimes a significant implementation consideration at the speeds at which real-time CFAR calculations often must be done.

Many additional variations on the approaches described previously can be used. For example, censoring can be combined with any of the CA, SOCA, or GOCA techniques. A more elaborate approach attempts to examine the behavior of the interference in the lead and lag windows and then choose an appropriate CFAR algorithm. One version of these ideas computes the mean and the variance in each of the lead and lag windows. If the variance in a window exceeds a certain threshold, it is assumed that the data in that window are not homogeneous Rayleigh interference, most likely due to target contamination. A series of logical decisions then determines whether to combine the windows for a CA CFAR using the data from both windows, use CA CFAR using only one window of data, or use GOCA or SOCA CFAR (Smith and Varshney, 2000). For example, if the means differ by less than a specified threshold and the variances in each window are less than the variance threshold, conventional CA CFAR is used to set the detection threshold. If instead the means do differ by more than that threshold, GOCA CFAR is applied. If the variance in one window exceeds the variance threshold, but in the other window does not, CA CFAR based only on the low-variance window is applied. If both windows have variances exceeding the variance threshold, SOCA CFAR is applied.

Another recent attempt to develop a CFAR algorithm that provides good performance in the presence of clutter edges and target masking while maintaining performance near to that of CA CFAR in homogeneous clutter is the *switching CFAR* (S-CFAR) (Van Cao, 2004). In this approach, the CFAR reference window is divided into two groups, not necessarily contiguous: those cells above a threshold set as a fraction of the test cell value, and those below. If the number of cells in the low amplitude group exceeds some threshold N_t , typically set to about one half of the total number N of reference cells, all N cells are used in a cell-averaging calculation. If the number of low amplitude cells. The principal advantage appears to be reduced losses compared to order statistic CFAR (OS CFAR, described in the next section) due to masking targets and somewhat improved clutter-edge performance, while avoiding the need for sorting required by OS CFAR.

Yet another approach that has been proposed to combat masking is the use of alternate detector laws (i.e., not linear or square-law). By far the most common is log CFAR, which applies conventional cell-averaging CFAR logic to the logarithm of the received power samples.¹⁶ There appears to be no simple closed form analysis equivalent to Eqs. (6.124) through (6.129) for determining the relationship between an average of the log-detected data and the interference power σ_{w}^{2} . However, motivated by considering the logarithm of the threshold computation for a square law detector seen in Eq. (6.130), the log CFAR threshold is computed by adding an offset to the averaged logarithmic data:

$$\hat{T}_{\log} = \frac{1}{N} \sum_{i=1}^{N} 10 \log_{10}(x_i) + \alpha_{\log}$$
(6.156)

In general, applying a logarithmic transformation to the data compresses its numerical dynamic range. This was an important implementation advantage in older systems built using analog or fixed-point digital hardware, but is less of a consideration with more modern processors. However, averaging the logarithmic data has the additional advantage that isolated interferers in the reference window do not have as great an influence on the numerical value of the estimated interference mean, thus reducing target masking effects. This effect is clearly shown in Fig. 6.28, which shows the same data set used previously containing closely spaced targets with 15 and 20 dB signal-to-clutter ratios and

a 10-dB clutter edge. The two targets are easily detected. Unfortunately, log CFAR exhibits poor performance at clutter edges, in particular an increased vulnerability to false alarms at clutter edges. In this example, not only does a false alarm occur at the clutter edge, but the target near the edge is not detected.



FIGURE 6.28 Comparison of CA CFAR and log CFAR on the same data as <u>Fig.</u> <u>6.27</u>.

No explicit expression is known for finding the required threshold offset α_{\log} as a function of \overline{P}_{FA} . Results have been obtained for \overline{P}_D and \overline{P}_{FA} that can be solved numerically to find suitable values of the threshold (Novak, 1980). To create Fig. 6.28, a Monte Carlo simulation was instead used to determine the required value by trial and error. The result for N = 20 reference cells and $\overline{P}_{FA} = 10^{-3}$ is $\alpha_{\log} = 11.85$ dB.

The CFAR loss of a log CFAR detector can also be estimated using Monte Carlo simulation techniques (Hansen and Ward, 1972). The log CFAR increases the CFAR loss relative to the linear detector. In homogenous clutter the number of log CFAR reference window cells N_{log} required to achieve the same CFAR loss as an *N*-cell conventional CA CFAR using a linear detector is approximately:

$$N_{\log} \approx 1.65N - 0.65$$
 (6.157)

Thus, the use of the log detector increases the required CFAR window size by

about 65 percent to avoid increasing the CFAR loss. Furthermore, for N > 8 the log CFAR loss in decibels is about 65 percent more than the loss for a CA CFAR with the same value of N.

6.5.6 Order Statistic CFAR

An alternative to cell-averaging CFAR is the class of *rank-based* or *order statistic* CFARs (OS CFAR). Proposed primarily for combating masking degradations, OS CFAR retains the one-dimensional or two-dimensional sliding window structure of CA CFAR, including guard cells if desired, but does away entirely with averaging of the reference window contents to explicitly estimate the interference level. Instead, OS CFAR rank orders the reference window data samples $\{x_1, x_2, ..., x_n\}$ to form a new sequence in ascending numerical order, denoted by $\{x_{(1)}, x_{(2)}, ..., x_{(N)}\}$. The *k*th element of the ordered list is called the *k*th *order statistic*. For example, the first order statistic is the minimum, the *N*th order statistic is the maximum, and the (*N*/2)th order statistic is the median of the data $\{x_1, x_2, ..., x_N\}$. In OS CFAR, the *k*th order statistic is selected as representative of the interference level and a threshold is set as a multiple of this value:

$$\hat{T} = \alpha_{\rm OS} x_{(k)} \tag{6.158}$$

The interference is thus estimated from only one actual data sample, instead of an average of all of the data samples. Nonetheless, the threshold in fact depends on all of the data since all of the samples are required to determine which will be the *k*th largest.

It will be shown that this algorithm is in fact CFAR (i.e., does not depend on the interference power σ_w^2), and the threshold multiplier required to achieve a specified \overline{P}_{FA} will be determined. The analysis follows (Levanon 1988). To simplify the notation, consider the square-law detected output x_i normalized to its mean, $y_i = x_i/\sigma_w^2$; this will have an exponential PDF with unit mean. The rankordered set of reference samples $\{y_i\}$ are denoted by $\{y_{(i)}\}$. For a given threshold *T*, the probability of false alarm will be

$$P_{FA}(T) = \int_{T}^{+\infty} e^{-y} \, dy = e^{-T}$$
(6.159)

The average P_{FA} will be computed as

$$\overline{P}_{FA} = \int_{0}^{+\infty} P_{FA}(T) p_{T}(T) dT$$
(6.160)

where $p_T(T)$ is the PDF of the threshold. Because *T* is proportional to the *k*th-ranked reference sample $y_{(k)}$, it is necessary to find the PDF of $y_{(k)}$, the *k*th largest of *N* i.i.d. random variables. The result can be found in many textbooks and is

$$p_{y_{(k)}}(y) = k \binom{N}{k} P_{y_i}^{k-1}(y) \left[1 - P_{y_i}(y)\right]^{N-k} p_{y_i}(y)$$
(6.161)

where $P_{y_i}(y)$ is the *cumulative distribution function* (CDF) of y_i . For complex WGN, the PDF and CDF of a single square-law detected and normalized reference sample y_i are

$$p_{y_i}(y) = e^{-y}$$

$$P_{y_i}(y) = \int_0^y p_{y(k)}(y') dy' = 1 - e^{-y}$$
(6.162)

Using Eq. (6.162) in Eq. (6.161) gives the PDF of the *k*th ranked sample:

$$p_{y(k)}(y) = k \binom{N}{k} [e^{-y}]^{N-k+1} [1 - e^{-y}]^{k-1}$$
(6.163)

The threshold is $\hat{T} = \alpha_{OS} y_{(k)}$, so the PDF of \hat{T} is $p_{\hat{T}}(\hat{T}) = (1/\alpha_{OS}) p_{y(k)}(\hat{T}/\alpha_{OS})$

$$p_{\hat{T}}(\hat{T}) = \frac{k}{\alpha_{\rm OS}} {N \choose k} \left[e^{-\hat{T}/\alpha_{\rm OS}} \right]^{(N-k)+1} \left[1 - e^{-\hat{T}/\alpha_{\rm OS}} \right]^{k-1}$$
(6.164)

Inserting this result into Eq. (6.160) gives

$$\overline{P}_{FA} = \int_{0}^{+\infty} e^{-\hat{T}} \frac{k}{\alpha_{\rm OS}} {N \choose k} [e^{-\hat{T}/\alpha_{\rm OS}}]^{N-k+1} [1 - e^{-\hat{T}/\alpha_{\rm OS}}]^{k-1} d\hat{T}$$
$$= \frac{k}{\alpha_{\rm OS}} {N \choose k} \int_{0}^{+\infty} e^{-(\alpha_{\rm OS}+N-k+1)\hat{T}/\alpha_{\rm OS}} [1 - e^{-\hat{T}/\alpha_{\rm OS}}]^{k-1} d\hat{T}$$
(6.165)

With the change of variable $T' = T/\alpha_{OS}$ this becomes the slightly more convenient form:

$$\overline{P}_{FA} = k \binom{N}{k} \int_0^{+\infty} e^{-(\alpha_{CS} + N - k + 1)T'} [1 - e^{-T'}]^{k-1} dT'$$
(6.166)

Utilizing integral 3.312(1) in Gradshteyn and Ryzhik (1980) gives

$$\overline{P}_{FA} = k \binom{N}{k} B(\alpha_{OS} + N - k + 1, k)$$
$$= k \binom{N}{k} \frac{\Gamma(\alpha_{OS} + N - k + 1)\Gamma(k)}{\Gamma(\alpha_{OS} + N + 1)}$$
(6.167)

where $B(\cdot, \cdot)$ is the beta function which in turn can be expressed in terms of the gamma function $\Gamma(\cdot)$ as shown. For integer arguments, $\Gamma(n) = (n - 1)!$ and Eq. (6.167) reduces for integer α_{os} to

$$\overline{P}_{FA} = k \frac{N!}{k!(N-k)!} \frac{(k-1)!(\alpha_{OS} + N - k)!}{(\alpha_{OS} + N)!}$$
$$= \frac{N!(\alpha_{OS} + N - k)!}{(N-k)!(\alpha_{OS} + N)!} \qquad (\alpha_{OS} \text{ integer})$$
(6.168)

Figure 6.29 plots \overline{P}_{FA} as a function of α_{OS} for two choices of OS windows, one with N = 20 and one with N = 50. In the first case, the k = 15th order statistic is chosen to set the threshold, while in the second the k = 37th order statistic is selected. Plots such as these can be used to determine the threshold multiplier needed to achieve a specified \overline{P}_{FA} for a given OS CFAR window configuration. For example, with N = 20 and k = 15, a multiplier of $\alpha_{OS} = 6.857$ gives $\overline{P}_{FA} = 10^{-3}$.



FIGURE 6.29 \overline{P}_{FA} versus threshold scale factor α_{OS} in order statistics CFAR. The selected order statistic is chosen as approximately 0.75*N*.

Figure 6.30 shows the performance of the OS CFAR on the same data set considered previously using N = 20 again and choosing the k = 15th order statistic to set the threshold. The use of the ordered statistic instead of a mean estimate makes the detector almost completely insensitive to masking by closely spaced targets so long as the number of cells contaminated by interfering targets does not exceed N - k. In this example, both closely spaced targets are readily detected. Although they were used in Fig. 6.30 for consistency, guard cells are less important in an OS CFAR since the rank ordering process will not be affected by targets spreading outside of the test cell. The effect of the window length and choice of order statistic k on the behavior at clutter edges are discussed in (Rohling, 1983). If $k \le N/2$, there will be extensive false alarms at clutter edges. Thus, k is usually chosen to satisfy N/2 < k < N. Typically, k is on the order of 0.75N (Nathanson, 1991).



FIGURE 6.30 Comparison of CA CFAR and OS CFAR on the same data as <u>Fig.</u> <u>6.27</u>.

To determine the OS CFAR loss, it is necessary to determine the SNR required to obtain a specified \overline{P}_D for a given \overline{P}_{FA} and compare that result to the ideal threshold or to CA CFAR in homogeneous interference. By repeating the analysis above but using the Swerling 1 or Swerling 2 model (Rayleigh amplitude PDF) for the test cell, it can be shown that the average probability of detection is also given by Eq. (6.167), but with α_{OS} replaced by (Levanon, 1988)

$$\alpha_{\rm OS}^D = \frac{\alpha_{\rm OS}}{1 + \overline{\chi}} \tag{6.169}$$

where \bar{x} is the average SNR of the target. By varying α_{OS} or α_{OS}^{D} , curves of \bar{P}_{D} versus \bar{P}_{FA} can be developed and the CFAR loss calculated. An alternative analysis approach is given by Blake (1988). In the absence of interfering targets, the OS-CFAR suffers a small additional loss over CA CFAR. The value depends on both k and N but is typically on the order of 0.3 to 0.5 dB. If interfering targets are present, the OS CFAR loss increases only very slowly until the number of interferers exceeds N - k, the number of "ignored" high-rank cells. In contrast, the loss in CA CFAR increases very rapidly in this case due to the elevated estimate of the interference power (Blake, 1988). Thus, OS CFAR losses are lower than CA CFAR losses in the presence of interferers. Additional results on the performance of OS CFAR, including the effect of

noncoherent integration and its behavior in Weibull clutter, are available in Shor and Levanon (1991).

6.5.7 Additional CFAR Topics

"Adaptive CFAR" is the name given to a growing class of CFAR algorithms designed to improve performance in nonhomogeneous clutter.¹⁷ Generally, they dispense with the fixed CFAR window structure (usually with half of the cells in each of the lead and lag windows). Instead they typically construct a statistical test to determine if the reference cells span one clutter field or two, i.e., whether or not the reference cell data are homogeneous. If not homogeneous, the algorithms estimate not only the clutter statistics in each field, but at which cell the transition from one to the other occurs (and therefore which type of clutter is competing with the target in the CUT).

The basic approach to adaptive CFAR for nonhomogeneous clutter was described in Finn (1986). The algorithm assumes a CFAR window of total length *N* cells, including the CUT, that spans two clutter fields, i.e., two regions with different statistical parameters. The clutter edge is presumed to occur between samples *M* and *M* + 1; however, *M* is not known. Initially, it is also assumed that the clutter follows the usual square-law detected, exponential distribution. Let $\sigma_{w1}^2(M)$ denote the estimate of σ_w^2 obtained by averaging the *M* samples 1 through *M* forming the first clutter region, and $\sigma_{w2}^2(M)$ be the average obtained from the N - M samples M + 1 through *N* forming the second region. The algorithm starts by setting M = 1 and computing $\sigma_{w1}^2(1)$, which is the "average" of only cell #1, and $\sigma_{w2}^2(1)$, the average of the remaining N - 1 cells. This process is repeated for M = 2, ..., N = 1. Thus, a pair of sample means $\sigma_{w1}^2(M)$ and $\sigma_{w2}^2(M)$ are computed for each of the N - 1 possible transition points between the two-clutter regions.

The next step is to choose the most likely transition point M_t . The maximum likelihood estimate of this transition point is the value M_t of M that maximizes the log-likelihood function (Finn, 1986):

$$\ln \ell(M) = -\{M \ln \sigma_{w1}^2(M) - (N - M) \ln \sigma_{w2}^2(M)\}$$
(6.170)

Once M_t is identified it is also known if the CUT is in the first or second clutter region. Standard CA CFAR, using the appropriate mean estimate and the number of cells with which it is estimated, can then be applied. For example, if the CUT is in the first region the threshold would be set according to

$$\alpha = M_t \left(\overline{P}_{FA}^{-1/M_t} - 1 \right)$$
$$\hat{T} = \alpha \sigma_{w1}^2(M_t)$$
(6.171)

Note that this procedure is effectively SOCA CFAR when the CUT is in the low clutter region, and GOCA CFAR when it is in the high clutter region (assuming the transition point is correctly located).

The discussion above does not allow for the possibility that the clutter is uniform. To address this an additional likelihood test is conducted to compare $\ln \Lambda(M_i)$ as computed in Eq. (6.170) with $M = M_i$ to the corresponding metric under the assumption of uniform clutter, $\ln \ell(0) = -N \ln \sigma_{w2}^2(0) = -N \ln \sigma_{w2}^2$. If $\ln \Lambda(0) > \ln \Lambda(M_i)$ the clutter is assumed homogeneous and conventional CA CFAR is applied.

When the two-region clutter hypothesis is accepted, the estimate of the transition point M_t can of course be incorrect. If the target is in fact in the low clutter region but is incorrectly determined to be in the high clutter region, the high clutter statistics will be used to set the threshold. There will then be an increased probability that the high clutter will mask the target. If the target is in the high clutter region but is incorrectly determined to be in the low clutter region, the low clutter statistics will be used to set the threshold, which will then be too low. While the target will have an enhanced probability of detection, the false alarm probability will rise, possibly dramatically. For this reason, the adaptive CFAR algorithm is modified to bias the decision in favor of the hypothesis that the target is in the high clutter region. While this will increase masking effects somewhat, it avoids the generally more damaging problem of large increases in the false alarm rate. Details are given by Finn (1986).

Again, many of the extensions to CA CFAR can also be applied to the adaptive CFAR. It can be applied to log normal or Weibull clutter by computing both sample means and variances in each region. The data in each region can be censored prior to estimating the statistics. Order statistic rather than cell-averaging rules can be used to set the threshold. Many such variations are available in the literature, as are algorithms building on the basic adaptive concept but applying different statistical estimators and decision logics.

All of the results discussed in this chapter so far have assumed exponential power (equivalently, Rayleigh voltage) interference, which is the appropriate model when the primary interference is WGN, whether the source is low-level receiver noise or high-level noise jamming. Only one parameter, the mean power σ_w^2 , is required to completely specify the PDF. However, as discussed in Chap. 2, many types of clutter are best modeled by more complicated PDFs such as the log-normal or Weibull PDF. Unlike the exponential PDF, these are two-parameter distributions and estimates of both the mean and variance (or a related parameter such as skewness) must be estimated in order to characterize them. Any threshold control mechanism must be based on estimates of both parameters if it is to exhibit CFAR behavior.

An example of a CFAR algorithm for log-normal clutter is given in Schleher (1977). The receiver uses a log detector so the detected samples $\{x_i\}$

are normally distributed. The CFAR structure is a conventional cell-averaging approach on the log data. The threshold is computed as follows:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

$$\hat{s} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^2}$$

$$\hat{T} = \hat{\mu} + \alpha \hat{s}$$
(6.172)

This CFAR threshold calculation could clearly be combined with many of the embellishments discussed earlier for CA CFAR, such as SOCA or GOCA rules or censoring.

Because of the need to estimate two parameters, the CFAR loss is greater with two parameter distributions than with the exponential distribution and in fact can be very large, especially for small numbers of reference cells. For example, with $\overline{P}_D = 0.9$, $\overline{P}_{FA} = 10^{-6}$, and N = 32 reference cells, the CFAR loss using Eq. (6.172) in log-normal interference is approximately 13 dB (Schleher, 1977). From Fig. 6.21, a conventional CA CFAR in exponential interference with the same detection statistics and window size has a CFAR loss of just under 1 dB.

The same calculations are used to set the CFAR threshold in Weibull clutter, though the specific values of α needed vary from the log-normal case. Two proposed Weibull detectors, the so-called log-*t* detector and another based on maximum likelihood estimates of the Weibull PDF parameters, have been shown to be equivalent to Eq. (6.172) (Gandhi et al., 1995).

Order statistic CFARs have also been proposed for two-parameter clutter. One example combines OS CFAR in each of the lead and lag windows with a greatest-of logic to estimate the interference mean, and then uses the single parameter Eq. (6.154) to set the threshold. Since the second (skewness) parameter of the PDF is not estimated implicitly or explicitly, the multiplier α must be made a function of the skewness, implying in turn that the skewness parameter must be known to correctly set the threshold. Performance results again suggest that choosing the order statistic k to be about 0.75N provides the best performance against interference and uncertainty in the skewness parameter (Rifkin, 1994).

In <u>Chap. 5</u> the technique of clutter mapping for detection of stationary or slowly moving targets by ground-based, fixed-site radars when the competing zero-Doppler clutter was not too strong was discussed. The threshold for each range-angle cell was computed as a multiple of the measured clutter in the same cell. The clutter measurement was obtained as a simple first-order recursive filter of the form:

$$\hat{x}[n] = (1 - \gamma)\hat{x}[n - 1] + \gamma x[n]$$
(6.173)

where $\hat{x}[n]$ is the estimate of the clutter reflectivity at time *n* (*n* usually indexes complete radar scans of an area) and x[n] is the currently measured clutter sample at time *n*. The factor γ controls the relative weight of the current measurement versus the preceding measurements. This equation is applied separately to each range-angle cell of interest. The detection threshold is then set for each range-angle cell as

$$\hat{T}[n] = \alpha \hat{x}[n-1] \tag{6.174}$$

Note that the threshold for testing for a target on the current scan is based on the clutter estimate from the previous scan. The current data are *not* included because, if the CUT does contain a target, it would distort the clutter measurement and raise the threshold, creating a self-masking effect. Even with this precaution, self-masking can cause CFAR losses of several dB if slow-moving targets are present, so that they persist in a map cell for more than one scan (Lops and Orsini, 1989). If the target persists in the map cell for a number of scans approaching the number of scans integrated to form the clutter map, detectability is essentially lost entirely.

The first-order difference equation (6.173) corresponds to an IIR filter with the impulse response

$$h[n] = \gamma (1-\gamma)^n u[n]$$
(6.175)

where u[n] is the unit step function. Consequently, the output of the filter can also be expressed as a convolution with h[n]:

$$\hat{x}[n] = x[n] * h[n] = \gamma \sum_{m=0}^{\infty} (1 - \gamma)^m x[n - m]$$
(6.176)

This equation shows that, similar to CA CFAR, the threshold will be based on an average of clutter measurements, but there are several important differences. In CA CFAR, clutter samples taken during the same pulse or dwell and from cells adjoining the CUT in spatial position, Doppler, or both are used to estimate the clutter level. The clutter must be assumed spatially homogeneous so that these adjoining cells represent the interference in the CUT accurately. The various CA CFAR extensions discussed previously are all motivated by realworld violations of this assumption. In clutter mapping, the threshold is based on clutter samples taken from the CUT but on previous time intervals. Thus, spatial homogeneity is not required, but homogeneity in time, i.e., statistical stationarity, is. Both types of CFAR require that the clutter samples be uncorrelated (in space or Doppler for CA CFAR, in time for clutter mapping) for the analyses given here to be valid. Finally, Eq. (6.176) shows that the threshold is based on an infinite, weighted sum of previous samples rather than the finite sums of CA CFAR and its variants.

It is also possible to estimate the clutter mapping threshold using a simple average of a finite number of previous measurements. One version of the *moving target detector* (MTD) is said to have used an eight-scan average, covering about 32 seconds of data (Skolnik, 2001). In this case, the CA CFAR analyses would apply to the clutter map as well. However, for computational simplicity the recursive filter of Eq. (6.173) is most often used. The average probabilities of false alarm and detection when the threshold is computed with the recursive scheme are (Nitzberg, 1986)

$$\overline{P}_{FA} = \prod_{m=0}^{\infty} \{1 + \alpha \gamma (1 - \gamma)^m\}^{-1}$$
(6.177)

$$\overline{P}_{D} = \prod_{m=0}^{\infty} \left\{ 1 + \frac{\alpha}{1 + \overline{\chi}} \gamma (1 - \gamma)^{m} \right\}^{-1}$$
(6.178)

These formulas are slow to converge in practice; a more rapidly converging variation is given in Levanon (1988). Again, α can be varied to generate curves of \overline{P}_D versus \overline{P}_{FA} and CFAR loss can then be determined by comparing these curves to the case where the interference is known exactly. The case $\gamma = 0$ in fact corresponds to the ideal case. For every increase of 0.2 in γ , the CFAR loss increases approximately 3 dB. A crude approximate formula for the CFAR loss is (Taylor, 1990)

$$\operatorname{Loss}\left(\mathrm{dB}\right) = -5.5 \log_{10} \overline{P}_{FA} \left(1 + \frac{2}{\gamma}\right)^{-1}$$

$$(6.179)$$

This formula is of limited accuracy, especially for large γ , but may be useful for rough calculations. Another implementation of the MTD used a recursive filter as described previously with a feedback coefficient of $\gamma = 7/8$ (Nathanson, 1991). A rational value of γ with a denominator that is a power of two was particularly amenable to the fixed point implementations used in early versions of the MTD, since the division by the denominator value (eight in this

case) can be implemented by simply right-shifting the binary data.

Some clutter map systems combine multiple range cells at a given azimuth direction to form a single, larger map cell. This allows the introduction of any of several standard CFAR techniques to combine the range cells and improve the map cell clutter estimate. Either cell-averaging or order statistics CFAR can be applied to the range cells to form the basic clutter power estimate, and the basic CA or OS approaches can be extended as appropriate for the environment with any of the techniques discussed previously for conventional CFAR: censoring, guard cells, smallest-of or greatest-of (SOCA and GOCA) detectors, log detectors, and two-parameter estimation algorithms (Lops, 1996; Conte and Lops, 1997).

The CFAR processors discussed so far assume a specific form of the PDF of the interference in order to determine the value of the threshold (equivalently, the value of the threshold multiplier α). For instance, the particular form of Eq. (6.134), and the formula for α in Eq. (6.135) is a result of having assumed an exponential distribution for the square-law detected interference-only samples. If noise is the dominant interference source, this is not a significant constraint. However, for a system operating in a clutter limited environment or in an environment where the dominant interference varies between clutter of various distributions, noise, and jamming, a threshold setting algorithm based on a particular interference PDF may produce large errors in the threshold setting algorithms that do not depend on the particular PDF of the interference are of interest. Such techniques are called *distribution free* or *non-parametric* CFAR algorithms (DF CFAR).

Historically, DF CFAR has been based most often on a two stage "doublethreshold" approach, in which the first stage threshold converts the raw data into binary detection/no detection decisions. This decision is repeated over multiple pulses or scans, and the individual detection decisions for a given resolution cell are combined using an "*M* out of *N*" rule as described in Sec. <u>6.4</u>. The PDF of the output of the first stage is binomial since the data are binary at that point, independent of the input PDF. Four variations on this idea are discussed in Barrett (1987). Two of them, the "double threshold detector" and the "rank order detector," use conventional cell-averaging or OS CFAR in the first stage to set a specific \overline{P}_{FA} and therefore require knowledge of the interference PDF to set the first stage threshold. These are therefore not truly distribution-free.

The "modified double threshold detector" replaces the deterministically computed first stage threshold with a feedback circuit that monitors \overline{P}_{FA} at the first stage output and adjusts the threshold to approximate the desired value. This technique requires large amounts of data to estimate the observed \overline{P}_{FA} but will work for any input distribution. In the *rank sum double quantizer*, the first stage does not actually threshold the data, but instead computes the rank of the

test cell compared to the reference cells and passes the rank number, instead of a binary detection/no detection decision, to the second stage. The second stage integrates this rank number over multiple pulses or scans, producing a random variable with a distribution that depends only on the correlation between the data samples from successive pulses or scans (Barrett, 1987). A two-parameter cell-averaging CFAR computation is applied to this variable to obtain the final detection decision.

A more modern approach based on rank order ideas is given in Sarma and Tufts (2001). Consider a set of *N* reference samples $\{y_i\}$ and the *k*th order statistic for this set, $y_{(k)}$. The PDF for $y_{(k)}$ was given in Eq. (6.163). The *coverage C* is defined as the probability that a reference sample y_i is greater than $y_{(k)}$. Note that $0 \le C \le 1$. Since the cumulative distribution function $P_{y_i}(y)$ is the probability that y_i is *less* than some value y, it follows that

$$C = 1 - P_{y_i}(y_{(k)}) \tag{6.180}$$

and the PDF of *C* becomes

$$p_{\rm C}(C) = k \binom{N}{k} (1-C)^{k-1} C^{N-k}$$
(6.181)

Furthermore, the expected value of C is

$$E\{C\} = \int_0^1 C \cdot p_C(C) dC = \frac{N+1-k}{N+1}$$
(6.182)

Now consider a system that uses the *k*th rank order statistic of the reference cell data as the threshold value for the CUT. This differs from a conventional OS CFAR, which computes a threshold as a multiple of the *k*th rank order statistic, where the multiplier α_{OS} is a function of the PDF of the interference data. If instead $y_{(k)}$ is itself the threshold, no multiplier α is required. Under these conditions, *C* will be the probability of false alarm. Equations (6.181) and (6.182) show that the PDF of *C*, and thus also its expected value $\overline{C} = \overline{P}_{FA}$, do not depend on the PDF of the raw data $\{y_i\}$. Thus, the detector is a distribution-free CFAR.

A limitation of this DF CFAR is that only certain values of \overline{P}_{FA} are achievable, and they depend on the number *N* of reference cells. \overline{P}_{FA} takes on the discrete set of values given by Eq. (6.182) as *k* varies from 1 to *N*. The minimum possible value is obtained with k = N and is

$$\overline{P}_{FA\min} = \frac{1}{N+1} \tag{6.183}$$

Equation (6.183) illustrates another limitation of this technique: small values of \overline{P}_{FAmin} require large values of N. More realistically, practical limitations on the reference window size N limit this method to relatively high values of \overline{P}_{FA} . For a given value of N, the design value of the probability of false alarm, say \overline{P}_{FAd} , must be chosen to be greater than or equal to \overline{P}_{FAmin} . Assuming this condition is satisfied, the rank order to be used as a threshold is the one that produces a value of \overline{P}_{FA} as close to \overline{P}_{FAmin} as possible without exceeding it. That rank is given by

$$k = \left[(N+1) \left(1 - \overline{P}_{FAd} \right) \right]$$
(6.184)

While the false alarm probability does not depend on the PDF of the interference, the detection probability does. For exponentially distributed interference and target (Swerling 1 target in complex WGN), the average probability of detection is (Sarma and Tufts, 2001)

$$\overline{P}_{D} = \prod_{i=0}^{k-1} \frac{N-i}{N-i+(1+\gamma)^{-1}}$$
(6.185)

As with the other CFAR detectors, Eqs. (6.184) and (6.185) can be used to determine the CFAR loss of the DF CFAR. The additional loss over a CA CFAR for the Swerling 1 case is typically less than about 0.4 dB.

6.6 System-Level Control of False Alarms

It has been seen in this chapter that achieving good detection performance (high \bar{P}_{D} , low \bar{P}_{FA}) requires a signal-to-interference ratio on the order of 15 dB or better at the point of detection. For a given target RCS, the SIR is determined in part by basic radar system design choices reflected in the radar range equation: transmitter power, antenna gain, operating frequency, and noise figure. Furthermore, the fundamental goal of many of the techniques of radar signal processing discussed in other chapters of this text is to improve the SIR before the point of detection. Examples include matched filtering, pulse compression, MTI, pulse Doppler processing, and space-time adaptive processing. Once the SIR has been maximized, the detector, whether fixed or adaptive threshold, sets the actual threshold value and thus the false alarm probability. The SIR then determines the detection probability.

In some cases the detection probability may still be lower than required. In this event, the threshold may be lowered, increasing \overline{P}_{D} but also increasing \overline{P}_{FA} . Additional techniques can then be applied at other stages of the overall system processing in order to reduce \overline{P}_{FA} back to an acceptable level. Several options are discussed in (Nathanson, 1991); their applicability depends on the particular system involved. If jamming is present, a sidelobe blanker or sidelobe canceller can be used to further improve the SIR before the detector (assuming more advanced techniques such as STAP have not already been applied). After the detector, a clutter map may be used in some systems to reject false alarms due to fixed clutter discretes or known radio frequency interference (RFI) sources. If valid targets can be expected to extend over more than one range, azimuth, or Doppler cell, apparent detections that occupy only a single cell can be rejected as false alarms after the detector, lowering the system \bar{P}_{FA} . Finally, an apparent target can be tracked to make sure it recurs over multiple scans; if it does not, it is rejected as a false alarm. If it does, its kinematics can be tracked over multiple scans. If the target track violates reasonable bounds on velocity and acceleration, it can be assumed to be a false alarm, quite possibly due to the presence of electronic countermeasures, and the track can be rejected. Thus, control of the overall system false alarm rate can be spread over virtually all stages of the system.

References

- Albersheim, W. J., "Closed-Form Approximation to Robertson's Detection Characteristics," *Proceedings of IEEE*, vol. 69, no. 7, p. 839, Jul. 1981.
- Barrett, C. R., Jr., "Adaptive Thresholding and Automatic Detection," Chap. 12 in J. L. Eaves and E. K. Reedy (eds.), *Principles of Modern Radar*. Van Nostrand Reinhold, New York, 1987.
- Blake, S., "OS-CFAR Theory for Multiple Targets and Nonuniform Clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-24, no. 6, pp. 785–790, Nov. 1988.
- Cantrell, P. E., and A. K. Ojha, "Comparison of Generalized *Q*-function Algorithms," *IEEE Transactions on Information Theory*, vol. IT-33, no. 4, pp. 591–596, Jul. 1987.
- Conte, E., and M. Lops, "Clutter-Map CFAR Detection for Range-Spread Targets in Non-Gaussian Clutter, Part I: System Design," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-33, no. 2, pp. 432–442, Apr. 1997.
- Di Vito, A., and G. Moretti, "Probability of False Alarm in CA-CFAR Device Downstream From Linear-law Detector," *Electronics Letters*, vol. 25, no. 5, pp. 1692–1693, Dec. 1989.
- DiFranco, J. V., and W. L. Rubin, Radar Detection. Artech House, Dedham,

MA, 1980.

- Finn, H. M., "A CFAR Design for a Window Spanning Two Clutter Fields," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 2, pp. 155–169, Mar. 1986.
- Gandhi, P. P., E. Cardona, and L. Baker, "CFAR Signal Detection in Nonhomogeneous Weibull Clutter and Interference," *Record of the IEEE International Radar Conference*, pp. 583–588, 1995.
- Gradshteyn, I. S., and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, A. Jeffrey (ed.). Academic Press, New York, 1980.
- Hansen, V. G., and H. R. Ward, "Detection Performance of the Cell-Averaging LOG/CFAR Receiver," *IEEE Transactions On Aerospace and Electronic Systems*, vol. AES-8, no. 5, pp. 648–652, Sep. 1972.
- Hansen, V. G., and J. H. Sawyers, "Detectability Loss Due to 'Greatest Of Selection in a Cell-Averaging CFAR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-16, no. 1, pp. 115–118, Jan. 1980.
- Hmam, H., "Approximating the SNR Value in Detection Problems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-39, no. 4, pp. 1446–1452, Oct. 2003.
- *IEEE Standard Radar Definitions*. IEEE Std. 686-2008, Institute of Electrical and Electronics Engineers, 2008.
- Johnson, D. H., and D. E. Dudgeon, *Array Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- Kay, S. M., Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory. Prentice Hall, Upper Saddle River, NJ, 1993.
- Kay, S. M., Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory. Prentice Hall, Upper Saddle River, NJ, 1998.
- Keel, B. M., "Constant False Alarm Rate Detectors", Chap. 16 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.
- Levanon, N., Radar Principles. Wiley, New York, 1988.
- Lops, M., "Hybrid Clutter-Map/L-CFAR Procedure for Clutter Rejection in Nonhomogeneous Environment," *IEE Proceedings of Radar, Sonar, and Navigation*, vol. 143, no. 4, pp. 239–245, Aug. 1996.
- Lops, M., and M. Orsini, "Scan-by-Scan Averaging CFAR," *IEE Proceedings*, Part F, vol. 136, no. 6, pp. 249–253, Dec. 1989.
- Meyer, D. P., and H. A. Mayer, *Radar Target Detection*. Academic Press, New York, 1973.
- Nathanson, F. E., (with J. P. Reilly and M. N. Cohen), *Radar Design Principles*, 2d ed. McGraw-Hill, New York, 1991.
- Nitzberg, R., "Clutter Map CFAR Analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 4, pp. 419–421, Jul. 1986.
- Novak, L. M., "Radar Target Detection and Map-Matching Algorithm Studies," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-16,

no. 5, pp. 620-625, Sep. 1980.

- Olver, F. W. J., et al., editors,, *NIST Handbook of Mathematical Functions*. National Institute of Standards and Technology, U.S. Dept. of Commerce, and Cambridge University Press, 2010.
- Pace, P. E., and I. L. Taylor, "False Alarm Analysis of the Envelope Detection GO-CFAR Processor," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-30, no. 3, pp. 848–864, Jul. 1994.

Raghavan, R. S., "Analysis of CA-CFAR Processors for Linear-Law Detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-28, no. 3, pp. 661–665, Jul. 1992.

Rifkin, R., "Analysis of CFAR Performance in Weibull Clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-30, no. 2, pp. 315–329, Apr. 1994.

Ritcey J. A., "Performance Analysis of the Censored Mean-Level Detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, no. 4, pp. 443–454, Jul. 1986.

Ritcey, J. A., and J. L. Hines, "Performance of Max-Mean-Level Detector with and without Censoring," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-25, no. 2, pp. 213–223, Mar. 1989.

Rohling, H., "Radar CFAR Thresholding in Clutter and Multiple Target Situations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-19, no. 4, pp. 608–620, Jul. 1983.

- Robertson, G. H., "Operating Characteristic for a Linear Detector of CW Signals in Narrow Band Gaussian Noise," *Bell System Technical Journal*, vol. 46, no. 4, pp. 755–774, Apr. 1967.
- Sarma, A., and D. W. Tufts, "Robust Adaptive Threshold for Control of False Alarms," *IEEE Signal Processing Letters*, vol. 8, no. 9, pp. 261–263, Sep. 2001.
- Schleher, D. C., "Harbor Surveillance Radar Detection Performance," *IEEE Journal of Oceanic Engineering*, vol. OE-2, no. 4, pp. 318–325, Oct. 1977.

Shnidman, D. A., "Binary Integration for Swerling Target Fluctuations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-34, no. 3, pp. 1043–1053, Jul. 1998.

Shnidman, D. A., "Determination of Required SNR Values," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-38, no. 3, pp. 1059–1064, Jul. 2002.

Shor, M., and N. Levanon, "Performance of Order Statistics CFAR," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-27, no. 2, pp. 214–224, Mar. 1991.

Skolnik, M. I., *Introduction to Radar Systems*, 3d ed. McGraw-Hill, New York, 2001.

Smith, M. E., and P. K. Varshney, "Intelligent CFAR Processor Based on Data

Variability," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-36, no. 3, pp. 837–847, Jul. 2000.

- Taylor, J. W., "Receivers," Chap. 3 in M. Skolnik (ed.), *Radar Handbook*, 2d ed. McGraw-Hill, New York, 1990.
- Tufts, D. W., and A. J. Cann, "On Albersheim's Detection Equation," *IEEE Transactions on Aerospace and* Electronic Systems, vol. AES-19, no. 4, pp. 643–646, Jul. 1983.
- Van Cao, T.-T., "A CFAR Thresholding Approach Based on Test Cell Statistics," *Proceedings of the 2004 IEEE Radar Conference*, pp. 349– 354, Apr. 26–29, 2004.
- Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear* Modulation *Theory*. Wiley, New York, 1968.
- Weiner, M. A., "Binary Integration of Fluctuating Targets," *IEEE Transactions* on Aerospace and Electronic Systems, vol. AES-27, no. 1, pp. 11–17, Jan. 1991.
- Weiss, M., "Analysis of Some Modified Cell-Averaging CFAR Processors in Multiple-Target Situations," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-18, no. 1, pp. 102–114, Jan. 1982.

Problems

- Consider a detection problem where under hypothesis H₀ the PDF of the signal *x* is p_x(x | H₀) = α ⋅ exp(-x/a), 0 ≤ x ≤ ∞, while under hypothesis H₁ the PDF of the signal *x* is p_x(x | H₁) = β ⋅ exp(-x/β), 0 ≤ x ≤ ∞, with β > α. What are the likelihood ratio and log likelihood ratio for this problem?
- 2. Consider Neyman-Pearson detection of a constant A = 0.5 in uniform (not Gaussian) white noise. Specifically, the PDF of the noise w[n] is

$$p_w(w) = \begin{cases} 1, & -0.5 < w < +0.5 \\ 0, & \text{otherwise} \end{cases}$$

The measured signal is x[n]. Under hypothesis H_0 (noise only), x[n] = w[n]. Under hypothesis H_1 , x[n] = A + w[n].

- a. On one graph, sketch carefully the PDF of x[n] under each hypothesis, $p_x(x | H_0)$ and $p_x(x | H_1)$. Label all significant values.
- b. What is the likelihood ratio $\Lambda(x)$ (not the log-likelihood ratio or the likelihood ratio test) for this problem? It may be necessary to express this in more than one region, i.e. " $\Lambda(x) =$ expression 1 for $\alpha < x < b$," etc.

c. Regardless of the LR found in part (b), suppose the rule H_0 is chosen as the detection test. Sketch the resulting curves for P_{FA} vs. T and P_D vs. T as T varies from -1.5 to +1.5. Label all important values.

 H_1 $x \ge T$

Problems 3 to 6 are a related group exploring how changes in the values of m, σ_w^2 , and T affect detection and false alarm performance in the real constant-in-WGN example.

- 3. Consider detection of a real-valued constant in zero-mean real-valued Gaussian noise. Let the noise variance $\sigma_w^2 = 2$, the number of samples N = 1, and the constant m = 4. What is the SNR χ for this case? Sketch the distributions $p(y | H_0)$ and $p(y | H_1)$; label appropriate numerical values on the axes. Write the likelihood ratio and log-likelihood ratio for this problem. Simplify the expressions.
- 4. Continuing with the same parameters given in the previous problem, suppose P_{FA} = 0.01 (1 percent) is required. What is the required value of the threshold *T*? What is the resulting value of P_D? Lookup tables or MATLAB[®] can be used to calculate the values of functions such as erf(·), erfc(·), erf⁻¹(·), or erfc⁻¹(·) that may be needed.
- 5. Suppose *m* in Prob. 3 is increased so as to double the SNR; $\sigma_w^2 = 2$ and N = 1 still. What is the new value of *m*? Sketch and label the distributions $p(y | H_0)$ and $p(y | H_1)$ with this new value of *m*. If the same threshold *T* found in Prob. 4 is retained, does P_{FA} change and, if so, what is the new value? If that same threshold *T* is retained, does P_D change and, if so, what is the new value?
- 6. Go back to the case of m = 4, but now reduce the noise variance to $\sigma_w^2 = 1$. N = 1 still. What is the SNR χ now? Sketch and label the distributions $p(y|H_0)$ and $p(y|H_1)$ with this value of m and σ_w^2 . If the threshold value used in Probs. 4 and 5 is still retained, what is the value of P_{FA} ? What is the value of P_D ?
- 7. Consider detection of a constant in complex Gaussian noise. Let the total noise variance $\sigma_w^2 = 2$, the number of samples N = 1, and the constant m = 4. What is the SNR *c* for this case? Suppose $P_{FA} = 0.01$ (1 percent) is required. What is the required value of the threshold *T*? What is the resulting value of P_D ? Lookup tables or MATLAB[®] can be used to calculate the values of functions such as $erf(\cdot)$, $erfc(\cdot)$, $erf^{-1}(\cdot)$, or $erfc^{-1}(\cdot)$ that may be needed.
- 8. Compute the threshold T and probability of detection P_D for the case of a constant in zero-mean complex Gaussian noise, but now with unknown
phase. Use m = 4, $\sigma_w^2 = 2$, N = 1, and $P_{FA} = 0.01$ again. It will be necessary to numerically evaluate the Marcum Q function Q_M . MATLAB[®] can be used with either the marcumq.m function available in the Signal Processing ToolboxTM or Communications Systems ToolboxTM or the marcum.m function available in the "MATLAB Supplements" area of the website <u>http://www.radarsp.com</u>.

- 9. Use Albersheim's equation to estimate the single-sample SNR χ_1 required to achieve $P_{FA} = 0.01$ and P_D equal to the same value obtained in the previous problem. Compare to the actual SNR used in Prob. 8.
- 10. Repeat Prob. 9 using Shnidman's equation in place of Albersheim's equation.
- 11. Rearrange Albersheim's equation to derive a set of equations for P_{FA} in terms of P_D , N, and single-pulse SNR in dB, χ_1 .
- 12. Use Shnidman's equation to estimate the single-pulse SNR χ_1 in dB required to achieve $P_{FA} = 10^{-8}$ and $P_D = 0.9$ when noncoherently integrating N = 10 samples. Do this for all four Swerling cases and for the nonfluctuating case. Figure 6.14 can be used to check the results.

Problems 13 to 16 are a related group comparing the effectiveness of coherent and noncoherent integration of measurements and showing how to compute noncoherent integration gain.

- 13. Coherently integrating *N* samples of signal-plus-noise produces an integration gain of *N* on a linear scale; that is, if the SNR of a single sample y_i is χ , the SNR of $z = \sum_{i=1}^{N} y_i$ is $N\chi$. In Probs. 13 through 16, Albersheim's equation will be used to investigate the relative efficiency of noncoherent integration for one example case. Throughout these problems, assume $P_D = 0.9$ and $P_{FA} = 10^{-6}$ are required, and that a linear detector is used. Start by assuming detection is to be based on a single sample, N = 1. Use Albersheim's equation to determine the SNR χ_1 in dB needed for this single sample to meet the specifications above.
- 14. Now consider noncoherent integration of 100 samples to achieve the same P_D and P_{FA} as in the previous problem. Each individual sample can then have a lower SNR. Use Albersheim's equation again to determine the SNR χ_{nc} of each pulse in dB needed to achieve the required detection performance.
- 15. Now consider coherent integration of 100 pulses. What is the SNR χ_c in dB required for each pulse such that the coherently integrated SNR will be equal to the value χ_1 found in Prob. 13?
- 16. Finally, the noncoherent integration gain is the ratio χ_1/χ_{nc} . Find α such that

 $\chi_1/x_{nc} = N^{\alpha}$. What would be the value of α for coherent integration? Which is more efficient (obtains more gain for the same number of samples integrated), coherent or noncoherent integration?

- 17. Consider 3-out-of-5 (M = 3, N = 5) binary integration. Determine the required values of the single-trial probabilities P_D and P_{FA} such that the cumulative probabilities are $P_{CFA} = 10^{-8}$ and $P_{CD} = 0.99$. A small-probability approximation can be used to solve for P_{FA} , but finding P_D will require some numerical trial-and-error; the estimate of P_D should be accurate to two decimal places. (*Hint:* The correct answer lies in the range $0.87 \le P_D \le 0.92$.)
- 18. A single noncoherently detected sample of a nonfluctuating target in complex Gaussian noise with power $\sigma_w^2 = 1$ is to be tested for the presence of a target. A square-law detector is used. Assuming the interference power level exactly, what ideal value of threshold *T* is required to obtain $P_{FA} = 10^{-4}$? If the SNR is $\chi = 10$ dB, what is P_D ? MATLAB® and one of the computer functions mentioned in Prob. 8 or their equivalent will be needed to evaluate the Marcum *Q* function Q_M .
- 19. Now assume that the interference level is not known a priori, so a cellaveraging CFAR is used to perform the detection test. Choose N = 30reference cells. What will be the threshold multiplier α such that the average false alarm probability \overline{P}_{FA} remains at 10⁻⁴? It turns out that if the SNR is $\chi = 10$ dB, the value of P_D using the ideal threshold in the previous problem is 0.616. Assuming the SNR is still $\chi = 10$ dB, what will be the average detection probability \overline{P}_D using the CA-CFAR?
- 20. Suppose the threshold in a standard (non-CFAR) threshold detector designed using the Neyman-Pearson approach is chosen to give $P_{FA} = 10^{-6}$. If the interference power level increases by 6 dB, what will be the new value of P_{FA} ?
- 21. Consider a detector designed to give an average false alarm probability of $\bar{P}_{FA} = 10^{-8}$. What is the SNR χ_{∞} in dB required to achieve $\bar{P}_D = 0.9$ when using the ideal Neyman-Pearson threshold? What SNR χ_N in dB is required when using a cell-averaging CFAR with N = 16 reference cells? What is the CFAR loss for this case, in dB?
- 22. Consider a CA-CFAR with a single interfering target (the "target masking" problem). Assume the SNR χ of the target of interest in the cell under test is 15 dB. What is the approximate "target masking loss" in dB if the SNR χ_i of the interferer is 10 dB? Repeat for $\chi_i = 15$ dB. Assume the number of averaging cells is N = 20. Compute the results numerically and show all work. Figure 6.23 can be used as an approximate check of the results.
- 23. Consider an order statistic CFAR in exponentially-distributed interference

with N = 20 and a threshold scale factor of $\alpha_{OS} = 10$. Compute the average false alarm probability P_{FA} for the order statistic k = 15, 16, ..., 20.

24. For the OS-CFAR in the previous problem with k = 15, how many outliers (e.g., other targets) could be tolerated in the lead and lag windows without serious degradation of the detection and false alarm performance?

 $\frac{3}{2}$ The exception occurs if points are added to or subtracted from 1 for which $p_Y(\mathbf{y}|H_0)$, $p_{\mathbf{y}}(\mathbf{y}|H_1)$, or both are zero. In that case the corresponding probability is unchanged.

 $\frac{4}{100}$ Some subtleties that can arise if the PDFs are noncontinuous are being ignored. See Johnson and Dudgeon (1993) for additional detail.

 $\frac{5}{2}$ A monotone decreasing operation would simply invert the sense of the threshold test.

⁶ All of the following development is fairly easily generalized for the case when m is negative or is of unknown sign. It will be seen later that radar detection generally involves working with the magnitude of the signal, thus it is sufficient to work with a positive value of m.

⁷ The definitions of Eqs. (6.18) and (6.19) are the same as those used in MATLABTM.

⁸ The erf⁻¹(·) function will often be used here, even when $erfc^{-1}(\cdot)$ gives a slightly more compact

expression because of the wider availability of $erf^{-1}(\cdot)$ functions than $erfc^{-1}(\cdot)$ in MATLAB[®] and similar computational software packages.

 $\frac{9}{2}$ Here *T* refers to receiver temperature, not the detection threshold. Which meaning of *T* is intended should be clear from context throughout this chapter.

 $\frac{10}{10}$ An alternative approach called the *generalized likelihood ratio test* (GLRT), in which the unknown parameter(s) are replaced by their maximum likelihood estimates, is discussed in many detection theory texts (e.g., Kay, 1998).

¹¹ For instance, the earlier discussion of unknown signal energy corresponds to choosing A = 1/E.

 $\frac{12}{12}$ The MATLAB® function gammainc is consistent with this definition.

 $\frac{13}{N-1}(x)$ is the modified Bessel function of the first kind and order N-1, not to be confused with the incomplete gamma function I(u, M).

¹⁴ Values of N > 100 are outside the range for which Albersheim's equation claims good accuracy. Nonetheless, the general trend of the proportionality of G_{nc} to \sqrt{N} does tend to hold for larger N and also for fluctuating target models.

 $\frac{15}{15}$ Unless otherwise stated, this same arrangement of lead, lag, and guard cells is used in all examples in this chapter.

 $\frac{16}{16}$ The base of the logarithm affects the specific offset needed to set the threshold but is otherwise unimportant. In Eq. (6.156) it is assumed that the log data are on a decibel scale.

¹⁷ This terminology is unfortunate because all CFARs are adaptive in that they estimate the detection threshold from the measured data. "Adaptive CFARs" are those for which one or more of the parameters of the estimation algorithm itself are varied depending on data characteristics.

 $[\]frac{1}{2}$ In some detection problems, a third hypothesis is allowed: "don't know." Most radar systems, however, force a choice between "target present" and "target absent" on each detection test.

² A fourth probability can be defined, that of choosing H_0 and thus declaring a target not present when in fact the test sample is due to interference only. This probability, equal to $1 - P_{FA}$, is not normally of direct interest.

CHAPTER 7

Measurements and Tracking

It has been seen in earlier chapters that resolution in range, angle, and Doppler shift is determined by the temporal and spatial bandwidth and the waveform duration of the radar; but what about the precision of location in these dimensions? To clarify the question, consider the notional output from the radar receiver for a given range bin as a radar system scans in angle past a single, isolated point target. Assume a high pulse repetition frequency relative to the antenna scan rate so that the angle samples are closely spaced. In the absence of noise, one would expect to measure an output voltage proportional to the two-way antenna voltage pattern, as illustrated in Fig. 7.1*a* for a sinc-squared two-way voltage pattern and a linear detector. The angular position of the target can be determined exactly simply by finding the angle at which the peak output voltage occurs. Thus, the target is located in angle to a precision much better than the resolution.



FIGURE 7.1 Received voltage from an angle scan of a single point target: (a) no noise, (b) 30 dB signal-to-noise ratio. The inset expands the area around the boresight angle.

Now add receiver noise to the problem. The receiver output will consist of the sum of the target echo weighted by the antenna pattern and the noise. The noise will cause the observed peak to occur at an angle other than the true target location, as seen in Fig. 7.1*b*; the actual peak in this realization (expanded in the inset) is at -0.033 Rayleigh widths. The larger the noise variance, the greater the likely deviation of the measurement from the noise-free case.

Because of the noise, the measured peak value is now a random variable (RV). If the peak measurement is repeated on noisy data many times, the PDF of that RV can be estimated. Figure 7.2 shows two histograms for the observed peak location when complex Gaussian noise at two values of peak SNR is added to the received signal prior to the detector. The black curves are zero-mean Gaussian PDFs with the same variance as the simulated data. In Fig. 7.2*b*, the SNR is 20 dB lower (a factor of $100 \times$ in power) than in 7.2*a*, resulting in a wider angle error distribution than in the higher SNR case. In this example, the variance of the distribution in Fig. 7.2*b* is 9.54 times that of the distribution in Fig. 7.2*a*. This factor is approximately the square root of the $100 \times$ change in SNR.



FIGURE 7.2 Histogram of angle error using peak power method: (*a*) 30 dB SNR at boresight, (*b*) 10 dB SNR at boresight. Note the wider spread of errors at lower SNR.

Measurement precision limitations due to additive noise are the primary focus in this chapter, but noise is not the only limitation on measurement accuracy. Others include sampling density, target reflectivity fluctuations, other interference sources such as clutter and jamming, and hardware limitations. For instance, the scanning antenna's received power would be measured only at discrete angles determined by the radar's PRF and the antenna scan rate rather than on the dense sampling grid used in the preceding example. In a search mode, there might be only one to two samples per beamwidth and the peak estimator used here would not be practical. The peak estimator is also a poor choice because it takes no advantage of multiple samples to average out the effects of the noise. More practical estimators will be considered shortly.

7.1 Estimators

7.1.1 Estimator Properties

Suppose that one has a vector of *N* measured data samples $\mathbf{x} = \{x_i, i = 0, ..., N - 1\}$ that depends on a deterministic but unknown parameter Θ . Θ might be the actual target angle, time delay (equivalent to range), or Doppler shift. An *estimator f*(\mathbf{x}) is an algorithm for computing an estimate $\hat{\Theta}$ of the actual value of Θ from the data \mathbf{x} :

$$\hat{\Theta} = f(\mathbf{x}) \tag{7.1}$$

If the data is noisy, $\hat{\Theta}$ will be an RV with its own PDF and moments. Because the quality of different estimators will vary it is reasonable to ask what properties characterize a good estimator?

Before answering that question, it is useful to define two important metrics of estimator quality, *accuracy* and *precision*. Figure 7.3 illustrates these two ideas heuristically with the example of target shooting. The hits on the left target are centered roughly on the bull's-eye (center of the target) but are widely dispersed. Because the average location is near the center of the target, this set of hits is said to exhibit good accuracy. However, because they are widely dispersed, they are said to exhibit poor precision. In the middle target, the hits are tightly clustered but the center of the cluster is far from the bull's-eye. These hits are precise but inaccurate. In the target on the right the hits are both tightly clustered on the bull's-eye: both precise and accurate.



FIGURE 7.3 Distinction between precision and accuracy: (a) hits on the target are accurate but not precise, (b) precise but not accurate, (c) both accurate and precise.

To make these ideas more formal, define the accuracy of an estimate $\hat{\Theta}$ of a parameter Θ as the *bias* of the estimator, that is, the expected value of the difference between the mean of the estimate and the actual parameter value:

accuracy =
$$B_{\hat{\Theta}}(\Theta) \equiv \mathbf{E}\{\hat{\Theta}\} - \Theta = \mathbf{E}(\hat{\Theta} - \Theta)$$
(7.2)

The precision is defined as the standard deviation of the estimate:

precision
$$\equiv \sigma_{\hat{\Theta}} = \mathbf{E} \left\{ \sqrt{\left(\hat{\Theta} - \overline{\hat{\Theta}}\right)^2} \right\}$$
(7.3)

(The overbar also indicates expected value.) In most cases it will be convenient to work with the variance of $\hat{\Theta}$ (square of precision), $\sigma_{\hat{\Theta}}^2$.

Three desirable properties of an estimator are that it be *unbiased*, *consistent*, and *minimum variance*. The first two of these are defined by the behaviors

$$\mathbf{E}\{\hat{\Theta}\} = \Theta \quad \text{(unbiased)}$$
$$\lim_{N \to \infty} \{\sigma_{\hat{\Theta}}^2\} \to 0 \quad \text{(consistent)}$$
(7.4)

The unbiased property states that the estimate matches the true parameter value "on average." The consistency property states that as more data is made available for the estimate, the precision improves and asymptotically approaches zero (perfect precision). The third property, minimum variance, expresses the goal that of all possible unbiased estimators, the one having the minimum variance (finest precision) should be selected. Such an estimator is called a *minimum variance unbiased* (MVU) estimator.

As an example of these properties, consider a simple estimation problem. The data \mathbf{x} is *N* independent observations of a real constant *A* in additive real stationary white Gaussian noise (WGN),

$$x[n] = A + w[n], \quad w[n] \sim N(0, \sigma_w^2) \implies x[n] \sim N(A, \sigma_w^2)$$
(7.5)

The goal is to estimate the constant *A* from the noisy data. In this example *A* is the parameter Θ on which the measured data $\mathbf{x} = [x[0], x[1], ..., x[N-1]]^T$ depends. An obvious estimator is the sample mean:

$$\hat{A} = f(\mathbf{x}) = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$
$$= \frac{1}{N} \sum_{n=0}^{N-1} (A + w[n]) = A + \frac{1}{N} \sum_{n=0}^{N-1} w[n]$$
(7.6)

Clearly the estimate is unbiased:

$$\mathbf{E}\{\hat{A}\} = \mathbf{E}\left\{A + \frac{1}{N}\sum_{i=0}^{N-1} w[n]\right\} = A + \frac{1}{N}\sum_{i=0}^{N-1} \mathbf{E}\{w[n]\}$$
$$= A$$
(7.7)

Consistency is tested by considering the variance of A:

$$\sigma_{\hat{A}}^{2} = \mathbf{E}\left\{\left(\hat{A} - \overline{\hat{A}}\right)^{2}\right\} = \mathbf{E}\left\{\left(\hat{A} - A\right)^{2}\right\} = \mathbf{E}\left\{\left(\frac{1}{N}\sum_{m=0}^{N-1}w[m]\right)\left(\frac{1}{N}\sum_{n=0}^{N-1}w[n]\right)\right\}$$
$$= \frac{1}{N^{2}}\mathbf{E}\left\{\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}w[m]w[n]\right\} = \frac{1}{N^{2}}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\mathbf{E}\left\{w[m]w[n]\right\}$$
(7.8)

Because the noise samples are independent, $\mathbf{E}\{w[m]w[n]\} = \mathbf{E}\{w[m]\} \cdot \mathbf{E}\{w[n]\}$. Because the noise is white the mean of w[n] is necessarily zero. Therefore, $\mathbf{E}\{w[m]w[n]\}$ equals zero whenever $m \neq n$ but equals σ_w^2 when m = n. The latter case occurs *N* times during the double summation. Equation (7.8) reduces to

$$\sigma_{\hat{A}}^{2} = \frac{1}{N^{2}} \left(N \sigma_{w}^{2} \right) = \frac{\sigma_{w}^{2}}{N} \implies \lim_{N \to \infty} \sigma_{\hat{A}}^{2} = 0$$
(7.9)

so the sample mean estimator is also consistent.

7.1.2 The Cramèr-Rao Lower Bound

Equation (7.9) established the variance of the sample mean estimator, but is it the estimator with the minimum possible variance? This question is answered by the famous *Cramèr-Rao Lower Bound* (CRLB). Denote the joint PDF of **x** given Θ as $p_x(\mathbf{x}|\Theta)$. The CRLB states that the variance of any unbiased estimate $\hat{\Theta}$ of Θ is lower bounded by

$$\sigma_{\hat{\Theta}}^{2} \ge \frac{1}{\mathbb{E}\left[\left\{\partial \ln\left[p_{\mathbf{x}}(\mathbf{x}|\Theta)\right]/\partial\Theta\right\}^{2}\right]}$$
(7.10)

An alternate form of the CRLB is also common. If $p_x(\mathbf{x}|\Theta)$ is twice differentiable and obeys some other mild regularity conditions, then it can be shown that

$$\sigma_{\hat{\Theta}}^{2} \ge \frac{-1}{\mathbb{E}\left[\partial^{2} \ln\left\{p_{\mathbf{x}}(\mathbf{x}|\Theta)\right\}/\partial\Theta^{2}\right]}$$
(7.11)

The choice between Eq. (7.10) or (7.11) is a matter of convenience. The CRLB is derived in <u>App. A</u>, which also details its extensions to data dependent on multiple parameters and to complex data and parameters. The CRLB can be generalized to biased estimators, but the unbiased case is sufficient here.

Because the sample mean estimator proved to be unbiased, the CRLB can be applied to evaluate its variance. Identifying $\Theta = A$, the required PDF $p_x(\mathbf{x}|\Theta)$ is

$$p_{\mathbf{x}}(\mathbf{x}|A) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma_{w}^{2}}} \exp\left[-(x[n] - A)^{2}/2\sigma_{w}^{2}\right]$$
(7.12)

Considering the CRLB in the form of Eq. (7.10), the partial derivative with respect to A of $\ln\{p_x(\mathbf{x}|A)\}$ is

$$\frac{\partial}{\partial A} \ln \left\{ p_{\mathbf{x}}(\mathbf{x}|A) \right\} = \frac{\partial}{\partial A} \left\{ -\frac{N}{2} \ln \left(2\pi \sigma_{w}^{2} \right) - \sum_{n=0}^{N-1} (x[n] - A)^{2} / 2\sigma_{w}^{2} \right\}$$
$$= 2 \sum_{n=0}^{N-1} (x[n] - A) / 2\sigma_{w}^{2} = \frac{1}{\sigma_{w}^{2}} \sum_{n=0}^{N-1} (x[n] - A)$$
$$= \frac{1}{\sigma_{w}^{2}} \sum_{n=0}^{N-1} w[n]$$
(7.13)

The expected value of the square of this quantity is

$$\mathbf{E}\left[\left(\frac{\partial}{\partial A}\ln\left\{p_{\mathbf{x}}(\mathbf{x}|A)\right\}\right)^{2}\right] = \mathbf{E}\left(\frac{1}{\sigma_{w}^{2}}\sum_{n=0}^{N-1}w[n]\right)^{2} = \frac{1}{\sigma_{w}^{4}}\mathbf{E}\left[\left(\sum_{m=0}^{N-1}w[m]\right)\left(\sum_{n=0}^{N-1}w[n]\right)\right]$$
$$= \frac{1}{\sigma_{w}^{4}}\mathbf{E}\left[\sum_{n=0}^{N-1}w^{2}[n]\right] = \frac{1}{\sigma_{w}^{4}}\left(N\sigma_{w}^{2}\right) = \frac{N}{\sigma_{w}^{2}}$$
(7.14)

The second line follows because the noise is white so that the expected value of the cross-products is zero. Finally, using Eq. (7.14) in Eq. (7.10) gives the CRLB for this problem:

$$\sigma_{\lambda}^2 \ge \frac{\sigma_w^2}{N} \tag{7.15}$$

which is exactly the variance of the sample mean estimator obtained in Eq. (7.9). The CRLB guarantees that the sample mean estimator can be used with the assurance that no other unbiased estimator will have a lower variance. An unbiased estimator that achieves the CRLB is called an *efficient* estimator.

The case of a signal in additive white Gaussian noise (AWGN) is important enough to warrant further attention. Assume the measurement vector \mathbf{x} is composed of *N* real-valued signal + noise samples:

$$x[n] = s[n;\Theta] + w[n], \quad n = 0, ..., N-1$$
(7.16)

where Θ is the parameter to be estimated and the variance of w[n] is σ_{w}^{2} . The following result is derived in <u>App. A</u> by applying <u>Eq. (7.11)</u> to the PDF for **x**:

$$\sigma_{\hat{\Theta}}^{2} \geq \frac{\sigma_{w}^{2}}{\sum_{n=0}^{N-1} \left(\frac{\partial s[n;\Theta]}{\partial \Theta}\right)^{2}} \quad \text{(real-valued signal in AWGN)}$$

(7.17)

Applying this to the constant-in-AWGN example, the parameter $\Theta = A$ and the signal $s[n; \Theta] = s[n;A] = A$ also, so $\partial s[n;\Theta]/\partial \Theta = \partial A/\partial A = 1$, again giving the result of Eq. (7.15).

If the signal and noise are complex-valued and the parameter to be estimated is real, the CRLB becomes (see <u>App. A</u>)

$$\sigma_{\Theta}^{2} \geq \frac{\sigma_{w}^{2}}{2\sum_{i=0}^{N-1} \left| \frac{\partial s[n;\Theta]}{\partial \Theta} \right|^{2}}$$
(complex signals in AWGN, real parameter)
(7.18)

The factor of two in the denominator of Eq. (7.18) is significant. It means that the real-valued CRLB is not a special case of the complex CRLB. Rather, the complex CRLB is smaller by a factor of two.

The case where the parameter to be estimated is complex can be handled by treating it as two real parameters, $\Theta = \Theta_R + j\Theta_I$. Equation (7.18) establishes the CRLB for each of the real and imaginary parts. It is easy to show that the variance of the complex parameter Θ is the sum of the variances of Θ_R and Θ_I (see Prob. 9). The CRLB for complex Θ is then the sum of the two component CRLBs and therefore is also given by Eq. (7.17) or (7.18).

7.1.3 The CRLB and Signal-to-Noise Ratio

A relationship between the CRLB and SNR can be found for the case of a signal in AWGN. Consider a complex sampled signal s[n] that is dependent on some real parameter Θ (not necessarily its amplitude) and write it as the product of its real-valued peak amplitude A and a normalized function s[n] having a peak magnitude of one,

$$s[n] = A \cdot \tilde{s}[n] \tag{7.19}$$

The *peak SNR* for the noisy signal x[n] = s[n] + w[n] is $\chi = A^2/\sigma_w^2$. Suppose the parameter of interest, Θ , is *not A*. Then

$$\left|\frac{\partial s[n]}{\partial \Theta}\right|^2 = A^2 \left|\frac{\partial \tilde{s}[n]}{\partial \Theta}\right|^2$$
(7.20)

and Eq. (7.18) becomes

$$\sigma_{\hat{\Theta}}^{2} \geq \frac{\sigma_{w}^{2}}{2A^{2}\sum_{n=0}^{N-1} \left|\frac{\partial \tilde{s}[n]}{\partial \Theta}\right|^{2}}$$
(7.21)

The quantity $\sum_{n=0}^{N-1} |\partial \tilde{s}[n] / \partial \Theta|^2$ depends on the signal shape \tilde{s} , but for a given signal it is some scalar *k* so that Eq. (7.21) is of the form:

$$\sigma_{\hat{\Theta}}^{2} \geq \begin{cases} \frac{1}{2k\chi} & \text{(complex-valued signals)} \\ \frac{1}{k\chi} & \text{(real-valued signals)} \end{cases}$$
(7.22)

The second line of Eq. (7.22) would result for the real-valued case beginning with Eq. (7.17). Equation (7.22) states that the CRLB is inversely proportional to SNR, and that the constant of proportionality depends on the rate of change of \vec{s} with respect to Θ . A waveform that changes more rapidly with Θ produces a smaller CRLB.

The case where the parameter of interest is the peak amplitude A must be handled separately. In this event $\Theta = A$, giving

$$\frac{\partial s[n]}{\partial A} = \frac{\partial (A \cdot \tilde{s}[n])}{\partial A} = \tilde{s}[n]$$
(7.23)

The CRLB is then

$$\sigma_{\hat{A}}^{2} \ge \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{N-1} \left|\tilde{s}[n]\right|^{2}} = \frac{\sigma_{w}^{2}}{2E_{\tilde{s}}} = \frac{1}{2\left(E_{\tilde{s}}/\sigma_{w}^{2}\right)}$$
(7.24)

where the sum term in the denominator of Eq. (7.24) has been recognized as the energy in \tilde{s} , resulting in a CRLB that is one-half the inverse of the energy SNR based on the normalized signal.¹ However, when estimating the amplitude, the variance relative to the actual amplitude is more likely to be of interest than the absolute variance. It is easy to see that the CRLB for the normalized amplitude estimation variance is the same as Eq. (7.24) but with the energy of the unnormalized signal *s*. The result is (including the equivalent real-valued result as well)

$$\sigma_{A/A}^{2} \geq \begin{cases} \frac{1}{2(E_{s}/\sigma_{w}^{2})} & \text{(complex-valued signals)} \\ \frac{1}{(E_{s}/\sigma_{w}^{2})} & \text{(real-valued signals)} \end{cases}$$
(7.25)

7.1.4 Maximum Likelihood Estimators

The CRLB establishes the minimum variance of an unbiased estimator. As discussed in App. A, it is sometimes but not always possible to find the minimum variance unbiased estimator. If the data can be modeled as a linear function of the parameters Θ with additive WGN, the MVU estimator can be found in a straightforward fashion, and is further guaranteed to be efficient. In other cases one may have to settle for a suboptimal estimator. Kay (1993) gives a thorough discussion of the hierarchy of techniques and achievable results in classical estimation.

Maximum likelihood estimation (MLE, also maximum likelihood estimator or estimate) is by far the most common approach to finding practical estimators, and for several good reasons:

- The MLE can be found for most problems, including many complicated ones, by a straightforward procedure.
- While it is not the MVU estimator in general, as the number of data samples $N \rightarrow \infty$ it becomes asymptotically unbiased and efficient. The PDF of $\hat{\Theta}$ also becomes Gaussian.
- If an efficient estimator in fact exists, the maximum likelihood procedure will produce it.
- The MLE of a function Φ = g(Θ) of Θ can be found by applying that same function to Θ̂, Φ̂ = g(Θ̂).
- Finally, various numerical methods exist for finding the MLE when a closed form expression cannot be obtained.

The MLE of Θ is that estimate $\hat{\Theta}$ that maximizes the likelihood function for the problem. The likelihood function is just the PDF of the data **x** given Θ , but viewed as a function of Θ with **x** fixed (see <u>App. A</u>):

$$\hat{\Theta} = \arg\max_{\Theta} \{\ell(\Theta|\mathbf{x})\} = \arg\max_{\Theta} \{p_{\mathbf{x}}(\mathbf{x}|\Theta)\}$$
(7.26)

The maximization is performed only over allowed values of Θ , e.g., $\Theta \ge 0$ if the parameter is the power of some signal. Note that maximizing a monotonically

increasing function of the likelihood is the same as maximizing the likelihood itself. When the data are i.i.d. and the noise is Gaussian it is often the case that the likelihood function is the product of N scalar PDFs, each containing exponential terms. Consequently, it is common to maximize the log-likelihood function because the algebra is greatly simplified in these cases:

$$\tilde{\Theta} = \underset{\Theta}{\arg\max\{\ln[\ell(\Theta|\mathbf{x})]\}} = \underset{\Theta}{\arg\max\{\ln[p_{\mathbf{x}}(\mathbf{x}|\Theta)]\}}$$
(7.27)

To illustrate the procedure, consider yet again the real constant in real additive WGN example with N samples of data. The PDF is an N-dimensional joint Gaussian, so the likelihood function is

$$\ell(A|\mathbf{x}) = \frac{1}{\left(2\pi\sigma_w^2\right)^{N/2}} \exp\left[-\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right]$$
(7.28)

The log-likelihood function is

$$\ln[\ell(A|\mathbf{x})] = -\frac{1}{2\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A)^2 - \frac{N}{2} \ln(2\pi\sigma_w^2)$$
(7.29)

Taking its derivative with respect to A and setting the result equal to zero to find the maximum gives

$$\frac{\partial}{\partial A} \left\{ \ln \left[\ell(A|\mathbf{x}) \right] \right\} = \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} (x[n] - A) = \frac{1}{\sigma_w^2} \left(-NA + \sum_{n=0}^{N-1} x[n] \right) = 0 \quad \Rightarrow$$
$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \tag{7.30}$$

The MLE for this problem is the sample mean estimator. It was seen earlier that this estimator is unbiased and achieves the CRLB. Thus, the MLE did produce the efficient estimator for this problem.

7.2 Range, Doppler, and Angle Estimators

7.2.1 Range Estimators

CRLB for Time Delay and Range Estimation

Consider a complex-valued continuous-time signal $x_i(t)$ at the input to a radar receiver that is the sum of an echo of a complex transmitted signal $s_i(t)$ and complex additive WGN $w_r(t)$ with a power spectral density (PSD) of $\sigma_{w_r}^2$ W/Hz. The echoed signal component is delayed by an unknown amount t_0 to form the received signal component $s_r(t - t_0) = \alpha s_i(t - t_0)$, where α is a complex scalar whose magnitude represents attenuation due to the radar range equation and whose phase represents the unknown $-4\pi R/\lambda$ radian phase shift. It is assumed that demodulation has already removed the carrier so that $x_r(t)$ is at baseband. The receiver input signal is therefore $x_r(t) = s_r(t - t_0) + w_r(t)$, as shown in Fig. 7.4. The goal is to estimate the unknown real parameter $\Theta = t_0$. The result can be scaled by c/2 to estimate the corresponding target range R_0 .



FIGURE 7.4 Receiver model for time delay and range estimation. See text for details.

The pulse $s_t(t)$, and therefore its scaled echo $s_r(t - t_0)$, is assumed to be τ seconds long and effectively bandlimited to β Hz. The radar receiver's frequency response is assumed to be a unit-gain bandpass characteristic over $\pm \beta/2$ Hz. Consequently, the pulse echo component of the output is essentially identical to the input pulse, while the total matched filter output noise power is $\sigma_{w_r}^2 \beta W$. This will be the variance of the output noise signal $w_o(t)$. The receiver output signal is therefore $x_o(t) = s_r(t - t_0) + w_o(t)$.

The receiver output is sampled at the Nyquist rate of β samples per second to produce the observed data $x[n] = s[n - n_0] + w[n]$, where $n_0 \approx t_0/T_s$ and $T_s = 1/\beta$ is the sampling interval. The maximum time delay of interest is *T*; this might be the radar's PRI, for example. It is assumed that $T > 2\tau$, the duration of a matched filter output for a single pulse at its input. $N = \lfloor T/T_s \rfloor$ is the total number of samples required to cover the maximum time interval of interest and $M = \lfloor \tau/T_s \rfloor$ is the number required to cover the signal duration τ . Let n_0 be the first sample that occurs within the received pulse so that $n_0T_s \approx t_0$. The discrete-time data is

$$x[n] = \begin{cases} w[n], & 0 \le n \le n_0 - 1\\ s_i(nT_s - t_0) + w[n] \equiv s[n - n_0] + w[n], & n_0 \le n \le n_0 + M - 1\\ w[n], & n_0 + M \le n \le N - 1 \end{cases}$$

The sampled noise will be white (see <u>App. A</u>). The PSD of the sampled noise and the variance of w[n] will both be $\sigma_w^2 = \sigma_{w_\sigma}^2 = \sigma_{w_r}^2 \beta$.

Before discussing a specific estimator, consider the CRLB for this problem. Applying the form for a complex signal in additive complex Gaussian noise in Eq. (7.18) gives

$$\sigma_{\hat{t}_{0}}^{2} \geq \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{N-1} \left| \frac{\partial s[n;t_{0}]}{\partial t_{0}} \right|^{2}} = \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{N-1} \left| \frac{\partial s(nT_{s}-t_{0})}{\partial t_{0}} \right|^{2}} = \frac{\sigma_{w}^{2}}{2\sum_{n=n_{0}}^{n_{0}+M-1} \left| \frac{\partial s(nT_{s}-t_{0})}{\partial t_{0}} \right|^{2}}$$
$$= \frac{\sigma_{w}^{2}}{2\sum_{n=n_{0}}^{n_{0}+M-1} \left| \frac{ds(t)}{dt} \right|_{t=nT_{s}-t_{0}}^{2}} = \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{M-1} \left| \frac{ds(t)}{dt} \right|_{t=nT_{s}-t_{0}}^{2}} \approx \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{M-1} \left| \frac{ds(t)}{dt} \right|_{t=nT_{s}-t_{0}}^{2}} \approx \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{M-1} \left| \frac{ds(t)}{dt} \right|_{t=nT_{s}-t_{0}}^{2}} \approx \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{M-1} \left| \frac{ds(t)}{dt} \right|_{t=nT_{s}}^{2}} \approx \frac{\sigma_{w}^{2}}{2\sum_{n=0}^{M-1} \left| \frac{ds(t)}{dt} \right|_{t=0}^{2}} \approx \frac{\sigma_{w}^{2}}$$

The last step used $n_0 Ts \approx t_0$. Recall that the integral of a function f(t) can be approximated by a sum using the rectangle rule, $\int_0^{\tau} f(t) dt \approx \sum_{n=0}^{\lfloor \tau/T_s \rfloor} f(nT_s) \cdot T_s$. Applying this to Eq. (7.32) and using $T_s = 1/\beta$ gives the CRLB for time delay estimation in complex signals as

$$\sigma_{\hat{t}_{0}}^{2} \geq \frac{\sigma_{w}^{2}}{\frac{2}{T_{s}} \int_{0}^{r} \left| \frac{ds(t)}{dt} \right|^{2} dt} = \frac{\sigma_{w}^{2} / \beta}{2 \int_{0}^{r} \left| \frac{ds(t)}{dt} \right|^{2} dt} = \frac{\sigma_{w_{r}}^{2}}{2 \int_{0}^{r} \left| \frac{ds(t)}{dt} \right|^{2} dt} \quad \text{(complex signals)}$$
(7.33)

Equation (7.33) can be put in a more useful form that reveals the role of the waveform bandwidth. The root-mean-square (RMS) bandwidth of s(t) is defined as

$$\beta_{\rm rms} \equiv \sqrt{\frac{\int_{-\infty}^{\infty} F^2 |S(F)|^2 dF}{\int_{-\infty}^{\infty} |S(F)|^2 dF}} \quad \text{Hz}$$
(7.34)

The derivative property of Fourier transforms states that the transform of ds(t)/dt is $j2\pi F \cdot S(F)$. Combining this with Parseval's theorem, the numerator can be replaced by $(1/4\pi^2) \int |ds(t)/dt| dt$. Therefore

$$\beta_{\rm rms}^2 = \frac{\frac{1}{4\pi^2} \int_0^\tau \left| \frac{ds(t)}{dt} \right|^2 dt}{\int_0^\tau |s(t)|^2 dt} = \frac{1}{4\pi^2 E_s} \int_0^\tau \left| \frac{ds(t)}{dt} \right|^2 dt$$
(7.35)

where the integral in the denominator is the energy E_s of s(t). Equation (7.33) can now be rewritten as

$$\sigma_{\hat{t}_0}^2 \ge \frac{\sigma_{w_r}^2}{8\pi^2 E_s \beta_{\rm rms}^2} = \frac{1}{8\pi^2 (E_s / \sigma_{w_r}^2) \beta_{\rm rms}^2} s^2$$
(complex signals)
$$\sigma_{\hat{K}_0}^2 \ge \frac{(\Delta R_{\rm rms})^2}{8\pi^2 (E_s / \sigma_w^2)} m^2$$
(7.36)

The first line of Eq. (7.36) shows that the lower bound on time delay estimation precision (square root of variance) becomes smaller when the mean-square signal bandwidth or the energy SNR E_s/σ_w^2 is increased. The second line scales the time delay variance by $(c/2)^2$ and defines $\Delta R_{\rm rms} = c/2\beta_{\rm rms}$, showing that the precision is proportional to RMS range resolution divided by the square root of the energy SNR.

The RMS bandwidth is not commonly used in radar signal processing. If it exists for a given waveform it is proportional to more common metrics such as the Rayleigh or 3-dB bandwidth. For example, both the Rayleigh and 3-dB bandwidths of a rectangular spectrum are simply the spectrum width β . It is easy to compute from Eq. (7.34) that $\beta_{\rm rms}$ for this spectrum is $\beta/\sqrt{12} \approx 0.29\beta$. The RMS bandwidth does not exist for a rectangular pulse because the defining integral Eq. (7.34) does not converge; see Prob. 23.

Practical Range Estimators

The above results establish the minimum variance of a time delay or range estimator, but they do not show how to actually obtain an estimate, let alone one that meets the bound. Two approaches that might come to mind based on earlier chapters are to estimate the time delay by detecting when the leading edge of the received pulse echo crosses some threshold, or finding the time at which the peak of the matched filter output occurs. The results of this chapter, however, suggest starting by seeking the MLE.

The signal model of Eq. (7.31) shows that the appropriate likelihood function is a zero-mean complex Gaussian for the noise-only samples, and a nonzero mean complex Gaussian for the signal-plus-noise samples. Because the

noise is i.i.d., the resulting joint likelihood function is

$$\ell(n_{0}|\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x}|n_{0}) = \left\{ \prod_{n=0}^{n_{0}-1} \frac{1}{\sqrt{2\pi\sigma_{w}^{2}}} \exp\left[-\frac{1}{2\sigma_{w}^{2}}|x[n]|^{2}\right] \right\}.$$

$$\left\{ \prod_{n=n_{0}}^{n_{0}+M-1} \frac{1}{\sqrt{2\pi\sigma_{w}^{2}}} \exp\left[-\frac{1}{2\sigma_{w}^{2}}|x[n]-s[n-n_{0}]|^{2}\right] \right\}.$$

$$\left\{ \prod_{n=n_{0}+M}^{N-1} \frac{1}{\sqrt{2\pi\sigma_{w}^{2}}} \exp\left[-\frac{1}{2\sigma_{w}^{2}}|x[n]|^{2}\right] \right\}$$

$$= \frac{1}{\left(2\pi\sigma_{w}^{2}\right)^{N/2}} \exp\left[-\frac{1}{2\sigma_{w}^{2}}\sum_{n=0}^{N-1}|x[n]|^{2}\right].$$

$$\left\{ \prod_{n=n_{0}}^{n_{0}+M-1} \exp\left[-\frac{1}{2\sigma_{w}^{2}}\left(-2\operatorname{Re}\{x[n]s^{*}[n-n_{0}]\}+|s[n-n_{0}]|^{2}\right)\right] \right\}$$
(7.37)

Because n_0 appears only in the product of exponentials in the last line of Eq. (7.37), maximizing $\ell(n_0 | \mathbf{x})$ with respect to n_0 is equivalent to minimizing the exponent of that term, which can be rewritten as

$$\exp\left[-\frac{1}{2\sigma_w^2}\sum_{n=n_0}^{n_0+M-1} \left(-2\operatorname{Re}\{x[n]s^*[n-n_0]\} + \left|s[n-n_0]\right|^2\right)\right]$$
(7.38)

With the change of variables $n' = n - n_0$ the sum involving $|s|^2$ is recognized as $\sum_{n'=0}^{M-1} |s[n']|^2 = E_s$, which does not actually depend on n_0 . Therefore, minimizing the exponent of Eq. (7.38) with respect to n_0 reduces to maximizing the quantity $\operatorname{Re}\left\{\sum_{n=n_0}^{n_0+M-1} x[n]s^*[n-n_0]\right\}$. Because $s[n - n_0]$ is zero outside of the summation interval, the summation is the same as the correlation $z[n_0] = \sum_{n=\infty}^{\infty} x[n]s^*[n-n_0]$, which is simply the output of the (noncausal) matched filter with impulse response $h[n] = s^*[-n]$ when the input is x[n]. The MLE estimate of n_0 becomes

$$\widehat{n_0} = \underset{n_0}{\operatorname{arg\,max}} \left\{ \operatorname{Re}\left[\sum_{n=-\infty}^{\infty} x[n]s^*[n-n_0]\right] \right\} = \underset{n_0}{\operatorname{arg\,max}} \{z[n_0]\}$$
(7.39)

Equation (7.39) states that the maximum likelihood estimate of time delay is obtained by passing the sampled data at the receiver bandpass output through a filter matched to the received waveform and then finding the sample at which

the real part of the output is maximum. To interpret this result further, recall that $x[n] = s[n - n_0] + w[n]$ and evaluate (7.39) at an estimated delay of $\overline{n_0}$:

$$\operatorname{Re}\left\{\sum_{n=-\infty}^{\infty} x[n]s^{*}[n-\widehat{n_{0}}]\right\} = \operatorname{Re}\left\{\left[\sum_{n=-\infty}^{\infty} s[n-n_{0}]s^{*}[n-\widehat{n_{0}}] + w[n]s^{*}[n-\widehat{n_{0}}]\right]\right\}$$
(7.40)

When n_0 equals the true delay n_0 this becomes

$$\operatorname{Re}\left\{\sum_{n=-\infty}^{\infty} x[n]s^{*}[n-\widehat{n_{0}}]\right\}\Big|_{\widehat{n_{0}}=n_{0}} = \operatorname{Re}\left\{\left[\sum_{n=-\infty}^{\infty} \left|s[n-n_{0}]\right|^{2} + w[n]s^{*}[n-n_{0}]\right]\right\}$$
$$= E_{s} + \operatorname{Re}\left\{\sum_{n=-\infty}^{\infty} w[n]s^{*}[n-n_{0}]\right\}$$
(7.41)

Equation (7.41) shows that at the true time delay, the output will consist of a real-valued component from the echoed signal, plus the real part of the filtered noise. This is consistent with the discussion of matched filters for waveforms in Chap. 4, where it was seen that the peak of the matched filter output occurred at the time delay of the target echo (increased by the delay T_M if a causal version of the impulse response is used) and that the SNR is maximized at that time. Thus, in the absence of noise the estimation rule of Eq. (7.39) will correctly identify the time delay n_0 . The presence of the noise can perturb the location of the peak real output, giving rise to the uncertainty in estimating delay quantified by the CRLB. Nonetheless, when the SNR is reasonably high it is likely that the matched filter output will be at its maximum or near-maximum at $n = n_0$ and will be primarily real-valued so that the time delay is estimated with good precision. When the SNR is not large, other effects dominate as will be seen shortly.

The analysis to this point assumes that the *received* pulse $s_r(t)$ is known and is used to define the matched filter. In practice, only the *transmitted* pulse $s_t(t)$ will be known, so the matched filter impulse response must be selected as $h[n]=s_t^*(-nT_s)=s[n]/\alpha$. The difference is the complex scale factor α . The magnitude of α is of no concern because it scales the signal and noise components equally and will not affect where the maximum occurs. However, the phase of α is the two-way propagation phase shift $-4\pi R/\lambda$ and must be considered uniform random over $[-\pi,\pi]$. It affects only the signal x[n] in Eq. (7.41). Consequently, it will rotate the signal component E_s of Eq. (7.41) to the unknown non-real value $E_s = E_s/\alpha$ so that the Re{ \cdot } operator is no longer appropriate. In practice, Eq. (7.39) is modified to produce the final time delay estimation rule:

$$\widehat{n_0} = \arg\max_{n_0} \left\{ \left\| \sum_{n=-\infty}^{\infty} x(n) s^*[n-n_0] \right\| \right\} = \arg\max_{n_0} \left\{ \left| (x*h)[n] \right| \right|_{n=n_0} \right\}$$
(7.42)

<u>Figure 7.5</u> modifies <u>Fig. 7.4</u> to illustrate this structure by adding the matched filter and maximization of the magnitude.



FIGURE 7.5 Maximum likelihood estimator time delay and range estimation.

It was shown in <u>Chap. 4</u> that the peak SNR at the output of a matched filter is the input signal energy divided by the input noise PSD, $\chi_{out} = E_s / \sigma_{w_i}^2$. Equation (7.36) can therefore be rewritten as

$$\sigma_{\hat{t}_0}^2 \ge \frac{1}{8\pi^2 \chi_{\text{out}} \beta_{\text{rms}}^2} \, \text{s}^2 \quad \text{(complex-valued signals)}$$
(7.43)

Equation (7.43) was obtained assuming that the signals involved were complexvalued, a case that includes real-valued signals in complex-valued noise. If both signal and noise are real-valued, then as seen in <u>App. A</u> the CRLB will be twice the value begun with in <u>Eq. (7.32)</u>. In this case <u>Eq. (7.43)</u> is replaced by

$$\sigma_{\hat{t}_0}^2 \ge \frac{1}{4\pi^2 \chi_{\text{out}} \beta_{\text{rms}}^2} \, \text{s}^2 \quad \text{(real-valued signals)}$$
(7.44)

The complex case is normally of interest in coherent radars.

An equivalent equation applies in discrete-time units. Converted to units of samples, Eq. (7.43) becomes

$$\sigma_{\widehat{n_0}}^2 \ge \frac{1}{8\pi^2 \chi_{\text{out}} \beta_{\text{rms}}^2 T_s^2} \text{ samples}^2 \quad \text{(complex-valued signals)}$$
(7.45)

Using a tilde to distinguish quantities in discrete-time units from those in continuous-time units, the sampled signal energy, noise PSD, matched filter output SNR, and RMS bandwidth are related to the corresponding analog quantities according to

$$\begin{split} E_{\rm s} &= \int_{-\infty}^{\infty} \left| s(t) \right|^2 dt \approx T_{\rm s} \sum_{n=-\infty}^{\infty} \left| s(nT_{\rm s}) \right|^2 = T_{\rm s} \sum_{n=-\infty}^{\infty} \left| s[n] \right|^2 = T_{\rm s} \tilde{E}_{\rm s} \implies \tilde{E}_{\rm s} = \frac{1}{T_{\rm s}} E_{\rm s} \\ \sigma_{\tilde{w}}^2 &= \frac{1}{T_{\rm s}} \sigma_{w_{\rm r}}^2 \quad \text{(effect of sampling on spectrum)} \\ &\Rightarrow \quad \tilde{\chi}_{\rm out} = \tilde{E}_{\rm s} / \sigma_{\tilde{w}}^2 = E_{\rm s} / \sigma_{w_{\rm r}}^2 = \chi_{\rm out} \\ \tilde{\beta}_{\rm rms} = T_{\rm s} \beta_{\rm rms} \quad \text{(conversion to normalized frequency in cycles per sample)} \end{split}$$

Applying these conversions to Eq. (7.45) and considering the equivalent real-valued case gives

$$\sigma_{\hat{n}_0}^2 \ge \frac{1}{4\pi^2 \tilde{\chi}_{out} \tilde{\beta}_{rms}^2} \text{ samples}^2 \quad \text{(real-valued signals)}$$

$$\sigma_{\hat{n}_0}^2 \ge \frac{1}{8\pi^2 \tilde{\chi}_{out} \tilde{\beta}_{rms}^2} \text{ samples}^2 \quad \text{(complex-valued signals)} \quad (7.47)$$

(7.46)

which is exactly the same form as Eqs. (7.43) and (7.44).

As an example, consider a complex-valued linear FM pulse with a timebandwidth product of $\beta \tau = 100$. Because the spectrum will be approximately rectangular, $\beta_{\rm rms} = \beta/\sqrt{12}$ as noted earlier. The signal is sampled at a very high rate of 20× the Nyquist bandwidth β , which is necessary if it is to be possible to measure peak location errors that are a small fraction of the pulse width. The data record has a total length of $T = 2\tau$ seconds (the duration of the matched filter output). The gray circles in Fig. 7.6 plot the precision (standard deviation) of the error in the observed peak location,² normalized to the Rayleigh time resolution of $\Delta t = 1/\beta$ seconds, versus the matched filter output SNR. The precision is estimated from 1000 random trials. At an SNR of 20 dB the precision is about 4 percent of the time resolution $\Delta t = 1/\beta$ seconds.



FIGURE 7.6 Behavior of simulated time delay maximum likelihood estimation precision in comparison to various bounds for a complex LFM pulse in complex noise. See text for details.

Also shown in the figure is the CRLB computed from Eq. (7.43). For SNRs of 16 dB or larger, the observed precision follows the CRLB very closely,³ confirming that the matched filter output peak is in fact a minimum variance estimator for large SNR. However, below 16 dB the estimate is much worse than suggested by the CRLB. Heuristically, reasonable performance of the MLE for time delay requires that the matched filter output peak be well above the filtered noise level. The value of the peak SNR at the output peak location is $E_s/\sigma_{w_i}^2$. High values of $E_s/\sigma_{w_i}^2$, say 20 dB or more, ensure that the peak of the signal component of the matched filter output is well above the filtered noise peaks so that the peak output, which is the MLE of delay, is in fact at the location of the signal peak or very close to it with high probability. For low values of $E_s/\sigma_{w_i}^2$, say 0 dB or less, the signal + noise peak is no larger than the noise fluctuations, so the peak output is likely to be anywhere in the complete output signal interval of (0, N - 1) samples or (0, T) seconds with approximately equal probability. In this event the matched filter output peak conveys no information. The upper bound on delay estimation variance becomes simply the a priori (low SNR) bound of $N^2/12$ samples squared, equivalent to $N^2T_s^2/12 = T^2/12 s^2$. This result applies to both real and complex signals. For the example of Fig. 7.6 where the record length T is only $2\tau = 200\Delta t$, the precision $\sigma_{\hat{h}}$ is about 58× the LFM time resolution or 58 percent of the pulse length since

 $\beta \tau = 100$. For a record length of $100\tau = 10,000\Delta t$ the "precision" would be bounded at 2890× the resolution! The low-SNR error can be limited if a tracking gate or other a priori knowledge is available to limit the range of possible time delays.

An early attempt to combine the CRLB performance bound for high SNR with the a priori upper bound for low SNR to obtain a more complete description of estimator performance was the "correlator performance estimate" described in Scarbrough et al. (1983) and attributed to Ianniello (1982). A number of other estimator bounds have been derived in the literature which also capture this behavior, extending the CRLB. The Ziv-Zakai bound (ZZB), applied to time delay estimation in the particular form given in Bell et al. (1997) and with the definitions used here for SNR and the $erfc(\cdot)$ function, is given by

$$\sigma_{\hat{t_0}}^2 \ge \frac{T^2}{12} \operatorname{erfc}\left(\sqrt{\frac{\chi_{\text{out}}}{2}}\right) + \frac{1}{8\pi^2 \chi_{\text{out}} \beta_{\text{rms}}^2} \Gamma\left(\frac{3}{2}, \frac{\chi_{\text{out}}}{2\sqrt{2}}\right) \quad \text{(complex signals)}$$

$$\sigma_{\hat{t_0}}^2 \ge \frac{T^2}{12} \operatorname{erfc}\left(\sqrt{\frac{\chi_{\text{out}}}{4}}\right) + \frac{1}{4\pi^2 \chi_{\text{out}} \beta_{\text{rms}}^2} \Gamma\left(\frac{3}{2}, \frac{\chi_{\text{out}}}{4\sqrt{2}}\right) \quad \text{(real signals)} \quad (7.48)$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function in the form defined in <u>Chap. 6</u> and $\Gamma(\cdot, \cdot)$ is the incomplete gamma function.⁴ The ZZB is also shown in <u>Fig.</u> 7.4. It matches the CRLB at high SNR and the a priori upper bound at low SNR and thus provides a better representation of achievable precision than the CRLB alone. The Weiss-Weinstein analysis provides another set of bounds that incorporate the a priori and Cramèr-Rao bounds and also provide tighter bounds in the transition regions between them (Weiss, 1986).

Comparison of Eq. (7.48) to Eqs. (7.43) and (7.44) and to the discussion above of the a priori bound shows that Eq. (7.48) can be rewritten as

$$\sigma_{\hat{t_0}}^2 \ge APB \cdot \operatorname{erfc}\left(\sqrt{\frac{\chi_{out}}{2}}\right) + CRLB \cdot \Gamma\left(\frac{3}{2}, \frac{\chi_{out}}{2\sqrt{2}}\right) \quad \text{(complex signals)}$$

$$\sigma_{\hat{t_0}}^2 \ge APB \cdot \operatorname{erfc}\left(\sqrt{\frac{\chi_{out}}{4}}\right) + CRLB \cdot \Gamma\left(\frac{3}{2}, \frac{\chi_{out}}{4\sqrt{2}}\right) \quad \text{(real signals)}$$
(7.49)

where *APB* and *CRLB* are the applicable a priori and Cramèr-Rao bounds, respectively. The erfc(·) and $\Gamma(\cdot, \cdot)$ functions provide the weightings that transition between these two bounds. This observation will be used to quickly write the ZZB for estimating sinusoid parameters in the next section.

Figure 7.7 compares the CRLB and actual delay estimation precision for two different waveforms. The gray curve and bound are for a complex-valued

LFM pulse with $\beta \tau = 100$ as considered above. The black curve and bound are for a real-valued trapezoidal pulse with an arbitrary but fixed complex phase shift and 20 percent rise and fall times (see Fig. 7.9*a*) of the same length. The RMS bandwidth of the LFM waveform is then 49.1 times that of the trapezoidal pulse. The noise is complex in both cases. In this case, both are normalized to the pulse length, making the normalized variance $100 \times$ smaller for the LFM pulse than in Fig. 7.6, but allowing direct comparison of the two waveforms. The ZZBs are not shown to avoid cluttering the graph. As expected, the precision achieved with the LFM for a given SNR in the high-SNR region exceeds that of the trapezoidal pulse by about 50× due to the greater RMS bandwidth.



FIGURE 7.7 Comparison of maximum likelihood delay estimation error for a real trapezoidal pulse and a complex LFM pulse of the same length in complex noise. The time-bandwidth product of the LFM pulse is 100.

Another limiting factor at high SNR that has not yet been considered is the quantization of the time delay axis. In the SNR regime where it is effective, the MLE locates the target time delay to sample n_0 , equivalent to time n_0T_s . Even if n_0 is the correct location of the highest amplitude sample of the signal component at the filter output, the implied time estimate n_0T_s can differ from the actual peak location t_0 by up to $\pm T_s/2$. Modeling this difference as an independent uniformly distributed error source establishes a lower sampling

bound (SB) of 1/12 squared samples or $T_s^2/12 s^2$ on the error variance. This will generally be higher than the CRLB unless the data is highly oversampled or the estimator is an analog (continuous-time) implementation. Suppose the matched filter output is oversampled by a factor k_{os} compared to the Nyquist rate so that $T_s = 1/k_{os}\beta$. The SB on time delay estimation variance for real or complex signals is then

$$\sigma_{\hat{t}_0}^2 \ge \frac{1}{12k_{os}^2\beta^2} \, \mathrm{s}^2 \tag{7.50}$$

Equating (7.50) with Eq. (7.43) or (7.44) shows the degree of oversampling needed to achieve the CRLB is $k_{os} = \pi(\beta_{ms}/\beta)\sqrt{\chi_{out}/3}$ in the real case and $\pi(\beta_{ms}/\beta)\sqrt{2\chi_{out}/3}$ in the complex case. For a complex-valued LFM waveform the required oversampling factor is about 7.4 at an SNR of 20 dB and 23.4 at 30 dB.

Figure 7.8 illustrates this effect for the same complex LFM pulse, MLE, and normalization to the time resolution used for Fig. 7.6. The matched filter output is oversampled by a factor $k_{os} = 2.5$ so that the sampling bound on precision is $0.116/\beta = 0.116T_s$. The analysis above shows (and the figure confirms) that the SB dominates the CRLB for $\chi_{out} > 10.1$ dB, but as can also be seen from the figure, in this example the SB does not dominate the ZZB until $\chi_{out} > 14$ dB. For values of SNR in this regime the gray MLE curve labeled "no interpolation" hews closely to the SB.



FIGURE 7.8 Effect of sampling rate limitations and interpolation techniques on high-SNR time delay estimation error. See text for details.

Time delay measurement accuracy could be improved by oversampling to a greater degree, but the rate required to achieve the CRLB at SNRs greater than about 20 dB can be quite large, as seen above. A common alternative approach is to interpolate between the samples in the vicinity of a measured peak to refine the estimated peak location, as discussed in each of <u>Chaps. 3–5</u> for reducing straddle errors. For the LFM waveform, the peak of the magnitude of the signal component at the matched filter output closely approximates a sinc function, which in turn can be approximated near its peak at $t = t_0$ by a parabola. Denote the magnitude of the matched filter output as y(t) with a peak occurring at time t_0 . The largest sample actually measured is at sample n_0 corresponding to time n_0T_s . Denote the value y at that time as $y[n_0] = y_0$, and the value of the two adjacent samples as $y[n_0 - 1] = y_{-1}$ and $y[n_0 + 1] = y_{+1}$. The actual peak is at location $t_0 = (n_0 + \Delta n)T_s$, where Δn is a fractional sample value between -1 and +1. It is straightforward but tedious to find the coefficients of a quadratic equation that passes through the three data points y_{-1} , y_0 , and y_{+1} corresponding to times $(n_0 - 1)T_s$, n_0T_s , and $(n_0 + 1)T_s$. This equation can then be differentiated with respect to Δn and set to zero to find the value of Δn at which the peak occurs. The resulting estimate of the time delay t_0 is

$$\Delta n = \frac{1}{2} \cdot \frac{y_{+1} - y_{-1}}{2y_0 - y_{-1} - y_{+1}}$$
$$\hat{t_0} = (n_0 + \Delta n)T_s$$
(7.51)

The curve marked "quadratic interpolation" in Fig. 7.8 shows the result of applying this technique to the coarsely sampled data that produced the "sampling bound" curve. The interpolation provides precision close to the CRLB up to at least 30 dB with only $2.5 \times$ oversampling.

Other interpolation techniques can be used. One example is the centroid of the apparent peak and its two nearest neighbors, which is given by

$$n_{0} + \Delta n = \frac{\sum_{n=n_{0}-1}^{n_{0}+1} n \cdot y[n]}{\sum_{n=n_{0}-1}^{n_{0}+1} y[n]} = \frac{(n_{0}-1)y_{-1} + n_{0}y_{0} + (n_{0}+1)y_{+1}}{y_{-1} + y_{0} + y_{+1}}$$
(7.52)

The result of centroid interpolation is also shown in <u>Fig. 7.7</u>. While it improves the precision below the sampling bound, it does not approach the CRLB. Still another is the two-sample estimator, which uses only the peak sample and the larger of its two nearest neighbors (Macleod, 1998):

$$n_0 + \Delta n = n_0 + \operatorname{sgn}(y_{+1} - y_{-1}) \left[\frac{\max(y_{+1}, y_{-1})}{y_0 + \max(y_{+1}, y_{-1})} \right]$$
(7.53)

where $sgn(\cdot)$ is the sign or signum function. The two-sample estimator is surprisingly accurate on noise-free data but is ineffective on noisy data, as seen in Fig. 7.8.

There are many additional issues to be considered in using peak interpolators. For example, parabolic interpolator coefficients should be modified to obtain the best performance if the data is windowed for range sidelobe control (which alters the peak shape) or if interpolation is applied to complex data instead of real data (Agrež, 2002; Lyons, 2011; Jacobsen and Kootsookos, 2007). In addition, both the quality of interpolation generally and the relative effectiveness of various interpolators varies widely with oversampling rate. Another factor is that these techniques are designed for isolated peaks not contaminated by other nearby peaks. Also, most interpolators are nonlinear and introduce bias (reduced accuracy) in the estimate, although the bias is generally small in the high-SNR region where interpolation is useful. Taken as a whole, interpolation procedures can be effective in improving precision but must be carefully selected and their details optimized for each particular data acquisition protocol.

Another common time delay estimator frequently used with approximately rectangular pulses is a leading edge threshold detector. This is a noncoherent technique that operates on the magnitude of the receiver output and is suitable only for high-SNR conditions. Consider a rectangular pulse filtered through a receiver filter of nominal bandwidth β Hz. The filter output can be approximated by a trapezoidal pulse with a rise time of $t_r \approx 1/\beta$ seconds, as shown in Fig. 7.9*a*. In a high-SNR environment, time delay could be estimated by measuring the time at which the received noisy pulse first crosses an amplitude threshold such as 50 percent of the peak amplitude. The addition of the filtered noise to the filtered pulse will perturb the time at which the threshold crossing occurs.



FIGURE 7.9 Leading edge detection for a trapezoidal pulse. (*a*) Trapezoidal pulse with rise time $t_r = 20\%$ of pulse length τ ; threshold is set at 50 percent of peak noise-free pulse amplitude. (*b*) Close up of threshold crossing region for noise-free and noisy pulse.

Consider the behavior of the noise-free and noisy output pulses of <u>Fig.</u> <u>7.9a</u>. The correlation interval of the noise will approximately equal the pulse

rise time t_r , which is expressed as a fraction α of the total pulse length, $t_r = \alpha \tau$; in this example, $\alpha = 0.2$. This shape is the result of convolving a square pulse of original duration $(1 - 2\alpha)\tau$ with a filter having a rectangular impulse response of width $\alpha \tau$ seconds, and therefore a bandwidth of approximately $1/\alpha \tau$ Hz.

Suppose the threshold crossing time in the absence of noise is t_0 . As suggested in the expanded view of the region around the amplitude threshold (here set at 50 percent of peak) in Fig. 7.9b, the slope of the noisy pulse is similar to that of the noise-free pulse so that the change in threshold crossing time, δt , is related to the noise amplitude at t_0 approximately according to

$$\frac{|n(t_0)|}{\delta t} = \frac{A}{t_r} = \frac{A}{\alpha \tau} \implies \delta t = \frac{\alpha \tau}{A} |n(t_0)|$$
(7.54)

If the bias in the estimate can be assumed small, the mean-square error in the time delay estimate, $E\{(\delta t)^2\}$, can be taken as an estimate of the variance of the delay error:

$$\sigma_{\delta t}^{2} \approx \frac{\alpha^{2} \tau^{2}}{A^{2}} \sigma_{w}^{2} = \frac{\alpha^{2} \tau^{2}}{\chi_{\text{out}}}$$
(7.55)

Equation (7.55) is referred to as the *mean-square error* (MSE) *bound*. The SNR χ_{out} is the peak output voltage SNR A^2/σ_w^2 , more appropriate than the waveform energy SNR since a matched filter is not used here.

It is interesting to compare the MSE bound to the general complex signal CRLB for time delay estimation of Eq. (7.43). It can be shown (see Prob. 22) that the RMS bandwidth of the trapezoidal pulse is $1/\pi\tau\sqrt{2\alpha(1-4\alpha/3)}$. It follows that the ratio of the complex CRLB precision to the MSE bound a factor of $\sqrt{(1-4\alpha/3)/4\alpha}$, which varies from 1.47 to 0.41 as α varies from 0.1 (nearly square pulse) to its maximum of 0.5 (triangular pulse). Thus in some cases, the CRLB is *larger* than the MSE bound! (For the value $\alpha = 0.2$ used in these examples, the two are nearly equal, with the MSE bound slightly smaller at 96 percent of the CRLB.) Since the CRLB is the minimum variance of any unbiased estimator, the leading edge estimator must be biased. This can be confirmed by considering the exact PDF of the threshold crossing time. In the notation used here, the MSE and bias in estimating t_0 can be shown to be (Bar-David and Anaton, 1981)

$$E\{(\delta t)^2\} = \frac{\alpha^2 \tau^2}{\chi_{out}}, \qquad E\{\delta t\} = -\frac{\alpha \tau}{\chi_{out}}$$
(7.56)

The MSE matches the earlier heuristic MSE bound of Eq. (7.55). Equation (7.56) confirms that the estimate is biased and shows that the bias asymptotically approaches zero at high SNR.

Figure 7.10 shows the precision, normalized to pulse length τ , for a simulated leading edge estimator with $\alpha = 0.1$ and the threshold at 50 percent of the noise-free peak amplitude. The total data record length was 3τ seconds. Also shown are the a priori bound, the MSE bound of Eq. (7.55), and the CRLB. The leading edge estimate approaches the MSE bound for SNRs of 20 dB or more. The MSE bound is below the CRLB for small α as discussed above. Also shown again in the solid gray line is the MLE for this problem, which approaches the CRLB but does not achieve the MSE bound.



FIGURE 7.10 Time delay estimation error for leading-edge detection of a trapezoidal pulse. Circles indicate leading edge estimate. The solid gray line is the maximum likelihood estimate. See text for details.

One might expect that at low SNR neither the CRLB or MSE bound would apply because the a priori bound would become dominant. Figure 7.10 shows that the a priori bound does not control the low-SNR behavior of the leadingedge estimator. As the SNR decreases below 20 dB, the variance of the leading edge estimate starts to rise as in earlier examples, but then decreases rapidly. This behavior occurs because at low SNR the noise alone is likely to cross the amplitude threshold in the first few moments of the data record, regardless of the actual pulse position. The "arrival time" is then consistently reported as very near to beginning of the data. While this estimate is badly biased, it varies little so the variance becomes small. This behavior can be avoided if the approximate position of the echo is known from a tracking loop so that such anomalous results can be ignored. Changing the level of the threshold alters somewhat the SNR at which the transition to CRLB-limited performance occurs. A higher threshold moves the "hump" in the data to the left by a few decibels, for instance. It is clear, however, that this estimator is only suitable for high-SNR signals.

Another common class of time delay or range estimators are various *split-gate* or *early-late gate* techniques. These attempt to find an estimated delay such that the energy in a finite window to either side of the estimated delay is approximately equal. As an example, for a rectangular pulse the matched filter output is a triangle. If the estimated delay coincides with the center of the triangle, then the sum of the noisy triangle voltages (or their squares) over one pulse length to either side is likely to be nearly equal. An analysis of various implementations of this idea is given in Peebles (1998).

7.2.2 Doppler Signal Estimators

CRLB for Sinusoid Parameters

A general uniformly sampled complex sinusoid s[n] in additive complex WGN w[n] of power σ_w^2 , using normalized cyclical frequency units is of the form

$$\begin{aligned} x[n] &= s[n] + w[n] \\ &= A \exp[j(2\pi f_0 n + \phi)] + w[n] \\ &= \tilde{A} \exp(j(2\pi f_0 n)) + w[n], \quad 0 \le n \le N - 1 \end{aligned}$$
(7.57)

where $\overline{A} = A \exp(j\phi)$ is the complex amplitude of the sinusoid. This signal is a good model for the slow-time signal produced by a target with a constant radial velocity (Doppler shift). Note that the SNR for this problem is $\chi = A^2/\sigma_w^2$.

There are three simultaneous parameters to be estimated: A, f_0 , and ϕ . Consequently, finding the CRLB for each requires computing the Fisher information matrix in the form of Eq. (A.84) in App. A. The (i, j)th element of the matrix is

$$[\mathbf{I}(\boldsymbol{\Theta})]_{ij} = \frac{2}{\sigma_w^2} \operatorname{Re} \left\{ \sum_{n=0}^{N-1} \left[\frac{\partial s[n; \boldsymbol{\Theta}]}{\partial \boldsymbol{\Theta}_i} \right]^* \left[\frac{\partial s[n; \boldsymbol{\Theta}]}{\partial \boldsymbol{\Theta}_j} \right] \right\}$$
(7.58)

where Θ is the three-element parameter vector $[A, f_0, \phi]^T$. The derivatives

needed are easily obtained:

$$\frac{\partial s[n;\Theta]}{\partial A} = \frac{\partial}{\partial A} \{A \exp[j(2\pi f_0 n + \phi)]\} = \exp[j(2\pi f_0 n + \phi)]$$

$$\frac{\partial s[n;\Theta]}{\partial f_0} = j2\pi nA \exp[j(2\pi f_0 n + \phi)]$$

$$\frac{\partial s[n;\Theta]}{\partial \phi} = jA \exp[j(2\pi f_0 n + \phi)]$$
(7.59)

The (1,1) element of $I(\Theta)$ is, from Eq. (7.58),

$$\begin{aligned} \left[\mathbf{I}(\Theta)\right]_{11} &= \frac{2}{\sigma_w^2} \operatorname{Re}\left\{\sum_{n=0}^{N-1} \left[\frac{\partial s[n;\Theta]}{\partial A}\right]^* \left[\frac{\partial s[n;\Theta]}{\partial A}\right]\right\} \\ &= \frac{2}{\sigma_w^2} \sum_{n=0}^{N-1} \left|\frac{\partial s[n;\Theta]}{\partial A}\right|^2 = \frac{2}{\sigma_w^2} \sum_{n=0}^{N-1} \left|e^{j(2\pi f_0 n + \phi)}\right|^2 = \frac{2N}{\sigma_w^2} \end{aligned} \tag{7.60}$$

It is easy to see that the (1, 2), (1, 3), (2, 1), and (3, 1) elements are all zero because of the Re $\{\cdot\}$ operator in Eq. (7.58) and the fact that the derivatives with respect to f_0 and ϕ are purely imaginary. The remaining elements are easily found from the derivatives above, giving the complete matrix as

$$\begin{bmatrix} \mathbf{I}(\Theta) \end{bmatrix}_{i,j} = \frac{2}{\sigma_w^2} \begin{bmatrix} N & 0 & 0 \\ 0 & A^2 \sum_n (2\pi n)^2 & A^2 \sum_n 2\pi n \\ 0 & A^2 \sum_n 2\pi n & NA^2 \end{bmatrix}$$
$$= \frac{2}{\sigma_w^2} \begin{bmatrix} N & 0 & 0 \\ 0 & 4\pi^2 A^2 \frac{N(N-1)(2N-1)}{6} & \pi A^2 N(N-1) \\ 0 & \pi A^2 N(N-1) & NA^2 \end{bmatrix}$$
(7.61)

The last step uses the identities

$$\sum_{n=0}^{N-1} n = \frac{N(N-1)}{2}, \quad \sum_{n=0}^{N-1} n^2 = \frac{N(N-1)(2N-1)}{6}$$
(7.62)

The CRLB for each parameter is the corresponding diagonal element of I-

 $^{1}(\Theta)$. The calculation is straightforward, if tedious, and gives the results

$$\sigma_{\hat{A}}^{2} \geq \frac{\sigma_{w}^{2}}{2N} \longrightarrow \sigma_{\hat{A}/A}^{2} \geq \frac{\sigma_{w}^{2}}{2NA^{2}} = \frac{1}{2N\chi}$$

$$\sigma_{\hat{f}_{0}}^{2} \geq \frac{6}{(2\pi)^{2}\chi N(N^{2}-1)} \xrightarrow{N \text{ large}} \frac{6}{(2\pi)^{2}N^{3}\chi}$$

$$\sigma_{\hat{\phi}}^{2} \geq \frac{2N-1}{\chi N(N+1)} \xrightarrow{N \text{ large}} \frac{2}{N\chi}$$

$$(7.63)$$

As usual, the CRLBs for frequency and phase are inversely proportional to SNR. The CRLB for amplitude is not, but if follows from Eq. (7.62) that the CRLB for relative amplitude $\overline{A/A} = \sigma_w^2/2N\chi$ and is also inversely proportional to SNR as shown above. The "large *N*" approximations shown are accurate to within 10 percent for $N \ge 10$ (frequency) and $N \ge 13$ (phase). It is interesting to note that the CRLBs for amplitude and phase decrease asymptotically as 1/N while that for frequency improves at the much faster rate of $1/N^3$ as the amount of data increases.

Two alternative forms for the frequency CRLB merit mention. The frequency in hertz is $F_0 = f_0/T_s$, where T_s is the sampling interval of interest. The data duration in seconds is NT_s , so the expected frequency resolution of the DTFT in hertz is $\Delta F = 1/NT_s$. Applying these in the asymptotic form of Eq. (7.63) gives

$$\sigma_{\bar{F}_0}^2 \ge \frac{6}{(2\pi)^2 N^3 T_s^2 \chi} = \frac{6(\Delta F)^2}{(2\pi)^2 N \chi}$$
(7.64)

The last form makes it clear that the N^3 improvement in the frequency estimation CRLB is due largely to the implied improvement in frequency resolution, which accounts for an N^2 term. The remaining factor of N represents the usual SNR improvement due to coherent integration that is also seen in the phase and amplitude CRLBs.

Having multiple simultaneously unknown parameters can result in CRLBs that are larger than they would be if all but one of the parameters is known. The interaction between parameters is manifested in nonzero off-diagonal elements of the Fisher information matrix. For the sinusoidal parameter estimation problem, it can be seen (see Prob. 30) that if the other two parameters are known, the CRLB of the remaining unknown parameter is equal to or tighter than the results of Eq. (7.63). Specifically, the CRLB for amplitude is the same, while in the limits of large *N* the CRLBs for frequency and phase are each $4 \times$ smaller.

MLE for Sinusoid Parameters

The next step is to develop the MLE for the sinusoid parameters. This requires maximizing the likelihood function:

$$\ell(A,\phi,f_0|\mathbf{x}) = \ell(\tilde{A},f_0|\mathbf{x}) = \frac{1}{\left(\pi\sigma_w^2\right)^N} \exp\left[-\frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} |x[n] - \tilde{A}\exp(j2\pi f_0 n)|^2\right]$$
(7.65)

with respect to all three parameters. In Eq. (7.65) the complex amplitude \tilde{A} has been reintroduced for convenience. Equivalently, the exponent term

$$J(\tilde{A}, f_0) \equiv \sum_{n=0}^{N-1} \left| x[n] - \tilde{A} \exp(j2\pi f_0 n) \right|^2$$
(7.66)

must be minimized.

Begin by minimizing *J* with respect to $\tilde{A} = \tilde{A}_R + j\tilde{A}_I$ with f_0 temporarily assumed known. The partial derivative with respect to \tilde{A}_R is (leaving x[n] and $\exp(j2\pi f_0 n)$ in complex form)

$$\frac{\partial J}{\partial \tilde{A}_{R}} = \sum_{n=0}^{N-1} \frac{\partial}{\partial \tilde{A}_{R}} \left\{ \left[x[n] - (\tilde{A}_{R} + j\tilde{A}_{I})e^{j2\pi f_{0}n} \right] \cdot \left[x^{*}[n] - (\tilde{A}_{R} - j\tilde{A}_{I})e^{-j2\pi f_{0}n} \right] \right\} \\
= \sum_{n=0}^{N-1} \left\{ \left[x[n] - (\tilde{A}_{R} + j\tilde{A}_{I})e^{j2\pi f_{0}n} \right] \left[-e^{-j2\pi f_{0}n} \right] + \dots \\
\dots \left[x^{*}[n] - (\tilde{A}_{R} - j\tilde{A}_{I})e^{-j2\pi f_{0}n} \right] \left[-e^{j2\pi f_{0}n} \right] \right\}$$
(7.67)

Setting Eq. (7.67) equal to zero leads to

$$0 = \frac{\partial J}{\partial \tilde{A}_R}$$

= $\sum_{n=0}^{N-1} \left\{ 2\tilde{A}_R - x[n]e^{-j2\pi f_0 n} - x^*[n]e^{j2\pi f_0 n} \right\} = 2N\tilde{A}_R - \sum_{n=0}^{N-1} 2\operatorname{Re}\{x[n]e^{-j2\pi f_0 n}\}$
(7.68)

and finally, along with a similar development for minimization with respect to \tilde{A}_i (see Prob. 32),
$$\left. \begin{aligned} \widehat{\tilde{A}_{R}} &= \operatorname{Re}\left\{ \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_{0}n} \right\} \\ \widehat{\tilde{A}_{I}} &= \operatorname{Im}\left\{ \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_{0}n} \right\} \end{aligned} \Rightarrow \quad \widehat{\tilde{A}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_{0}n}$$
(7.69)

where \hat{A} denotes the estimated value of \hat{A} . Notice that if f_0 is indeed known, then the MLE for \hat{A} is simply the DTFT of the data evaluated at f_0 , a very sensible result.

However, with f_0 not known, J must now be minimized with respect to f_0 . Expanding the definition of J in Eq. (7.66) gives

$$J(\tilde{A}, f_0) = \sum_{n=0}^{N-1} \left[x[n] - \tilde{A} \exp(j2\pi f_0 n) \right] \left[x^*[n] - \tilde{A}^* \exp(-j2\pi f_0 n) \right]$$
$$= \sum_{n=0}^{N-1} \left| x[n] \right|^2 - \tilde{A}^* \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_0 n} - \tilde{A} \sum_{n=0}^{N-1} x^*[n] e^{+j2\pi f_0 n} + \sum_{n=0}^{N-1} \left| \tilde{A} \right|^2$$
(7.70)

Recognizing the sums in the two middle terms as $N \cdot \hat{A}$ and $N \cdot \hat{A}^*$ [see Eq. (7.69)] reduces this to

$$I(\tilde{A}, f_0) = \sum_{n=0}^{N-1} |x[n]|^2 - N \left| \hat{\tilde{A}} \right|^2$$
$$= \sum_{n=0}^{N-1} |x[n]|^2 - \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f_0 n} \right|^2$$
(7.71)

The first term of Eq. (7.71) does not depend on f_0 . Therefore, *J* is minimized by maximizing the second term, which is just the magnitude-squared of the DTFT of the data.⁵ Again very sensibly, $\overline{f_0}$ is found by computing the DTFT of the data and finding the frequency at which the squared-magnitude of the DTFT is largest.

To summarize, the MLEs of the amplitude, frequency, and phase of a complex sinusoid in complex WGN are found by the following sequence of operations:

- 1. Compute the DTFT X(f) of the data x[n].
- 2. $\widehat{f_0}$ is the frequency at which the peak magnitude of the DTFT occurs, $\widehat{f_0} = \underset{f}{\arg \max} \{ |X(f)|^2 \}$

3. The MLE of the complex amplitude is 1/N times X(f) evaluated at \hat{f}_0 , $\hat{A} = X(\hat{f}_0)/N$. The MLEs \hat{A} and $\hat{\phi}$ of the magnitude and phase are the magnitude and angle of \hat{A} .

Figures 7.11 to 7.13 illustrate the behavior of the error in the MLEs for frequency, phase, and relative amplitude, respectively. All three figures are based on simulations of the DFT-based MLE applied to complex sinusoids of length M = 40 samples with random frequencies and initial phases. A large DFT of size K = 1000 was used with the frequency and phase estimates, while K = 80 for the amplitude estimation in order to have a sampling bound high enough to have an impact within the SNR range shown. As with range estimation, in all three cases the measurement precision follows the CRLB only at relatively high SNRs. At low SNR an a priori bound dominates the estimates as the signal becomes essentially all noise, while at high SNR the precision is limited by a sampling bound based on sampling density.



FIGURE 7.11 Behavior of simulated maximum likelihood normalized frequency estimation error. A K = 1000 point DFT was applied to sinusoids with M = 40 time samples.



FIGURE 7.12 Behavior of simulated maximum likelihood phase estimation error in radians. A K = 1000 point DFT was applied to sinusoids with M = 40 time samples.



FIGURE 7.13 Behavior of simulated maximum likelihood relative amplitude estimation error. A K = 80 point DFT was applied to sinusoids with M = 40 time samples.

In the low-SNR case the signal peak will be swamped by the noise, so the MLE for frequency will return simply the location of the peak of the DFT of the noise. The peak is equally likely to occur anywhere in the normalized frequency interval [-0.5, +0.5].

Modeling the estimated normalized frequency $\hat{f_0}$ as a uniform random variable over that interval, the a priori bound on its variance is simply 1/12 and the precision is the square root of this, $1/\sqrt{12}$. Similarly, the phase at the DFT peak is uniform over the full range $[-\pi, \pi]$ so that the a priori bound on phase precision is $2\pi/\sqrt{12}$.

To establish the a priori bound on amplitude, consider that each sample of the *K*-point DFT of x[n] is the sum of *M* samples of complex WGN with variance σ_w^2 weighted by the DFT kernel samples $\exp(-j2\pi k/K)$. Because the magnitudes of those weights are all unity, the DFT samples are also complex Gaussian with variance $M\sigma_w^2$. The magnitude of the DFT samples is therefore a Rayleigh-distributed RV with variance $(1-\pi/4)M\sigma_w^2/2$. The MLE scales the DFT by 1/M so the variance becomes $(1-\pi/4)\sigma_w^2/2M$. The MLE is the maximum of the *K* DFT samples, so the estimate variance will be the variance of that quantity. Denote the PDF of the magnitude of a scaled DFT sample as $p_{|X|}(|X|)$ and the corresponding cumulative distribution function (CDF) as $P_{|X|}(|X|)$. For the scaled DFT of complex WGN these are

$$p_{|X|}(|X|) = \frac{2M|X|}{\sigma_w^2} \exp\left[-\frac{M|X|^2}{\sigma_w^2}\right], \quad |X| > 0$$

$$P_{|X|}(|X|) = 1 - \exp\left[-\frac{M|X|^2}{\sigma_w^2}\right], \quad |X| > 0$$

$$(7.72)$$

Now let z be the maximum of the K samples of the DFT of the M-point input. The PDF of the maximum of N independent samples of |X| is

$$p_{z}(z) = N \cdot P_{|X|}(|X|) \left[p_{|X|}(|X|) \right]^{N-1}$$
(7.73)

Equation (7.72) can be used in Eq. (7.73) to compute $p_z(z)$. Unfortunately, there is no simple formula for the variance of z, but it can be computed numerically

from $p_z(z)$ to establish the a priori bound on amplitude precision.

For K = M, the DFT samples are uncorrelated. It is easy to show that DFT samples are correlated, even for a white noise input, when K > M (see Prob. 33). In fact, for K > M the effective number of independent samples remains approximately equal to M. Consequently, a good estimate of the low-SNR PDF of the MLE is obtained by choosing N = M in Eq. (7.73) regardless of the size of K.

The SNR in Figs. 7.11 to 7.13 is that of the input sinusoidal data. The a priori bounds for all three parameters are controlling for input SNRs below about -10 dB. Since the integration gain for the DFT of a 40-point sequence is $10 \log_{10}(40) = 16$ dB, this corresponds to integrated SNRs of about +6 dB for M = 40.9

When the SNR is sufficiently high and the DFT size is relatively small, the DFT frequency sampling interval limits the precision of all three parameter estimates. Most obvious is the sampling bound on frequency estimation precision. A *K*-point DFT samples the normalized frequency axis in intervals of 1/K cycles per sample. For large SNR this frequency quantization becomes the dominant factor. The frequency error is then modeled as a uniform RV over an interval of width 1/K, so the sampling bound on the precision of the frequency estimate is simply $1/\sqrt{12K}$. This bound is also shown in Fig. 7.11.

The errors in the amplitude and phase estimates at high SNR are induced by the frequency estimation error. Recall that the DTFT of an *M*-point sinusoid like that of Eq. (7.57) with frequency f_0 and initial phase ϕ_0 in the absence of noise is

$$X(f) = Ae^{j\phi_0} \frac{\sin(2\pi f_0 M/2)}{\sin(2\pi f_0 / 2)} \exp[-j\pi f(M-1)]$$
(7.74)

When evaluated at the actual frequency f_0 , $X(f_0) = MA\exp(i\phi_0)$. Dividing by M and taking the magnitude and phase of $X(f_0)$ returns the correct amplitude and phase of the original sinusoid. However, if the DTFT is evaluated at a frequency $f_0 + \delta f$, a different amplitude and phase will be measured. At high SNR, the range of δf is $\pm 1/2K$ (one DFT bin). The resulting range of phase variations 1)/2K, and for amplitude is $\mp \pi(M)$ _ is -A[M] $\sin(\pi M/2K)/\sin(\pi/2K)]/M$. The phase error is expected to be uniform over its interval because it arises from a linear mapping of frequency error to phase error. This is not the case for the amplitude error, but when K is reasonably large compared to M, the amplitude error can be approximately modeled as uniformly distributed over its interval. The sampling bounds on phase and then $\pi (M-1)/K\sqrt{12}$ estimation precision are and amplitude $A[M-\sin(\pi M/2K)/\sin(\pi/2K)]/M\sqrt{12}$ The precision bound for relative

amplitude is $[M - \sin(\pi M/2K)/\sin(\pi/2K)]/M\sqrt{12}$.

The problem of estimating sinusoid frequency and the MLE that resulted is very similar in nature to that of maximum likelihood estimation of time delay. Both involve finding the peak of a relatively narrow pulse in noise. Consequently, the ZZB in the generalized form of Eq. (7.49) can be applied to this problem as well. The equivalent of χ_{out} for the frequency estimation MLE is $M\chi$. The Ziv-Zakai bound computed using this change and the CRLBs and a priori bounds computed above for frequency estimation is plotted on Figs. 7.11 and 7.12. It is not shown in Fig. 7.13 because the particular arguments in the erfc(·) and $\Gamma(\cdot)$ weighting functions of Eq. (7.49) do not match well to the transition between the CRLB and a priori bounds in the amplitude estimation problem.

As with time delay estimation, interpolation techniques can be used to try to improve precision beyond the sampling bound at high SNR. In fact, many of the interpolation techniques have been developed primarily for the purpose of frequency estimation. Once an improved frequency estimate is obtained, the same logic used above to establish the sampling bounds can be used to infer improved phase and amplitude estimates. Figure 7.14 shows the results of quadratic interpolation for the case M = 40 and K = 120. All three parameters show significant precision improvements. The two-sample and centroid interpolation results are not shown because, similar to the time delay case in Fig. 7.8, they improve precision only slightly. These two interpolators are more effective for more coarse sampling, e.g., K = M, but even then the quadratic interpolator produces better results. All of the previously discussed caveats regarding the effects of windowing, real versus complex data, oversampling rates, and bias on interpolation apply in sinusoid parameter estimation as well.





FIGURE 7.14 Improvement in precision using quadratic interpolation. K = 120 point DFT and M = 40. (*a*) Normalized frequency, (*b*) phase (radians), (*c*) relative amplitude.

7.2.3 Angle Estimators

The remaining parameters to be estimated in order to locate a target in threedimensional space and Doppler shift are the elevation and azimuth angles. It is sufficient to consider estimation of one angle; the results apply to both. There are two fundamental methods for estimating target angle relative to the antenna orientation in radar. The first uses phase measurements at multiple antenna phase centers in what is essentially phase interferometry, while the second uses multiple amplitude measurements in a process known as *lobing* or *lobe switching*. The choice between them depends on the type of antenna available and the data collection protocol. Some techniques are blends of the two basic approaches.

Phase-Based Angle Measurement

Consider the horizontally oriented uniform linear array shown in Fig. 7.15. Each of the gray triangles represents an antenna phase center having its own independent receiver. These could represent either subarrays or individual elements of a phased array antenna. The number of phase centers *N* is assumed to be at least two. They are separated by a distance *d*. An electromagnetic plane wave of wavelength λ (frequency $\Omega = 2\pi/\lambda$ radians per second) arrives at the

array from an angle of arrival (AOA) of θ radians relative to the array normal.



FIGURE 7.15 Plane wave impinging on a uniform linear array.

Two successive planar phase fronts are shown, spaced by λ meters in the direction of propagation.

Consider a phase front arriving at phase center #0 at some time *t*. That phase front must travel $d\sin\theta$ meters in the direction of propagation in order to arrive at phase center #1, and that will take $d\sin\theta/c$ seconds. If the voltage signal observed at phase center #0 is of the form $\overline{y}_0(t) = A \exp\{j[\Omega t + \phi]\}$, the signal at phase center #1 will be $\overline{y}_1(t) = A \exp\{j[\Omega(t - d\sin\theta/c) + \phi]\}$. More generally, the signal at the *n*th phase center will be

$$\overline{y}_n(t) = A \exp\{j[\Omega(t - nd\sin\theta/c) + \phi_0]\}, \quad 0 \le n \le N - 1$$
(7.75)

Now consider the set of *N* voltage samples obtained by measuring the voltage at each phase center simultaneously at some time $t = t_0$ and assembling the results into a column vector **y**:

$$y[n] \equiv \overline{y}_n(t_0) = A \exp\{j[\Omega(t_0 - nd\sin\theta/c) + \phi_0]\}$$

= $\hat{A} \exp(-j\Omega nd\sin\theta/c), \quad n = 0, ..., N-1$
= $\hat{A} \exp(-j2\pi nd\sin\theta/\lambda)$
$$\mathbf{y} = [y[0] \quad y[1] \quad \cdots \quad y[N-1]]^T$$
(7.76)

The spatial phase history vector **y** is called a *spatial snapshot* of the signal arriving at the array. Define the normalized spatial frequency $k_{\theta} = 2\pi d \sin\theta / \lambda$. k_{θ} is the spatial frequency in radians per meter projected onto the face of the array,

 $(2\pi/\lambda)\sin\theta$, times the array sampling interval (the element spacing *d*), and has dimension of radians per sample. The spatial snapshot becomes

$$\mathbf{y} = \hat{A} \begin{bmatrix} 1 & e^{-jk_{\theta}} & \cdots & e^{-j(N-1)k_{\theta}} \end{bmatrix}^T$$
(7.77)

Equation (7.77) shows that the snapshot is a sampled complex sinusoid with normalized radian spatial frequency $-k_{\theta}$.

The radar's operating wavelength λ and the phase center spacing *d* are presumably known, so the AOA θ can be determined if k_{θ} can be measured; and since the snapshot is a sampled sinusoid, the results of the previous section on frequency estimation can be applied. Specifically, scaling Eq. (7.63) for a normalized radian frequency instead of the cyclical frequency f_0 , the CRLB for estimating k_{θ} given a spatial snapshot with complex AWGN is

$$\sigma_{\hat{k_{\theta}}}^2 \ge \frac{6}{\chi N(N^2 - 1)} \xrightarrow{N \text{ large}} \frac{6}{N^3 \chi}$$
(7.78)

The CRLB for the AOA can be obtained using the transformation of parameters rule in Sec. A.3.2 of App. A. Strictly speaking, a generalized version of this rule is needed for estimation of a transformed vector parameter; see Kay (1993) for details. However, in this sinusoid estimation problem the transformation from a CRLB on k_{θ} to one on θ simplifies to the scalar transformation case given in Eq. (A.71). Defining $\alpha = \lambda/2\pi d$ to temporarily simplify the notation,

$$\sigma_{\hat{\theta}}^{2} \geq \left[\frac{\partial\theta}{\partial k_{\theta}}\right]^{2} \sigma_{\hat{k}_{\theta}}^{2} = \left[\frac{\partial}{\partial k_{\theta}} \sin^{-1}(\alpha k_{\theta})\right]^{2} \sigma_{\hat{k}_{\theta}}^{2}$$
$$= \frac{\alpha^{2}}{1 - (\alpha k_{\theta})^{2}} \sigma_{\hat{k}_{\theta}}^{2} = \frac{\alpha^{2}}{1 - \sin^{2}\theta} \sigma_{\hat{k}_{\theta}}^{2} = \frac{\alpha^{2}}{\cos^{2}\theta} \sigma_{\hat{k}_{\theta}}^{2}$$
$$= \frac{\lambda^{2}}{(2\pi)^{2} d^{2} \cos^{2}\theta} \sigma_{\hat{k}_{\theta}}^{2}$$
(7.79)

Combining Eqs. (7.78) and (7.79) gives the CRLB on AOA,

$$\sigma_{\hat{\theta}}^2 \ge \frac{6\lambda^2}{(2\pi)^2 d^2 \cos^2 \theta \chi N(N^2 - 1)} \xrightarrow{N \text{ large}} \frac{6\lambda^2}{(2\pi)^2 d^2 \cos^2 \theta \chi N^3}$$
(7.80)

Note that the accuracy depends on the AOA. In particular, signals arriving

normal to the array ($\theta = 0$) are estimated with the finest precision, while the "precision" of the estimate of the AOA of those arriving from the end of the array ($\theta = \pm \pi/2$) becomes infinite. In this case, there is no difficulty estimating k_{θ} with the variance given in Eq. (7.78). However, the derivative of the mapping $\theta = \sin^{-1} (\lambda k_{\theta}/2\pi d) \rightarrow \infty$ as the argument tends to ± 1 so that a finite precision in k_{θ} becomes arbitrarily large in θ .

In practice, the AOA and therefore the precision cannot become unbounded. Ignoring signals arriving from behind the antenna, the AOA must be in the interval $[-\pi/2, +\pi/2]$. The worst-case variance is then $\pi^2/12$. This is also the a priori bound for low-SNR angle estimation.

Several substitutions put Eq. (7.80) in a more useful form. First, define the length of the array D = (N - 1)d. Next, recall that the 3-dB beamwidth in the direction normal to the face of the array is $k\lambda/D$ for some k in the range of 0.89 to perhaps 2 or 3, depending on aperture weighting for sidelobe control. For an electronically scanned array antenna, the 3-dB beam-width in a direction θ from the array normal then becomes $\theta_3 = k\lambda/D\cos\theta$ due to projection of the array length to the shorter effective length $D\cos\theta$ in the direction θ . Finally, note that χ is the SNR at an individual phase center; the coherently integrated SNR will be $\chi_{out} = N\chi$. With these substitutions, Eq. (7.80) becomes

$$\sigma_{\hat{\theta}}^2 \ge \frac{6\theta_3^2}{(2\pi)^2 \chi_{\text{out}} \left(\frac{N+1}{N-1}\right) k^2} \xrightarrow{N \text{ large}} \frac{6\theta_3^2}{(2\pi)^2 \chi_{\text{out}} k^2}$$

$$(7.81)$$

This equation shows that the precision (square root of $\sigma_{\hat{\theta}}^2$) in estimating AOA is a fraction, determined by the integrated SNR, of its angular resolution. That fraction is independent of AOA. However, it must be remembered that θ_3 depends on the angle of arrival. The precision still becomes unbounded in theory for end-fire AOAs because the beamwidth $k\lambda/D\cos\theta$ becomes unbounded in theory, but in practice the beamwidth is limited to π radians (2π if allowing for signals from behind the array) and the estimate is still limited to $\pm \pi/2$.

Figure 7.16 plots the precision bound relative to the minimum 3-dB beamwidth (at AOA $\theta = 0^{\circ}$) for a case where k = 1.34, $d = 5\lambda$ (implying a subarrayed antenna), N = 10, and the single-subarray (nonintegrated) SNR = 0 dB so that the integrated SNR is 10 dB. In this case the minimum 3-dB beamwidth is $k\lambda/(N-1)d \approx 30$ mrad or 1.7°. For signals arriving within about $\pm 55^{\circ}$ of the array normal, the precision is better than two-tenths of the minimum 3 dB beamwidth. The low-SNR precision bound is $\sqrt{\pi^2/12} \approx 0.9$ radian or 52°. This is 30.6 times the minimum beamwidth and is well above the range of the plot.



FIGURE 7.16 CRLB for relative precision of angle of arrival measured with a subarrayed antenna. See text for details.

If the phase center spacing $d > \lambda/2$, the Nyquist spatial sampling criterion of <u>Chap. 3</u> will not be met. This will be the case if each phase center is a multielement subarray. In this event, $|k_{\theta}| > \pi$ for some range of AOAs, resulting in aliasing of the array factor portion of the antenna pattern (grating lobes) and making the estimate of θ ambiguous. For example, if $d = 2\lambda$ then AOAs of 0°, $\pm 30^{\circ}$, and $\pm 90^{\circ}$ result in $k_{\theta} = 0$, $\pm 2\pi$, and $\pm 4\pi$, all of which alias to $k_{\theta} = 0$. If $k_{\theta} = 0$, the AOA could be any of those five angles. This concern is eased, if not eliminated, by the subarray antenna pattern, which attenuates some of the ambiguous lobes; see Bailey (2010) for more information.

Lobing-Based Angle Measurement

Consider the scenarios shown in Fig. 7.17. The radar antenna is pointed at an angle θ_0 relative to some axis. Now suppose the radar reports a target detection at some range. It is tempting to assume the angle to the target is θ_0 as shown in Fig. 7.17*a*, but in fact the target might actually be at some other angle $\theta_0 + \delta\theta$ as shown in Fig. 7.17*b*. It can reasonably be assumed that $\delta\theta$ is in the range of $\pm \theta_3/2$ so that the target is in the mainbeam and has sufficient SNR to be detected. Nonetheless, a method is needed to determine the target location within the main-beam so that the angle can be measured to an accuracy finer than the beamwidth.



FIGURE 7.17 Effect of antenna gain pattern on received target echo amplitude: *(a)* target on boresight of antenna, *(b)* target off boresight.

This problem can be solved by noting that the target echo strength (power or voltage) measured in these two situations will not be the same because of the different gains of the two-way antenna pattern on and off the boresight. Suppose that once a target is detected at radar pointing angle θ_0 the procedure illustrated in Fig. 7.18 is followed. First, the antenna is steered to one side of the boresight by some amount $\Delta\theta$ radians and the target echo amplitude is measured again. The antenna is then steered $\Delta\theta$ radians to the opposite side of the original boresight and the echo amplitude measured once more. Either mechanical or electronic steering can be used, as appropriate for the particular antenna. The offset $\Delta\theta$ should be limited so that the target remains in the mainlobe in most cases. This issue is discussed in slightly more detail shortly.



FIGURE 7.18 Sequential lobing concept: (*a*) new target measurements are taken with the antenna displaced from the nominal look direction, (*b*) relative amplitude of resulting measurements.

By examining the pattern of target amplitude versus angle, the direction of the target relative to the original boresight can be determined. In the example of Fig. 7.18*a*, the target is to the left of the original boresight. When the antenna is shifted counterclockwise, the target is nearer the boresight and is therefore viewed with a higher antenna gain, resulting in a higher amplitude measurement. When the antenna is shifted clockwise the antenna bore-sight moves away from the target, resulting in a lower amplitude measurement. This pattern of measurements, shown in Fig. 7.18*b*, indicates that the target must be counterclockwise from the original look direction. With knowledge of the antenna pattern, analysis of the relative amplitudes can be used to determine by how much. It should be clear that if the target were clockwise from the original boresight the pattern of the two measurements would be reversed, and if it was exactly on the boresight, the two measurements would produce equal amplitudes.

This process of making multiple measurements of target strength at offset angles relative to the nominal pointing angle and analyzing the results to indicate target angle is called *lobing*. When the measurements are made sequentially in time as described above it is called *sequential lobing*. Alternatively, antennas can be constructed that form multiple beams at once, providing multiple output signals corresponding to the original centered beams and to pairs of squinted beams in two orthogonal angle coordinates. These antennas are called *monopulse antennas* because they can form all of the required signals for target angle measurement on a single pulse. A brief introduction to monopulse antennas is given in Bailey (2010). A thorough discussion of monopulse radar, covering a variety of antenna structures and processing methods, is given in Sherman (1984).

Essentially the same approach can be used for analyzing the bias and precision of two-beam lobing for both the sequential and monopulse cases. An important modeling detail is whether the antenna pattern is the same for the squinted beams and the original beam, differing only in the boresight direction. This depends on antenna type and scanning mechanism. If squinted beams are formed by mechanically tilting a conventional antenna, all the beam shapes are the same. If a phased array antenna is electronically scanned to form squinted beams, they are not because the pattern broadens as the scan angle from normal increases. Monopulse antennas typically form identical squinted beams and then form the unsquinted beam from their sum, so that the unsquinted pattern is not the same as the squinted patterns. In this analysis it is assumed that the sum pattern

is formed from two identically shaped squinted beams either in a monopulse or sequential lobing approach. The results obtained will still be generally applicable to other designs, though some details may change. The reader is referred to Sherman (1984) and Howard (2008) for much more information on antenna patterns for lobing systems.

Denote the complex amplitudes of the voltages after demodulation and matched filtering for a range bin containing a single point target of interest in each of the two beams of Fig. 7.18*a* as v_L and v_R . The symbols *L* and *R* refer to the antenna being shifted counterclockwise ("left" or *L*) and clockwise ("right" or *R*) by $\Delta\theta$ radians from the nominal pointing angle θ_0 , respectively. The nominal pointing direction of the antenna can be set to $\theta_0 = 0$ for the time being without loss of generality.

Ignore for now any effects of noise or target fluctuations. Let $G(\theta - \Delta \theta)$ be the one-way antenna voltage pattern when the beam is squinted $\Delta \theta$ radians clockwise (right). The gain on the boresight $\theta = \Delta \theta$ is normalized to G(0) = 1. The two-way "right" voltage pattern is then $G^2(\theta - \Delta \theta)$. Similarly, the two-way "left" voltage pattern is $G^2(\theta + \Delta \theta)$. The complex amplitude of the target *L* and *R* echo voltages will be $v_R = A\exp(j\phi)G^2(\delta\theta - \Delta\theta)$ and $v_L = A\exp(j\phi)G^2(\delta\theta - \Delta\theta)$ when the target is located at an angle $\delta\theta$ relative to the original pointing direction. Another important detail of the modeling is whether for a stationary non-fluctuating point target in the absence of noise the phases of v_L and v_R are identical. The answer depends on the design of the antenna and, in the nonmonopulse case, the scanning mechanism used to form the squinted beams. In this analysis, it is assumed that they are identical for a single nonfluctuating point target with no noise. This is the intended result for many designs, although it is not always realized in practice.

The most common way by far to utilize v_L and v_R begins by computing the sum and difference of the received voltages. The left-minus-right *difference signal* is

$$v_{\Delta}(\delta\theta \mid \Delta\theta) = v_L - v_R = A \exp(j\phi)[G^2(\delta\theta + \Delta\theta) - G^2(\delta\theta - \Delta\theta)]$$
$$\equiv A \exp(j\phi)G_{\Delta}^2(\delta\theta \mid \Delta\theta)$$
(7.82)

Because this depends on the target amplitude and not just its angular location, a *sum signal* is also defined as

$$v_{\Sigma}(\delta\theta | \Delta\theta) = v_L + v_R = A \exp(j\phi)[G^2(\delta\theta + \Delta\theta) + G^2(\delta\theta - \Delta\theta)]$$
$$\equiv A \exp(j\phi)G_{\Sigma}^2(\delta\theta | \Delta\theta)$$
(7.83)

The Δ/Σ ratio is then independent of target amplitude:

$$v_{\Delta/\Sigma}(\delta\theta \,|\, \Delta\theta) = \frac{v_L - v_R}{v_L + v_R} = \frac{G_{\Delta}^2(\delta\theta \,|\, \Delta\theta)}{G_{\Sigma}^2(\delta\theta \,|\, \Delta\theta)} \equiv G_{\Delta/\Sigma}^2(\delta\theta \,|\, \Delta\theta)$$
(7.84)

The arguments $(\delta \theta | \Delta \theta)$ in all three of these voltages emphasizes that they depend on both the target location relative to the original boresight and the squint angle used in the lobing measurement system. For brevity, these arguments will usually be dropped in the remainder of this section.

Figure 7.19 illustrates the antenna lobing patterns involved when the basic two-way antenna voltage pattern is the sinc² pattern of Fig. 7.1*a*. Figure 7.19*a* shows the central portion of the *L* and *R* patterns when each is shifted by 25 percent of the Rayleigh beamwidth away from the original boresight at $\theta = 0$. Figure 7.19*b* shows the difference pattern $G_{\Delta}^2 = L - R$, while 7.19*c* shows the sum pattern $G_{\Sigma}^2 = L + R$. Figure 7.19*d* is the pattern $G_{\Delta/\Sigma}^2$ of Eq. (7.84).



FIGURE 7.19 Two-way antenna voltage patterns for sequential lobing: (*a*) pattern $G^2(\theta)$ of Fig. 7.1*a* shifted left and right by $\Delta \theta = 0.25$ Rayleigh widths in each direction, (*b*) $G_{\Delta}^2 = L - R$ difference pattern, (*c*) $G_{\Sigma}^2 = L + R$ sum pattern, (*d*) normalized difference pattern $G_{\Delta/\Sigma}^2$. The gray dashed line is a linear

approximation to the central region with slope $k_{\Delta\Sigma}$.

Notice that the normalized difference pattern is approximately linear in the angular region corresponding to the central 40 percent or so of the original antenna mainlobe. This suggest a simple algorithm for estimating the target angle $\delta\theta$ relative to the original boresight. For a given antenna, the Δ/Σ pattern $G_{\Delta/\Sigma}^2$ can be measured and its central portion approximated by a straight line with slope $k_{\Delta\Sigma}$, $G_{\Delta/\Sigma}^2(\theta) \approx k_{\Delta/\Sigma} \cdot \theta$, as shown by the gray dashed line in Fig. 7.19*d*. When a target is detected, the lobing measurements v_L and v_R are collected and the quantity $v_{\Delta/\Sigma}$ computed. The estimated target angular position then becomes

$$\hat{\theta} = \theta_0 + \widehat{\delta\theta} = \theta_0 + \frac{1}{k_{\Delta/\Sigma}} \cdot v_{\Delta/\Sigma}$$
(7.85)

where θ_0 is the boresight angle of the sum pattern, which is also the axis of odd symmetry of the difference pattern.

Now suppose that the noise-free values of v_L and v_R are each contaminated by i.i.d. complex WGN samples having variance σ_{w}^2 . Because they are the sum and difference of v_L and v_R , v_{Δ} and v_{Σ} will each be contaminated with noise having a variance of $2\sigma_{w}^2$. The Δ/Σ ratio voltage will be

$$\frac{v_{\Delta} + n_{\Delta}}{v_{\Sigma} + n_{\Sigma}} \approx \frac{v_{\Delta} + n_{\Delta}}{v_{\Sigma}} = v_{\Delta/\Sigma} + \frac{n_{\Delta}}{v_{\Sigma}}$$
(7.86)

The second step assumes high SNR so that $v_{\Sigma} >> n_{\Sigma}$. The assumption that the phases of v_L and v_R are identical implies that the phases of v_{Δ} and v_{Σ} also have the same value, so that $v_{\Delta/\Sigma}$ is real. Consequently, the target angle information is contained entirely in the real part of the Δ/Σ ratio under ideal conditions; the imaginary part contains only noise. For this reason, the estimation algorithm is modified to its final form

$$\hat{\theta} = \theta_0 + \widehat{\delta\theta} = \theta_0 + \frac{1}{k_{\Delta/\Sigma}} \operatorname{Re}(v_{\Delta/\Sigma})$$
(7.87)

This is the form most commonly used in practice.

While this estimation approach seems reasonable, it is so far only heuristically based. A more systematic approach would again derive the CRLB for angle estimation and the MLE given the measurements v_L and v_R . Begin with the CRLB. The signals v_L and v_R will be nonzero mean (due to the target component) independent complex Gaussians. They will not have the same mean due to the different *L* and *R* antenna weights, but will have the same variance σ_w^2 . The sum and difference voltages v_{Δ} and v_{Σ} will also be nonzero mean complex Gaussians, but with different means and variance $2\sigma_w^2$. The target angle $\delta\theta$ is real, so Eq. (7.18) applies. A slight extension is required since the antenna patterns for v_L and v_R differ due to the different squints. The result is

$$\sigma_{\widehat{\delta\theta}}^{2} \geq \frac{\sigma_{w}^{2}}{2A^{2} \left\{ \left| \frac{\partial G^{2}(\theta + \Delta\theta)}{\partial \theta} \right|^{2} + \left| \frac{\partial G^{2}(\theta - \Delta\theta)}{\partial \theta} \right|^{2} \right\}}$$
$$= \frac{1}{2\chi \left\{ \left| \frac{\partial G^{2}(\theta + \Delta\theta)}{\partial \theta} \right|^{2} + \left| \frac{\partial G^{2}(\theta - \Delta\theta)}{\partial \theta} \right|^{2} \right\}}$$
(7.88)

Further progress requires assuming a specific antenna pattern.

Derivation of the MLE proves difficult. Several versions of the result for various assumptions regarding the relative phase of the left and right signals, whether multiple samples are integrated, and so forth are given in Hofstetter and DeLong (1969). For the simple case assumed here, the MLE $\delta\theta$ of $\delta\theta$ is the value that satisfies

$$\frac{G^2(\widehat{\delta\theta} + \Delta\theta)}{G^2(\widehat{\delta\theta} - \Delta\theta)} = \frac{v_L}{v_R}$$
(7.89)

i.e., the value of $\delta\theta$ such that the ratio of the left and right squinted patterns at that angle matches the ratio of the measured left and right voltages. Simple algebraic manipulations can restate $\delta\theta$ in terms of sum and difference voltages and patterns as the value that satisfies

$$\frac{G^{2}(\delta\theta + \Delta\theta) - G^{2}(\delta\theta - \Delta\theta)}{G^{2}(\delta\theta + \Delta\theta) + G^{2}(\delta\theta + \Delta\theta)} = \frac{v_{L} - v_{R}}{v_{L} + v_{R}} \implies G^{2}_{\Delta/\Sigma} \left(\delta\theta \middle| \Delta\theta \right) = \operatorname{Re}(v_{\Delta/\Sigma})$$
(7.90)

The Re{·} operator was again added in the last step because the Δ/Σ pattern is ideally real, so that in the presence of noise the imaginary part of the Δ/Σ ratio can be ignored. If the linear approximation to the central portion of the difference pattern as illustrated in Fig. 7.19*d* holds, $G_{\Delta/\Sigma}^2(\theta) \approx k_{\Delta/\Sigma} \cdot \theta$, then $\delta \theta = v_{\Delta/\Sigma}/k_{\Delta/\Sigma}$ and Eq. (7.87) is in fact the maximum likelihood estimate of the target angle.

Equation (7.89) implies that the target angle could be estimated based on the ratio of v_L and v_R rather than the ratio of v_{Δ} and v_{Σ} . This is possible but is usually avoided because it would involve use of the pattern $G_{L/R}^2(\theta | \Delta \theta) \equiv G^2(\delta \theta + \Delta \theta)/G^2(\delta \theta - \Delta \theta)$ which is much less linear than $G_{\Delta/\Sigma}^2$, not symmetric around $\theta = 0$, and not usable over as wide a range of θ . This pattern is shown in Fig. 7.20 for the same basic sinc² antenna pattern and ± 0.25 beamwidth squint used earlier.



FIGURE 7.20 Left/right ratio pattern $G_{L/R}^{2}(\theta)$ for same antenna used in Fig. 7.19.

The PDF of the ratio voltage $v_{\Delta/\Sigma}$ is needed to determine the mean and bias of the estimator [Eq. (7.90)]. It has been derived in Kanter (1977), but the derivation is difficult and does not result in a tractable form. The result shows that the MLE exhibits a bias of the form $v_{\Delta/\Sigma} \exp(-\chi)$. This bias is very small for a target near θ_0 for any SNR, and decreases rapidly with increasing SNR for any target angle. For SNR values adequate to provide quality tracking it is usually not significant.

The variance of the MLE is infinite in theory, but a simple approximation applies for high SNR, the main case of interest. Again suppose that noise-free values of v_{Δ} and v_{Σ} are each contaminated by i.i.d. complex WGN samples n_{Δ} and n_{Σ} having variances $2\sigma_{w}^2$. The error in the estimate of $v_{\Delta/\Sigma}$ will be

$$\varepsilon_{\Delta/\Sigma} = \frac{v_{\Delta} + n_{\Delta}}{v_{\Sigma} + n_{\Sigma}} - \frac{v_{\Delta}}{v_{\Sigma}} = \frac{n_{\Delta} - v_{\Delta/\Sigma} n_{\Sigma}}{v_{\Sigma} + n_{\Sigma}}$$
$$\approx \frac{n_{\Delta} - v_{\Delta/\Sigma} n_{\Sigma}}{v_{\Sigma}} = \frac{n_{\Delta}}{v_{\Sigma}} - v_{\Delta/\Sigma} \frac{n_{\Sigma}}{v_{\Sigma}}$$
(7.91)

In this equation, v_{Σ} represents the measured signal in the absence of noise, so only the n_{Δ} and n_{Σ} terms are random.

Because the noise components are zero mean, the mean of $\varepsilon_{\Delta/\Sigma}$ in this analysis is zero. The variance is

$$\operatorname{var}(\varepsilon_{\Delta/\Sigma}) = \overline{\varepsilon_{\Delta/\Sigma}^{2}} = \frac{\overline{n_{\Delta}^{2}}}{|v_{\Sigma}|^{2}} + v_{\Delta/\Sigma}^{2} \frac{\overline{n_{\Sigma}^{2}}}{v_{\Sigma}^{2}}$$
$$= \frac{2\sigma_{w}^{2}}{|v_{\Sigma}|^{2}} \left(1 + v_{\Delta/\Sigma}^{2}\right) = \frac{1}{\chi_{\Sigma}} \left(1 + v_{\Delta/\Sigma}^{2}\right)$$
(7.92)

where the sum-channel SNR has been defined as $\chi_{\Sigma} \equiv |v_{\Sigma}|^2 / 2\sigma_w^2$. The ideal lobing angle estimation processor selects the real part of $v_{\Delta/\Sigma}$. Ideally, this contains all of the target power but only half of the noise power. The expected variance of the real part of the error is then

$$\operatorname{var}(\operatorname{Re}\{\varepsilon_{\Delta/\Sigma}\}) = \frac{1}{2\chi_{\Sigma}} \left(1 + v_{\Delta/\Sigma}^{2}\right) = \frac{1}{2\chi_{\Sigma}} \left(1 + k_{\Delta/\Sigma}^{2} \cdot \theta^{2}\right)$$
(7.93)

This is the *lobing bound* on the variance of $v_{\Delta/\Sigma}$. The variance of the angle error is probably of more direct interest. Using $v_{\Delta/\Sigma} = G_{\Delta/\Sigma}^2(\theta) \approx k_{\Delta/\Sigma} \cdot \theta$ gives $\varepsilon_{\theta} = \varepsilon_{\Delta/\Sigma}/k_{\Delta/\Sigma}$. Equation (7.93) then becomes the lobing bound on angle precision in radians:

$$\operatorname{var}(\operatorname{Re}\{\varepsilon_{\theta}\}) = \frac{1}{2k_{\Delta/\Sigma}^{2} \cdot \chi_{\Sigma}} \left(1 + k_{\Delta/\Sigma}^{2} \cdot \theta^{2}\right)$$
$$= \frac{\theta_{3}^{2}}{2k_{m}^{2} \cdot \chi_{\Sigma}} \left(1 + k_{m}^{2} \cdot \left(\frac{\theta}{\theta_{3}}\right)^{2}\right)$$
(7.94)

Equation (7.94) is the desired result. In the second line, a normalized Δ/Σ slope $k_m \equiv k_{\Delta/\Sigma}\theta_3$ has been introduced. This or a closely similar definition is used in many common references, so it is included here for convenience of comparison. Several extensions to Eq. (7.94) to account for unequal *L* and *R* noise powers or correlated noise (both possibly due to jamming) and higher-order approximations are given in Sherman (1984).

Equation (7.94) is a reasonably good model of the error statistics for high SNR. However, because of the ratio calculation the error can occasionally become very large. In the absence of noise v_L and v_R are in phase, but when significant noise is present, their relative phase can take on any value. When it

happens that v_L and v_R are nearly equal in magnitude but close to 180° out of phase with one another, v_{Σ} can become very small while v_{Δ} does not. The ratio voltage $v_{\Delta/\Sigma}$ can then become very large, resulting in large outliers well outside the range $\pm \pi/2$ for the estimated angle and in turn resulting in large error variances. This behavior is consistent with the infinite variance of the exact theoretical PDF mentioned previously.

Unrealistically large angle errors can be essentially eliminated by one of several techniques, such as clipping the allowed magnitude of $v_{\Delta/\Sigma}$ to some limit or simply discarding those measurements that exceed a limit. The limit could be based, for instance, on the sum pattern beamwidth or the width of the linear region of the Δ/Σ pattern.

Figure 7.21 shows the precision observed in a simulation using Eq. (7.87) to estimate the target angle with the sinc² antenna patterns. Figure 7.21*a* shows the variation in precision versus SNR when the beams are squinted by ± 0.2 Rayleigh beamwidths and the actual target angle is ± 0.1 Rayleigh beamwidth from the nominal boresight. Also shown are the lobing bound and CRLB of Eqs. (7.94) and (7.88), respectively, as well as the a priori bound for a uniform random phase over the interval $\pm \pi/2$. Note that the lobing bound is very close to, but slightly less than, the CRLB. This is possible only because the estimator is biased.



FIGURE 7.21 Angle estimation precision using lobing: (*a*) precision versus SNR, (*b*) precision versus target angle. See text for details.

Two curves are shown for the simulated precision. The dashed line shows

the results obtained without any clipping of the magnitude of $v_{\Delta/\Sigma}$, while the gray dots show the result when it is (very loosely) clipped to a value twenty times that corresponding to the beam squint angle. This is sufficient to result in estimated angles that approach the a priori bound at low SNR while giving much more consistent behavior. Both curves follow the lobing bound and CRLB closely at high SNR.

In Fig. 7.21*b* the variation of precision versus target angle and the various bounds are shown for SNR = 0 dB and 15 dB. For the 0 dB case, the result is shown both with and without clipping of the $v_{\Delta/\Sigma}$ measurement. The lobing bound in the 15 dB case is obscured by the data. Again, the lobing bounds are slightly less than the CRLB, but this figure shows that the difference is greater at larger target angles and that neither bound is particularly tight at low SNR and larger angles

There is a tradeoff between precision and width of the linear region around the bore-sight when choosing the squint angle. Equation (7.94) makes clear the effect of the Δ/Σ slope on estimate precision. Since $k_{\Delta/\Sigma} \cdot \theta \approx v_{\Delta/\Sigma}$ is less than one, often significantly so, the dominant term in the angle measurement precision is $1/k_{\Delta/\Sigma} \cdot \sqrt{2\chi_{\Sigma}}$, which is inversely proportional to the Δ/Σ slope. Consequently, a larger slope results in a finer precision for a given SNR. For a given basic squint antenna pattern $G^2(\theta \pm \Delta\theta)$, the slope increases if the squint angle $\Delta\theta$ is increased. On the other hand, larger slopes maintain the approximate linearity of $G_{\Delta/\Sigma}^2$ over a smaller angular region about the boresight. Furthermore, if the squint angle is too large the sum pattern loses both gain and resolution near the boresight. The gain reduction decreases χ_{Σ} , degrading the precision. All of these effects are evident in Fig. 7.22, which shows the sum pattern and normalized difference pattern for various squint angles.



FIGURE 7.22 Effect of lobing squint angle $\Delta \theta$ on the sum and difference patterns: (*a*) normalized difference pattern $G_{\Delta/\Sigma}^2(\theta)$, (*b*) sum pattern $G_{\Sigma}^2(\theta)$.

The optimum squint depends on the antenna patterns and what is considered "optimum." The angle precision near boresight will be finest when the quantity $k_{\Delta/\Sigma} \sqrt{2\chi_{\Sigma}}$ is largest. For the sinc² pattern used here, this occurs for $\Delta \theta = 0.42$ Rayleigh beamwidths or about $0.65\theta_3$. The cost of this finest precision is a relatively narrow linear region of about ± 0.15 Rayleigh beamwidths in the difference pattern.

The error analysis so far has considered only nonfluctuating targets and a single measurement. In short, if N independent angle measurements are averaged, the error variance will decline by the factor 1/N. Target fluctuations

make the SNR a random variable. Their effect on the variance depends on the fluctuation model and whether or not a threshold is used to eliminate lower-SNR returns; the result can be either an increase or decrease in variance, depending on threshold setting, for correlated target echoes. For uncorrelated echoes, the variance approaches the nonfluctuating case with N measurements if N is sufficiently large. More information on both issues is available in Sherman (1984).

Angle measurement suffers an important additional source of error known as *glint*. Glint occurs when there are multiple point targets in the antenna beam or with multiple-scatterer targets. Consider two point targets in the tracking antenna beam at angles θ_1 and θ_2 relative to boresight. By superposition, the output of the *L* and *R* channels, and therefore of the sum and difference channels, is just the complex sum of the respective outputs. In the absence of noise, $v_{\Sigma 1}$ will be in phase with $v_{\Delta 1}$ and $v_{\Sigma 2}$ will be in phase with $v_{\Delta 2}$ as discussed previously. However, $v_{\Sigma 1}$ and $v_{\Sigma 2}$ will not be of equal amplitude nor in phase with one another in general. Write $v_{\Sigma 2} = \alpha \cdot v_{\Sigma 1}$ for some complex scalar $\alpha = \rho$ $\exp(j\phi)$. Assuming operation in the linear slope region, $v_{\Delta 1} = k_{\Delta/\Sigma}v_{\Sigma 1}\theta_1$ and $v_{\Delta 2} = k_{\Delta/\Sigma}v_{\Sigma 2}\theta_2 = \alpha k_{\Delta/\Sigma}v_{\Sigma 1}\theta_2$. The estimated angle becomes

$$\hat{\theta} = \operatorname{Re}\left\{\frac{1}{k_{\Delta/\Sigma}} \frac{v_{\Delta 1} + v_{\Delta 2}}{v_{\Sigma 1} + v_{\Sigma 2}}\right\}$$
$$= \operatorname{Re}\left\{\frac{\theta_1 + \alpha \theta_2}{1 + \alpha}\right\} = \frac{\theta_1 + (\theta_1 + \theta_2)\rho\cos\phi + \rho^2\theta_2}{1 + 2\rho\cos\phi + \rho^2}$$
(7.95)

Ideally, $\hat{\theta}$ always lies between θ_1 and θ_2 , is halfway between them when $\rho = 1$, and is closer to the stronger of the two targets for other values of ρ .

It is easy to see from Eq. (7.95) that in fact $\hat{\theta} \rightarrow \hat{\theta}_1$ or $\hat{\theta}_2$ when $\rho \rightarrow 0$ or ∞ , corresponding to target 1 or 2 dominating the other. When $\rho = 1$, indicating equal-strength targets, $\hat{\theta}$ will be the average of the two angles, $(\theta_1 + \theta_2)/2$. However, for some values of α it is possible for $\hat{\theta}$ to fall outside of the angular interval defined by θ_1 and θ_2 . The problem occurs when ϕ is close to 180° and ρ is close to 1 (see Prob. 37). Figure 7.23 shows the estimated angle $\hat{\theta}$ when $\theta_1 = -2^\circ$ and $\theta_2 = -1^\circ$. The two gray dashed lines mark these two angles, so it is desirable for $\hat{\theta}$ to fall between them. The figure confirms that the estimate equals the average for $\rho = 1$ and tends to θ_1 or θ_2 as appropriate when ρ differs from 1 by a large factor, i.e., $\rho = 0.1$ or 10. However, when ρ is relatively close to 1 and ϕ is close to 180°, $\hat{\theta}$ can fall well outside of θ_1 and θ_2 . If the two scatterers represent two scattering centers on a single target, this means the estimated AOA will be outside of the error. If $\rho = 1.1$ in this example, $\hat{\theta}$ will be 9°, a full 10° beyond θ_2 and 10.5° from the average.



FIGURE 7.23 Estimated angle for two unresolved scatterers.

Real complex targets have many scattering centers, not just two. Statistical analyses suggest that a reasonable rule of thumb for the standard deviation of $\hat{\theta}$ due to glint is of the form

$$\sigma_{\hat{\theta}_{\text{glint}}} \approx k\theta_w = k \frac{L_w}{R}$$
(7.96)

where θ_w is the angular width of the target as viewed from the radar and the scale factor *k* is variously reported to be between 0.15 and 0.35. If the width of the target as viewed from the radar is L_w , its angular extent will be $\theta_w = L_w/R$. Glint error is therefore a function of range. The combined effect of glint and noise on the angle precision is obtained by adding the variances or, equivalently, computing the RMS combination of the two precisions. Because glint error decreases with range while noise error increases with range due to declining SNR, their combination tends to produce an optimum range for a given system where the total error is minimum.

The discussion of angle measurement by lobing has been confined so far to one dimension, but angle tracking requires estimation of the target location in two angular dimensions, azimuth and elevation. Monopulse antennas provide difference outputs in both planes, while sequential lobing antennas can usually generate offset angles in both planes as well. The processing is the same in each plane. Another means of 2D angle tracking is *conical scan* or *conscan*. In this technique, an antenna feed offset from a reflector antenna axis is mechanically rotated around that axis, producing a squinted beam that continually rotates around the antenna nominal boresight. This is suggested by Fig. 7.24*a*, which shows a series of beam positions at various points in the rotation. Two point targets are also shown, one on the centerline of the rotating squinted beam and displaced vertically (conscan rotation angle of 90°), the other much nearer to the nominal boresight (one-tenth of the centerline radius) and at a conscan rotation angle of 225°. Analysis of the variation in amplitude and the angular location of the peak response reveals the location of the target. For example, the upper half of Fig. 7.24*b* suggests a target halfway between 80° and 100°, while the lower half suggests a target close to 220°. Estimating the degree of squint from the amplitude variation requires use of the two-way voltage antenna pattern (see Prob. 38).



FIGURE 7.24 Conical scan: (*a*) eight beam positions in a continuous conical scan. The gray dashed line shows the loci of the beam centers, rotating counterclockwise. Two target locations are shown; (*b*) relative amplitude of signals received for each target.

More pulses would enable finer precision in both rotation angle and degree of squint. The number of pulses per rotation is determined by the rotation rate and the radar PRF; 18 pulses are shown in <u>Fig. 7.24</u> before the cycle repeats on the 19th pulse, but real systems may have more or less than this.

The lobing techniques discussed so far are primarily for use in dedicated tracking radars or tracking modes of multipurpose radars. Scanning search radars may also utilize angle estimation algorithms as part of the search process. Consider a rotating search radar. As the beam passes over a target, that target may be detected in several sequential beam positions as suggested in Fig. 7.25*a*, and the resulting relative power measurements P[n] may appear similar to those in Fig. 7.25*b*. The centroid of the measurements is



FIGURE 7.25 Centroiding of scanning radar angle measurements: (a) the target is detected on three successive beam positions, (b) relative power of the signals received at each position and their centroid.

For the measured powers and beam indices shown, $C_P = 2.16$. This value is then translated into an estimated AOA based on the angular orientation and spacing of the measurements. Other variations on this basic idea, such as a thresholding and binary-weighted centroiding techniques having some similarities to the split-gate range tracking method, are discussed in Blair et al (2010).

7.3 Introduction to Tracking

The result of each measurement of the coordinates of a target in range, angle, or Doppler shift is an estimate of the true value of the target corrupted due to the noise in the measurement process. The estimate therefore has an associated uncertainty. More information can be obtained by collecting a series of estimates over time and combining them with some (hopefully reasonable) assumptions about the target's kinematic motion to provide an improved estimate of the true coordinates of the target over time. The true coordinates of the target are called its *state*. The state is usually considered to include the velocity and sometimes the acceleration in each spatial coordinate and may also include other aspects of the target, such as its pose relative to the radar. Usually the state components will be real-valued. *Tracking* is the process of developing the best estimate in some sense of the target state over time based on noisy measurements.

Several issues in addition to measurement noise make tracking a challenging problem. False alarms can cause tracks to deviate from the true target trajectory. Missed detections can allow the uncertainty in the track to grow. If multiple targets are present in the same vicinity, unresolved measurements or crossing paths may make it difficult to associate new measurements with the correct track. In addition, radar naturally measures position in spherical coordinates relative to the antenna boresight, whereas it is usually desirable to track a target in Cartesian coordinates relative to the radar position or a fixed reference frame such as an earth-centered coordinate system. The nonlinear coupled transformation between the two coordinate systems complicates processing.

Tracking is a subject that supports many complete texts on its own, so only a brief treatment of the basic issues is sketched here. Borrowing heavily from Kay (1993), the fairly simple idea of sequential least-squares estimation (LSE) is developed first. Next the common α - β filter for radar tracking is presented heuristically. Both the sequential least squares estimator and the α - β filter will be seen to share a similar prediction-correction structure. The Kalman filter is then presented as the optimal sequential least-mean-square estimator for a dynamically evolving system. While the extensive details required for its derivation are left to other texts, its structure is seen to be a generalization of the simpler sequential LSE problem, while the α - β filter is shown to be a special case.

Two excellent, concise introductory general references for tracking applied specifically to radar are Blair (2010) and Ehrman (2013). Another good tutorial text is Mahafza (2008). More in-depth radar-specific treatments are given in Bar-Shalom and Fortmann (1988) and Blackman (1986).

7.3.1 Sequential Least Squares Estimation

Consider a deterministic one-dimensional data sequence x[n] dependent on some parameter Θ . Suppose a set of N possibly perturbed observations z[n] of x[n] is available. The perturbation could be additive noise in the data but could also be due to other sources such as inaccuracies in the assumed data model. The *least-squares estimate* (LSE) $\hat{\Theta}$ of Θ given z[n], n = 0, 1, ..., N - 1 is defined as the value of Θ that minimizes

$$\varepsilon^{2}(\Theta) = \sum_{n=0}^{N-1} (z[n] - x[n])^{2}$$
(7.98)

 ε^2 depends on Θ through x[n]. If $x[n] = \Theta$, straightforward minimization of ε^2 by differentiation with respect to Θ leads to the sample mean as the general LSE (see Prob. 39).

The linear LSE problem is of special interest. It assumes that the data x are linearly related to Θ , $x[n] = f[n] \cdot \Theta$ for some known system function f[n] (which could be just a constant). Using this model in Eq. (7.98) and minimizing ε^2 gives (see Prob. 40)

$$\hat{\Theta} = \frac{\sum_{n=0}^{N-1} z[n] f[n]}{\sum_{n=0}^{N-1} f^2[n]}$$
(7.99)

with minimum error

$$\varepsilon^{2}(\hat{\Theta}) = \sum_{n=0}^{N-1} (z[n] - \hat{\Theta}f[n])^{2}$$

= $\sum_{n=0}^{N-1} z[n] (z[n] - \hat{\Theta}f[n]) - \hat{\Theta} \sum_{n=0}^{N-1} f[n] (z[n] - \hat{\Theta}f[n])$
= $\sum_{n=0}^{N-1} z^{2}[n] - \hat{\Theta} \sum_{n=0}^{N-1} f[n]z[n]$
(7.100)

The last step relies on the fact that the last sum in the second step is identically zero, as can be verified by substituting $\hat{\Theta}$ from Eq. (7.99) into the sum.

As a trivial example, consider least squares estimation of a scalar constant *A*. Assume f[n] = 1 for all *n* so that the observations z[n] are just perturbed measurements of *A*. Then $\Theta = A$ and from Eq. (7.99) the LSE is the sample mean $\hat{A} = \sum_{n=0}^{N-1} z[n]/N$.

The procedure described so far is called batch processing. All of the

measurements are collected and then the parameter estimate is generated in a single calculation. In tracking, new measurements are expected to occur on a regular basis and it is desirable to generate and update estimates of the target parameters as data arrives rather than wait until when (if ever) all of the measurements are collected. A procedure that allows this is *sequential least squares* estimation.

Continue the example above of estimating a constant. Let Let $\hat{A}[N-1]$ be the estimate based on all observations z[n] up to and including time N - 1, $\hat{A}[N-1] = \sum_{n=0}^{N-1} z[n]/N$. Now observe a new measurement z[n]. The new LSE of A must be

$$\hat{A}[N] = \frac{1}{N+1} \sum_{n=0}^{N} z[n]$$

$$= \frac{N}{N+1} \hat{A}[N-1] + \frac{1}{N+1} z[N]$$

$$= \hat{A}[N-1] + \frac{1}{N+1} (z[N] - \hat{A}[N-1])$$
(7.101)

The estimate can be computed recursively using either the second or third form of Eq. (7.101). The third form is particularly interesting. It shows that the updated estimate is the previous estimate augmented with a correction term based on the difference between that previous estimate and the new observation. The weight accorded the correction term declines with time, since the previous estimate incorporates more data as time proceeds.

The LSE for this problem can also be updated recursively:

$$\varepsilon^{2}(\hat{A} \mid N) = \sum_{n=0}^{N} (z[n] - \hat{A}[N])^{2}$$

$$= \sum_{n=0}^{N-1} \left[z[n] - \hat{A}[N-1] - \frac{1}{N+1} (z[N] - \hat{A}[N-1]) \right]^{2} + (z[N] - \hat{A}[N])^{2}$$

$$= \varepsilon^{2} (\hat{A} \mid N-1) - \frac{2}{N+1} \sum_{n=0}^{N-1} (z[n] - \hat{A}[N-1]) (z[N] - \hat{A}[N-1]) + \dots$$

$$\dots + \frac{N}{(N+1)^{2}} (z[N] - \hat{A}[N-1])^{2} + (z[N] - \hat{A}[N])^{2}$$

$$= \varepsilon^{2} (\hat{A} \mid N-1) + \frac{N}{N+1} (z[N] - \hat{A}[N-1])^{2}$$
(7.102)

The last line takes advantage of the fact that the term $\sum_{n=0}^{N-1} (z[n] - \hat{A}[N-1])$ is zero.

The LSE grows with time because as more data is obtained, the number of squared error terms increases.

No assumptions have yet been made about the nature of the perturbations in the observations z[n], but going forward it will now be assumed that the observations are linear functions of the parameter Θ corrupted by additive WGN, $z[n] = f[n] \cdot \Theta + w[n]$. In particular, z[n] = A + w[n] for the problem of estimating a constant in noise.

A useful extension to the sequential LSE estimator is to allow weighting of the data in computing the MSE. Weighting would be appropriate for the constant-in-WGN model if the noise samples are uncorrelated but may have a time-varying variance $\sigma^2[n]$. Choose the weight on the *n*th term of the MSE defined in Eq. (7.98) as $w_n = 1/\sigma^2[n]$. This choice places greater weight on the least noisy measurements. Under these conditions it can be shown that the variance of the estimate \hat{A} is

$$\operatorname{var}(\hat{A}[N]) = \frac{1}{\sum_{n=0}^{N} w_n} = \frac{1}{\sum_{n=0}^{N} (1/\sigma^2[n])}$$
(7.103)

The estimate update for this problem is

$$\hat{A}[N] = \hat{A}[N-1] + \frac{1/\sigma_N^2}{\sum_{n=0}^N (1/\sigma^2[n])} (z[N] - \hat{A}[N-1])$$
$$= \hat{A}[N-1] + K[N] (z[N] - \hat{A}[N-1])$$
(7.104)

where the gain K[N] is

$$K[N] = \frac{1/\sigma^{2}[N]}{\sum_{n=0}^{N} (1/\sigma^{2}[n])} = \frac{\operatorname{var}(\hat{A}[N-1])}{\operatorname{var}(\hat{A}[N-1]) + \sigma^{2}[N]}$$
(7.105)

It follows that

$$\operatorname{var}(\hat{A}[N]) = (1 - K[N]) \operatorname{var}(\hat{A}[N-1])$$

(7.106)

Finally, the LSE update can be put in the form

$$\varepsilon^{2}(\hat{A} \mid N) = \varepsilon^{2}(\hat{A} \mid N-1) + \frac{\left(z[N] - \hat{A}[N-1]\right)^{2}}{\operatorname{var}(\hat{A}[N-1]) + \sigma^{2}[N]}$$

(7.107)

Equations (7.104) to (7.107) define the sequential LSE for a constant in additive WGN with non-stationary power. In particular, Eqs. (7.104) and (7.105) generalize Eq. (7.101), and Eq. (7.107) generalizes (7.102).

To execute this estimator, it is necessary to specify the initial values $\hat{A}[0]$ and $\operatorname{var}(\hat{A}[0])$. The iteration can be initialized by choosing $\hat{A}[0] = z[0]$, which is the MLE of *A* based on the single measurement. The initial variance estimate is chosen as $\operatorname{var}(\hat{A}[0]) = \sigma^2[0]$, implying that the initial noise power is known. Equations (7.105), (7.104), and (7.106) are then exercised in that order to update to time step N = 1, and the process repeated. If no estimate of $\sigma^2[0]$ is known, $\operatorname{var}(\hat{A}[0])$ can be set to a "large" value. This in effect states that the confidence in the initial value $\hat{A}[0]$ is low so that the estimator places little weight on it. This will result at N = 1 in $K[1] \approx 1$, $\hat{A}[1] \approx z[1]$, and a reduced variance estimate going forward. Another method is to use the first several observations in a batch process to develop an estimate of \hat{A} and $\operatorname{var}(\hat{A})$ and then switch to the sequential estimator for subsequent updates.

Figure 7.26 illustrates the behavior of this sequential LSE for A = 10, $\sigma^2 \sigma N$] = $\sigma^2 = 1$ for all N, $\hat{A}[0] = z[0]$, and $\operatorname{var}(\hat{A}[0])$. It is easily seen in this case (see Prob. 41) that K[N] = 1/(N+1) and $\operatorname{var}(\hat{A}[N]) = \sigma^2/(N+1) = 1/(N+1)$ also. Part *a* of the figure shows the gain and estimate variance versus the time step *N*. The curves overlay one another. Part *b* shows the estimate $\hat{A}[N]$ of the actual value A = 10.



FIGURE 7.26 Sequential least squares estimation of a constant in additive WGN with constant variance: (a) estimator gain and variance, (b) estimated value.

Figure 7.27 presents a similar example but with time-varying noise power. Specifically, the same specific sequence of random noise samples is used, except that the noise samples are multiplied by 10 (a 100× variance increase) from time steps 80 to 160 and by $\sqrt{1/10}$ (a 10× variance decrease) from time steps 240 through 320. Equation (7.105) shows that the sudden increase in noise

variance will decrease the gain, while a decrease in noise variance will increase the gain. These effects are very evident in Fig. 7.27*a*. During the period of low estimator gain, the estimate variance changes very little; essentially, the new data is being discounted because it is so noisy. From steps 161 to 239 the noise in the data returns to its original value, the gain increases, and the estimate variance resumes its decrease. At step 240, the data noise is decreased and the gain increases to take advantage, allowing the estimate variance to decrease more rapidly. Figure 7.27*b* shows the actual estimate. It is clear that the estimate changes little when the gain is reduced from steps 80 to 160. It follows the data more actively from 161 to 239, and then again becomes more stable at 240 when the gain is lowered.


FIGURE 7.27 Sequential least squares estimation of a constant in additive WGN with time-varying variance: (a) estimator gain and variance, (b) estimated value. See text for details.

The generalization of the sequential LSE for a $P \times 1$ vector parameter is

$$\hat{\Theta}[n] = \hat{\Theta}[n-1] + \mathbf{k}[n](z[n] - \mathbf{f}^{T}[n]\hat{\Theta}[n-1])$$

$$\mathbf{k}[n] = \frac{\mathbf{M}[n-1]\mathbf{f}[n]}{\mathbf{f}^{T}[n]\mathbf{M}[n-1]\mathbf{f}[n] + \sigma^{2}[n]}$$

$$\mathbf{M}[n] = (\mathbf{I} - \mathbf{k}[n]\mathbf{f}^{T}[n])\mathbf{M}[n-1]$$
(7.108)

Here **M** is the $P \times P$ covariance matrix of the estimate $\hat{\Theta}$. With the notation change $n \to N$ and the equivalences $\hat{\Theta} \to \hat{A}$, $\mathbf{f} \to 1$, $\mathbf{k}[n] \to K[n]$, and $\mathbf{M}[n] \to \operatorname{var}(\hat{A}[N])$, the similarity between Eqs. (7.108), (7.104), (7.106) is obvious. Similar to the scalar example above, one simple initialization is to choose $\hat{\Theta}[0] = 0$ and $\mathbf{M}[0] = \kappa \mathbf{I}$ for some large value of κ , effectively discounting the initial covariance estimate.

7.3.2 The α - β Filter

Track filtering is readily viewed as an estimation problem. The radar makes noisy coordinate measurements at a series of discrete times t_n . The track filter has two goals: to *smooth* the measurements to provide an improved estimate of the coordinates at time n, and to use those smoothed estimates to *predict* the target position at the time t_{n+1} of the next planned measurement. A variety of filters are commonly used in the radar tracking community: the α - β filter, the α - β - γ filter, the Kalman filter, and the extended Kalman filter. These are all closely related to the sequential LSE methods described above. New techniques such as "particle filters" are continuously emerging.

Two further generalizations of the sequential LSE results given so far are important in developing effective sequential estimation procedures for track filtering. The first is that the target model is dynamic, i.e., the target parameters (position, velocity, etc.) change with time according to a defined model such as constant-velocity or constant-acceleration motion. A more fundamental generalization is to model the target parameters at a given time as random variables with a prior PDF instead of as deterministic quantities. The actual target parameters over time are then a single realization of a random process rather than a deterministic time series. This approach allows for uncertainty in the target model. The goal of the estimator is to minimize the expected value of squared error between the estimate and the realization, averaged over the joint PDF of the data and the target parameters. This modeling approach, called Bayesian estimation, is subtly different from the classical estimation considered so far. In either viewpoint the data z are random, typically due to additive noise in the measurements. In classical estimation the target parameters are not random, while in Bayesian estimation they are. More discussion and examples of the difference in philosophies are given in Kay (1993).

Before stating the general solution to the sequential Bayesian LSE problem, consider the problem of tracking a target in a single dimension x with the simplest commonly used track filter, the α - β filter. The α - β filter is designed for a constant-velocity target. Denoting the actual target position and velocity at time n as x[n] and x[n], a true constant-velocity model is

$$x[n] = x[n-1] + \dot{x}[n-1]T$$
(7.109)

where *T* is the interval between track update times.² In practice, it is useful to include a *process noise* u[n] with variance σ_{π}^2 that allows for deviations from perfectly constant-velocity motion due primarily to target maneuvers, but also for such effects as turbulence or cross-winds affecting an aircraft. Process noise is not a measurement noise; rather, it is part of the model of the target state evolution. A typical model is to consider u[n] to be a piecewise-constant acceleration between each time step. The contribution of a constant acceleration to the position and velocity then gives the more complete "constant velocity" model:

$$x[n] = x[n-1] + T\dot{x}[n-1] + \frac{T^2}{2}u[n]$$
$$\dot{x}[n] = \dot{x}[n-1] + Tu[n]$$
(7.110)

The available measurements are the observations of only the target position,

$$z[n] = x[n] + w[n]$$
(7.111)

where w[n] is the measurement noise with variance σ_{w}^{2} . The noises u[n] and w[n] are usually modeled as independent zero mean WGN processes.

Define the following estimated quantities:

- Predicted position at time *n* given all measured data up through $\hat{x}[n|n-1]$ time n-1. Does not include knowledge of the new measurement z[n].
 - Estimated (smoothed or corrected) position at time *n* given all
- $\hat{x}[n|n]$ measured data up through time *n*. This is the updated estimate based on the new information z[n].
- $\hat{x}[n|n-1]$ Predicted velocity at time *n* given all measured data up through time n-1.
- Estimated (smoothed or corrected) velocity at time *n* given all measured data up through time *n*.

The α - β filter equations can be grouped in two stages, the *prediction* stage and the *innovation* stage. The prediction stage uses data through time n - 1 and the assumed kinematic model to predict the target's position at time n. The innovation stage then applies the new measurement z[n] at time n to correct or smooth the prediction. The equations are

Prediction stage:

$$\hat{x}[n \mid n-1] = \hat{x}[n-1 \mid n-1] + T \cdot \hat{x}[n-1 \mid n-1]$$
$$\hat{x}[n \mid n-1] = \hat{x}[n-1 \mid n-1]$$

Correction stage:

$$e[n] = z[n] - \hat{x}[n|n-1]$$

$$\hat{x}[n|n] = \hat{x}[n|n-1] + \alpha \cdot e[n]$$

$$= (1-\alpha)\hat{x}[n|n-1] + \alpha \cdot z[n]$$

$$\hat{x}[n|n] = \hat{x}[n|n-1] + \beta \frac{e[n]}{T}$$

$$= (1-\beta)\hat{x}[n|n-1] + \beta \left(\frac{z[n] - \hat{x}[n-1|n-1]}{T}\right)$$
(7.112)

The constant velocity assumption is evident in the first equation of the prediction stage. Notice that the first form of the correction equation for $\hat{x}[n|n]$ has the same structure as the sequential LSE of Eq. (7.104) with a constant gain α taking the role of the more general time-varying gain K[n]. A typical practical initialization collects the first two measurements, z[0] and z[1], forms

$$\hat{x}[1|1] = z[1]$$

$$\hat{x}[1|1] = (z[1] - z[0])/T$$
(7.113)

and begins iterating Eq. (7.112) at n = 2.

The quantity e[n] is called the *residual* or the *innovation*. It represents new information available as a result of the new measurement z[n]. The second form of the last two equations makes the role of α and β clear. α controls the relative weight assigned to new data versus prediction from prior data in updating the target position. Smaller values of α place more weight on the prediction and less on the new data and are appropriate when the new data has higher noise. β plays a similar role for updating velocity, though there is a subtle difference in the new data term.

Figure 7.28 illustrates the α - β filter in one dimension. The target is presumed to start at x = 1000 meters at time n = 0 and move in the +x direction with a constant velocity $\dot{x} = 50$ m/s for the time steps n = 0, ..., 30. For the next nine steps the target decelerates at a constant rate such that at n = 39 the velocity reaches zero. The target then remains at x = 2750 meters and $\dot{x} = 0$ m/s through n = 69. While the deceleration violates a strict constant-velocity model, it is permitted by the introduction of acceleration process noise in Eq. (7.110).



FIGURE 7.28 Target tracking in one dimension with an α - β filter: (*a*) effect of varying α on position estimate, (*b*) effect of varying β . See text for details.

The noisy measurements z[n] are shown in Fig. 7.28*a* by the diamond shapes and the actual position by the gray line in the background. The measurement noise is stationary additive WGN with $\sigma_w = 100$ m. The filter was initialized with $\hat{x}[1|1] = z[0]$ and $\hat{x}[1|1] = (z[1] - z[0])/T$. The black dashed line is the smoothed track $\hat{x}[n|n]$ obtained with $\alpha = 0.15$ and $\beta = 0.1$, while the solid black line is the track $\hat{x}[n|n]$ obtained by filtering exactly the same data samples with the same $\beta = 0.1$ but $\alpha = 0.85$. As expected, small values of α de-emphasize the current measurement in favor of the prediction, producing smoother tracks but requiring longer to correct deviations. The larger value of *a* puts more emphasis on the current measurement, producing a noisier "smoothed" track but correcting more quickly.

Figure 7.29*b* shows the effect of β on the position estimate for a fixed $\alpha = 0.2$. Small values of β put more emphasis on the previous estimate of velocity, smoothing the velocity estimate more strongly and contributing to a smoother track. Larger values of β put more emphasis on the velocity estimate based on the innovation. Combined with a relatively small α , this tends to produce oscillatory behavior in the position estimate and is generally to be avoided. Although not shown here, the estimates of velocity become less smooth as β increases.



FIGURE 7.29 Comparison of α - β and Kalman filters on same data: (*a*) smoothed position estimate, (*b*) smoothed velocity estimate. See text for details.

It is useful to express the α - β filter in a state space representation. Define the state vector for tracking in one dimension as $x[n] = [x[n] \ \dot{x}[n]]^T$. The target kinematic and measurement models of Eqs. (7.110) and (7.111) are then

$$\mathbf{x}[n] = \mathbf{F} \cdot \mathbf{x}[n-1] + \mathbf{g} \cdot u[n]$$

$$= \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \mathbf{x}[n-1] + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} u[n]$$

$$\mathbf{z}[n] = \mathbf{h}^T \mathbf{x}[n] + w[n]$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \mathbf{x}[n] + w[n]$$
(7.114)

The filter equations [Eq. (7.112)] become

Prediction stage:

$$\hat{\mathbf{x}}[n|n-1] = \mathbf{F} \cdot \hat{\mathbf{x}}[n-1|n-1]$$

Correction stage:

$$\hat{\mathbf{x}}[n|n] = \hat{\mathbf{x}}[n|n-1] + \mathbf{k} \left(\mathbf{z}[n] - \mathbf{h} \cdot \hat{\mathbf{x}}[n|n-1] \right)$$
$$= \hat{\mathbf{x}}[n|n-1] + \begin{bmatrix} \alpha \\ \beta/T \end{bmatrix} \left(\mathbf{z}[n] - \mathbf{h} \cdot \hat{\mathbf{x}}[n|n-1] \right)$$
(7.115)

In these equations, z, u, and w are scalars; \mathbf{x} , \mathbf{h} , \mathbf{v} , $\hat{\mathbf{x}}$, \mathbf{g} , and \mathbf{k} are $P \times 1$ vectors; and \mathbf{F} is a $P \times P$ matrix. Again, the similarity between the first version of the correction stage in Eq. (7.115) and the correction operation in Eq. (7.108) is clear.

7.3.3 The Kalman Filter

The new element in the α - β filter as compared to the sequential LSE is the prediction stage, which allows the state variables to evolve over time according to the kinematic model instead of being fixed as in the constant-in-WGN example. However, it is still a fixed-gain filter. The idea of sequential least squares estimation can be extended to apply to a dynamic model of the target motion by introducing a time-varying gain that minimizes the MSE at each step. The result is the *Kalman filter* (KF). The KF derivation details are beyond the scope of this text, but for a scalar observation the KF equations are

Kinematic and observation models:

$$\mathbf{x}[n] = \mathbf{F} \cdot \mathbf{x}[n-1] + \mathbf{g} \cdot u[n]$$
$$z[n] = \mathbf{h}^{T}[n] \cdot \mathbf{x}[n] + w[n]$$

Prediction and prediction MSE:

$$\hat{\mathbf{x}}[n|n-1] = \mathbf{F} \cdot \hat{\mathbf{x}}[n-1|n-1]$$
$$\mathbf{M}[n|n-1] = \mathbf{F} \cdot \mathbf{M}[n-1|n-1] \cdot \mathbf{F}^{T} + \mathbf{g} \cdot \mathbf{g}^{T} \sigma_{u}^{2}$$

Kalman gain:

$$\mathbf{k}[n] = \frac{\mathbf{M}[n|n-1]\mathbf{h}[n]}{\sigma_w^2 + \mathbf{h}^T[n]\mathbf{M}[n|n-1]\mathbf{h}[n]}$$

Correction and estimate MSE:

$$\hat{\mathbf{x}}[n|n] = \hat{\mathbf{x}}[n|n-1] + \mathbf{k}[n](\mathbf{z}[n] - \mathbf{h}^{T}[n]\hat{\mathbf{x}}[n|n-1])$$
$$\mathbf{M}[n|n] = (\mathbf{I} - \mathbf{k}[n]\mathbf{h}^{T}[n])\mathbf{M}[n|n-1]$$

In addition to a time-varying gain, notice that the KF also allows for a timevarying observation noise by allowing **h** to vary. **M** is a $P \times P$ matrix that describes the MSE of the predicted and corrected signals. The filter can be initialized by choosing $\hat{x}[0|0] = 0$ and **M** $[0|0] = \kappa \mathbf{I}$ for some large κ , or by using the same procedure described above for the α - β filter with the additional step of setting an initial value for **M**:

$$\hat{\mathbf{x}}[0|0] = \begin{bmatrix} z[0] \\ 0 \end{bmatrix}, \quad \hat{\mathbf{x}}[1|1] = \begin{bmatrix} z[1] \\ (z[1] - z[0])/T \end{bmatrix}$$
$$\mathbf{M}[1|1] = \begin{bmatrix} \sigma_w^2 & \sigma_w^2/T \\ \sigma_w^2/T & 2\sigma_w^2/T^2 \end{bmatrix}$$
(7.117)

This choice for **M** results from noting that the variance of z[n] is σ_{w}^{2} for all *n*, but that values of z[n] at different values of *n* are uncorrelated. Computing the covariance matrix of $\hat{x}[1|1]$ then gives **M** [1 | 1] as shown.

The scalar observation Kalman filter has a number of interesting properties. These include

• No matrix inversions are required for its execution. This would not be the case for batch estimation.

- The KF is a time-varying linear filter.
- The KF provides its own performance metric in the form of the MSE matrix M[n | n]. Furthermore, its computation does not depend on the actual observations so it can be precomputed, as can the Kalman gain k[n]. The availability of M is helpful in gating and association, to be discussed in the next section.
- The KF exhibits the same prediction-correction structure seen in both the nondynamic sequential estimator and the α - β filter.
- If the driving noises *u* and *w* are stationary, the KF will asymptotically approach a steady state in which it becomes a time-invariant linear filter and can be viewed as a whitening filter.
- As will be seen, the α - β filter is an example of a steady-state KF.
- The same KF equations apply if the system model is extended to allow time-varying versions of F, g, and ^{o²}. This makes it possible to adapt to variable update times, missed updates, and other complications.

The equations above apply to a scalar observation, i.e., a one-dimensional measurement. Normally a target is tracked in two or three dimensions using measurements of position in each. The KF can be further generalized to allow for an $M \times 1$ vector observation $\mathbf{z}[n]$ and an $R \times 1$ vector process noise $\mathbf{u}[n]$. The equations are

Kinematic and observation models:

 $\mathbf{x}[n] = \mathbf{F} \cdot \mathbf{x}[n-1] + \mathbf{G} \cdot \mathbf{u}[n]$ $\mathbf{z}[n] = \mathbf{H}[n] \cdot \mathbf{x}[n] + \mathbf{w}[n]$

Prediction and prediction MSE:

$$\hat{\mathbf{x}}[n|n-1] = \mathbf{F} \cdot \hat{\mathbf{x}}[n-1|n-1]$$
$$\mathbf{M}[n|n-1] = \mathbf{F} \cdot \mathbf{M}[n-1|n-1] \cdot \mathbf{F}^{T} + \mathbf{GS}_{\mathbf{u}}\mathbf{G}^{T}$$

Kalman gain:

$$\mathbf{K}[n] = \mathbf{M}[n|n-1]\mathbf{H}^{T}[n] (\mathbf{S}_{w}[n] + \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^{T}[n])^{-1}$$

Correction and estimate MSE:

$$\hat{\mathbf{x}}[n|n] = \hat{\mathbf{x}}[n|n-1] + \mathbf{K}[n] (\mathbf{z}[n] - \mathbf{H}[n] \hat{\mathbf{x}}[n|n-1])$$
$$\mathbf{M}[n|n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{H}[n])\mathbf{M}[n|n-1]$$
(7.118)

F and **M** are $P \times P$; **G** is $P \times R$; **H** is $M \times P$; **K** is $M \times P$; **x** and $\hat{\mathbf{x}}$ are $P \times 1$; and **z** and **w** are $M \times 1$. $\mathbf{S}_{\mathbf{u}}$ and $\mathbf{S}_{\mathbf{w}}$ are the $R \times R$ and $M \times M$ covariance matrices of the noise processes **u** and **w**, which are assumed to each be zero mean uncorrelated Gaussian, independent of one another. All of the previously listed properties still apply, except that a matrix inversion is now required to compute **K**.

Return now to the scalar observation case of Eq. (7.116). Over time, the Kalman filter will asymptotically approach a steady-state condition in which the gain **k** is constant. The prediction MSE $\mathbf{M}[n|n-1]$ and smoothed estimate MSE $\mathbf{M}[n|n]$ will also approach constants, with the former larger than the latter. The values of the steady-state gain can be determined in terms of the *tracking index* Γ (Kalata, 1984), also called the *maneuverability index*:

$$\Gamma \equiv \frac{\sigma_u T^2}{\sigma_w} \tag{7.119}$$

 Γ is a measure of the target position uncertainty due to the acceleration noise relative to that due to the measurement noise. It can be shown that for the single-coordinate case considered here, the steady-state KF gains $\mathbf{k} = [k_x \ k_t]$ obey the following relationships:

$$\Gamma^{2} = \frac{k_{x}^{2}}{1 - k_{x}}$$

$$k_{x} = 2(2 - k_{x}) - 4\sqrt{1 - k_{x}}$$
(7.120)

These equations can be solved to find the values of k_x and k_x for a given value of Γ (Blair, 2010):

$$k_{x} = -\frac{\Gamma^{2}}{8} - \Gamma + \frac{1}{8}(\Gamma + 4)\sqrt{\Gamma^{2} + 8\Gamma}$$

$$k_{\dot{x}} = +\frac{\Gamma^{2}}{4} + \Gamma - \frac{1}{4}\Gamma\sqrt{\Gamma^{2} + 8\Gamma}$$
(7.121)

Not surprisingly, the α - β filter is a special case of the Kalman filter, obtained by choosing the gain vector to be a fixed value $\mathbf{k} = [\alpha \ \beta/T]^T$. A

common way to choose α and β is to use the steady-state Kalman gains, i.e., $\alpha = k_x$ and $\beta = k_x T$.

Figures 7.29 and 7.30 show an example similar to the one of Fig. 7.28. Both a Kalman filter initialized according to Eq. (7.117) and an α - β filter were run on the same data samples. The process noise standard deviation was set at $\sigma_v^2 = a^2/2$, where *a* is the maximum acceleration (-5.5566 m/s² in this example). The resulting value is $\sigma_v = 3.93$ m/s². The values of $\alpha = 0.2443$ and $\beta = 0.0342$ were chosen based on Eq. (7.121).



FIGURE 7.30 (*a*) α - β and Kalman filters gains for example of Fig. 7.29: (*b*) kalman filter mean-square error estimate for position, M[1,1]. See text for details.

Figure 7.29*a* compares the smoothed trajectory estimates from the two filters, while Fig. 7.29*b* shows the velocity estimates. Both filters track the first two measurements exactly due to the initialization process. This particular example was chosen to emphasize the disadvantage of the fixed-gain α - β filter. The first two samples produce an initial velocity estimate that differs greatly from the actual target velocity. Consequently, the α - β estimate exhibits very large position errors over the first 20 time steps. The Kalman filter's variable gain recovers much more quickly and exhibits much lower errors. This is also seen in the velocity estimates, which track the correct values of 50 or zero in the appropriate regions. However, the KF converges toward the initial velocity of 50 m/s much sooner than does the α - β filter. If another random realization was used where the first two observations produced an initial velocity estimate closer to the true value, the results of the two filters would be much more comparable.

As seen in Fig. 7.30*a*, by time step 20 the Kalman filter gains have reached steady state. The fixed gains α and β were chosen to match these steady-state values, so both filters behave virtually identically from this point on, including the overshoot and recovery during the target deceleration from time steps 31 to 39. Finally, Fig. 7.30*b* shows the variation of M[1, 1], which is the KF computation of the variance of the estimated position, for the first 30 time steps needed to attain steady-state behavior. Half-integer indices show the value of the prediction MSE M[n|n - 1], while integer indices are the values of the smoothed MSE M[n|n]. As expected, the MSE is always increased by the prediction step, and then decreased by the correction step when new data is incorporated.

The failure of the KF to respond to the deceleration in steps 31 to 39 more quickly than the α - β filter can be remedied by a number of techniques that combine multiple filters designed for different target dynamics. One popular approach is *interacting multiple models* (IMM). Multiple model techniques run multiple track filters in parallel, each with different target dynamics represented by different process noise variances. At each step, the model yielding the smallest innovation is used to update the track. IMM is a particular technique in this class that attempts to assign relative probabilities to each model and blend their track updates. An introduction to the IMM approach and others is given in Ehrman (2013).

Just as the choice of α and β plays an important role in the tradeoff of smoothing and convergence time in the α - β filter, the choice of the process noise

variance σ_v^2 plays a similar role in the Kalman filter. Figure 7.31 illustrates this effect on the position gain $\mathbf{k}[1] = k_x$ for an example similar to those preceding. A s σ_v^2 increases, the steady-state gain rises and the filter converges more quickly, resulting in less smoothing and more rapid adaptation, the same effect observed as α is increased in the α - β filter. Selection of the process noise variance is discussed in Blair (2012).



FIGURE 7.31 Effect of process noise variance on position gain. See text for details.

A technique called *gain scheduling* can be applied to improve the performance of the α - β filter during the startup transient without resorting to the more complex KF. Let α_{ss} and β_{ss} be the steady-state values of α and β obtained from the steady-state KF gains of Eq. (7.121). Similar to the 1/(N + 1) gain of the least squares estimator of a constant in stationary WGN (Fig. 7.26 and Prob. 42), the gains for least squares estimation of position and velocity for the constant-velocity target model as a function of the number of data samples *n* can be computed and used until the steady-state value is reached (Blair, 2010):

$$\alpha[n] = \max\left\{\frac{2(2n+1)}{(n+1)(n+2)}, \alpha_{ss}\right\}$$

$$\beta[n] = \max\left\{\frac{6}{(n+1)(n+2)}, \beta_{ss}\right\}$$

(7.122)

Notice that $\alpha[n]$ is approximately proportional to 1/n, similar to the earlier sequential LSE analysis, while $\beta[n]$ is approximately proportional to $1/n^2$.

The Kalman filter is based on a linear model for both the target dynamics and the sensor observation. As noted in <u>Chap. 1</u>, many radars naturally measure three-dimensional target position in the modified spherical coordinate system relative to the radar antenna of range, azimuth angle θ , and elevation angle ϕ . Some radars, particularly planar phased array systems, measure range and two positions along the *sine space* coordinates *u* and *v*, where $u = \cos\phi \cdot \cos\theta$ and *v* $= \cos\phi \cdot \sin\theta$ are the direction cosines from the antenna center to the target. On the other hand, targets are usually tracked in a stabilized Cartesian *x-y-z* system referenced to the radar system or possibly to a fixed coordinate system such as earth-centered. Because the transformation from spherical to Cartesian coordinates is nonlinear, the observation equation, state equation, or both can be nonlinear.

Consider a two-dimensional example where the radar measures the polar coordinates of range *R* and azimuth angle θ . It is desired to track the target in the Cartesian coordinates *x* and *y*, so the state vector is $\mathbf{x} = [x \ y]^T$. At the *n*th time step the radar makes noisy measurements:

$$\hat{R}(n) = R[n] + w_R[n] = \sqrt{x^2[n] + y^2[n]} + w_R[n]$$
$$\hat{\theta}[n] = \theta[n] + w_{\theta}[n] = \arctan(y[n]/x[n]) + w_{\theta}[n]$$
(7.123)

Clearly the measurement vector $\mathbf{z} = [\hat{R} \quad \hat{\theta}]^T$ cannot be written as a linear function of the state vector, $\mathbf{z} = \mathbf{F} \cdot \mathbf{x} + \mathbf{w}$, so the KF cannot be applied as described so far.

What if the target were tracked in the more natural polar coordinates? In this case the measurement equation will be linear. Assume the target travels at constant velocity with components v_x and v_y and ignore any process noise. The target motion in *x*-*y* coordinates satisfies

$$x[n] = x[n-1] + v_x T$$

$$y[n] = y[n-1] + v_y T$$
(7.124)

The range becomes

$$R[n] = \sqrt{x^{2}[n-1] + y^{2}[n-1] + 2v_{x}Tx[n-1] + 2v_{y}Ty[n-1] + (v_{x}^{2} + v_{y}^{2})T^{2}}$$

= $\sqrt{R^{2}[n-1] + 2R[n-1]T\{v_{x}\cos(\theta[n-1]) + v_{y}\sin(\theta[n-1])\} + (v_{x}^{2} + v_{y}^{2})T^{2}}$
(7.125)

The state update equation is now a nonlinear function of *R* and θ , so again the KF cannot be applied.

The *extended Kalman filter* (EKF) is a common technique for dealing with this problem. The basic idea of the EKF is to linearize a nonlinear state update equation around the current state estimate, and linearize a nonlinear observation equation around the current state prediction. The linear KF is then applied to the resulting approximate system equations. This is a dynamic linearization that must be repeated at every time step as the state estimates evolve. The EKF has no optimality properties, but is nonetheless a common technique as well as the basis for numerous extensions. More information on the EKF is available in the references cited at the beginning of this section.

Many other approaches have been proposed to address nonlinear Bayesian estimation problems, including the "unscented" KF, particle filters, nonlinear recursive filters, batch filtering methods, and numerical solution of the Fokker-Planck equation. A good overview of their philosophies, advantages, and disadvantages and a good initial bibliography is given in Daum (2005).

7.3.4 The Tracking Cycle

Track filtering is just one element of the overall target tracking cycle depicted in Fig. 7.32. The radar produces a new set of detections at every CPI or dwell. A "detection" generally consists of the estimated position in range, angle, and possibly velocity; an estimate of the measurement error in each dimension; and a time stamp. Other metadata may also be included, such as estimated SIR of the detection. Since more than one target may be detected and tracked at a time, the first step is to *associate* detections with existing tracks, i.e., determine which detection is used to update that track. In dense target scenarios this can be a daunting task. It will be discussed shortly, after consideration of the other elements of the tracking cycle.



FIGURE 7.32 The tracking cycle.

Any detections that cannot be associated with existing tracks are candidates for new targets and are therefore sent to a *track initiation* process. An electronically scanned radar will typically transmit a confirmation CPI or dwell to verify the detection and ensure it is not a false alarm. This is generally not practical with rotating radars. In either case, an M-of-N logic is applied over the next N dwells to decide whether to promote a tentative track to a confirmed track. More sophisticated sequential probability analysis or detection pattern recognition can also be used for this function (Blackman, 1986). Once a new track is established, it is entered into the radar database and its initial state and covariance established using an initialization procedure appropriate for the track filter that will be used.

The measurement data associated with each track is next passed into the track filter for that target. The filter first predicts the state at the current measurement time and then smoothes it using the innovation based on the new measurement. A track quality score is often maintained for each track based on the measurement-to-prediction distance for each association, normalized by the covariance. If the score exceeds a preestablished upper threshold it may indicate incorrect association or a poor target model. If it falls below a lower threshold, it may indicate that the target model process noise is too large.

Tracks must eventually be dropped, for instance when a target moves outside the detection range of the radar. For a single missed detection the track filter can be used to *coast* the track by simply performing another update without the correction step, or recomputing a single update from the last detection using a longer update time *T*. In practice, the track filter may be allowed to coast for a small number of misses. However, failure to associate a new detection with a track for some minimum number of updates will trigger deletion of the track. A more sophisticated approach than a hard threshold on the number of misses would again apply sequential probability techniques or threshold the track quality metric. When a track is deleted the track database must be updated. A notification may be sent to the end user consumer of the track data as well.

Before discussing the data association stage, it is useful to introduce the idea of statistical distance. Figure 7.33 illustrates the statistics of two measurements that might be associated with a predicted value of x = 30. One has a mean $\mu_{x1} = 20$ and a standard deviation $\sigma_{x1} = 5$, while the other has $\mu_{x2} = 50$ and $\sigma_{x2} = 20$. Which is closer to the predicted value and should be used to smooth the prediction? μ_{x1} is only 10 units away from the prediction while μ_{x2} is 20 units away. However, the first measurement is $d_1 = \mu_{x1}/\sigma_x l = 2$ standard deviations away from the prediction while the second is $d_2 = \mu_{x2}/\sigma_{x2} = 1$

standard deviation away. Consequently, the likelihood that the predicted value of 30 could have arisen from the estimated distribution of the first measurement is much less than the likelihood that it could have arisen from the second. The second measurement is said to be statistically closer to the track and should probably be associated with the track.



FIGURE 7.33 Statistical distance.

A generalization of this simple calculation for tracking in multiple dimensions is the *Mahalanobis distance* $d_M[n]$ applied to the innovation $\mathbf{i}[n]$ in the vector observation KF, given by the calculations:

$$\mathbf{i}[n] = \mathbf{z}[n] - \mathbf{H}[n]\hat{\mathbf{x}}[n|n-1]$$

$$\mathbf{S}_{\mathbf{i}}[n] = \mathbf{S}_{\mathbf{w}}[n] + \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^{T}[n]$$

$$d_{M}^{2}[n] = \mathbf{i}^{T}[n](\mathbf{S}_{\mathbf{i}}[n])^{-1}\mathbf{i}[n]$$
(7.126)

In the scalar observation case this reduces to the heuristic measure used above, $d_M[n] = i[n]/[i[n]]$. Notice that the expression for Si[n] is the denominator of the Kalman gain calculation in Eq. (7.118). It combines a term S_w[n] for the covariance of the measurements with the covariance of the state prediction M[n|n-1] as observed through H[n].

Another distance measure is the log-likelihood function for the hypothesis that the prediction and measurement stem from the same object. Assuming

Gaussian statistics, this is given by

$$\ln \Lambda = \frac{1}{2} \ln[2\pi \det(\mathbf{S}_i)] + \frac{1}{2} \mathbf{i}^T \mathbf{S}_i^{-1} \mathbf{i} = \frac{1}{2} \ln[2\pi \det(\mathbf{S}_i)] + \frac{1}{2} d_M$$
(7.127)

where the time index *n* has been temporarily dropped from the notation for clarity. This measure is said to be more robust against "track stealing" by young tracks, which tend to exhibit pessimistic (overly large) covariances when only a few updates have been made (Ehrman, 2013). The determinant term increases Λ for large covariances, reducing this effect.

Consider an arbitrary covariance matrix S in a multidimensional space. Suppose the eigenvectors of the $P \times P$ covariance matrix S are $\{e_p\}$ with corresponding eigenvalues $\{\lambda_{\nu}\}$. Assuming Gaussian statistics, these describe a P-dimensional joint Gaussian PDF. Because S will be symmetric the eigenvectors will be mutually orthogonal. The surface describing the standard deviation of the PDF is an ellipsoid having its axes in the directions $\{e_p\}$ with one-sided lengths $\{\sqrt{\lambda_t}\}$. Figure 7.34*a* illustrates this for P = 2 dimensions. Figure 7.34b is a notional illustration of the covariance ellipses of a predicted state and the measurements corresponding to two different detections. The innovation is the vector connecting the prediction to whichever measurement is associated with it. The statistical distance between the measurement and a track prediction depends on the orientation and size of the covariance ellipsoids of both the measurements and the track prediction. In this notional example, both measurements are within the 1- σ ellipse of the prediction. The prediction is within the 1- σ ellipse of $\mathbf{z}_1[n]$ but not of $\mathbf{z}_2[n]$. Consequently, $\mathbf{z}_1[n]$ is statistically closer to $\hat{\mathbf{x}}[n|n-1]$ than is $\mathbf{z}_2[n]$. The Mahalanobis distance formalizes this calculation.



FIGURE 7.34 (*a*) Two dimensional ellipsoid defined by the eigenstructure of a 2D covariance matrix, (*b*) covariance ellipsoids for measurements and

prediction determine the statistical distance.

Statistical distance ideas are important in developing the algorithms for *gating* and *assignment* in the "associate detections" stage of the track cycle. Different classes of radars operating in different scenarios may have a number of detections at a given time step that is less than (sometimes much less than), equal to, or greater than the number of tracks. The first step in determining which detections will be used to update which tracks (assignment) is usually gating, which seeks to reject highly implausible pairings so as to reduce the number of combinations that must be considered in the assignment process. Gating rejects any pairing for which the measurement exceeds some threshold distance from the track update, but accepts any measurements within that distance as potentially valid. Common approaches are spherical, rectangular, and ellipsoidal gates:

Spherical gate:

$$\sqrt{(x_p - x_m)^2 + (y_p - y_m)^2 + (z_p - z_m)^2} < d_{\text{spherical}}$$

Rectangular gate:

$$\begin{vmatrix} x_p - x_m \\ d_x \end{vmatrix} < d_R, \text{ or } \begin{vmatrix} y_p - y_m \\ d_y \end{vmatrix} < d_y \begin{vmatrix} z_p - z_m \\ d_z \end{vmatrix} < d_z$$
(7.128)

Ellipsoidal gate:

 $\ln \Lambda < d_{\text{statistical}}$

In this equation, x, y, and z are the three spatial components of the state vector \mathbf{x} ; R is range. The subscripts m and p refer to measured and predicted values. Gating is often performed hierarchically. The surveillance space can be partitioned into regions so that a measurement falling in one region need not be tested against tracks in other regions. Similarly, it may be possible to group tracks within a region into mutually exclusive partitions, again reducing the number of comparisons to be made. The next step is often a relatively coarse but computationally efficient rectangular gating. The radar may gate only in range, or in all three spatial coordinates. Finally, a tighter ellipsoidal gating may be

applied using either the Mahalonobis distance or the log-likelihood.

The last major step in data association is assignment of the candidate detections after gating to the tracks. This process begins by constructing a measurement-to-track cost matrix. The elements of the matrix are some measure of the statistical cost assigning a particular measurement to a particular track, usually the negative of the log-likelihood of Eq. (7.127). One of a variety of algorithms is then used to decide the final assignments based on the cost matrix. Common algorithms include

- *Nearest neighbor* (*NN*) Assigns the lowest cost measurement to each track, independent of the other tracks. As a result, one measurement could be used to update more than one track in some situations. A common modification removes each measurement from consideration when it is assigned to one track so that it cannot also be assigned to another.
- *Strongest neighbor (SN)* Assigns to each track the validated measurement with the highest SNR. Like NN, can assign the same measurement to more than one track unless modified to disallow this.
- *Global nearest neighbor (GNN)* Assigns tracks such that the total cost is minimized, subject to the constraint that each measurement can be assigned to only one track (or to no track).
- *Auction methods* These globally optimum methods typically compare a prediction to the two or more closest measurements at each step. A bidding process assigns the closest measurement to each track. Costs are adjusted based on assignments made so far, allowing for reassignment of some measurements. A popular, computationally efficient version is the JVC algorithm (Malkoff, 1997).
- *Probabilistic data association (PDA)* Updates tracks with a weighted average of all candidate measurements. The weights are based on an estimated probability that a given measurement is the correct one and are related to the likelihood function.

Figure 7.35*a* is a notional example of a cost matrix for an environment in which two tracks are being maintained, but three measurements have passed the gating stage. The numbers in each cell represent the log-likelihood cost of the corresponding assignment. The cost matrix includes a column for the possibility that a track receives no update, and a row for the possibility that a measurement is assigned to no track. Figure 7.35*b* shows the results of a simple NN assignment. In this case, Measurement 1 is assigned to both tracks 1 and 2 at respective costs of 5 and 10. Measurements 2 and 3 are then assigned to "none," i.e., not used, since assignments have already been made to tracks 1 and 2. These measurements may be used to initiate new tracks. The total cost of the set of assignments is 5 + 10 + 20 + 20 = 55, but two tracks have been updated with

the same measurement.

			Measu	iremen	it				Measu	iremer	nt
		1	2	3	None			1	2	3	None
T r a c k	1	5	10	15	20	T r a c k	1	5	10	15	20
	2	10	30	25	20		2	10	30	25	20
	None	20	20	20	00		None	20	20	20	~~~~
						_					
)	(a)		_			9	(b)		
)	(a) Measu	iremen	it		1		(b) Measu	iremer	nt
		1	(a) Measu 2	iremen 3	it None			1	(b) Measu 2	iremer 3	nt None
Ţ	1	1	(a) Measu 2 10	iremen 3 15	t None 20	Ţ	1	1	(b) Measu 2 10	iremer 3 15	nt None 20
Trac	1	1 5 10	(a) Measu 2 10 30	iremen 3 15 25	t None 20 20	Tra	1	1 5 10	(b) Measu 2 10 30	iremer 3 15 25	None 20 20
Track	1 2 None	1 5 10 20	(a) Measu 2 10 30 20	17emen 3 15 25 20	t None 20 20 ∞	T r a c k	1 2 None	1 5 10 20	(b) Measu 2 10 30 20	17emer 3 15 25 20	nt None 20 20 ∞

FIGURE 7.35 (*a*) Track-measurement assignment cost matrix; (*b*) result of simple NN assignment, Cost = 55; (*c*) modified NN assignment, cost = 65; (*d*) GNN assignment, cost = 40.

A modified NN that avoids this problem removes a measurement from further consideration once it is assigned to a track. Applying this approach and starting with track 1 assigns measurement 1 to track 1; measurement 1 is then no longer eligible for assignment to any other track. No measurement is assigned to track 2, meaning that it will be coasted to the next measurement update. Measurement 3 is again not used. The result, shown in Fig. 7.35*c*, incurs a higher total cost of 65 but does not assign the same measurement to multiple tracks. Finally, a GNN algorithm will find the solution shown in Fig. 7.35*c* wherein measurement 2 is assigned to track 1, measurement 1 to track 2, and measurement 3 is again not used. The total cost is 40, lower than the sequential NN method while still avoiding dual use of measurement 1. The JVC algorithm would be expected to find the same solution.

Tracking, especially in multi-target scenarios, is a rich area of ongoing research. When there is little difference in assignment costs, *feature-aided tracking* may be used to augment the decision variables with estimates of target amplitude, pose, or image information such as high range resolution profiles. In especially dense scenarios, *multiple hypothesis tracking* methods may be used to improve track integrity. Introductory comments and references for these techniques are given in (Ehrman, 2013).

References

- Agrež, D., "Weighted Multipoint Interpolated DFT to Improve Amplitude Estimation of Multifrequency Signal," *IEEE Trans. Instrumentation and Measurement*, vol. 51, no. 2, pp. 287–292, Apr. 2002.
- Bailey, C. D., "Radar Antennas," Chap. 9 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.
- Bar-David, I., and D. Anaton, "Leading Edge Estimation Errors," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-17, no. 4, pp. 579–584, Jul. 1981.
- Bar-Shalom, Y., and T. E. Fortmann, *Tracking and Data Association*. Academic Press, Orlando, Florida, 1988.
- Bell, K. L. et al., "Extended Ziv-Zakai Lower Bound for Vector Parameter Estimation," *IEEE Trans. Information Theory*, vol. 43, no. 2, pp. 624– 637, Mar. 1997.
- Blackman, S. S., *Multiple-Target Tracking with Radar Applications*. Artech House, Dedham, Massachusetts, 1986.
- Blair, W. D., "Radar Tracking Algorithms," Chap. 19 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.
- Blair, W. D., "Design of Nearly Constant Velocity Filters for Radar Tracking of Maneuvering Targets," *Proceedings IEEE 2012 Radar Conference*, pp. 1008–1013, 2012.
- Blair, W. D., M. A. Richards, and D. G. Long, "Radar Measurements," Chap.
 18 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, 2010.
- Daum, F., "Nonlinear Filters: Beyond the Kalman Filter," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, Part 2: Tutorials II, pp. 57–69, 2005.
- Ehrman, L. M., "Multitarget, Multisensor Tracking," Chap. 15 in W. L. Melvin and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Topics*. SciTech Publishing, 2013.
- Hofstetter, E. M., and D. F. DeLong, Jr., "Detection and Estimation in an Amplitude-Comparison Monopulse Radar," *IEEE Trans. Information Theory*, vol. IT-15, no. 1, pp. 22–30, Jan. 1969.
- Howard, D. D., "Tracking Radar," Chap. 9 in M. I. Skolnik (ed.), *Radar Handbook*, 3d ed. McGraw-Hill, New York, 2008.
- Ianniello, J. P., "Time Delay Estimation via Cross-Correlation in the Presence of Large Estimation Errors," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-30, no. 6, pp. 998–1003, Dec. 1982.
- Jacobsen, E., and P. Kootsookos, "Fast, Accurate Frequency Estimators," *IEEE Signal Processing Magazine*, pp. 123–125, May 2007.

- Kalata, P. R., "The Tracking Index: A Generalized Parameter for α - β and α - β - γ Target Trackers," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-20, no. 2, pp. 174–182, Nov. 1984.
- Kanter, I., "The Probability Density Function of the Monopulse Ratio for N Looks at a Combination of Constant and Rayleigh Targets," *IEEE Trans. Information Theory*, vol. IT-23, no. 5, pp. 643–648, Sep. 1977.
- Kay, S. M., Fundamentals of Statistical Signal Processing, Vol. 1: Estimation Theory. Prentice-Hall, Upper Saddle River, New Jersey, 1993.
- Lyons, R. G., *Understanding Digital Signal Processing*, 3d ed., Sec. 13.15. Prentice-Hall, New York, 2011.
- Macleod, M. D., "Fast Nearly ML Estimation of the Parameters of Real or Complex Single Tones or Resolved Multiple Tones," *IEEE Trans. Signal Processing*, vol. 46, no. 1, pp. 141–148, Jan. 1998.
- Mahafza, B. R., *Radar Systems Analysis and Design Using MATLAB*, 3d ed. Chapman and Hall/CRC Press, 2008.
- Malkoff, D. B., "Evaluation of the Jonker-Volgenant-Castanon (JVC) Assignment Algorithm for Track Association," *Proc. SPIE 3068, Signal Processing, Sensor Fusion, and Target Recognition VI*, pp. 228–239, Jul. 28, 1997.
- Peebles, P. Z., Jr., Radar Principles. Wiley, New York, 1998.
- Rife, D. C., and R. R. Boorstyn, "Single-Tone Parameter Estimation from Discrete-Time Observations," *IEEE Trans. Information Theory* vol. IT-20, no. 5, pp. 591–598, Sep. 1974.
- Scarbrough, K., R. J. Tremblay, and G. C. Carter, "Performance Predictions for Coherent and Incoherent Processing Techniques of Time Delay Estimation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 5, pp. 1191–1196, Oct. 1983.
- Sherman, S. M., *Monopulse Principles and Techniques*. Artech House, Norwood, Massachusetts 1984.
- Weiss, A. J., "Composite Bound on Arrival Time Estimation Errors," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-22, no. 6, pp. 751–756, Nov. 1986.

Problems

- 1. Re-derive the CRLB of Eq. (7.15) for the real-valued constant-in-AWGN problem starting from the special form of the CRLB for AWGN given in Eq. (7.17).
- 2. Use the CRLB for transformed parameters (see <u>App. A</u>) and <u>Eq. (7.15)</u> to determine the CRLB for the signal power *A*².
- 3. Since the sample mean was an efficient estimator for a constant A in

additive WGN, it seems plausible that an efficient estimator for A^2 might be the square of the sample mean, $\overline{A^2} = (\hat{A})^2 = (\sum_{n=0}^{N-1} x[n])^2$. Show that this estimator is not efficient even though the estimator for A was efficient. *Hint:* consider the estimator bias.

- 4. Show that as $N \to \infty$ the estimator of Prob. 3 becomes efficient. *Hint:* note that the estimate \hat{A} is Gaussian with mean A and variance σ_w^2/N and use known results for the moments of a Gaussian RV.
- 5. Determine whether an alternative power estimator for the real-valued constant-in-AWGN problem that forms the sample mean of $x^2[n]$, $\widehat{A^2} = (1/N)\sum_{n=0}^{N-1} x^2[n]$, is efficient. If not, is it asymptotically efficient?
- 6. Suppose *N* samples are available of the real-valued signal $x[n] = A\cos(2\pi f_0 n + \phi) + w[n]$, where w[n] is i.i.d. white Gaussian noise with variance σ_w^2 . Starting with Eq. (7.17), find the CRLB for the variance σ_A^2 of an estimate \hat{A} of *A*. *Hint* #1: $\cos^2(2\pi f_0 n + \phi) = \frac{1}{2} + \frac{1}{2}\cos(4\pi f_0 n + 2\phi)$. *Hint* #2: Assume $f_0 \neq 0$ or $\pi/2$ and that *N* is "reasonably large." In that event it can be shown that $\sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) = 0$.
- 7. Consider a complex signal x[n] = s[n] + w[n]. w[n] is i.i.d. complex WGN with variance $\sigma_{w}^2 s[n]$ is the complex echo series from a Swerling 2 target, which means that it is also i.i.d. complex Gaussian but with variance σ_s^2 . Write the joint PDF of *N* samples of x[n]. Use this and Eq. (7.11) to find the CRLB for estimating the signal power σ_s^2 .
- 8. Verify that the same result is obtained for the CRLB as in the previous problem if the CRLB form of Eq. (7.10) is used instead. *Hint*: use standard results for the moments of N(0, σ^2) RVs.
- 9. Show that the variance of a complex parameter $\Theta = \Theta_R + j\Theta_I$ is the sum of the variances of the real and imaginary parts, $\sigma_{\Theta}^2 = \sigma_{\Theta_R}^2 + \sigma_{\Theta_I}^2$.
- 10. Starting with Eq. (7.17), verify the real-valued signal form (the second line) of Eq. (7.22).
- 11. Consider a signal $x[n] = s[n;\theta] + w[n] = Ar^n + w[n], n = 0,..., N-1$. w[n] is white Gaussian noise with variance σ_w^2 . The decay rate r > 0 is known. All signals and parameters are real-valued. Find the CRLB for estimating *A*. Use the geometric sum formula $\sum_{n=0}^{N-1} \alpha^n = (1-\alpha^N)/(1-\alpha)$ to put the result in a closed form.
- 12. If $r \to 1$ In the previous problem the data model becomes just a constant *A* in AWGN. Show that the CRLB found in that problem approaches the CRLB for a constant in AWGN in the limit as $r \to 1$.
- 13. Find the log-likelihood function for the data in Prob. 11. Use the fact that

the noise is WGN and the signal portion Ar^n is a nonrandom constant for a given *n* to write the PDF of a single sample of x[n]. Then write the joint PDF of all *N* samples and take the natural logarithm to get the log-likelihood function.

- 14. Find the maximum likelihood estimator of *A* in the scenario of Prob. 11. To do this, find the value *Â* for *A* that maximizes the log-likelihood function found in Prob. 13. Put any geometric sums in closed form.
- 15. Show that the MLE for *A* in Prob. 14 is unbiased.
- 16. For the same signal model used in Prob. 11, now assume that *A* is known but *r* is not. Find the CRLB for estimating *r*. Hint: It is not necessary (or advisable) to convert geometric sums or similar terms to a closed form in this problem; they can be left in summation form.
- 17. Consider the signal model x[n] = A + Bn + w[n], n = 0, ..., N 1. w[n] is white Gaussian noise with variance σ_{w}^{2} . All signals and parameters are real-valued. Finding *A* and *B* amounts to fitting a straight line to the data. Find the Fisher information matrix for this problem (see Eq. (7.58) or App. <u>A</u>). The identities of Eq. (7.62) may be helpful.
- 18. Find the CRLBs for the slope B and intercept A in the previous problem. Which parameter estimate improves more rapidly as the amount of data N is increased? The following identities may be helpful:

$$\sum_{n=0}^{N-1} n = \frac{N(N-1)}{2}, \quad \sum_{n=0}^{N-1} n^2 = \frac{N(N-1)(2N-1)}{6}, \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

19. Suppose *N* i.i.d. samples are available from a random process described by the Gaussian PDF

$$p_x(x;\mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-\mu)^2\right]$$

where μ is an unknown parameter. Find the maximum likelihood estimator of μ .

20. Find the maximum likelihood estimator for the parameter λ in the exponential PDF

$$p_x(x;\lambda) = \begin{cases} \lambda \exp(-\lambda x), & x > 0\\ 0, & x < 0 \end{cases}$$

Discuss how this estimator relates to the mean of the random variable x having this PDF.

21. Using the definition of Eq. (7.34), compute the RMS bandwidth of the following ideal lowpass spectrum S(F):

$$S(F) = \begin{cases} 1, & -B/2 < F < B/2 \\ 0, & \text{otherwise} \end{cases}$$

22. Compute the RMS bandwidth of a trapezoidal pulse defined in the time domain as follows:



Hint: Use Eq. (7.35) instead of Eq. (7.34). How will time delay estimation precision vary with the steepness of the pulse edges?

- 23. Use the result from the previous problem to show that the RMS bandwidth of a rectangular pulse is unbounded.
- 24. Show that the time delay estimate of Eq. (7.42) is unchanged when an unknown gain and phase α are added to the data, i.e., $x[n] \rightarrow \alpha x[n]$.
- 25. Consider a coherent (therefore complex-valued signals) radar using an LFM pulse with a "large" (greater than 100) BT product. What is the minimum SNR required to achieve a time delay estimation precision of 5 percent of the Rayleigh resolution in time? Figure 7.6 can be used as an approximate check on the result. If the bandwidth is $\beta = 60$ MHz, what is the range estimation precision in meters?
- 26. Suppose the output of the matched filter for a complex time delay MLE is oversampled by the factor $k_{os} = 10$. At what output SNR χ_{out} will the sampling bound become the limiting factor in the measurement precision? Assume the pulse is a large-BT product LFM so that $\beta_{\rm rms} \approx \beta/\sqrt{12}$.
- 27. A time delay MLE produces a peak value of $y_0 = 22.47$ at sample $n_0 = 14$. The two neighboring output samples are $y_{-1} = 21.42$ at sample 13 and $y_{+1} = 16.47$ at sample 15. The SNR for this data was $\chi_{out} = 30$ dB. The true peak location in the absence of noise corresponds to sample 13.8. Estimate the peak location $n_0 + \Delta n$ using each of the techniques in Eqs. (7.51) to (7.53) and compare.
- 28. Verify that the (1, 2) and (1, 3) elements of the multi-parameter complex

sinusoid estimation Fisher information matrix of Eq. (7.58) are zero.

- 29. Verify the CRLBs of Eq. (7.63) by explicitly computing the diagonal elements of the inverse of $I(\Theta)$ of Eq. (7.61).
- 30. Compute the CRLB for estimation of the frequency of a complex sinusoid in AWGN when the amplitude and phase are presumed known. Verify that this CRLB is smaller by a factor of four in the limit of large *N* than the case when all three parameters were unknown given in Eq. (7.63).
- 31. What are the minimum input SNR values χ in dB required to estimate the parameters of a complex sinusoid to the following precision? In each case, give the SNR for N = 40 and N = 400 samples.
 a. Relative amplitude: 10 percent
 b. Normalized frequency: 0.001 cycles per sample
 c. Phase: 0.01 radians (0.573°)
 Some of the results may be approximately checked using Figs. 7.11 to 7.13.
- 32. Verify the result given for \widehat{A}_{t} in Eq. (7.69).
- 33. Suppose w[n] is M samples of a complex white Gaussian noise process with variance σ_w^2 and W[k] is its K-point DFT. Compute the correlation $\mathbf{E}\{W[k_1]W^*[k_2]\}$ of two arbitrary DFT samples $W[k_1]$ and $W[k_2]$ as a function of M and K. Use the result to show that two different DFT samples $(k_1 \neq k_2)$ are correlated if K > M, and also that they are uncorrelated if K = M.
- 34. Consider angle of arrival estimation with a phased array radar having $d = \lambda/2$ and N = 10 elements. What is the SNR χ required to obtain a precision of 1° in estimating the AOA when the actual AOA $\theta = 0^{\circ}$ (broadside) and $\theta = 45^{\circ}$? Repeat for N = 100 elements.
- 35. What is the integrated SNR χ_{out} required to obtain a precision of 10 percent of the 3-dB beamwidth of a phased array radar? What is the equivalent input SNR χ if the array consists of N = 10 elements, and again for N = 100 elements?
- 36. Derive the MLE for lobing-based angle measurement in terms of the Σ and Δ signals [Eq. (7.90)] from the MLE in terms of the L and R signals [Eq. (7.89)].
- 37. Verify that in Eq. (7.95) the angle estimate $\hat{\theta} \rightarrow \theta_1$ as $\rho \rightarrow 0$ and $\hat{\theta} \rightarrow \theta_2$ as $\rho \rightarrow 0$. Develop an expression for $\hat{\theta}$ when $\phi = 180^\circ$ and $\rho = 1 + \varepsilon$ and show that the result can be well outside the range of $[\theta_1, \theta_2]$ when ε is small.
- 38. Suppose the two-way antenna pattern of a conical scan tracking system is modeled as a circularly symmetric Gaussian shape, $E^2(\theta) = G \cdot \exp(-\theta^2/\theta_0^2)$, where θ is the angular displacement from the boresight in any direction.

Assume that the conical scan boresight locus is displaced from the center of the rotation pattern by the 3 dB point of the pattern, which is approximately $\theta_3 = 0.59\theta_0$ radians off the boresight. Derive an expression for estimating the angular displacement θ_i of a target in terms of the maximum and minimum echo amplitudes during the scan, G_{max} and G_{min} , and θ_3 . Ignore noise and assume that the maximum and minimum response occur 180° apart in the rotation. Consider both $\theta_i < \theta_0$ and $\theta_i > \theta_0$.

- 39. Show that $\varepsilon^2(\Theta)$ is minimized by choosing the estimate $\hat{\Theta}$ as the sample mean of the data z[n].
- 40. Show that $\varepsilon^2(\Theta)$ of Eq. (7.98) is minimized in the linear model by choosing the estimate $\hat{\Theta}$ as the weighted sample mean of the data z[n] given in Eq. (7.99).
- 41. Using the initial conditions stated for Fig. 7.26, verify that the sequential LSE gain and variance obey K[N] = 1/(N+1) and $var(\hat{A}[N]) = 1/(N+1)$.
- 42. Using the **F**, **g**, and **h** that define the α - β filter and assuming the process and observation noise variances are $\sigma_w^2 = 1$ and $\sigma_w^2 = 5$, compute the prediction MSE **M**[*n*|*n*] and Kalman gain **k**[*n*] for *n* = 2, ..., 6. Initialize according to Eq. (7.117) with *T* = 1. MATLAB® or other computational tools may be used to aid in the calculation.
- 43. With T = 1, what is the tracking index for the previous problem? What will be the steady-state Kalman gain for the previous problem? Do the computed gains of the previous problem appear to be converging to this value?
- 44. Suppose two real-valued random variables x₁ and x₂ are jointly Gaussian but independent with zero means, σ²_{x1}=1, and σ²_{x2}=3. What is the covariance matrix S of the random vector x = [x₁ x₂]^T? Now consider a measured sample of the random process generating x represented by the vector y = [1 1]^T. Because the mean of x = [0 0]^T, the innovation is i = y-x̄ = y. Find the distance between y and x (i.e., "length" of i) using the following three distance measures:
 - a. Euclidean
 - b. Mahalanobis, d_M
 - c. Log-likelihood $\ln\Lambda$ [Eq. 7.127)]

Explain why these take on their particular relative values, i.e., why one is largest and another smallest.

45. Consider the track-to-measurement assignments in Fig. 7.35*c*, which corresponds to the modified nearest neighbor algorithm where a measurement is removed from consideration once it is assigned to a track. Recompute this case beginning by assigning a measurement to track 2 first. Are the results the same?

 2 In this and subsequent plots of estimate variance or precision in this chapter, the quantity plotted is the variance or standard deviation of the estimate error, rather than of the estimate itself. The result is the same, but computing and plotting the error variance has some simulation advantages.

 $\frac{3}{2}$ The measured precision falls below the CRLB for SNR = 29 and 30 dB because the sampling density in the simulation is too coarse to reliably measure such small variations.

 $\frac{4}{2}$ The definitions of both functions in this equation are consistent with those used in MATLAB®.

 $\frac{5}{2}$ This function is often called the *periodogram* of the data x[n].

 $\frac{6}{2}$ While the a priori bound is shown for higher SNRs on the frequency and phase plots, it is not shown for the amplitude precision plot because the calculation required produces unstable results at higher SNRs.

 $\frac{1}{2}$ Constant update intervals are assumed here, but many of the results can be generalized to nonconstant update intervals.

¹ Energy SNR was defined in <u>Chap. 4</u>.

CHAPTER 8

Introduction to Synthetic Aperture Imaging

When it was first developed, radar had two primary functions, detection and tracking. To these have been added fine-resolution radar imaging in two and, more recently, three dimensions. The technique of fine-resolution two-dimensional radar imaging is called *synthetic aperture radar* (SAR). SAR is most often applied to imaging of static ground scenes; thus, the "target" in SAR operation is the ground clutter. Civilian applications of radar imagery include cartography, land use analysis, oceanography, forestry, agriculture, natural disaster assessment, and more (Henderson and Lewis, 1998). Equally numerous military applications include reconnaissance, surveillance, battle damage assessment, ground target classification and identification, navigation, and more. SAR maps are routinely generated from both airborne and spaceborne platforms, and with resolutions ranging from several tens of meters down to a few inches.

Figure 8.1*a* is an example of a SAR image produced in the mid-1990s. Collected by the Sandia National Laboratories' K_u band radar, this image obtains a resolution of 3 m at ranges of tens of kilometers. Figure 8.1*b* is an aerial photograph of the same scene; close examination reveals many similarities as well as many significant differences in the appearance of the scene at radar and visible wavelengths. Figure 8.2 is another example, a SAR image of the National Mall area of Washington, DC.



(a)



(b)

FIGURE 8.1 Comparison of optical and SAR images of the Albuquerque airport: (*a*) K_u band (15 GHz) SAR image, 3-m resolution, (*b*) aerial photograph. (Images courtesy of Sandia National Laboratories.)



FIGURE 8.2 Synthetic aperture radar image of the Capitol Mall area in Washington, DC. (Image courtesy of Sandia National Laboratories.)

Despite the impressive quality of the SAR images, a human observer would likely prefer photographs for purposes of understanding and analyzing the scene. Though printed here in black and white, the original photograph in Fig. 8.1b is in color, whereas the SAR image is monochrome since SAR measures only the scalar quantity of reflectivity.¹ The photograph has finer resolution than the SAR image. The SAR image exhibits a granular speckle, often referred to as "salt-and-pepper" noise, typical of coherent imaging systems (including holograms, for instance) but absent in the noncoherent optical image. Close inspection reveals differences in phenomenology that can confound image analysis. For example, in the bottom center of the photograph there is a large concrete pad area on which there are three rectangular buildings; the concrete appears light in color, the buildings dark. In the SAR image the contrast is reversed and the building outlines are indistinct on the bottom side of the image. Another example is the painted stripes visible in the photograph at the ends of the runways (top center and right); these are entirely absent in the SAR image.

Why then is SAR of interest? The answer becomes apparent if the comparison of Fig. 8.1 is repeated on a cloudy night. The photograph would become a solid black, since the ground would not be visible through the clouds and the sun would not be present to provide illumination. The SAR image would be unchanged because the SAR is an active system that provides its own illumination, and because microwaves pass through clouds and other weather with little attenuation. Thus, radar provides a means for surveillance at any time of day or night and in a much wider range of weather conditions. Figure 8.3 compares two images taken from the space shuttle of the Manhattan and Long

Island areas of New York. Both were taken from an altitude of approximately 233 km in April 1994 at 3:00 A.M., though on different days. The lower image is a SAR map formed from three bands of radar data collected by the *shuttle imaging radar-C* (SIR-C) instrument.² The upper image is a photograph of the Manhattan and Long Island areas of New York city. The overlaid polygonal outline shows the approximate region of the SAR image. The radar image shows the full outline of Long Island. The photograph only shows those portions of the island that are well illuminated at 3:00 A.M., which does not include the northern edges of the island.



FIGURE 8.3 Comparison of optical (*top*) and radar (*bottom*) images of the Manhattan and Long Island, New York area as viewed from the space shuttle at 3:00 A.M. (Image courtesy NASA/JPL-CalTech.)

The history of SAR is described briefly in Sherwin et al. (1962) and Wiley (1985). The SAR concept was first described and demonstrated by Carl Wiley of Goodyear Aircraft in 1951. The technique discussed in his patent (Wiley, 1965) would now be categorized as *Doppler beam sharpening* (DBS) (see <u>Sec.</u> <u>8.3.1</u>). Since then, SAR has undergone several significant technology

development phases. The late 1950s and early 1960s developed the original concept and implementations of what is now known as stripmap SAR. In these pre-Moore's law days, SAR data were collected on photographic film and image formation was performed using remarkably elegant optical processing systems (Cutrona, 1966; Brown and Porcello, 1969; Harger, 1970; Ausherman, 1980; Elachi, 1988). The 1960s saw the development of spotlight SAR, generally credited to Jack Walker of the Environmental Research Institute of Michigan (ERIM) (Ausherman et al., 1984).³ In the 1970s, digital processing for SAR image formation was developed (Kirk, 1975), while the range-Doppler algorithm (Wu, 1982) significantly improved the attainable resolution and image size. In the 1970s, David Munson of the University of Illinois published the connection between spotlight SAR imaging and certain forms of computerized tomography ("CAT scanning") (Munson et al., 1983). This observation was the first of several that significantly expanded the capabilities of SAR algorithms by moving them beyond the range-Doppler viewpoint adopted in their early development and adapting techniques from other fields, such as tomography and seismic prospecting (Munson and Visentin, 1989; Cafforio et al., 1991).

SAR is the first of two advanced radar signal processing techniques to which this text provides an introductory overview. (The other, space-time adaptive processing, is the subject of <u>Chap. 9</u>.) Beginning in the mid-1990s, a number of excellent textbooks on SAR have become available (Curlander and McDonough, 1991; Carrara et al., 1995; Jakowatz et al., 1996; Soumekh, 1999; Franceschetti and Lanari, 1999; Cumming and Wong, 2005). Modern, concise introductory references are provided in the SAR chapters in the Principles of Modern Radar series (Showman, 2010; Showman, 2013; Cook, 2013; Richards, 2013). The reader is referred to these for in-depth discussion of SAR processing. This chapter begins with a heuristic overview of the SAR concept from two points of view: synthetic antenna apertures, and Doppler resolution. These are sufficient to derive many of the fundamental equations describing SAR resolution, coverage, and sampling requirements, and to describe the nature of the SAR data set. The signal processing required for SAR image formation is then addressed more directly, describing a basic SAR data model and introducing three of the most common algorithms for SAR image formation. Finally, the concept of the interferometric SAR approach to three-dimensional radar imaging is introduced.

8.1 Introduction to SAR Fundamentals

8.1.1 Cross-Range Resolution in Radar

To be useful, a radar map must provide adequate resolution for its intended use. Requirements may range from tens of meters to fractions of a meter.
Furthermore, this resolution should be available in both the range and crossrange dimensions, since there is no reason to prefer one over the other in most operational scenarios. Finally, the resolution should be maintained throughout the imaged scene. Sufficient range resolution for radar mapping is relatively easy to achieve using the pulse compression techniques discussed in <u>Chap. 4</u>. Comparable cross-range⁴ resolution, however, is not possible in conventional operation, often termed *real-beam imaging*.

Figure 8.4 illustrates cross-range resolution in a real-beam, forward-looking radar. The antenna scans in azimuth angle. It has an azimuth beamwidth of θ_{az} radians; thus, at a range R_0 the width of the beam is $R_0\theta_{az}$ meters to a good approximation. The cross-range dimension is the direction orthogonal to range. As discussed in Chap. 2, the receiver output for a fixed range as a function of azimuth scan angle is the range-averaged reflectivity convolved with the two-way antenna voltage pattern. In Fig. 8.4*a*, the two scatterers are separated in cross range by less than one beamwidth, so the receiver output will blur the response to the two scatterers together (see Fig. 2.26). In Fig. 8.4*b*, they are separated by more than the beamwidth so that the receiver output for the appropriate range bin as a function of scan angle will show two distinct peaks. By convention, the two scatterers are therefore considered just resolvable if they are separated by the width of the antenna beam. Assuming narrow azimuth beamwidths, the cross-range resolution ΔCR is well approximated by



FIGURE 8.4 Resolution in cross range of two scatterers at the same range: (a) unresolved in cross range, (b) resolved in cross range.

$$\Delta CR = 2R_0 \sin\left(\frac{\theta_{az}}{2}\right) \approx R_0 \theta_{az}$$
(8.1)

The beamwidth θ_{az} in this equation is usually taken as the two-way 3-dB beamwidth, but it is sometimes taken as the Rayleigh (peak-to-null) or even the null-to-null beamwidth.

As shown in <u>Chap. 1</u>, the azimuth beamwidth of a conventional antenna is of the form

$$\theta_{az} = k \frac{\lambda}{D_{az}} \quad \text{rad}$$
(8.2)

where D_{az} is the width of the antenna in the azimuth dimension. The scale factor k depends on the antenna design. It is as little as 0.89 for an ideal aperture antenna with uniform illumination, but more often is on the order of 1.2 to 2.0 for practical antenna designs. In general, the lower the antenna sidelobes, the higher is k. Here it is sufficient to say that k is on the order of one and ignore it in subsequent calculations for compactness. Combining Eqs. (8.2) and (8.1) gives the cross-range resolution of a real beam radar as

$$\Delta CR_{\text{real beam}} = \frac{\lambda R}{D_{az}}$$
(8.3)

The resolution indicated by Eq. (8.3) is not acceptable for imaging purposes. Unlike range resolution, the cross-range resolution degrades in proportion to range instead of being constant through the image. Far more important, the cross-range resolution is too coarse for useful images. Consider some typical numbers. An airborne tactical X band radar (10 GHz) with a 1-m antenna width would achieve a cross-range resolution of 300 m at 10 km range, or roughly three American football fields. A satellite in *low earth orbit* (LEO, for example around 770 km altitude) operating in C band (5 GHz) with a 10-m antenna would exhibit a cross-range resolution of 4.6 km. These numbers are too coarse for useful imagery.

Equation (8.3) suggests that cross-range resolution can be improved by restricting the operating range, using higher frequencies, or using larger antennas. Considering the airborne example, a change of two orders of magnitude is required to improve the resolution from 300-m to the 3-m resolution of Fig. 8.1*a*. This requires a change to either a 1-THz radar frequency, limitation to only 100 m operating range, an increase in the antenna size to 100 m, or some combination of less drastic but still very large changes in

these parameters. Such large changes are impractical. For instance, <u>Fig. 8.5</u> compares the approximate relative size of a 100-m phased array antenna and a typical fighter aircraft. It seems unlikely that such a large antenna could be constructed and flown on that aircraft.



FIGURE 8.5 Relative size of a fighter aircraft and a phased array antenna large enough to achieve $\Delta CR = 3$ m at X band and 10 km range.

8.1.2 The Synthetic Aperture Viewpoint

In fact, Fig. 8.5 does suggest a way in which fine cross-range resolution could be achieved. Rather than constructing a large physical phased array antenna to meet the requirements of Eq. (8.3), consider implementing only a single array element of the antenna, and then utilizing the platform motion to move that element through successive element positions to form the complete array. At each element position a pulse is transmitted and the fast-time data collected. When the element has traversed the length of the complete array, the data from each position is coherently combined in the signal processor to create the effect of a large phased array antenna with elements at each of the positions. The individual "element" can be the conventional antenna on the platform. In effect, the usual combining of phased array element signals in microwave hardware is performed instead in the signal processor. The system thus "synthesizes" a large phased array antenna aperture by operating a single element from multiple locations in space; hence the name, "synthetic aperture radar" (in some older literature, "synthetic array radar"). Put another way, the phased array antenna data is collected serially, one element at a time, rather than in parallel, all at once. This process is suggested by Fig. 8.6, which shows an aircraft that has collected data at four positions along the array; data are still to be collected at several more positions. While good range resolution is obtained via pulse compression, the cross range resolution of the fast-time samples is wide and increases with range. After processing, the data have fine resolution in both range and cross-range. Because data from multiple pulses are combined to form the effective fine-resolution beam, the scene being imaged should not change while the data are collected so that each pulse represents data from the same scenario. This again emphasizes that SAR is intended primarily for imaging static scenes.



FIGURE 8.6 The concept of synthetic aperture operation.

In practice the radar does not usually collect data over just a single synthetic aperture length sufficient to obtain the desired narrow effective beamwidth. Instead, it operates continuously during flight, producing an ongoing sequence of fast-time data vectors from different positions along the flight path. The effective synthetic aperture size D_{SAR} is determined by selecting the number of spatial positions from which data will be combined to form a narrow effective beam. The amount of data combined is expressed in terms of D_{SAR} or equivalently the *aperture time* T_a , which is related to D_{SAR} by the platform velocity:

$$D_{SAR} = v T_a \tag{8.4}$$

By combining a sliding window of data collected over the most recent T_a seconds or D_{SAR} meters of the flight path, a series of narrow effective beams can be formed at successive cross-range positions. At each such position the signal processor forms a set of range bins with fine resolution in both range and cross-range centered at the cross-range position corresponding to the center of the synthetic aperture as illustrated in Fig. 8.7.⁵ This series of range traces forms a two-dimensional radar image of the scene. So long as the radar remains in this mode and the signal processor can keep up with the data influx, the system can generate a continuous strip of imagery, rather like unrolling a long scroll.



FIGURE 8.7 Forming multiple synthetic apertures by combining sliding subsets of continuously collected data.

The mode of operation implied by Fig. 8.7 is called *sidelooking stripmap SAR*. Because the synthetic aperture is formed by the forward motion of the radar platform, the array face is naturally oriented perpendicular to the flight path. The effective antenna pattern is then oriented orthogonal to the velocity vector, a configuration referred to as sidelooking radar. In stripmap SAR, the physical antenna that serves as the array element is not actively scanned. It is instead locked into the sidelooking position, and its antenna pattern therefore moves across the ground as the aircraft flies forward. Conventional phase steering combined with mechanical steering of the physical antenna can be used to orient the effective SAR beam at an angle to the velocity vector other than 90°; this is referred to as a *squinted SAR*. For simplicity, only the sidelooking case is considered in this chapter.

In a conventional phased array antenna, all elements are present and active at once on both transmit and receive for each pulse, so the phase center of the antenna is in the middle of the physical structure for both transmit and receive. For a synthetic phased array, this is not the case: only one element at a time is active, so the phase center for transmit and receive moves across the face of the synthetic aperture as the data are collected. The resulting antenna pattern for a synthetic array differs somewhat from that of a physical array. To see this, consider the geometry of Fig. 8.8, which shows a scatterer at a range R and angle θ from the center of an array. Assuming that *R* is large compared to the array size, the range from the *n*th element is well approximated as



FIGURE 8.8 Geometry of synthetic array.

$$R_n = R - nd\sin\theta \tag{8.5}$$

where *d* is the element spacing. Suppose the signal $\exp(j\Omega t)$ is transmitted from the *n*th element. The received echo will be, ignoring scale factors,

$$y_n(t) = \exp\left\{j\Omega\left[t - \frac{2}{c}(R - nd\sin\theta)\right]\right\}$$
(8.6)

In synthetic array operation, each element is operated separately and the outputs combined. The total output is

$$y(t) = \sum_{n=-M}^{+M} y_n(t) = \sum_{n=-M}^{+M} \exp\left\{j\Omega\left[t - \frac{2}{c}(R - nd\sin\theta)\right]\right\}$$
$$= e^{j\Omega t} e^{-j4\pi R/\lambda} \sum_{n=-M}^{+M} \exp\left(j2\Omega nd\sin\theta\right)$$
$$= e^{j\Omega t} e^{-j4\pi R/\lambda} \left\{\frac{\sin\left[(2M+1)\Omega d\sin\theta/c\right]}{\sin(\Omega d\sin\theta/c)}\right\}$$
(8.7)

The term in brackets is the array factor of the synthetic array. It determines the amplitude of the received signal and therefore specifies the two-way voltage pattern of the antenna. The Rayleigh beamwidth is the angle to the first null, which occurs when the argument of the sine function in the numerator equals π . Defining the total aperture size as $D_{SAR} = (2M + 1)d$, this occurs when

$$\sin\theta_{SAR} \approx \theta_{SAR} = \frac{\lambda}{2D_{SAR}}$$
(8.8)

In comparison, the array factor of a conventional phased array, given in Eq. (1.14), gives a Rayleigh beamwidth of $\theta = \lambda/D$. Thus, the synthetic array has a beamwidth half that of a conventional array of the same aperture size.

It is now possible to determine the cross-range resolution obtained by the synthetic array. Combining Eqs. (8.4) and (8.8) gives

$$\Delta CR = R\theta_{SAR} = \frac{\lambda R}{2D_{SAR}} = \frac{\lambda R}{2vT_a}$$
(8.9)

Equation (8.9) is a fundamental result that relates the amount of data combined, represented through the aperture time T_a or size D_{SAR} , and the SAR cross-range resolution. Solving for T_a or D_{SAR} gives the design equation

$$T_a = \frac{\lambda R}{2v\Delta CR}$$
 or $D_{SAR} = \frac{\lambda R}{2\Delta CR}$
(8.10)

Equation (8.10) also gives the first hint of one of the major complications in SAR signal processing: the aperture time required to obtain a constant cross-range resolution is proportional to range, implying that the required processing is different at different ranges. SAR is a linear process, but is not shift-invariant in general.

One of the major advantages of SAR operation over a fixed physical antenna is that the effective aperture size is determined by the amount of data along the flight path that is combined to form any one pixel in the SAR image. The cross-range resolution can be changed by changing T_a (equivalently, D_{SAR}) according to Eq. (8.9). Consequently, a single radar can achieve several different resolution modes with the same antenna hardware. As an example, Table 8.1 lists the various resolution modes available in the Canadian RADARSAT-2 space-borne SAR. Available resolutions vary by a factor of over $30\times$, while *swath lengths* (depth of the image in the range dimension) vary by $25\times$. Notice that the range resolution in most, but not all, modes is equal or nearly equal to the cross-range resolution, a condition referred to here as "square pixels." Table 8.1 also shows that finer resolution is generally associated with shorter swath lengths, implying lesser area coverage rates. The reason for this will be discussed in Sec. 8.1.4.

Beam Mode	Nominal Swath Length, km	Approximate Resolution, m	
		Range	Cross Range
Ultra fine	20	3	3
Multi-look fine	50	8	8
Fine quad-pol	25	12	8
Standard quad-pol	25	25	8
Fine	50	8	8
Standard	100	25	26
Wide	150	30	26
ScanSAR narrow	300	50	50
ScanSAR wide	500	100	100
Extended high	75	18	26
Extended low	170	40	26

Data from http://www.asc-csa.gc.ca/eng/satellites/radarsat2/inf_data.asp.

TABLE 8.1 Resolution Modes in RADARSAT-2

Equation (8.10) also suggests that arbitrarily fine resolution can be obtained, at least in principle, by letting T_a or D_{SAR} become large. However, the maximum practical value of aperture time is limited by the physical antenna on the platform. Consider Fig. 8.9. When the aircraft is at the position on the left, the target is just entering the mainbeam of the physical antenna; when the aircraft is at the position on the right, the target is just exiting the main-beam. For aircraft positions before or after this interval, the target is not in the physical antenna mainbeam and the data collected outside of this interval will have no significant contribution from the target. Consequently, any one scatterer contributes to the SAR data only over a maximum synthetic aperture size equal to the travel distance between the two points shown, which equals the width of the physical antenna beam at the range of interest, namely $R\theta_{az}$. The corresponding maximum effective aperture time is $R\theta_{az}/v$. Inserting this result into Eq. (8.9) and using Eq. (8.2) (with k = 1) gives a lower bound on sidelooking strip-map SAR cross-range resolution of



FIGURE 8.9 Limitation of the aperture time by the physical antenna beamwidth.

$$\Delta CR_{\min} = \frac{D_{az}}{2}$$
(8.11)

Equation (8.11) is a remarkable result. It states that the lower bound on stripmap SAR resolution is independent of range, a desirable result for imaging, and more importantly, is much smaller than the real-beam resolution of Eq. (8.3). Considering the same two examples given previously, the lower bound on cross-range resolution becomes 0.5 m instead of 300 m for the airborne case, and 5 m instead of 4.6 km for the spaceborne case. (In practice, this resolution will be degraded by 50 to 100 percent by the use of windows for sidelobe control in the processing.) Furthermore, the lower bound on cross-range resolution does not depend on the wavelength. Finally, note that Eq. (8.11) states that to improve the lower bound on cross-range resolution, the physical antenna size should be reduced! It is rare that improved performance requires reducing the antenna size. In SAR, this occurs because a smaller physical antenna will have a broader beamwidth θ_{az} , allowing a larger maximum synthetic aperture size. Of course, shrinking the antenna size reduces the gain and SNR and, as will be seen, also reduces the maximum area imaging rate. The real implication of Eq. (8.11) is that fine cross-range resolution requires a large synthetic aperture, not a large physical antenna. Finally, note that in many cases the lower bound of Eq. (8.11) will be finer than required. In this case, the aperture time is simply limited to whatever value is required to obtain the desired resolution, as given by Eq. (8.10).

Resolutions of 1 to 10 m at ranges of 10 to 50 km, cruise velocities of 100

to 200 m/s, and frequencies from 10 to 35 GHz are typical of many operational airborne SARs, and thus might be considered "mainstream" parameters. Figure 8.10*a* is a representative plot of aperture time versus radar frequency with ΔCR as a parameter using Eq. (8.10). While low-frequency, fine-resolution SARs can demand very long aperture times, for most systems in their mainstream modes T_a is in the range of a few tenths of a second to 1 or 2 seconds. Much finer resolutions are possible but require much longer aperture times, e.g. about 10 seconds for 4-inch resolution at 10 GHz. For spaceborne systems, frequencies are most often in the range of 1 to 5 GHz, with newer systems ranging up to 10 GHz. In LEO a typical velocity and range are about 7500 m/s and 770 km. Figure 8.10*b* repeats the calculation of T_a for this scenario. Again, "mainstream" aperture times are on the order of a few tenths of a second to 1 or 2 seconds. Finer resolutions of one or two meters are available with longer aperture times of a few seconds.



FIGURE 8.10 Aperture time versus cross-range resolution and radar frequency: (*a*) airborne platform, v = 150 m/s and R = 10 km, (*b*) spaceborne platform in low earth orbit, v = 7500 m/s and R = 770 km.

While the stripmap resolution lower bound of Eq. (8.11) is finer than needed in many cases, in some fine-resolution applications it may not be good enough. The solution is to construct a longer synthetic aperture. To keep a given scatterer within the physical antenna mainbeam, it then becomes necessary to abandon the restriction that the antenna is not scanning. As the radar flies the

synthetic aperture, inertial navigation data are used to actively scan the antenna so as to keep its boresight pointed at the center of a *region of interest* (ROI) on the ground. In so doing, the ability to map a continuous strip is sacrificed for a finer-resolution map of the ROI. Once enough data have been collected to image the ROI, the antenna can be re-steered to image another discrete region. This mode of operation, termed *spotlight SAR*, is illustrated in Fig. 8.11. Because the radar essentially rotates around the region being imaged, spotlight SAR resolution is often expressed in terms of the rotation angle γ of the radar boresight vector as the platform traverses the synthetic aperture length. Clearly, $D_{sar} = 2R \sin (\gamma/2) \approx R\gamma$ for small γ . Using this equivalence in Eq. (8.9) gives the alternate cross-range resolution expression



FIGURE 8.11 Spotlight mode SAR operation.

$$\Delta CR = \frac{\lambda}{2\gamma} \tag{8.12}$$

Figure 8.12 is a 1-m resolution spotlight SAR image of the Pentagon in Washington, DC, clearly showing remarkable details of the five rings of the building, individual trees, and the surrounding road network.



FIGURE 8.12 One-meter resolution spotlight SAR image of the Pentagon. (Courtesy of Sandia National Laboratories.)

8.1.3 Doppler Viewpoint

The fundamental SAR resolution Eq. (8.9) or (8.10) was derived above from a basic antenna properties point of view. It can also be derived from a Doppler processing point of view. In fact, this approach is more consistent with the original conception of SAR and serves as a better starting place for considering some SAR image formation algorithms. It also allows an easy generalization to squinted SAR.

Consider two scatterers at range *R* separated in cross range by ΔCR meters. The angular separation $\Delta \theta$ of the two scatterers satisfies

$$\Delta CR = 2R \sin\left(\frac{\Delta\theta}{2}\right) \approx R \cdot \Delta\theta \implies \Delta\theta = \frac{\Delta CR}{R}$$
(8.13)

The small angle approximation is accurate within 1 percent so long as $\Delta\theta < 14^{\circ}$. In <u>Chap. 3</u> it was seen that two scatterers separated in angle by $\Delta\theta$ radians around a nominal squint angle ϕ from forward looking have a difference in Doppler shift of $(4v/\lambda) \sin(\Delta\theta/2) \sin \phi \approx (2v\Delta\theta/\lambda) \sin \phi$ Hz.

Using this result and Eq. (8.13) gives the Doppler difference between two scatterers separated by ΔCR meters at a squint angle ϕ as

$$\Delta F_D = \frac{2v\Delta CR\sin\phi}{\lambda R} \tag{8.14}$$

These two scatterers could be resolved by Doppler processing of the slow-time data provided that the Doppler resolution is no larger than ΔF_D Hz. As discussed in <u>Chap. 1</u> and <u>App. B</u>, frequency resolution is inversely proportional to signal duration in time, $\Delta F = 1/T$. In this scenario the signal duration is the aperture time T_a . Thus, two scatterers separated in cross range by ΔCR meters can just be resolved if $T_a = 1/\Delta F_D$. This gives

$$T_a = \frac{\lambda R}{2v\Delta CR\sin\phi}$$
(8.15)

which generalizes Eq. (8.10) to include the squinted case. In the sidelooking case $\phi = 90^{\circ}$ and Eq. (8.15) is identical to Eq. (8.10).

The Doppler viewpoint suggests a starting point for SAR image formation algorithms. Figure 8.13 illustrates a two-dimensional view⁶ of a sidelooking stripmap SAR. A scatterer **P** is located at range and cross-range coordinates R_P and x_P relative to the antenna and is within the physical antenna mainbeam. **P** is therefore at an angle $\theta_P = \operatorname{atan}(x_P/R_P) \approx x_P/R_P$, where the approximation is accurate within 1 percent for $\theta \leq 9.9^\circ$. This scatterer will produce a Dopplershifted echo. Pulse Doppler processing of the slow-time data in the range bin corresponding to R_P will produce a peak in the Doppler spectrum at $F_{DP} =$ $(2v/\lambda) \cos(\phi_P) = (2v/\lambda) \sin(\theta_P)$. The squint angle ϕ_P of the scatterer from the forward-looking velocity vector has been used in previous chapters but the angle θ_P measured from sidelooking is more relevant here. The corresponding cross-range position x_P is approximately R_P times the angular offset from sidelooking



FIGURE 8.13 Geometry relating cross-range position to Doppler shift.

$$x_p = R_p \sin^{-1} \left(\frac{\lambda F_{DP}}{2v} \right) \approx \frac{\lambda R_p F_{DP}}{2v}$$
(8.16)

Equation (8.16) provides a mapping of the Doppler axis to cross-range position. Note that the mapping is range-dependent, and therefore different for each range bin. Some systems having a range swath that is a small fraction of the nominal range use only a single mapping based on the nominal range at the center of the imaged swath, since the percentage range change across the image is small.

An image is formed by collecting a fast-time/slow-time data set that covers the range window of interest and provides an aperture time adequate for the desired resolution. The Doppler spectrum is computed at each range of interest. The Doppler axis is then remapped to cross-range position using Eq. (8.16), resulting in a range/cross-range image. If the same aperture time (slow-time data set size) is used in each range bin, this particular SAR imaging algorithm is called *Doppler beam sharpening* (DBS). If the aperture time is increased proportional to range so as to maintain constant cross-range resolution, it becomes a simple form of the *range-Doppler algorithm*. There are several effects that limit the scene size and resolution obtainable with this simple algorithm, primarily the problems of *range migration* and *quadratic phase errors*. Before discussing these, it is useful to consider some additional aspects of SAR operation.

8.1.4 SAR Coverage and Sampling

To determine the amount of terrain which a radar can image, the size of the image in two dimensions must be considered. In stripmap SAR, the along-track extent of the image is unlimited; so long as the radar platform continues

collecting and processing data, the image is extended in cross range. What determines the swath length (range extent) of the image?

Consider Fig. 8.14, which illustrates a sidelooking SAR scenario viewed in the along-track direction. The nominal grazing angle is δ radians. Scatterers outside of the mainbeam elevation beamwidth θ_{el} will not produce significant echoes. Thus, the swath length L_s is upper bounded by the projection of the elevation beam onto the ground plane. This is the same geometry considered in discussing the beam-limited resolution cell size in Chap. 2; the maximum swath length is therefore



FIGURE 8.14 Illustration of swath length and unambiguous range.

$$L_{s_{\max}} = \frac{R\theta_{el}}{\sin\delta} = \frac{R\lambda}{D_{el}\sin\delta}$$
(8.17)

The swath length can be made less than this by simply collecting fast-time samples over a more limited extent. In the illustration, the swath length L_s is indicated by the row of range samples within the elevation beam and is shown at its maximum extent, $L_s = L_{max}$. Since the radar platform moves forward at v meters per second while imaging a swath L_s meters deep, the area coverage rate is simply

$$A = vL_{\rm s} \quad {\rm m}^2/{\rm s} \tag{8.18}$$

A technique called *scanSAR* can be used to increase area coverage by

increasing the swath length beyond the elevation beamwidth limit, though at the cost of degraded cross-range resolution. The antenna is periodically switched between two elevation angles to create two contiguous swaths. Careful attention to timing is required to avoid leaving gaps in the data in either swath. The method is described in detail in Cumming and Wong (2005).

To image the full swath L_s with the intended resolution, two conditions must be met by the received data. The first is that the full pulse echo must be received from all scatterers within the swath on each pulse. Any eclipsing will degrade the resolution of the eclipsed scatterers. The second is that clutter from other ranges must not be allowed to fold over onto the swath of interest. Clutter outside the mainbeam footprint is usually considered insignificant, so this requirement applies only over the range footprint of the antenna. Also, it is not necessary that the radar be unambiguous out to the full range *R*. Rather, the requirement is that echoes from the far edge of the mainbeam footprint on one pulse be received before echoes of the next pulse from the near edge of the swath of interest.

The eclipsing constraint requires that the time interval from receipt of the leading edge of the echo from the near edge of the swath to the trailing edge of the echo from the far edge be less than the time $T - \tau$ between pulses when the receiver is on. At a nominal grazing angle of δ radians the difference in delay between the swath edges is approximately $(2/c)L_{s}\cos\delta$, so the duration of the data from the swath including the tail of the pulse from the far edge is $(2/c)L_{s}\cos\delta + \tau$. The resulting requirement is the first relation in Eq. (8.19). The ambiguity constraint is tightest when the swath occupies the full range footprint of the antenna. In that case it requires that the trailing edge of the clutter from the far edge of the swath not overlap with the leading edge echo from the near end and the next pulse. This results in the second relation in Eq. (8.19) and is seen to be slightly more lax than the eclipsing constraint. The shaded bands within the antenna beam in Fig. 8.14 indicate successive pulses in flight, with a spacing just large enough to meet this requirement. In many but not all SAR systems the swath length is much larger than the equivalent length of the pulse, so that either constraint results in the simplified approximate upper bound on PRF shown in the third relationship in Eq. (8.19)

Eclipsing constraint:

$$\frac{2}{c}(L_s\cos\delta) + \tau < T - \tau \implies PRF < \frac{1}{\frac{2}{c}L_s\cos\delta + 2\tau}$$

Range ambiguity constraint:

$$\frac{2}{c}(L_s \cos \delta) + \tau < T \implies PRF < \frac{1}{\frac{2}{c}L_s \cos \delta + \tau}$$
(8.19)

Approximation for $L_s \gg c\tau/2$:

$$PRF < \frac{c}{2L_s \cos \delta}$$

Because it is the slow-time sampling rate of the radar, a second constraint on the PRF is that it exceed the Doppler bandwidth β_D , as discussed in <u>Chap. 3</u>. For the moment, it is not assumed that β_D necessarily is the maximum sidelooking Doppler bandwidth of $2v\theta_{az}/\lambda$ [Eq. (3.5)]. β_D is related to the crossrange resolution of a sidelooking SAR as follows. A resolution of ΔCR meters is equivalent to a resolution in time of $v \cdot \Delta t = \Delta CR/v$ seconds. In the absence of weighting for sidelobe control, the time resolution is the inverse of the processed bandwidth, $\Delta t = 1/\beta_D$. The PRF lower bound in terms of cross-range resolution is therefore

$$PRF \ge \beta_D = \frac{v}{\Delta CR}$$
(8.20)

Typical sidelooking SAR Doppler bandwidths vary from a few tens of hertz to a few kilohertz, while typical stripmap range swaths vary from a few ones of km to 100 km. The combination of these constraints results in SAR PRFs of several hundred to a few kilohertz in most cases. SAR is therefore a low PRF radar mode.

Combining the last form of Eq. (8.19) with Eq. (8.20) establishes a range on the allowable PRF,

$$\frac{v}{\Delta CR} < PRF < \frac{c}{2L_s \cos\delta}$$
(8.21)

Equation (8.21) is the basis for several interesting constraint equations in SAR processing. The first constraint follows directly from Eq. (8.21):

$$\frac{L_s \cos \delta}{\Delta CR} < \frac{c}{2v} \tag{8.22}$$

Notice that the left-hand side is proportional to the number of range resolution cells in the swath length, assuming square pixels ($\Delta CR = \Delta R$). The quantity c/2v is on the order of 20,000 for LEO satellites and the space shuttle and can

become a significant constraint. For airborne SAR, it is much larger at 300,000 to 750,000 and much less likely to constrain the system design. If a system is already operating near this limit, Eq. (8.22) shows that further improving resolution requires that the swath length and therefore the area coverage rate [Eq. (8.17)] must be reduced proportionally. Thus, there is a potential conflict between fine resolution and large area coverage.

Now suppose that the finest possible sidelooking stripmap resolution is desired so that the full Doppler bandwidth $\beta_D = 2v\theta_{az}/\lambda$ is processed. As seen earlier, with $\theta_{az} = \lambda/D_{az}$ this results in $\Delta CR = D_{az}/2$. Inserting this value into Eq. (8.22) and rearranging gives a bound on stripmap swath length:

$$L_{s} \leq \frac{cD_{az}}{4v\cos\delta} \,\mathrm{m}$$
(8.23)

Equation (8.21) shows that for a given platform velocity, the swath length is limited by the physical antenna width. It also shows that a wider antenna is needed to achieve a given swath length on a faster platform. This is one major reason that spaceborne SARs have much wider antennas than airborne SARs.

Using Eq. (8.23) in Eq. (8.18) gives a closely related area coverage rate constraint of

$$A \le \frac{cD_{az}}{4\cos\delta} \qquad \text{m}^2/\text{s} \tag{8.24}$$

While good SAR resolution encourages antennas that are narrow in the azimuth dimension [Eq. (8.11)], this equations states that high area coverage rates encourage wider azimuth antenna extents, again reflecting the conflict between fine resolution and high area coverage rates mentioned earlier.

Finally, if maximum area coverage is sought by setting the swath length to the full extent of the antenna range footprint while still maintaining minimum resolution, Eqs. (8.17) and (8.23) can be combined to derive the inequality:

$$D_{az}D_{d} \ge \frac{4vR\lambda}{c\tan\delta}$$
(8.25)

This result specifies a lower bound on the antenna area, not just its width. It shows that spaceborne SARs must have larger antennas than airborne SARs due to their much larger velocities and ranges. It also states that lower RFs tend to require larger antennas.

Equations (8.23) and (8.24) are often referred to as the stripmap SAR swath constraint and mapping rate constraint equations, while Eq. (8.25) is

called the *antenna area constraint*. A number of variations and extensions of these ideas for other cases are described in Freeman et al. (2000).

The previous equations apply to stripmap SAR. It is straightforward to estimate the mapping rate for spotlight SAR operation as well. From Eq. (8.12), an image with resolution ΔCR requires that the antenna LOS rotate through $\lambda/2\Delta CR$ radians. Assuming a straight flight path, the platform must travel $\lambda R/2\Delta CR$ meters; this will require $\lambda R/2\nu\Delta CR$ seconds. This is the time required for one spotlight SAR image. The number of images per unit time, N_{spot} , is therefore upper bounded by

$$N_{\text{spot}} \leq \frac{2v\Delta CR}{\lambda R} \qquad \text{images per second} \\ = \frac{7200v\,\Delta CR}{\lambda R} \qquad \text{images per hour}$$
(8.26)

In SAR, the PRF sets the cross-range sampling interval $\delta x = vT = v/PRF \le \Delta CR$, representing one pulse per cross-range pixel. Failure to use a high enough PRF results in cross-range ambiguities in SAR imaging. In order to minimize this problem, the maximum Doppler bandwidth $\beta_D = 2v\theta_{az}/\lambda$ is typically used as the lower bound on PRF even if the Doppler bandwidth needed for the desired cross-range resolution is less. Furthermore, θ_{az} is often interpreted conservatively as the null-to-null beamwidth, effectively increasing the PRF by about a factor of two compared to the result obtained if the 3-dB or Rayleigh beamwidth is used. Choosing a PRF greater than required by Eq. (8.20) tightens the swath length and area coverage constraints by the same factor.

8.2 Stripmap SAR Data Characteristics

Up to the point where the complex image is finally passed through a detector to convert it to pixel data suitable for display, the SAR signal processing system is nominally linear. (It is *not* shift invariant, as will become clear.) Consequently, the behavior of the SAR system, including such characteristics as resolution, sidelobes, and the response to complex multi-scatterer scenes, can be understood by considering the response to a single isolated point scatterer. In the nomenclature of linear systems analysis, the data set generated by a single point scatterer is called the *point scatterer response* (PSR) or the SAR system *impulse response*. The former term is preferred because the latter term is generally associated with shift invariant systems.

8.2.1 Stripmap SAR Geometry

Figure 8.15 defines the geometry for sidelooking stripmap SAR data acquisition. The *ground plane*, which is the surface to be imaged, is defined by

the x and y_g axes. The dimension y_s defines the *slant range* dimension, and the plane defined by the x and y_s axes is called the *slant plane*. Because the radar measures delay, this is the more natural plane in which to discuss imaging. Of course, real terrain varies in height. This aspect is discussed in <u>Sec. 8.6</u>. For the present, it is assumed that the ground plane scene is two dimensional.



FIGURE 8.15 Geometry for stripmap SAR data acquisition.

Consider a sidelooking radar platform flying at constant altitude *h* and velocity *v* along the +*x* direction such that the antenna phase center is located at coordinates (u = vt, 0, *h*) at time *t*. The antenna is therefore above the origin at t = 0. A scatterer **P** is located in the ground plane at coordinates (x_P , y_{gP} , 0). Assume the antenna is steered such that **P** is in the center of the antenna beam azimuth extent at t = 0. If the aperture time required to achieve the desired resolution is T_a seconds corresponding to a synthetic aperture size of D_{SAR} meters, the time interval of interest in imaging **P** is $t \in (-T_a/2, T_a/2)$, corresponding to platform positions along the *x* axis (the synthetic aperture) of $u \in (-D_{SAR}/2, D_{SAR}/2)$. This geometry implies that if $x_P = 0$ the radar is sidelooking, while if $x_P \neq 0$ it is squinted. In Fig. 8.15, a case is shown with the

radar squinted forward.

The range from the radar to **P** is the Euclidean distance

$$R = R(u) = \sqrt{(u - x_p)^2 + y_{gp}^2 + h^2} = \sqrt{(u - x_p)^2 + R_p^2}$$
$$= R_p \sqrt{1 + \frac{(u - x_p)^2}{R_p^2}}$$
(8.27)

Equation (8.27) shows that in a stripmap SAR the range between the radar and an arbitrary scatterer varies hyperbolically as the radar moves along the synthetic aperture. The form of the range variation is invariant to the scatterer's along-track position x_P , in the sense that the x_P dependence of R(u) involves only the position of the aircraft relative to the scatterer, $(u - x_P)$. More explicitly, if the range variation for a scatterer at x_{P1} is denoted by $R_1(u)$, the range variation for a scatterer at position $x_{P2} = x_{P1} + \Delta x$ is $R_2(u) = R_1(u - \Delta x)$. In contrast, the range variation does vary with the absolute slant or ground range of the target, R_P or y_{qP} (which appears through R_P).

In the analysis of low- to medium-resolution SAR systems it is common to further simplify Eq. (8.27) by using the series expansion of the square root $\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \cdots$ and keeping only the first two terms, giving

$$R(u) \approx \left[1 + \frac{(u - x_p)^2}{2R_p^2}\right] R_p = R_p + \frac{(u - x_p)^2}{2R_p}$$
$$= R_p + \frac{u^2}{2R_p} - u\frac{x_p}{R_p} + \frac{x_p^2}{2R_p}$$
(8.28)

This assumption requires $|u - x_p|/R_p \ll 1$. Since the maximum value of $|u - x_p|$ of interest is $D_{SAR}/2$, this is equivalent to stating that the synthetic aperture size is short compared to the nominal range. Equation (8.28) shows that the range from the radar to the target varies approximately quadratically as the data set is collected. Since the received phase of the target echo is shifted from the transmitted phase by an amount proportional to range, namely $\phi = -(4\pi/\lambda)R$, if follows that the absolute phase of the received echoes will also vary approximately quadratically. In Chap. 4, it was seen that a quadratic phase function corresponds to linear frequency modulation. LFM is used to obtain fine resolution in fast time by spreading the waveform bandwidth. Equation (8.28) shows that the slow-time data from a given scatterer will also have an approximately quadratic phase modulation that will spread the slow-time bandwidth and enable fine cross-range resolution. This slow-time modulation is induced by the changing geometry due to platform motion relative to the imaged scene.

It is useful to determine the equivalent of the LFM time-bandwidth product for this cross-range chirp. For a given scatterer position x_P , the received phase will vary with aperture position u as $\phi(u) = -(4\pi/\lambda)R(u)$. The instantaneous frequency corresponding to this phase modulation is obtained as usual as

$$K_{ui} = \frac{d\phi(u)}{du} = -\frac{4\pi}{\lambda R_p} (u - x_p) \qquad \text{radians per meter}$$
(8.29)

Since the independent variable *u* is a spatial coordinate in meters instead of time in seconds, K_{ui} is a spatial frequency. As *u* varies over the synthetic aperture length of D_{SAR} meters, the spatial frequency bandwidth will sweep over $(4\pi/\lambda R_p) D_{SAR}$ radians per meter = $2D_{SAR}/\lambda R_p$ cycles per meter. Defining the latter quantity as the spatial bandwidth β_u , the "space-bandwidth" product of the cross-range chirp is

$$\beta_{\mu}u = \frac{2D_{SAR}^2}{\lambda R_p}$$
(8.30)

Just as matched filtered LFM chirps have a time resolution equal to their duration divided by their BT product, it should be possible by proper processing to achieve a cross-range resolution in SAR that is the synthetic aperture size divided by the space-bandwidth product. Note that dividing D_{SAR} by Eq. (8.30) does in fact give the result of Eq. (8.9). Also, recall that the BT product $\beta\tau$ is the radar range equation signal processing gain for fast-time processing (pulse compression). Similarly, the signal processing gain for slow-time (SAR) processing will be the slow-time space-bandwidth product $\beta_u u$. The overall signal processing gain G_{sp} in the SNR of a point scatterer will be the product $(\beta\tau)(\beta_u u)$.

The variation in range to a point scatterer over the synthetic aperture is called *range migration*. It can be broken into two components, *range walk* ΔR_w and *range curvature* ΔR_c . Range walk is the difference between the range to P at the beginning and end of the synthetic aperture:

$$\Delta R_{w} = R(-D_{SAR}/2) - R(+D_{SAR}/2)$$

= -[(D_{SAR}/2) - (+D_{SAR}/2)] $\frac{x_{p}}{R_{p}}$
= $\frac{D_{SAR}}{R_{p}}x_{p} = \frac{vT_{a}}{R_{p}}x_{p}$ (8.31)

Note that this is just the change over the aperture time in the linear term in *u* of

Eq. (8.28). Range curvature is the variation in the quadratic term of R(u), which has its maximum at either extreme of the aperture position, $u = \pm D_{SAR}/2$, and its minimum at u = 0:

$$\Delta R_{c} = R(\pm D_{SAR}/2) - R(0)$$

$$= \frac{D_{SAR}^{2}}{8R_{p}} = \frac{v^{2}T_{a}^{2}}{8R_{p}}$$
(8.32)

Both the range walk and range curvature depend on D_{SAR} or T_a and R_P . If a constant aperture time is used, range walk and curvature both decrease as $1/R_P$. However, if constant cross-range resolution is desired, T_a must increase proportionally to range. In this case, ΔR_w is constant over range for a given x_P , while ΔR_c increases in proportion to R_P .

Range migration is significant only if it exceeds the range bin size. If so, echo samples from a single scatterer will start in different range bins on different pulses, complicating the process of combining them to form a fine-resolution image. Both range walk and range curvature can be predicted using only parameters which are known to the radar system: own-ship velocity, aperture time, and range.

8.2.2 Stripmap SAR Data Set

The fast-time/slow-time complex baseband data set used to form the SAR image is the two-dimensional SAR phase history. Figure 8.16 is the real part of the received data from a single point scatterer located at $x_P = 0$, $y_{gP} = 10$ km. The data correspond to a 5-GHz radar at an altitude of 5 km. The pulse is an LFM upchirp of bandwidth $\beta = 10$ MHz and duration $\tau = 5$ µs, sampled at a fast-time rate of 30 Msamples per second. The PRF is 500 Hz and the platform velocity is 250 m/s. The aperture time is $T_a = 13.5$ seconds. With these parameters, the expected resolutions are 15 m in range and only 0.1 m in cross range. While not very realistic, this example results in a data set with exaggerated curvatures for illustrating some general characteristics of stripmap SAR data.



FIGURE 8.16 (*a*) Real part of the received data from a point scatterer located at $x_P = 0$, $y_{gP} = 10$ km, (*b*) fast-time slice at azimuth position u = 500 m. See text for radar parameters.

Figure 8.16*a* shows the real part of the complete two-dimensional PSR. The range curvature is clearly evident. The complicated pattern of the data within the PSR reflects the changing amplitude of the real part of the chirp

pulse. Each fast-time slice through the data set is simply an echo of the transmitted chirp. For instance, the slice of data in the thin rectangular box at azimuth +500 m is shown in part *b* of the figure; the upchirp is clear. All other fast-time slices are identical except for their start time and initial phase offset, determined by the differing ranges to the target from each point along the radar platform's synthetic aperture.

Figure 8.17 shows a cross-range slice of the data in Fig. 8.16*a*, taken at ground range 10.4 km. Part *a* shows the real part of the complete cross-range slice, emphasizing the attenuation caused by the physical antenna pattern at the edges of the data set, when the point scatterer is near the edge of the antenna mainbeam. The segment inside the rectangular box is expanded in Fig. 8.17*b*, where it can be seen that the motion-induced phase modulation forms a chirp that passes through zero frequency at $u = x_p$ (broadside from the scatterer). Thus, the PSR is approximately a two-dimensional chirp function.



FIGURE 8.17 Cross-range slice of Fig. 8.16*a* at ground range $y_{gP} = 10.4$ km: (*a*) complete slice, (*b*) central portion of the slice.

The size of the PSR data set which must be processed by a stripmap SAR can be quite large. For a swath length L_s , range resolution ΔR , and pulse time-bandwidth product $\beta \tau$, the number of fast-time samples required is

$$L = \frac{L_s}{\Delta R} + (\beta \tau - 1) \tag{8.33}$$

This result assumes a fast-time sampling rate equal to the bandwidth β ; if oversampling is used, *L* increases accordingly. The number of slow-time samples equals the number of pulses within the aperture time:

$$M = T_{a} PRF$$
$$= \left(\frac{\lambda R}{2v\Delta CR}\right) \left(\frac{2v\theta_{az}}{\lambda}\right) = \frac{R\theta_{az}}{\Delta CR}$$
(8.34)

Notice that this is just the number of cross-range resolution cells in the physical beamwidth $R\theta_{az}$. Equation (8.34) assumes that the PRF equals the clutter bandwidth. Again, if oversampling is used, *M* increases accordingly.

As an example, consider the SIR-C, which operates at L band with a swath length of 15 km from an orbital range of 250 km, range and cross-range resolution of 15 m, and a 17-µs pulse with a 10-MHz bandwidth. The antenna size is 10 m, suggesting a beamwidth θ_{az} of about 0.03 radians = 1.72°. These parameters result in L = 1169 fast-time samples per pulse and M = 500 pulses in the aperture time. This large number of pulses contributing to the image of each point target differentiates SAR from the much simpler but related pulse Doppler processing. In pulse Doppler, a CPI is typically a few tens of pulses at most. Returning to the SIR-C example, (1170)(500) = 585,000 samples contribute to each point scatterer image. If properly coherently integrated, a 57.7-dB integration gain can be achieved.

8.3 Stripmap SAR Image Formation Algorithms

For a general scene, the stripmap SAR data set is the superposition of a large number of weighted and shifted replicas of the PSR. Recall that the PSR is, in general, a function of range but is independent of cross range. The goal of any SAR image formation algorithm is to compress the PSR of each scatterer into an impulse-like function with the appropriate location and amplitude, as shown notionally in <u>Fig. 8.18</u>.



FIGURE 8.18 The goal of SAR image formation algorithms is to transform the two-dimensional PSR in (a) into the point target image in (b).

A variety of stripmap SAR image formation algorithms exist. They vary in resolution capability, scene size capability, and computational complexity. In general, finer resolution, longer standoff range, lower radar frequency, squint geometries, and greater scene size require more elaborate and computationally expensive algorithms. In this section, two basic imaging algorithms are introduced, Doppler beam sharpening and the range-Doppler algorithm. More advanced techniques such as the *range migration algorithm* (also called the ω -*k* algorithm) and the *chirp scaling algorithm* are described in detail in Bamler (1992), Gough and Hawkins (1997), Franceschetti and Lanari (1999), Soumekh (1999), Cumming and Wong (2005), and Showman (2013).

8.3.1 Doppler Beam Sharpening

Doppler beam sharpening is the original form of SAR envisioned by Wiley (Wiley, 1965). It uses a constant aperture time for all ranges so that the cross-range resolution is proportional to range. It is the simplest SAR algorithm, and is suitable only for relatively coarse resolution imagery. Nonetheless, it provides a substantial improvement over real beam cross-range resolution and has relatively low computational requirements.

DBS presumes that the PSR is compressed in the fast time dimension by conventional pulse compression, whether implemented by direct convolution, fast convolution, or stretch processing. Ignoring fast-time sidelobes, the fast-time output of the matched filter for a pulse transmitted from aperture position u can be approximated as $A\exp(j\phi)\delta_D(t - 2R(u)/c)$, where $\delta_D(\cdot)$ is the continuous-time Dirac impulse function, the phase ϕ due to the round trip travel is $-(4\pi/\lambda)R(u)$, and the constant A absorbs all amplitude factors. Using Eq. (8.28), the slow-time phase variation is therefore

$$\phi(u) \approx -\left(\frac{4\pi}{\lambda}\right) \left(R + \frac{u^2}{2R} - u\frac{x}{R} + \frac{x^2}{2R}\right)$$
$$\approx -\left(\frac{4\pi}{\lambda}\right) \left(R - u\frac{x}{R} + \frac{x^2}{2R}\right)$$
(8.35)

where the last step assumes that the quadratic (in *u*) phase term can be neglected; this assumption will be revisited in Sec. 8.3.2. The subscript *P* has been dropped on the coordinates *x* and *R* for simplicity. The instantaneous cross-range spatial frequency K_u is

$$K_u = \frac{d\phi(u)}{du} \approx \frac{4\pi}{\lambda} \frac{x}{R}$$
 rads/m (8.36)

so that

$$x = \frac{\lambda R}{4\pi} K_u \tag{8.37}$$

Equation (8.37) relates scatterer cross-range position to spatial frequency. Alternatively, x can be expressed in terms of Doppler (temporal) frequency. Using u = vt in Eq. (8.35), the Doppler frequency F_D in hertz due to a scatterer at (x, R) is $(1/2\pi)(d\phi/dt)$, so that

$$x = \frac{\lambda R}{2v} F_D \tag{8.38}$$

a relation that is more natural to the DBS name. This is the same as Eq. (8.16) presented previously.

The algorithm implied by Eq. (8.38) is diagrammed in Fig. 8.19. It requires only one discrete Fourier transform per range bin, followed by a rescaling of the axes to express the result in range/cross-range coordinates. Typically, windowing for sidelobe suppression would be included in the processing in both dimensions.



FIGURE 8.19 Block diagram of Doppler beam sharpening algorithm.

Figure 8.20 is the result of applying a simple DBS algorithm to simulated data from a test array. The magnitude of the DBS image in decibels is shown. Point scatterers were simulated at all combinations of x and R equal to -1000, 0, or +1000 m relative to the center of a scene [the *central reference point* (CRP)] at a nominal range of 20 km. The position of the radar is above the image, with range increasing from top to bottom. The aperture time was 40 ms and the velocity was 150 m/s. The radar frequency was 10 GHz, and an LFM pulse with $\beta = 3$ MHz and $\tau = 5 \mu s$ ($\beta \tau = 15$) was used at a PRF of 7.5 kHz. Consequently, the image resolution is $\Delta R = \Delta CR = 50$ m. The Doppler frequency axis was scaled to cross-range position using Eq. (8.38) with R set equal to the nominal range $R_{CRP} = 20$ km. Each of the scatterers is well-focused with clearly defined range and cross-range sidelobes. Scatterers off the center line of x = 0, however, are imaged as slightly further away from the radar than those on the center line because of the contribution of x to the total range. In addition, the columns of scatterers bow inward slightly as range increases. This is because the spectrum in each range bin was remapped from Doppler frequency to crossrange position using a scale factor based on the same central range of 20 km. That factor is too small at longer ranges, and too large at shorter ranges.



FIGURE 8.20 Magnitude in dB of the DBS image of a test array. See text for parameters.

These geometric distortions are readily corrected after the basic image formation. The cross-range displacement is corrected by simply updating the cross-range scale for display of each row with the appropriate range R for each range bin, instead of using R_{CRP} for all of the range bins. The result of this operation is shown in Fig. 8.21*a*. The scatterers are now aligned vertically at the correct cross-range coordinates of -1, 0, or +1 km. Interpolation is required to provide samples on a common grid for display. However, the range shift for scatterers off the center line is still present. This can be corrected by computing for each cross-range bin the actual range $R' = \sqrt{R^2 + x^2}$ to that bin, and then shifting that cross-range column by the difference of R' - R meters. Interpolation will again be required to accommodate partial-range bin shifts. Figure 8.21*b* shows the result of this operation; the nine scatterers are now accurately centered on correct coordinates. The curvature of the cross-range sidelobes nicely reflects the range curvature that was present in the original DBS image and has now been corrected.



FIGURE 8.21 Correction of DBS geometric distortion: (*a*) application of range-varying cross-range scale factor, (*b*) application of cross-range-varying range shift.

The DBS algorithm assumes that the echo samples from a given scatterer remain in the same range bin over the aperture time so that the FFT of the slowtime data will integrate all of the scatterer samples. This can be assured by requiring that the range walk ΔR_w not exceed some fraction (typically one-half) of a range bin over the aperture. Using Eq. (8.31), this condition becomes

$$\Delta R_w = \frac{D_{SAR}}{R} x = \frac{vT_a}{R} x \le \frac{\Delta R}{2}$$
(8.39)

Substituting from Eq. (8.15) for T_a gives the constraint:

$$x \le \frac{\Delta R \cdot \Delta CR}{\lambda} = \frac{\Delta CR^2}{\lambda}$$
(8.40)

where the last step assumes "square pixels," that is, $\Delta R = \Delta CR$. Equation (8.40) is a constraint on the maximum value of *x* and thus the maximum scene width in Doppler beam sharpening. The total allowable scene width of 2*x* is plotted as a function of resolution and RF in Fig. 8.24.

The resolution performance of DBS is sometimes quantified in a Doppler *beam sharpening ratio* (BSR), defined as the ratio of the real-beam cross-range resolution to the DBS cross-range resolution. Using Eqs. (8.3) and (8.9), the BSR is

$$BSR = \frac{\Delta CR_{\text{real beam}}}{\Delta CR_{\text{DBS}}} = \frac{(\lambda R/D_{az})}{(\lambda R/2vT_{a})}$$
$$= \frac{2vT_{a}}{D_{az}} = 2\frac{D_{SAR}}{D_{az}}$$
(8.41)

where the last step used the relation $D_{SAR} = vT_a$. Thus, the BSR is just twice the ratio of the synthetic and real aperture sizes. The factor of two reflects the difference in synthetic and real array patterns discussed in <u>Sec. 8.1.2</u>.

8.3.2 Quadratic Phase Error Effects

The term $u^2 / 2R$ was neglected in simplifying Eq. (8.35). Inclusion of this term adds a quadratic phase component to the slow-time signal. Because the DBS algorithm associates a frequency component with a cross-range coordinate, the instantaneous frequency variation implied by the quadratic phase will tend to smear the cross-range response of the processor. The ideal response of the DBS processor to a point scatterer is simply the discrete time Fourier transform of a sinusoid. Figure 8.22*a* shows the DTFT of the signal $y[m] = \exp(j\phi_{\max} m^2/M^2)$, $-M \le m \le + M$, where ϕ_{\max} is the peak phase error. This sequence can be viewed as the product of a quadratic phase sequence and an ideal sinusoid of frequency zero, that is, a vector of ones. Increasing the maximum value of the quadratic component attenuates and spreads the mainlobe. This causes a loss of brightness and resolution, respectively, in the DBS image of a scatterer (Richards, 1993). As the maximum of the quadratic component approaches π radians, the well-defined mainlobe/sidelobe structure of the response is breaking down. Figure 8.22*b* repeats the experiment with a Hamming window applied to the data before the DTFT. The Hamming window greatly moderates the effects of the phase error. The gain and resolution *variations* are reduced, and the general shape of the response is better maintained. This increased robustness of response is yet another benefit of windowing data.



FIGURE 8.22 The effect of quadratic phase error on the DTFT of a 101-point sinusoid: (a) no window, (b) hamming window.

Figure 8.23 plots the loss in peak amplitude and the increase in 3-dB mainlobe width of the DTFT as the peak quadratic phase increases. The moderating effect of the Hamming window is again evident. These plots are for a signal length of 101 samples. The degree of degradation is a mild function of
sequence length. For shorter signals, the degradations are greater; for longer signals, they are somewhat less. For the example shown, limiting the maximum quadratic phase error to $\phi_{\text{max}} \leq \pi/2$ limits the loss in amplitude to 1 dB and the resolution increase to 7 percent, even without a window. This suggests that an uncompensated quadratic phase component can be ignored so long as it does not exceed $\pi/2$ radians over the aperture time. Tighter limits can be adopted if higher quality standards are desired.



FIGURE 8.23 The effect of a quadratic phase component on the DTFT of a 101point sinusoid: (*a*) reduction in peak amplitude, (*b*) percentage increase in 3-dB resolution.

The assumption that the $u^2/2R$ term in Eq. (8.35) can be neglected can now be revisited. This term contributes a quadratic phase component to $\phi(u)$. The maximum value of u of interest is $D_{SAR}/2 = vT_a/2$. Requiring that this phase term be limited to $\pi/2$ gives

$$\frac{4\pi}{\lambda} \frac{v^2 T_a^2}{8R} \le \frac{\pi}{2}$$
(8.42)

Using $T_a = \lambda R / 2v \Delta CR$ gives the constraint

$$\Delta CR \ge \frac{1}{2}\sqrt{\lambda R} \tag{8.43}$$

Thus, the quadratic phase term can be ignored in DBS processing so long as this constraint is observed. Figure 8.24 plots the constraint of Eq. (8.43) for four values of range appropriate to airborne radars, as well as the DBS scene size constraint of Eq. (8.40). Taken together, these two constraints show that while DBS processing can achieve resolutions much better than real beam systems, it is most appropriate for imaging relatively large scenes at moderate-to-low resolution, and is best used at higher radar frequencies.



FIGURE 8.24 (a) Quadratic phase error constraint, (b) range walk constraint.

Figure 8.25*a* shows a portion of the DBS image for the same target array, but imaged with a resolution goal of 10 m. The standoff range is now 50 km, while the radar frequency remains 10 GHz. This scenario meets the range migration constraint of Eq. (8.40) (Fig. 8.24*b*), but does not meet the quadratic phase error constraint of Eq. (8.41) (Fig. 8.24*a*). The slow-time quadratic phase

term represents a cross-range spatial frequency chirp. Because spatial frequency maps to cross-range position, the scatterer response is smeared in cross range. (This is the same effect illustrated in <u>Fig. 8.22</u>.) Range resolution is not affected.



FIGURE 8.25 (*a*) Effect of quadratic phase on DBS image, (*b*) DBS image with azimuth dechirp.

This smearing can be corrected by compensating the data to remove the quadratic phase term. The required correction is implemented as a range-bin-dependent phase multiplication in each slow-time row of the raw data y[l, m]:

$$y'[l,m] = y[l,m] \exp\left[j\left(\frac{4\pi}{\lambda}\right)\left(\frac{u^2}{2R}\right)\right]$$
$$= y[l,m] \exp\left\{j\left(\frac{4\pi}{\lambda}\right)\frac{\left[v\left(m-\frac{M-1}{2}\right)T\right]^2}{2\left(R_0 + \frac{c(l-1)T_f}{2}\right)}\right\}$$
(8.44)

where T_f = fast-time sampling interval

T = slow-time sampling interval (pulse repetition interval)

 R_0 = range corresponding to the first range bin

M = number of slow-time samples

Figure 8.25*b* shows the effect of this compensation, called *azimuth dechirp*, on the same data. The full cross-range resolution is restored.

Because of the short aperture times required for its modest resolution, DBS is often used in an actively scanning mode, unlike finer-resolution radar imaging modes (Stimson, 1998). As long as the region of interest stays within the area illuminated by the mainbeam over the aperture time the scanning is of little consequence. In keeping with the scanning operation, DBS is often used in squint mode rather than in a sidelooking configuration. In this event, range walk becomes large and must be compensated. Additional enhancements to DBS can also be implemented. For instance, *range migration correction* can be implemented to compensate for fractional-bin range curvature, and *secondary range compression* can improve focusing of targets at large cross-range displacements from the line of sight. These extensions are discussed in Schleher (1991) and Showman (2013).

8.3.3 Range-Doppler Algorithms

As resolution becomes finer, standoff range shorter, or radar frequency lower, both range curvature and azimuth quadratic phase tend to become more pronounced. For example, Fig. 8.26 illustrates the amount of range curvature as a function of cross-range resolution for an L-band (1 GHz) radar at a standoff range $R_0 = 50$ km and velocity v = 150 m/s. If it is assumed that the range resolution is set approximately equal to the cross-range resolution, the range curvature is completely insignificant when $\Delta CR = 50$ m and is still only about 1/5th of a range bin when $\Delta CR = 10$ m. Augmented DBS algorithms are effective for these cases. However, when ΔCR is reduced to 3 m the curvature rises to about five range bins, assuming $\Delta R = 3$ m also. DBS algorithms are unsuitable in this situation. To achieve well-focused finer-resolution imagery, it is necessary to develop an algorithm that can address significant range migration and quadratic phase. The *range-Doppler* (RD) algorithm is a family of algorithms in widespread use that offer this capability while maintaining separable range and cross-range processing and taking advantage of 2D FFT computational efficiency.



FIGURE 8.26 Increase in range curvature as cross-range resolution becomes finer.

The RD algorithm assumes that pulse compression is performed first. For a single point scatterer, the output of the fast-time matched filter (assuming that the filter delay is removed) for the pulse transmitted from aperture position u is a narrow peak at time delay 2R(u), surrounded by sidelobes. This can be approximated by an impulse at that range, $\delta_D [t - 2R(u)]$. The responses from the series of pulses follow the range migration curve of Eq. (8.27). The response of the stripmap SAR system to a scatterer at (x, R) after pulse compression is therefore (ignoring amplitude factors)

$$h(u,t;R) = \delta_D(t-R(u)) = \delta_D\left(t - \frac{2}{c}\sqrt{(u-x)^2 + R^2}\right)$$

(8.45)

The response is parameterized by *R* but not *x* because its shape varies with *R* but is invariant to *x*. To derive the range-Doppler algorithm, begin by expressing *R* in terms of differential range δR around a nominal standoff range R_0 :

$$R(u) = \sqrt{(u-x)^2 + R^2} = \sqrt{(u-x)^2 + (R_0 + \delta R)^2}$$

= $\sqrt{(u-x)^2 + R_0^2 + 2R_0\delta R + \delta R^2}$
 $\approx \sqrt{(u-x)^2 + R_0^2 + 2R_0\delta R}$
(8.46)

The last step assumes $|\delta R| \ll R_0$, that is, the swath length is small compared to the standoff range. Define $R_x^2 = (u-x)^2 + R_0^2$; then

$$R(u) = R_x \sqrt{1 + \frac{2R_0 \delta R}{R_x^2}}$$

$$\approx R_x \left(1 + \frac{R_0 \delta R}{R_x^2}\right) = R_x + \frac{R_0 \delta R}{R_x}$$

$$= R_x + \frac{R_0 \delta R}{\sqrt{(u-x)^2 + R_0^2}}$$

$$\approx R_x + \delta R = \sqrt{(u-x)^2 + R_0^2} + \delta R$$
(8.47)

where the last line assumes that $|u - x| \ll R_0$ (synthetic aperture small compared to the standoff range). Using Eq. (8.47) in Eq. (8.45) gives the form of the PSR used in the RD algorithm:

$$h(u,t;R_{0}) \approx \delta_{D} \left[\left(t - \frac{2}{c} \delta R \right) - \frac{2}{c} \sqrt{(u-x)^{2} + R_{0}^{2}} \right]$$
$$= \delta_{D} \left[\left(t - \frac{2}{c} R + \frac{2}{c} R_{0} \right) - \frac{2}{c} \sqrt{(u-x)^{2} + R_{0}^{2}} \right]$$
(8.48)

Unlike the general PSR, the shape of this form of the PSR is invariant to both x and R.

While there is a dependence of the PSR shape on R_0 , the key characteristics of the RD algorithm is that it uses a single fixed value of R_0 , typically the value at the middle of the imaged swath. The resulting PSR of Eq. (8.48) focuses scatterers at that range essentially perfectly. Scatterers at other

ranges are not focused as well, but if the change in range from R_0 is not too large the degradation is negligible. The maximum allowable range change before defocusing becomes apparent is quantified by the *depth of focus*, to be discussed in the next section.

The image could be formed by performing a two-dimensional convolution of $h^*(-u, -t; R_0)$ with the data set y(u, t) (matched filtering). However, the processing is usually done in the frequency domain by computing the twodimensional Fourier transforms of $h(u, t; R_0)$ and y(u, t), $H(K_u, \Omega; R_0)$ and $Y(K_u, \Omega)$, and forming the image as

$$f(u, t; R_0) = \mathbf{F}^{-1} \{ H^*(K_u, \Omega; R_0) Y(K_u, \Omega) \}$$

$$f(x, \delta R; R_0) = f(u, ct/2; R_0)$$
(8.49)

To proceed, an expression is needed for the RD SAR system transfer function $H(K_u, \Omega; R_0)$. The two-dimensional Fourier transform of $h(u, t; R_0)$ is

$$H(K_{u},\Omega;R_{0}) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} h(u,t;R_{0})e^{-j\Omega t} dt \right) e^{-jK_{u}u} du$$
(8.50)

 K_u is a spatial frequency in radians per meter and Ω is a temporal frequency in radians per second. Using Eq. (8.48) and performing the Fourier transform over the fast-time variable *t* gives the intermediate result, denoted H_t to indicate transformation only over *t*,

$$H_t(K_u, \Omega; R_0) = \int_{-\infty}^{+\infty} \exp\left\{-j\left[K_u u + \Omega\left(\delta R + \sqrt{(u-x)^2 + R_0^2}\right)\right]\right\} du$$
(8.51)

The slow-time transform can be completed using the principle of stationary phase (Bamler, 1992; Raney, 1992); the resulting 2D PSR is

$$H(K_{u}, \Omega; R_{0}) = \sqrt{\frac{\pi c R_{0}}{j\Omega} \exp\left\{+jR_{0}\left[\frac{2\Omega}{c} - \sqrt{\left(\frac{2\Omega}{c}\right)^{2} - K_{u}^{2}}\right]\right\}}$$
$$\approx A \exp\left\{+jR_{0}\left[\frac{2\Omega}{c} - \sqrt{\left(\frac{2\Omega}{c}\right)^{2} - K_{u}^{2}}\right]\right\}$$
(8.52)

The last step recognizes that most SAR systems, despite their fine resolution, are relatively narrowband. The fast-time frequency Ω varies over a limited range (usually less than 10 percent), so the amplitude factor is relatively

constant.

The range-Doppler algorithm is diagrammed in Fig. 8.27. The major advantages of the RD algorithm are its ability to process the entire range swath [provided it is small compared to the standoff range as assumed in Eq. (8.46)], compensate for both range migration and quadratic phase, and utilize the computational efficiency of the two-dimensional FFT. Since this version operates in the spatial frequency/temporal frequency domain, however, it is not clear why the technique is called the range-Doppler algorithm. In fact, the name range-Doppler derives from a version of the algorithm obtained by additional simplifications of Eq. (8.52).



FIGURE 8.27 Block diagram of range-Doppler algorithm.

Using a binomial expansion and assuming $2\Omega/c \gg K_u$, the argument of the complex exponential in *H* can be simplified as follows:

$$R_{0}\left(\frac{2\Omega}{c}-\sqrt{\left(\frac{2\Omega}{c}\right)^{2}-K_{u}^{2}}\right) = R_{0}\left(\frac{2\Omega}{c}-\frac{2\Omega}{c}\sqrt{1-\frac{K_{u}^{2}}{(2\Omega/c)^{2}}}\right)$$
$$\approx R_{0}\left[\frac{2\Omega}{c}-\frac{2\Omega}{c}\left(1-\frac{K_{u}^{2}}{2(2\Omega/c)^{2}}\right)\right] = \frac{cK_{u}^{2}R_{0}}{4\Omega}$$
(8.53)

Again, even fine-resolution SAR systems are narrowband in the sense that the waveform bandwidth $\Delta\Omega$ is much less than the RF Ω_0 . Expanding Ω about Ω_0 and using $\Delta\Omega \ll \Omega_0$ further simplifies Eq. (8.53) to

$$\frac{cK_u^2 R_0}{4\Omega} = \frac{cK_u^2 R_0}{4(\Omega_0 + \Delta\Omega)} = \frac{cK_u^2 R_0}{4\Omega_0(1 + \Delta\Omega/\Omega_0)}$$
$$\approx \frac{cK_u^2}{4\Omega_0} \left(1 - \frac{\Delta\Omega}{\Omega_0}\right) = \frac{cK_u^2 R_0}{4\Omega_0} - \frac{cK_u^2 R_0}{4\Omega_0^2} \Delta\Omega$$
(8.54)

To obtain the first expression in the second line, the term $(1 + \Delta\Omega/\Omega_0)^{-1}$ was expanded in a binomial series and only the first two terms retained. The simplified transfer function is

$$H(K_{u}, \Delta\Omega; R_{0}, \Omega_{0}) \approx A \exp\left(j\frac{cK_{u}^{2}R_{0}}{4\Omega_{0}}\right) \exp\left(-j\frac{cK_{u}^{2}R_{0}}{4\Omega_{0}^{2}}\Delta\Omega\right)$$
(8.55)

The approximate transfer function of Eq. (8.55) implies two distinct operations for performing the RD cross-range compression. For a given crossrange spatial frequency K_u , the second term is a linear phase in the fast-time frequency $\Delta\Omega$, corresponding to a shift in the fast time dimension of $dK_u^2 R_0/4\Omega_0^2$ seconds. This shift in the data "straightens out" the range curvature in the PSR. Because the shift is a function of K_u rather than u itself, it must be applied after a slow-time DFT, that is, to the range-Doppler data. The first phase term compensates the slow-time quadratic phase modulation of the data, essentially performing an LFM pulse compression in slow time. Figure 8.28 is a purely notional illustration of this version of the range-Doppler algorithm, beginning with the pulse-compressed fast-time/slow-time data matrix in step 1 and continuing to the compressed point spread response in step 4.



FIGURE 8.28 Notional illustration of sequence of operations in classic range-Doppler algorithm.

The fast-time shift can be implemented as a frequency domain multiplication or as an explicit shift and interpolation in the fast-time domain. In the latter case, the data are operated on in the time domain in one dimension (fast time, or range) and the frequency domain in the other (spatial frequency, or Doppler), hence the name "range-Doppler" algorithm.

Figure 8.29 illustrates the performance of the range-Doppler algorithm on two simulated scenarios using the full RD point spread response of Eq. (8.52). In Fig. 8.29*a*, the RD algorithm is applied to a collection of point scatterers in a 50×50 m area sensed with an X-band (9.5 GHz) radar from a standoff range of 7.5 km. While the range and cross-range resolutions are only 0.5 m, this example is relatively unchallenging due to the high RF frequency and long standoff range compared to the small scene size. All five scatterers are well-focused, with no visible degradations in the scene. Figure 8.29*b* is a more challenging case. Although the resolution is relaxed to 1.0 m in both dimensions, the image area is larger (100 m by 100 m), the standoff range is shorter (4.1 km), and the RF frequency is lowered to L band (1.5 GHz). All of these factors except the relaxed resolution increase the variability of the range curvature as a function of range. Although differences in sidelobe structure are evident, the scatterer at the center reference range R_0 remains well focused. However, the cross-range resolution of the other scatterers degrades in proportion to their

distance from the scene center.



FIGURE 8.29 Magnitude in dB of the range-Doppler image of a test array: (*a*) X band, long standoff range, small scene size, (*b*) L band, short standoff range, larger scene size. See text for parameters. (Images courtesy of Dr. Gregory A. Showman, GTRI.)

8.3.4 Depth of Focus

The exact stripmap PSR of Eq. (8.45) is range dependent. The range-Doppler algorithm linearizes the PSR about a particular range R_0 . However, if applied over a deep enough swath, the variation in the PSR with R will become significant. Failure to address this variation will result in increasingly poor focusing at increasing ranges from R_0 because of the increasing mismatch between a scatterer's actual PSR and the PSR at R_0 . Advanced algorithms such as the *range-migration algorithm* (Bamler, 1992; Gough and Hawkins, 1997) fully account for the variation in the PSR with range, but are not discussed in this introduction. Another approach is to break the desired swath into N subswaths centered on ranges R_{01}, \ldots, R_{0N} . The PSR is updated and the RD algorithm applied independently in each subswath. The resulting images are mosaiced together to form the complete image. The swath length over which the RD algorithm can be applied without significant defocusing is called the *depth of focus* (DOF).

The DOF is determined by finding the change in range R_0 which will cause a specified *change* in range curvature over the aperture time. The tolerable variation is often taken as $\lambda/8$, corresponding to a two-way variation of $\lambda/4$ and thus a phase change of $\pi/2$ radians. It was seen in Sec. 8.3.2 that a quadratic phase error limited to this amount would cause minimal degradation in the DFT of a sinusoid, a model that is applicable to cross-range matched filtering once the primary quadratic phase is removed.

Equation (8.32) gave a formula for range curvature. Differentiating with respect to range gives the change in curvature as nominal range changes; setting that quantity equal to $\lambda/8$ gives the maximum change in range as (see Prob. 20)

$$\delta R_{\max} = \frac{\lambda R_0^2}{D_{SAR}^2} = \frac{\lambda R_0^2}{(vT_a)^2}$$
(8.56)

Since this range change can occur in either direction from R_0 , the DOF is twice this:

$$DOF = 2\delta R_{\max} = \frac{2\lambda R_0^2}{D_{SAR}^2} = \frac{2\lambda R_0^2}{(vT_a)^2} = \frac{8(\Delta CR)^2}{\lambda}$$
(8.57)

Consider SAR imaging of a scene with a swath depth of 10 km. If imaged using an X-band (10 GHz) radar with a cross-range resolution of 3 m, the depth of focus is 2.4 km. To use the RD algorithm, the swath should be broken into at least four and possibly more subswaths, with the PSR updated for each. If the scene was imaged with an L-band (1 GHz) radar at 1 m resolution, the depth of

focus is only 240 m, requiring at least 42 subswaths to use the RD algorithm. The large number of subswaths needed indicates that the RD algorithm is poorly suited to this scenario and a more advanced algorithm such as range migration should be used instead. As another example, the scenario of Fig. 8.29*a* has a DOF of 63.33 m = ± 31.67 m. This exceeds the maximum 20 m range offset from the scene center to the most distant scatterer, so all scatterers are focused well. Figure 8.29*b* has a DOF of 40 = ± 20 m, only half the maximum range displacement of approximately 40 m in that scenario. Scatterers at ranges greater than 20 m from the scene center are significantly smeared in cross range.

8.4 Spotlight SAR Data Characteristics

The spotlight SAR scenario was illustrated in Fig. 8.11. Many spotlight SAR systems utilize a linear FM waveform with stretch processing. Assume an LFM pulse that sweeps from $F_0 - \beta/2$ to $F_0 + \beta/2$ Hz in the time interval from $-\tau/2$ to $+\tau/2$ seconds. Adapting to the notation of this chapter, it was shown in Chap. 4 (Eq. 4.110) that on a pulse taken from aperture position *u*, an ensemble of scatterers at ranges $\delta R_i = c \delta t_i/2$ relative to a central range R_0 results in a signal at the stretch mixer output:

$$y(t) = \sum_{i} \tilde{\rho}_{i} \exp\left(-j\frac{4\pi R_{i}}{\lambda}\right) \exp\left(-j2\pi \frac{\beta}{\tau} \delta t_{i}(t-t_{0})\right) \exp\left(j\pi \frac{\beta}{\tau} (\delta t_{i})^{2}\right)$$
$$= e^{-j4\pi R_{0}/\lambda} \sum_{i} \tilde{\rho}_{i} \exp\left[-j\left(\Omega_{0} + 2\pi \frac{\beta}{\tau}(t-t_{0})\right) \delta t_{i}\right] \exp\left(j\pi \frac{\beta}{\tau} (\delta t_{i})^{2}\right)$$
(8.58)

where $t_0 = 2R_0/c$ and \tilde{P}_i is the complex reflectivity of the echo from the *i*th scatterer,² located at range $R_0 + \delta R_i$. Extending this formula to a continuum of scatterers gives

$$y(t) = w(t)e^{-j4\pi R_0/\lambda} \int_{-\infty}^{+\infty} \tilde{\rho}(\delta t) \exp\left[-j\left(\Omega_0 + 2\pi\frac{\beta}{\tau}(t-t_0)\right)\delta t\right] \exp\left(j\pi\frac{\beta}{\tau}(\delta t)^2\right) d(\delta t)$$
(8.59)

In this equation, w(t) is a rectangular window function that limits the data extent to that of the reference chirp in the stretch mixer. Specifically

$$w(t) = \begin{cases} 1, & t_0 - \frac{L_s}{c} - \frac{\tau}{2} \le t \le t_0 + \frac{L_s}{c} + \frac{\tau}{2} \\ 0, & \text{otherwise} \end{cases}$$
(8.60)

Consider the Fourier transform of $\tilde{\rho}(\delta t)$

$$\tilde{P}(\Omega) = \int_{-\infty}^{+\infty} \tilde{\rho}(\delta t) \exp(-j\Omega \delta t) d(\delta t)$$
(8.61)

Comparing Eq. (8.61) to Eq. (8.59) and using the time limits established by Eq. (8.60) shows that

$$y(t) = w(t)e^{-j4\pi R_0/\lambda} \tilde{P}\left(\Omega_0 + \frac{2\pi\beta}{\tau}(t-t_0)\right), \quad t_0 - \frac{L_s}{c} - \frac{\tau}{2} \le t \le t_0 + \frac{L_s}{c} + \frac{\tau}{2}$$
(8.62)

provided that the *residual video phase* (RVP) term $\exp[j\pi\beta(\delta t)^2/\tau]$ can be ignored; this condition will be reconsidered later. Equation (8.62) shows that stretch processing of an LFM waveform has swapped the time and frequency domains in a sense: the mixer output is a time-domain voltage whose numerical values trace out the value of the Fourier transform $\tilde{P}(\Omega)$ of the scene reflectivity versus range, $\tilde{\rho}(\delta t)$. Put another way, the LFM stretch processor acts as a spectrum analyzer for the reflectivity distribution.

The range of frequencies evaluated by this spectrum analyzer is determined by evaluating the argument of $\tilde{P}(\cdot)$ for the maximum and minimum values of t allowed by the window. Converting to cyclical frequency units, the result is

$$y(t) = \tilde{P}(F), \quad F \in F_0 \pm \left(1 + \frac{L_s}{(c\tau/2)}\right) \frac{\beta}{2}$$
$$\approx \tilde{P}(F), \quad F \in F_0 \pm \frac{\beta}{2}$$
(8.63)

The second line assumes that $L_3 \ll c\tau/2$, that is, the swath length is small compared to the pulse duration. Stretch processing systems are usually designed such that this is the case. Equation (8.63) states the remarkable result that the *time* domain output of the stretch mixer traces out the Fourier transform $\tilde{P}(F)$ of the range profile $\tilde{\rho}(\delta t)$ over the frequency interval $[F_0 - \beta/2, F_0 + \beta/2]$. This is intuitively satisfying since this is exactly the frequency interval over which the LFM pulse sweeps.

 $\tilde{P}(F)$ can be rescaled into units of spatial frequency in radians per meter via the transformation $F \rightarrow cK_R/4\pi$. Using $\Delta R = c/2\beta$, the mixer output range becomes $K_R \in (4 \pi/\lambda_0) \pm (\pi/\Delta R)$ radians per meter. Figure 8.30 is a notional illustration of the portion of the range profile spectrum $\tilde{P}(F)$ or $\tilde{P}(K_R)$ measured on a single pulse by the stretch processor.



FIGURE 8.30 Portion of the range profile bandwidth measured by the stretch processor, in both temporal and spatial frequency units.

As discussed in Sec. 2.7.3, $\tilde{\rho}(\delta t)$ is a transformation of the scatterers in a two-dimensional scene into the one-dimensional range profile. This concept is illustrated in Fig. 8.31. The complex reflectivities of all of the scatterers on the isorange contour corresponding to a delay of $t_0 + \delta t_0$ are weighted by the antenna pattern and integrated to form the value of the range profile at time $t_0 + \delta t_0$, $\tilde{\rho}(\delta t_0)$. "CRP" again marks the central reference point at range $R_0 = ct_0/2$. If the width of the illuminated area is much less than the nominal range, $R_0\theta_{az} \ll R_0$, the lines of integration become nearly straight. The range averaging then approximates a *projection* in the tomographic sense of the two-dimensional scene into a one-dimensional function.



FIGURE 8.31 Projection of a two-dimensional scene into a one-dimensional angle-averaged range profile.

Consider the two coordinate systems shown in Fig. 8.32. The (p_{θ}, q_e) axes are rotated by θ radians with respect to the (u, R) axes. The coordinate systems are related by the equations



FIGURE 8.32 Coordinate systems for defining a two-dimensional projection.

$$u = p_{\theta} \cos\theta - q_{\theta} \sin\theta$$
$$R = p_{\theta} \sin\theta + q_{\theta} \cos\theta$$
(8.64)

A projection of a two-dimensional effective reflectivity scene $\rho'(u, R)$ into a one-dimensional cross-range-averaged reflectivity range profile $\tilde{\rho}_{\theta}(p_{\theta})$ is defined as

$$\tilde{\rho}_{\theta}(p_{\theta}) = \int_{-\infty}^{+\infty} \rho'(p_{\theta} \cos\theta - q_{\theta} \sin\theta, p_{\theta} \sin\theta + q_{\theta} \cos\theta) \, dq_{\theta}$$
(8.65)

The *projection-slice theorem* of Fourier analysis then states that the onedimensional Fourier transform of the projection is a slice of the twodimensional Fourier transform of the original function $\rho'(u, R)$ (Dudgeon and Mersereau, 1984)

$$\int_{-\infty}^{+\infty} \tilde{\rho}_{\theta}(p_{\theta}) e^{-jp_{\theta} U} dp_{\theta} = \mathbf{P}'(U\cos\theta, U\sin\theta)$$
(8.66)

Figure 8.33 illustrates the consequences of the LFM/stretch data acquisition and the projection-slice theorem for spotlight SAR. The scene $\rho'(u, R)$ on the left represents a patch of terrain, perhaps containing a road, building, stand of trees, and small lake. The drawing on the right represents the Fourier transform P'(K_u, K_R) of the scene in spatial frequency units. The radar views the scene from a particular angle. The radar and antenna effectively project the two-dimensional scene into a one-dimensional function. The LFM waveform and

stretch processing measure a portion of the spectrum of that one-dimensional function. The projection-slice theorem allows that spectral segment to be interpreted as a measurement of the two-dimensional Fourier transform of the scene along the same angle, as shown on the right side of the sketch. Because of the limited bandwidth of the stretch measurement, only the segment of the total slice highlighted in a lighter color is actually measured (see Fig. 8.30).



FIGURE 8.33 Spotlight SAR data acquisition model.

As the aircraft flies along the synthetic aperture, it continues to transmit pulses while steering the antenna to remain aimed at the CRP of the region of interest. Each successive pulse measures a segment of the spectrum of a projection of $\rho'(u, R)$ at a new angle and therefore a segment of a slice of P'(K_u , K_R) at that same angle. Over a series of pulses the system therefore collects data over an annular region of the two-dimensional spectrum of the image. This model of the spotlight SAR data set is depicted in Fig. 8.34. The extent of the annulus is $4\pi\beta/c$ radians per meter = $2\beta/c$ cycles per meter in the radial direction. The spatial range resolution is then $(2\beta/c)^{-1} = \Delta R$ meters as expected. The width of the annulus at its center $K_R = 4\pi/\lambda_0$ is $(4\pi/\lambda_0)\gamma$ radians per meter, so the spatial cross-range resolution is $\lambda/2\gamma$, again as expected. Finally, note that the spectral data are on a polar, rather than rectangular, grid in (K_u , K_R). For this reason, the data set is referred to as *polar format data*.



FIGURE 8.34 Spotlight SAR data acquisition maps an annular region in Fourier space.

An interesting question is the number of projections required to reconstruct the image; the answer determines the required PRF. Suppose that the final image p'(u, R) desired is $L_u \times L_s$ meters (cross-range \times range). To avoid aliasing artifacts over a region of this size, Nyquist sampling theory requires that samples of P'(K_u, K_R) be no more than $1/L_u$ cycles per meter $= 2\pi/L_u$ radians per meter apart in K_u ; similarly, samples should be no more than $1/L_s$ cycles per meter = $2 \pi/L_s$ radians per meter apart in K_R . Consider the K_u dimension. At the center of the annulus ($K_R = 4\pi/\lambda_0$), a sample spacing of $2\pi/L_u$ corresponds to an angular spacing between successive slices of $\lambda_0/2L_u$ radians. The number of slices required to span the total angular extent of the annulus is then

$$N_{\gamma} = \frac{\gamma}{\lambda_0/2L_u} = L_u \left(\frac{2\gamma}{\lambda_0}\right) = \frac{L_u}{\Delta CR}$$
(8.67)

which is simply the number of resolution cells in the cross-range extent. The linear distance traveled by the radar platform between slices and the resulting PRI are

$$\Delta u = R_0 \frac{\lambda_0}{2L_u} \implies PRI = \frac{\lambda_0 R_0}{2vL_u}$$
(8.68)

If L_u is set equal to its practical maximum of the illuminated beamwidth $R_0\theta_{az} = R_0\lambda_0/D_{az}$, the platform travel between pulses is $D_{az}/2$, the same minimum value required in stripmap SAR to avoid Doppler ambiguities. As explained earlier, a higher cross-range sampling rate is often used in stripmap and also in spotlight

SAR to minimize ambiguities.

The minimum number of range samples needed (equivalent to time samples of the stretch mixer output for each pulse) is the radial extent of the annulus divided by the sample spacing in K_R , which reduces to the number of range cells in the range extent:

$$N_R = \frac{(4\pi\beta/c)}{(2\pi/L_s)} = \frac{L_s}{\Delta R}$$
(8.69)

Since the stretch mixer output traces out the two-dimensional Fourier transform slice, the interval $1/L_R$ between samples in K_R specifies the time Δt between samples at the mixer output, using the mapping of time to spatial frequency $K_u = (4\pi\beta/c\tau)t$:

$$\Delta t = \frac{c\tau}{4\pi\beta} \frac{2\pi}{L_R} = \frac{c\tau}{2\beta L_R}$$
(8.70)

It is important to remember that this spotlight data model assumes an LFM waveform and a stretch receiver. In addition, two assumptions were made in deriving the results above. The first is that the RVP term in Eq. (8.59) can be ignored. The second is that the curvature in the isorange contours can be ignored in modeling the mixer output $\tilde{\rho}(\delta t)$ as a projection $\tilde{\rho}_{\sigma}(p_{\sigma})$. Both assumptions are valid provided the scene size is not too large. The limitation due to isorange curvature can be shown to be

$$L_{u} < 2\Delta CR \sqrt{\frac{2R_{0}}{\lambda}}$$

$$(8.71)$$

while the limitation due to RVP is

$$L_{u} < \frac{2\Delta CR \cdot F_{0}}{\sqrt{\beta/\tau}}$$
(8.72)

Both are derived in Jakowatz (1996, App. B). Figure 8.35 plots these two constraints for $R_0 = 10$ km, $\tau = 10$ µs, and $\beta/\tau = 0.75$ MHz/µs. The limit on scene size due to RVP is very loose; the limit due to isorange curvature is more constraining in this example.



FIGURE 8.35 Scene size limitations in the polar format algorithm: (a) limit due to isorange contour curvature, (b) limit due to residual video phase.

8.5 The Polar Format Image Formation Algorithm for Spotlight SAR

Given the spotlight SAR data model, the basic image formation algorithm is

fairly obvious: inverse Fourier transform the available two-dimensional spectral region of $P'(K_u, K_R)$ to estimate the image p'(u, R). However, a conventional inverse DFT algorithm assumes the spectral data are on a rectangular grid in (K_u, K_R) , while the spotlight data are on a polar grid. Consequently, an extra step is required to interpolate the data from a polar to a rectangular grid before the IDFT can be applied. In practice, the data will often be windowed in both dimensions after interpolation in order to reduce sidelobes in the final image. The resulting *polar format algorithm* (PFA) is diagrammed in Fig. 8.36.



FIGURE 8.36 Polar format image formation algorithm for spotlight SAR.

The key step in the PFA is the polar-to-rectangular interpolation of the Fourier data. While a two-dimensional interpolation would be theoretically ideal, separable schemes using successive one-dimensional interpolations in two different dimensions are generally preferred for simplicity. Two separable approaches are commonly used. In either approach, a rectangular grid of desired (K_{μ}, K_{R}) sample locations is first established based on scene size and resolution considerations. The rectangular region covered by this grid must be contained within the annular region of available data. The first approach, illustrated in Fig. 8.37a, appears to be the more commonly used of the two. The complex polar format spectral data are interpolated along each radial to the keystone grid depicted in the center of the figure. After this step, the data are evenly spaced in K_R but not in K_{u} . A second series of one-dimensional interpolations, now in the K_{μ} dimension, align the samples to the desired grid in K_{μ} . In principle, the keystone grid could be directly measured by the SAR system by varying the mixer output sample rate on successive pulses in a carefully controlled fashion.



FIGURE 8.37 Two schemes for separable polar-to-rectangular interpolation: (*a*) radial-keystone interpolation, (*b*) angular-range interpolation.

The second approach, depicted in Fig. 8.37*b*, interpolates first along constant-radius lines to align data on the desired K_u sample locations. The intermediate grid is then interpolated in the K_R dimension to the final rectangular grid (Munson et al., 1985).

In the radial-keystone scheme, the interpolation is based on a separable two-dimensional sinc kernel that arises from the bandlimited nature of the data. Practical implementations window the sinc kernel to both limit sidelobes and provide a finite-length interpolating kernel in each dimension. Interpolating kernels typically range from 7th to 11th order, and the total computational load for polar-to-rectangular interpolation often rivals that of the two-dimensional inverse FFT used for the final image formation. The angular-range scheme uses a periodic asinc kernel in the angular interpolation and a sinc kernel in the range interpolation. A comparison of several of these schemes is given in Munson et al. (1985). A detailed discussion of the implementation of the radial-keystone approach is given in Jakowatz et al. (1996, Chap. 3).

At higher radar frequencies the annular region of polar format spectral data are further from the origin in (K_u, K_R) space. If the resolution is relatively coarse the required bandwidth is also small, so the extent of the annulus in range and angle is small. In this case the polar grid may be nearly rectangular, in which case the computationally expensive and quality-limiting polar-to-rectangular interpolation can be avoided and the image formed with a simple two-dimensional IFFT. This approach is called the *rectangular format algorithm* (RFA). It is essentially the DBS algorithm, but with an inverse FFT required in

the range dimension to transform the stretch receiver output back to the fast-time domain. Consequently, the constraints to DBS processing apply to the RFA algorithm as well.

While the polar format algorithm is a commonly used technique in many operational spotlight SAR systems, other classes of algorithms can be applied to spotlight image formation. Application of the range migration and chirp scaling algorithms to spotlight SAR is described in Carrara et al. (1995). Interest is growing in *backprojection* methods for both strip-map and spotlight modes despite their greater computational cost due to their increased flexibility in addressing nonrectangular data formats, avoidance of interpolations (which are a major source of image degradations), and ability to form images from non-rectilinear flight paths. Backprojection methods are discussed in Desai and Jenkins (1992), Jakowatz et al. (2008), Frey et al. (2009), and Showman (2013).

8.6 Interferometric SAR

One of the newest developments in synthetic aperture radar is the ability to do fine-resolution imaging in three dimensions using interferometric techniques, commonly called IFSAR or InSAR. The basic approach employs two complex SAR images of a scene formed using two displaced apertures. The apertures may be physically separate receive apertures on a single antenna structure, often with a common transmit aperture. Alternatively, IFSAR can be implemented using images collected from a conventional single aperture system on multiple passes. A good introductory discussion is Richards (2013).

8.6.1 The Effect of Height on a SAR Image

The output of a SAR image formation algorithm is a complex-valued twodimensional image: both an amplitude and phase for each pixel. Conventional two-dimensional SAR imaging discards the phase of the final image, displaying only the magnitude information. In IFSAR, the pixel phase data are retained. Because SAR image formation is a nominally linear process, the complex amplitude $f(x, y_g)$ of a pixel at ground coordinates (x, y_g) and elevation of z = 0can be viewed as the product of four factors:

$$f(x, y_g) = A \cdot G \cdot e^{-j4\pi R_f(x, y_g)/\lambda} \rho(x, y_g)$$

= $A \cdot |G| \cdot |\rho(x, y_g)| \cdot \exp\left\{j\left[\phi_G + \phi_\rho(x, y_g) - \frac{4\pi}{\lambda}R_f(x, y_g)\right]\right\}$
(8.73)

In this equation, ρ is the complex reflectivity of the pixel, the complex exponential term is the phase shift due to the range to the pixel, *G* is the complex

gain of the receiver and SAR image formation algorithm, A contains all range equation factors, and R_f is the range from the aperture phase center to the ground coordinates (x, y_g) . The phase of the pixel is therefore

$$\phi_f(x, y_g) = \phi_G + \phi_\rho(x, y_g) - \frac{4\pi}{\lambda} R_f(x, y_g)$$
(8.74)

Now consider the two scatterers **P1** and **P2** illustrated in Fig. 8.38. Both are at ground range y_{g1} but one is at an elevation $z = h_0$ relative to some unknown reference plane, while the other is at an elevation $z = h_0 + \Delta h$. They are observed from two distinct radar apertures at an altitude $z = h_0 + Z$ and separated horizontally by a *baseline B*.⁸ Each aperture independently transmits a radar waveform, receives the data, and forms a complex SAR image of the scene; these images are denoted by $f(x, y_g)$ and $g(x, y_g)$. The baseline should be orthogonal to the flight path; therefore the direction of the aircraft motion is into the page.⁹ The radar range is great enough that the incoming wavefront can be considered planar. If the depression angle from the middle of the baseline to **P1** i s ψ , the difference in range to the two aperture phase centers is well approximated as $B \cos \psi$. The difference in received phase at the two apertures then becomes, using Eq. (8.74)



FIGURE 8.38 Geometry for interferometric height estimation.

$$\phi_{fg} \equiv \phi_f - \phi_g \approx -\frac{4\pi}{\lambda} B \cos \psi \tag{8.75}$$

 ϕ_{fg} is called the *interferometric phase difference* (IPD).

Now consider scatterer **P2**, also located at ground range y_{g1} but elevated by Δh meters. Because the radar measures time delay and thus slant range, the echo from **P2** will be indistinguishable from that of a scatterer located on the ground plane at the range where the wavefront impacting **P2** also strikes the ground. As shown in Fig. 8.38, this ground coordinate is approximately:

$$y_{g2} = y_{g1} - \Delta h \tan \psi \tag{8.76}$$

Because SAR images are two dimensional,¹⁰ P2 will be incorrectly imaged at range y_{g2} . The imaging of the elevated scatterer at an incorrect range coordinate is termed *layover* because the scatterer appears to have been shifted toward the radar. As illustrated, the layover is only in the range coordinate. In squinted operation, layover occurs in both range and cross-range; details are given in Sullivan (2000) and Jakowatz et al. (1996).

Elevating the scatterer at y_{g1} to height Δh will also change the depression angle from the center of the IFSAR baseline to the scatterer. This change can be found by differentiating Eq. (8.75) with respect to the grazing angle:

$$\frac{d\phi_{fg}}{d\psi} = \frac{4\pi}{\lambda} B \sin \psi \quad \Rightarrow \quad \Delta \psi = \frac{\lambda}{4\pi B \sin \psi} \Delta \phi_{fg}$$
(8.77)

This equation states that a change in the IPD of $\Delta \phi_{fg}$ implies a change in depression angle to the scatterer of $\Delta \psi$ radians. To relate this depression angle change to an elevation change, consider <u>Fig. 8.38</u> again, which shows that

$$\frac{\Delta h}{H} \approx \frac{\Delta \psi}{\tan \psi}$$
(8.78)

Eliminating y_g using a series approximation of the tangent function and assuming that $\Delta \psi$ is small, Eq. 8.78 can be reduced to

$$(Z - \Delta h) \tan \psi = Z \tan \psi - Z \Delta \psi \implies \Delta h = Z \cot \psi \Delta \psi$$
(8.79)

Finally, using Eq. (8.79) in Eq. (8.77) gives a measure of the relationship between the change in the IPD for a given pixel and a change in the scatterer

elevation (Carrara et al., 1995):

$$\Delta h = \frac{\lambda Z \cot \psi}{4\pi B \sin \psi} \Delta \phi_{fg}$$
(8.80)

Equation (8.80) is the basic result of IFSAR.

The IPD $\phi_{fg}(x, y_g)$ can be written as

$$\phi_{fg}(x, y_g) = \phi_{fg}(x_0, y_{g0}) + \Delta \phi_{fg}(x, y_g)$$
(8.81)

Multiplying Eq. (8.81) by the scale factor in Eq. (8.80) gives

$$h(x, y_g) = \frac{\lambda Z \cot \psi}{4\pi B \sin \psi} [\phi_{fg}(x_0, y_{g0}) + \Delta \phi_{fg}(x, y_g)]$$

$$\equiv h_{offset} + \Delta h(x, y_g)$$
(8.82)

Thus, given an IPD map $\phi_{fg}(x, y_g)$, multiplying by the scale factor of Eq. (8.80) gives a height map that gives the height variations from pixel-to-pixel but has an unknown overall offset. That is, the IFSAR technique provides a good measure of *relative* height versus spatial location.

Equation (8.82) describes height variations around a given depression angle ψ , implying a fixed ground plane range. If the terrain is perfectly flat, the IPD will still vary with range because ψ will vary. The variation of the IPD with range for flat terrain is called the *flat earth phase difference*, $\phi_{fg}^{FE}(x, y_g)$. This quantity must be subtracted from the measured IPD before it is scaled to height in order to obtain height relative to the same reference level at each range.

Because the radar signal processor can only measure the phase of a signal sample as the arctangent of the ratio of its imaginary and real parts, only the *wrapped* IPD $((\phi_{fg}))_{2\pi}$ can be measured, where the notation $((\cdot))_{2\pi}$ indicates arithmetic modulo 2π . Consequently, $\Delta \phi_{fg}$ is also measured modulo 2π .

The two SAR images required for IFSAR processing can be generated in one of two basic ways. In *one-pass IFSAR*, the approach described above, the radar platform has two displaced receive apertures so that the data for both images are collected on a single pass. This approach has the advantages of operational simplicity, much greater ease of aligning the trajectories of the two apertures, and no decorrelation of the scene between passes. *Two-pass IFSAR* uses a conventional single-receiver SAR system and flies two independent passes past the scene to be imaged. This approach requires careful offset alignment of the two flight paths to establish a constant baseline *B*, which can be difficult to achieve in airborne systems but is very feasible for orbiting radars. The chief advantage of two-pass IFSAR is that it can be implemented with existing conventional single-receiver SAR sensors.

8.6.2 IFSAR Processing Steps

Formation of an IFSAR image involves the following major steps:

- Formation of the two individual SAR images, f[l, m] and g[l, m]
- Registration of the two images
- Formation of the wrapped interferometric phase difference map $((\phi_{fg}[l, m]))_{2\pi}$
- Smoothing of $((\phi_{fg}[l, m]))_{2\pi}$ to reduce phase noise
- Two-dimensional *phase unwrapping* to obtain $\phi_{fg}[l, m]$ from $((\phi_{fg}[l, m]))_{2\pi}$
- Scaling of the unwrapped phase map $\phi_{fg}[l, m]$ to obtain the height map $\Delta h[l, m]$
- Orthorectification to correct layover based on the height information.

The images are formed using any SAR imaging algorithm appropriate to the collection scenario. Because the height estimation depends on the difference in phase of the echo from each pixel at the two apertures, it is important to ensure that like pixels are compared. The slightly different geometries of the two offset apertures will result in slight image distortions relative to one another, so an image registration procedure is used to warp one image to align well with the other. Many registration procedures have been developed in the image processing literature. One that is popular in IFSAR uses a series of correlations between small subimages of each SAR map to develop a warping function. This concept is illustrated in Fig. 8.39. The actual resampling is typically done with simple bilinear interpolators. The procedure is described in detail in Jakowatz et al. (1996). Once the two images are registered, the wrapped phase map is easily computed as



FIGURE 8.39 Generation of image pair warping function.

$$((\phi_{f_{g}}[l,m]))_{2\pi} = \arg\{f[l,m]g^{*}[l,m]\}$$
(8.83)

Local averaging, typically using a 3×3 , 5×5 , or 7×7 window, is often applied to the phase map at this point to smooth phase noise at the cost of a loss of spatial resolution in the phase map and thus ultimately in the height map.

The two-dimensional phase unwrapping step to recover $\phi_{fg}[l, m]$ from $((\phi_{fg}[l, m]))_{2\pi}$ is the heart of IFSAR processing. Unlike many two-dimensional signal processing operations such as FFTs, two-dimensional phase unwrapping cannot be decomposed into one-dimensional unwrapping operations on the rows and columns. Two-dimensional phase unwrapping is an active research area. A brief introduction is given in Richards (2013) and a very thorough discussion is given in Ghiglia and Pritt (1998). Most unwrapping techniques can be classified as either *path following methods* or *minimum norm methods;* the latter are often based on fast transform techniques. An example of a minimum norm algorithm is the *discrete cosine transform* (DCT) method given in Ghiglia and Romero (1994). This technique finds an unwrapped phase function such that, when rewrapped, it minimizes the mean squared error between the gradients of the rewrapped phase function and the original measured wrapped phase.

The method begins by computing the wrapped gradients in the range (l) and cross-range (m) dimensions of the raw wrapped phase history data; these are then combined into a "driving function" d[l, m]:

$$\Delta y_{g}[l,m] = \begin{cases} \left(\left(\left(\left(\phi_{fg}[l+1,m] \right) \right)_{2\pi} - \left(\left(\phi_{fg}[l,m] \right) \right)_{2\pi} \right) \right)_{2\pi}, \\ 0 \le l \le L-2, \ 0 \le m \le M-1 \\ 0, & \text{otherwise} \end{cases} \\ \Delta y_{g}[l,m] = \begin{cases} \left(\left(\left(\left(\phi_{fg}[l,m+1] \right) \right)_{2\pi} - \left(\left(\phi_{fg}[l,m] \right) \right)_{2\pi} \right) \right)_{2\pi}, \\ 0 \le l \le L-1, \ 0 \le m \le M-2 \\ 0, & \text{otherwise} \end{cases} \\ d[l,m] = \left(\Delta_{yg}[l,m] - \Delta_{yg}[l-1,m] \right) + \left(\Delta_{x}[l,m] - \Delta_{x}[l,m-1] \right) \end{cases}$$
(8.84)

Let D[k,p] be the two-dimensional DCT of the driving function. The estimate of the unwrapped phase is then obtained as the inverse DCT of a filtered DCT spectrum:

$$\hat{\phi}_{fg}[l,m] = DCT^{-1} \left\{ \frac{D[k,p]}{2\left[\cos\left(\frac{\pi k}{M}\right) + \cos\left(\frac{\pi p}{N}\right) - 2\right]} \right\}$$
(8.85)

This function is then used in Eq. (8.80) to estimate the terrain height map $\Delta h[l, m]$.

As a simple idealized demonstration of this algorithm, consider the simulated terrain profile of a hill shown in Fig. 8.40. This function was created as the outer product of two one-dimensional Hann window functions. A simulation of one-pass IFSAR data collection produces the wrapped interferometric phase function of Fig. 8.41*a*. In addition, noise has been added to a rectangular patch of the phase data to simulate a low-reflectivity or degraded area. Straightforward application of Eqs. (8.84) and (8.85) produces the unwrapped interferometric phase map estimate of Fig. 8.41*b*. The unwrapping is successful even in the noisy region. When converted back to terrain height, this phase map accurately reconstructs the hill height profile, although the height estimate is noisy in the degraded region.



FIGURE 8.40 Height profile of simulated hill.



FIGURE 8.41 Example of DCT-based two-dimensional phase unwrapping on "hill" height map with noisy data patch: (*a*) wrapped phase map $((\phi_{fg}))_{2\pi}$, (*b*) estimated unwrapped phase ϕ_{g} . (Images courtesy of Mr. Will Bonifant.)

IFSAR processing produces only relative height variations. Absolute height can be estimated by a variety of techniques. The most common is simply the use of surveyed reference points within an image; height relative to this point is then easily converted to absolute height. Another method, analogous to the use of multiple PRIs to resolve range and Doppler ambiguities, uses two interferometric phase maps having different scale factors and therefore different ambiguity intervals in height. Multiple maps with different scale factors have been obtained using at least two different methods. In systems having a waveform bandwidth in excess of that required for the desired range resolution, the fast-time bandwidth can be split in half, and IFSAR processing completed separately for each half of the data. The effective center wavelength λ will be different for the two data sets, giving two different scale factors in Eq. (8.80). Another technique applicable to radar systems having three or more receive apertures forms one IFSAR height map using the first and second apertures, and another using the first and third apertures. The baselines for these two pairs are different, again giving rise to two different scaling factors in Eq. (8.80).

The last step in IFSAR processing is orthorectification, which corrects the displacement of image pixels due to layover. For each pixel in the SAR image f[l, m], the corresponding height pixel $\Delta h[l, m]$ is used to estimate the layover – Δh tan ψ present in that pixel. The image is then resampled in the downrange dimension to shift each pixel by $+\Delta h$ tan ψ to its correct downrange position.¹¹

Figure 8.42 is an IFSAR image of the football stadium and surrounding area at the University of Michigan, Ann Arbor. The grayscale is related to height, with lighter shades representing taller features. The image clearly shows that the trees above the stadium in the image are taller than those to the left of the stadium, and that the stadium playing surface is actually below the level of the surrounding terrain.



FIGURE 8.42 Interferometric SAR height image of the football stadium area at the University of Michigan, Ann Arbor. (Image courtesy of General Dynamics Advanced Information Systems.)

An important metric in quantifying IFSAR quality and describing error sources is the *coherence* $\gamma[l, m]$ of the two images f[l, m] and g[l, m]. Coherence is a pixel-by-pixel correlation coefficient. A typical definition is

$$\gamma = \frac{\mathbf{E}\{f \cdot g^*\}}{\sqrt{\mathbf{E}\{|f|^2\}\mathbf{E}\{|g|^2\}}}$$
(8.86)

In the absence of errors, $|\gamma| = 1$ and $\arg\{\gamma\} = \phi_{fg}$ at each pixel. The actual value

of coherence is a random variable due to additive noise, scene temporal decorrelation, baseline decorrelation, and other effects. Generally, coherence can be factored into a product of terms reflecting different error sources. The higher the total coherence, the better the IFSAR height estimate quality. Coherence magnitude values of 0.6 or larger are usually desirable.

The coherence due to thermal noise is given by

$$\gamma_{\text{noise}} = \frac{1}{1 + \chi^{-1}}$$
(8.87)

This equation can be used for both thermal and quantization noise. An SNR of χ = 10 dB gives a noise coherence of $\chi_{noise} = 0.91$. A second major contributor to degraded coherence is *baseline decorrelation*. This refers to the effect of the slightly different viewing angles of the two sensors on the RCS of a given clutter cell. It is exactly the same RCS decorrelation effect discussed in connection with many-scatterer targets in <u>Chap. 2</u>. The angular decorrelation interval of Eq. (2.62) can be adapted to determine a *critical perpendicular baseline* $B_{c\perp}$ that would result in complete decorrelation of the two images (see Richards, 2013). "Perpendicular" refers to the baseline component perpendicular to the line of sight. The critical perpendicular baseline and the coherence factor that results from an actual perpendicular baseline B_{\perp} are

$$\gamma_{\text{baseline}} = 1 - \frac{B_{\perp}}{B_{c\perp}}, \qquad B_{c\perp} = \frac{\lambda R}{\Delta R \cdot \tan \psi}$$
(8.88)

It is desirable to design the system so that $B_{\perp} \ll B_{c\perp}$.

Another decorrelation source is *temporal decorrelation* in two-pass systems, which is primarily an issue for spaceborne IFSARs. Stable terrain regions can have decorrelation intervals as observed from satellites of several days, but forested areas may have coherences on the order of 0.2 due to leaf motion, while water surfaces decorrelate almost completely in milliseconds. Atmospheric conditions can also affect IPD coherence. Changes in water vapor content and distribution can alter propagation delays in the atmosphere between passes in two-pass systems. Delay changes map directly into phase changes. Weather such as rain both decreases the echo strength and SNR and increases the weather clutter interference, affecting both one- and two-pass systems.

The effects of low coherence can be combatted by averaging multiple looks. Many of the same averaging techniques used to combat speckle can be employed (see <u>Section 8.7.3</u>), such as subbanding the data and averaging multiple lower-resolution IFSAR images, or spatial averaging of the IPD as described earlier.

The relative phase of two SAR images can be used in other ways (Richards, 2013). IFSAR presumes that there is no change in the imaged scene, so phase differences are due only to height variations viewed from slightly different aspect angles. Another application of growing interest is coherent change detection (CCD). CCD compares two images taken from the same trajectory at different times, from a few minutes apart to many hours or days apart. The pixel-by-pixel correlation coefficient will be close to one for pixels with the same complex reflectivity in both maps, and close to zero for pixels whose complex reflectivity has changed between the two collections. An image formed from these correlation coefficients can provide a very sensitive indicator of activity in an observed area. Finally, terrain motion mapping also computes correlation coefficients between two images. However, it is assumed that the reflectivity is unchanged between the two collections, so any complex image changes are due to changes in the height of the scene. The time lag between collections may be days to years. Terrain motion mapping has been used to study such phenomena as glacier movement, land subsidence, volcanic activity, and seismic activity.

8.7 Other Considerations

8.7.1 Motion Compensation and Autofocus

By assuming the radar platform's position in the *x* coordinate is u = vt, it has been implicitly assumed that the radar platform is flying a straight, level, and constant velocity path. In practice, this is never quite true. For airborne systems especially, atmospheric turbulence, minor maneuvers and course corrections, crosswinds, vibration, antenna gimbal transient motions, and similar effects all cause deviations of the SAR antenna phase center from this ideal. However, this assumption is built into the design of the image formation processor, which is based on Eq. (8.27) or one of its various approximations. To the extent that the actual R(u) does not follow the model of R(u) used in the processor design, the actual SAR data will not match the expected PSR. These differences, caused by deviations from the ideal flight path, manifest themselves as phase errors in the processing.

Specifically, suppose the transmitted waveform is of the form $A\exp(j2\pi Ft)$ and the range to some scatterer is *R* on a particular pulse. Then the received signal is

$$y(t) = A' \exp\left[j2\pi F\left(t - \frac{2R}{c}\right)\right]$$
(8.89)

where A' absorbs all of the range equation factors. Now suppose that the phase

center of the radar antenna is, for whatever reason, displaced from the intended path by a distance ε along the line of sight to the scatterer; the received signal will instead be

$$\tilde{y}(t) = \tilde{A} \exp\left[j2\pi F\left(t - \frac{2(R-\varepsilon)}{c}\right)\right] = \frac{\tilde{A}}{A} \exp\left[-j4\pi\frac{\varepsilon}{\lambda}\right] y(t)$$
(8.90)

Because ε is small, the amplitude ratio in Eq. (8.90) will be very nearly unity, so the primary effect of motion errors is a phase rotation to the data. All of the fast-time samples for a given pulse are rotated by the same phase factor.

The job of motion compensation is to estimate ε and correct the data by a simple counter phase rotation:

$$\hat{y}(t) = \exp\left[+j4\pi\frac{\varepsilon}{\lambda}\right] \tilde{y}(t) \approx y(t)$$
(8.91)

The difficult part is estimating ε to the required accuracy. A displacement of only $\lambda/4$ meters (0.3 inch at X band) corresponds to a two-way range change of $\lambda/2$ meters and therefore a 180° phase reversal. Thus, path deviations must be estimated to a small fraction of a wavelength to minimize phase error effects. Two factors ease this challenge somewhat. First, a constant displacement of the entire flight path has no effect; the same factor is added to the phase of all data samples, contributing a complex constant with no effect on focusing quality. Only variations in ε are significant. Second, any one pixel receives contributions from data only over the aperture time, on the order of a few tenths of a second to a few seconds in most SAR systems. Path drifts over longer time periods do not affect image focus.¹²

Phase errors are frequently categorized as either low- or high-frequency, based on the variation of the phase error in the slow-time dimension (Lacomme et al., 2001). Low frequency errors are those that repeat on a time scale greater than the aperture time so that any periodicity of low frequency errors is not apparent during formation of an image. High frequency errors are subdivided into deterministic and noise-like errors. In the former, periodic structure is evident during the aperture time. Low frequency errors produce net phase tapers across the aperture, affecting the resolution, gain, and cross-range accuracy of the PSR. For example, a linear phase taper is equivalent to an uncompensated Doppler shift and can produce a cross-range displacement of scatterers according to Eq. (8.38). Examples of the resolution and gain loss effects of quadratic phase errors were shown in Figs. 8.23 and 8.25*a*. High-frequency errors affect primarily the sidelobes of the PSR. Table 8.2 describes the major effect of the most important phase errors.
Phase Error Class	Dominant Effect on PSR		
Low Frequency			
Linear	Cross-range displacement		
Quadratic	Mainlobe broadening, amplitude loss		
Cubic	Mainlobe asymmetry and cross-range displacement		
High Frequency			
Deterministic periodic	Discrete "paired echo" high sidelobes		
Random	Increased sidelobe level		

 TABLE 8.2
 Effects of Various Phase Errors on the SAR Point Spread Response

Phase errors are corrected by estimating the displacement $\varepsilon(u)$ at each aperture position *u* using a position measurement system, and in finer-resolution systems by also using *auto-focus* algorithms. Position estimates are developed using a motion compensation system with some or all of the elements in Fig. <u>8.43</u>. Data from the aircraft or spacecraft's *inertial navigation system* (INS) provide the first level estimate of deviations from a straight line, constantvelocity flight path. The INS tracks the platform centroid, not the antenna phase center. Higher precision systems mount an additional strapdown inertial measurement unit (IMU) onto the antenna structure (Kennedy, 1988a). The IMU accounts more accurately for motions of the antenna relative to the airframe; these relative motions include not only intentional scanning but also vibration and airframe flexure. Inertial antenna mounts can also be used to stabilize the antenna as much as possible. Additional independent kinematic estimates such as radar-derived Doppler velocity estimates and global positioning system (GPS) data may also be used. In this system, the INS provides an initial attitude reference to the IMU. IMU data are sent to the *radar data processor* (RDP), which corrects for the lever arm displacement between the INS and IMU. The difference in INS and corrected IMU position estimates is input to the Kalman transfer alignment filter, whose output is fed back to the IMU to update its state. The Kalman filter typically has on the order of 20 states. The updated IMU state is fed to the motion compensation computer, which uses these data to compute the position error ε and the associated phase correction. The motion compensation computer typically also computes antenna steering commands to stabilize the line of sight.



FIGURE 8.43 Generic motion compensation instrumentation package.

A motion compensation error budget can be developed for the various INS/IMU errors. Table 8.3 is an example of such an analysis (Kennedy, 1988b). The parameter values used were an aperture time of 2 seconds, velocity of 152 m/s, standoff range of 37 km, gravitational acceleration g = 9.8 m/s², and a line-of-sight acceleration A of 3 m/s². These parameters imply a cross-range resolution of 1.8 m at 10 GHz. In this example, the total error is 0.67 radian, well under the $\pi/2$ standard suggested earlier. Since the phase error contributions all scale as T_a^2/λ , the error margins get tighter as radar frequency increases and especially as resolution is made finer. Because of its dependence on range, the velocity error tends to dominate at shorter ranges; at longer ranges the other error effects become more significant. Another analysis of error contributions, in terms of longitudinal and latitudinal motions, is given in Lacomme et al. (2001).

Error Source	1σ Value				1
	INS	IMU	Root Sum Square (RSS) Total	Quadratic Phase Error Formula, rads	Quadratic Phase Error, rads
Cross-range velocity error, $v_e (m/s)$	0.24	0.06	0.247	$\frac{\pi v v_e T_e^2}{\lambda R_0}$	0.43
LOS accelerometer bias, $A_b (\mu g)$	70	80	106.3	$\frac{\pi A_b T_a^2}{2\lambda}$	0.22
Platform tilt, $lpha$ (mrad)	0.2	0.1	0.22	$\frac{\pi \alpha ag T_a^2}{2\lambda}$	0.45
Boresight accelerometer scale factor error, s (PPM)	150	150	212	$\frac{\pi s A T_a^2}{2\lambda}$	0.13
Total RSS quadratic phase	0.67				

Source: Kennedy (1988b).

TABLE 8.3 Motion Compensation Error Budget

In addition to estimates of the path deviation ε , motion compensation systems often provide two additional types of control data. The first is fine pulse-to-pulse PRI adjustments to provide uniform sample spacing along the velocity vector, i.e., in the x or u dimension. This technique eliminates the need to interpolate data in the slow-time dimension to obtain uniform sample spacing in finer-resolution systems. The second, applicable primarily to squinted SAR and spotlight SARs, computes the range walk relative to a reference point on the ground and varies the time delay from the time of pulse transmission to the time when the A/D converter at the receiver output is triggered to begin fast-time sampling. The delay is chosen such that the echo from some reference point, generally in the middle of the imaged scene, always occurs in the same range bin. This technique, called *motion compensation to a point*, reduces or eliminates the computation required for range walk correction. Both techniques require more complicated receiver control.

8.7.2 Autofocus

In fine-resolution SAR systems, the motion compensation systems discussed previously often may not provide sufficient phase correction to achieve good image quality. Autofocus algorithms attempt to estimate residual phase errors present in the data after motion compensation and compensate them to improve image quality. A number of techniques have been suggested; see Carrara et al. (1995) for an overview. Two common and representative methods are the map drift method and the phase gradient method. The most basic form of the map drift algorithm is designed specifically to correct quadratic phase errors. The algorithm divides the SAR data into two portions corresponding to the first and

second halves of the synthetic aperture and forms a lower-resolution image from each. If quadratic phase error is present, peaks in the images will be shifted away from their correct locations but in opposite directions. Cross-correlating the two images produces a peak whose offset from the origin is directly proportional to the amount of residual quadratic phase error. The full data set is then phase-corrected with the conjugate of the estimated error and the image reformed from the modified data. The algorithm is able to correct quadratic phase errors up to a few tens of radians. Extensions to the map drift algorithm which divide the data into more than two subapertures are capable of estimating higher-order polynomial phase error terms.

A very basic analysis of a simple map drift algorithm can be outlined as follows. Suppose a slow-time phase history sequence y[m] is contaminated with a quadratic phase error sequence

$$y'[m] = y[m] \exp\{j4\alpha [m - (M - 1)/2]^2 / (M - 1)^2\}, \qquad 0 \le m \le M - 1$$
(8.92)

The parameter α is the maximum value of the quadratic phase error, which occurs at the two ends of y'[m]. Assume for convenience that *M* is even, and define the two half-sequences:

$$y'_{1}[m] = y'[m], \qquad 0 \le m \le M/2 - 1$$

$$y'_{2}[m] = y'[m + M/2], \qquad 0 \le m \le M/2 - 1$$
(8.93)

The phase error in $y'_1[m]$ can be decomposed into a quadratic component that is symmetric about the midpoint of $y'_1[m]$ and a linear component, which is just the straight line connecting the first and last sample of the phase curve in $y'_1[m]$. It is straightforward to show that this linear phase component is

$$\phi_{\text{llin}}[m] = \alpha - \left(\frac{2}{M-2}\right) \left(1 - \frac{1}{(M-1)^2}\right) \alpha m$$
$$\equiv \alpha - \alpha' m \tag{8.94}$$

The linear phase term of $y'_2[m]$ has an identical quadratic component, but a linear component of the opposite sign (see Prob. 27). Because each sequence is half of the original sequence y'[m], their DFTs will be similar in shape. The linear phase terms, however, will shift the two DTFTs $y'_1[m]$ and $y'_2[m]$ by $-\alpha'K/2\pi$ and $+\alpha'K/2\pi$ samples, respectively, where *K* is the DFT size. A cross-correlation of $Y'_1[k]$ and $Y'_2[k]$ will therefore produce a peak at lag $k_0 = \alpha'K/\pi$ radians. The quadratic phase error coefficient α can then be estimated from the spectral correlation peak as

$$\alpha' = \frac{\pi k_0}{K}$$
$$\hat{\alpha} = \left(\frac{M-2}{2}\right) \left(1 - \frac{1}{(M-1)^2}\right)^{-1} \alpha'$$
(8.95)

In practice, the estimate is not exact due to nonquadratic terms in the phase error and noise in the data. A corrected data sequence is then formed as

$$\hat{y}[m] = y'[m] \exp\{-j4\hat{\alpha}[m - (M - 1)/2)^2/(M - 1)^2\}, \qquad 0 \le m \le M - 1$$
(8.96)

The map drift algorithm is applied iteratively, typically requiring on the order of three to six iterations. The iteration is stopped when the peak estimated phase error is less than $\pi/2$ (or some other appropriate quality threshold).

Figure 8.44 illustrates the map drift algorithm on a real scene. The original image in part *a* of the figure was transformed to a simulated range-compressed phase history domain by assigning a random phase to each image pixel and then performing a Fourier transform in the cross-range dimension. A simulated phase error consisting of a quadratic term with a maximum of 5π radians across the aperture was applied in the cross-range dimension and the image formed again, giving the severely blurred image in Fig. 8.44*b*. Part *c* of the figure is the result after applying map drift autofocus to the blurred image; it is visually indistinguishable from the original. Finally, Fig. 8.44*d* is the difference image between the original and the autofocused image. While some correlation between the difference image and the original is visible on close inspection, the nearly uniform difference image confirms the quality of the phase error correction.







Cross-range (b)



FIGURE 8.44 Illustration of map drift autofocus: (*a*) original image, (*b*) image formed from data with simulated noisy quadratic phase errors, (*c*) image after map drift autofocus, (*d*) error image. (Original image courtesy of Sandia National Laboratories.)

The phase gradient algorithm (PGA) (Carrara et al., 1995; Wahl et al.,

1994) does not assume a polynomial form for the phase error and appears to offer excellent, robust performance. PGA uses the (blurred) image of the strongest scatterers in the SAR scene to estimate the actual cross-range phase function, and then multiplies the data by a compensating phase function. The algorithm estimates the first derivative of the cross range phase error from a SAR image that is converted to (K_u , R) units by performing an FFT in cross range. The phase error derivative is then integrated to get the actual phase error function, and a correction applied to the data. As with map drift, the phase gradient algorithm is usually applied iteratively.

A simplified analysis of the PGA algorithm gives some insight into its operation. PGA assumes that each cross-range row in the complex blurred image is dominated by the image of a single point scatterer. Consider an arbitrary range bin and assume that the point scatterer in that range bin is located at $k = k_0$. The model of the complex image domain data in this range bin is then

$$f[k] = A\delta[k - k_0] + w[k]$$
(8.97)

where w[k] is a random process representing other clutter scatterers. The dominant scatterer amplitude *A* and location m_0 will vary in different range bins, but for notational simplicity the dependence on the range bin index *l* is not shown. If the cross-range compression algorithm is assumed to be approximately the Fourier transform of the range-compressed phase history data implemented using a *K*-point DFT, the corresponding phase history data y[m] will be, ignoring the clutter, the inverse DFT of the cross-range image:

$$y[m] = \text{IDFT}\{f[k]\} = \frac{A}{K} \exp\left[j\frac{2\pi}{K}k_0m\right]$$
(8.98)

Equation (8.98) represents the phase history in the absence of phase errors. When phase errors $\phi[m]$ are present the data y[m] are modulated by the phase error modulation function $\exp\{j\phi[m]\}$:

$$y'[m] = \exp\{j\phi[m]\}y[m]$$
$$= \frac{A}{K} \exp\left\{j\frac{2\pi}{K}k_0m\right\} \exp\{j\phi[m]\}$$
(8.99)

Denote the DFT of the phase error modulation function $\exp\{j\phi[m]\}\$ as e[k]. The DFT of the phase history data including the phase errors is the actual observed complex cross-range image slice. Applying the modulation property of the DFT gives

$$f'[k] = DFT\{y'[m]\}\$$

= $A \cdot e[k - k_0]$ (8.100)

Thus, the point target image is degraded into a replica of the DFT of the phase error function, centered at the point target location. That is, e[k] is the blurring function in the image domain due to the phase errors. Isolated point targets can be used to estimate e[k] and therefore the phase error function $\phi[m]$.

The algorithm begins by finding the peak amplitude of the blurred image data |f[k]|; suppose this occurs at $k = k_p$. The cross-range slice is then circularly shifted left by k_p samples, giving the new sequence

$$f'_{p}[k] = A \cdot E[k - k_{0} + k_{p}]$$
(8.101)

with corresponding IDFT

$$y'_{p}[m] = \frac{A}{K} \exp\{j\phi[m]\} \exp\{j2\pi(k_{0} - k_{p})m/K\}$$
(8.102)

The next step is to estimate the gradient of the phase of $y'_{r}[m]$. A simple maximum likelihood estimator is the first difference (Jakowatz and Wahl, 1993):

$$\Delta \phi[m] = \arg \left\{ y'_p^*[m-1]y'_p[m] \right\}$$

$$\approx \phi[m] - \phi[m-1] \qquad (8.103)$$

The phase error itself is estimated by integrating this gradient:

$$\hat{\phi}[m] = \begin{cases} 0, & m = 0\\ \sum_{q=1}^{m} \Delta \phi[q], & \text{otherwise} \end{cases}$$
(8.104)

The original slow-time phase history is then corrected by compensating it with the estimated phase:

$$\hat{y}[m] = y'[m] \exp\{-j\hat{\phi}[m]\}$$
(8.105)

Finally, the deblurred image slice is obtained by transforming back to the image domain:

$$f[k] = DFT{\hat{y}[m]}$$
(8.106)

If $\phi[m]$ is a good estimate of $\phi[m]$, the result will be

$$\hat{f}[k] \approx A\delta[k - k_p]$$
(8.107)

Note that if the blur function e[k] does not have its peak at the origin, $k_p \neq k_0$ and the corrected image slice will still be shifted by $(k_p - k_0)$ samples from the correct position. This does not hurt image quality but does impact geolocation accuracy.

Since the image has many rows, each degraded by the same actual phase error function, an independent phase error estimate can be obtained by processing the data for each range bin separately. The signal-to-noise ratio of the phase estimate can then be improved by averaging over the range bins. Reintroducing the range bin index l, the integrated phase gradient estimate is

$$\Delta \phi[l,m] = \arg\{y'^*[l,m-1]y'[l,m]\} \\ \Delta \phi[m] = \sum_{l=1}^{L-1} \Delta \phi[l,m]$$
(8.108)

Figure 8.45 illustrates this process on a simple synthetic image. The original image in part a of the figure is simply a collection of randomly located point scatterers of varying amplitude and phase in complex white Gaussian noise. A phase error consisting of quadratic, sinusoidal, and noise terms (see Fig. 8.46) was applied to the cross-range DFT of these data, and the image-reformed via a cross range IDFT. The result, seen in part b, is severely blurred in cross-range. Application of the PGA algorithm produces the corrected image of part c. The smearing has been completely removed.





FIGURE 8.45 Illustration of phase gradient algorithm autofocus: (*a*) synthetic image, (*b*) image blurred by non-polynomial phase error, (*c*) image after 1 iteration of PGA autofocus, (*d*) error image showing cross-range displacement.



FIGURE 8.46 Actual and estimated phase error for example of Fig. 8.45.

Close inspection, however, shows that all of the scatterers have migrated three pixels to the left. This occurs because for the particular phase error function $\phi[m]$ used, the resulting blur function e[k] has its peak at k = 3 rather than k = 0. Consequently, in the blurred image the peaks of each blurred scatterer are shifted three pixels right; that is, $k_p = k_0 + 3$. This adds a linear phase term in k corresponding to a +3 pixel shift to the phase function $\hat{\phi}[m]$ estimated by the PGA algorithm. When the phase of the original data is compensated, a -3 pixel shift results. This effect is clearly seen in the error image of Fig. 8.45d, which shows paired positive and negative differences three pixels apart for each scatterer.

The actual phase error function $\phi[m]$ applied in this example is shown in Fig. 8.46. The estimated phase error, corrected for the three pixel offset, and the difference between the two is also shown. Only one iteration of PGA was needed because this simple simulation matches the algorithm assumptions very closely. The residual phase error is well within the $\pm \pi/2$ bounds denoted by the horizontal lines.

A number of details in any practical implementation have not been addressed here. As mentioned earlier, the algorithm is typically applied iteratively. It is often applied to only a subset of the most energetic range bins, similar to the map drift algorithm. Furthermore, a window is used to select only a symmetric portion of the cross-range image slice around each peak magnitude pixel; this reduces the effects of noise and other bright pixels in the line. The window size is usually selected based on an estimate of the width of the blur function e[m] and is updated with each iteration. These and other details are addressed in Jakowatz et al. (1996).

8.7.3 Speckle Reduction

Like any coherent imaging system, SAR produces images contaminated by "speckle," an aptly-named multiplicative noise. Speckle is the natural result of the coherent combination of echoes from many different scatterers to form an image pixel. If the amplitude distribution of the real and imaginary parts (I and Q channels) of the received signal is Gaussian and the phase distribution is uniform, conditions assured by the law of large numbers when many scattering centers are involved, then the pixel amplitude will be Rayleigh distributed as shown in <u>Chap. 2</u>. Thus, pixels representing areas with the same average RCS (mean echo amplitude) can give rise to different pixel amplitudes. These variations are not due to thermal, quantization, or other noise sources, but are nonetheless considered "noise" because of their appearance and effect on image quality.

Speckle is reduced through various forms of filtering and averaging schemes (Oliver and Quegan, 2004). One of the most effective is noncoherently integrating multiple uncorrelated images, or *looks*, of the same scene. This process reduces the pixel variance, reducing the amplitude variations among pixels representing the same RCS. Uncorrelated looks can be obtained using multiple flights past the scene, called *passes*. Obtaining well-aligned passes is difficult for airborne systems due to atmospheric turbulence and limited navigation precision. For spaceborne systems, orbits are well known and well behaved, but considerable time can pass between repeat passes so that data acquisition time can be lengthy. In a single pass, transmitter frequency or multiple polarization channels can be used to obtain uncorrelated looks. Another method takes advantage of the fact that in many strip-map modes the maximum aperture time available exceeds that required to meet resolution goals. In this case the slow-time data can be divided into multiple subapertures, each long enough to form an image of the proper resolution (Cumming and Wong, 2005). Images are calculated for each subaperture and combined. Typically, 4 to 10 looks might be used to obtain effective speckle reduction. Figure 8.47a and \underline{b} demonstrates the image enhancement obtained by integration of 10 simulated looks.



FIGURE 8.47 Speckle reduction: (*a*) simulated single-look speckled image derived from Fig. 8.1*a*, (*b*) full resolution image obtained by integration of nine looks, (*c*) reduced-resolution image obtained with 3×3 spatial filtering, (*d*) image obtained with 3×3 median filter. (Original image courtesy of Sandia National Laboratories.)

If multiple looks are not available, another method averages adjacent pixels of a fine-resolution image, typically using a 3×3 or 5×5 window, to form one lower resolution but reduced-speckle pixel as illustrated in Fig. 8.47*c*.

Figure 8.47*d* shows the result of a 3×3 median filter applied to a single look image. This approach sacrifices less resolution than spatial averaging. Other techniques include a variety of adaptive filters and statistical methods (Oliver and Quegan, 2004; Richards, 2009).

References

- Ausherman, D. A., "Digital vs. Optical Techniques in Synthetic Aperture Radar (SAR) Data Processing," *Optical Engineering*, vol. 19, no. 2, pp. 157– 167, Mar./Apr. 1980.
- Ausherman, D. A., et al., "Developments in Radar Imaging," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-20, no. 4, pp. 363–400, Jul. 1984.
- Bamler, R., "A Comparison of Range-Doppler and Wavenumber Domain SAR Focusing Algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 4, pp. 706–713, Jul. 1992.
- Brown, W. M., and L. J. Porcello, "An introduction to Synthetic-Aperture Radar," *IEEE Spectrum*, vol. 6, no. 9, pp. 52–62, Sep. 1969.
- Cafforio, C., C. Prati, and F. Rocca, "SAR Data Focusing Using Seismic Migration Techniques," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-27, no. 2, pp. 194–207, Mar. 1991.
- Carrara, W. G., R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar*. Artech House, Norwood, MA, 1995.
- Cook, D. A., "Spotlight Synthetic Aperture Radar," Chap. 6 in W. L. Melvin and J. A. Scheer, (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, 2013.
- Cumming, I. G., and F. H. Wong, *Digital Processing of Synthetic Aperture Radar Data*. Artech House, Norwood, MA, 2005.
- Curlander, J. C., and R. N. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*. Wiley, New York, 1991.
- Cutrona, L. J. et al., "On the Application of Coherent Optical Processing Techniques to Synthetic-Aperture Radar," *Proceedings of the IEEE*, vol. 54, no. 8, pp. 1026–1032, Aug. 1966.
- Desai, M. D., and W. K. Jenkins, "Convolution Backprojection Image Reconstruction for Spotlight Mode Synthetic Aperture Radar," *IEEE Transactions on Image Processing*, vol. 1, no. 4, pp. 505–517, Oct. 1992.
- Dudgeon, D. E., and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1984.
- Elachi, C., Spaceborne Radar Remote Sensing: Applications and Techniques. IEEE Press, New York, 1988.
- Franceschetti, G., and R. Lanari, *Synthetic Aperture Radar Processing*. CRC Press, New York, 1999.

- Freeman, A., et al., "The Myth of the Minimum SAR Antenna Area Constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 1, pp. 320–324, Jan. 2000.
- Frey, O., C. Magnard, and M. Rüegg, "Focusing of Airborne Synthetic Aperture Radar Data from Highly Nonlinear Flight Tracks," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-47, no. 6, pp. 1844–1858, Jun. 2009.
- Ghiglia, D. C, and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*. Wiley, New York, 1998.
- Ghiglia, D. C., and L. A. Romero, "Robust Two-dimensional Weighted and Unweighted Phase Unwrapping That Uses Fast Transforms and Iterative Methods," *Journal of Optical Society of America*, vol. 11, no. 1, pp. 107– 117, Jan. 1994.
- Gough, P. T., and D. W. Hawkins, "Unified Framework for Modern Synthetic Aperture Imaging Algorithms," *International Journal of Imaging Systems* & *Technology*, vol. 8, pp. 343–358, 1997.
- Harger, R. O., Synthetic Aperture Radar Systems: Theory and Design. Academic Press, New York, 1970.
- Henderson, F. M., and A. J. Lewis (eds.), Manual of Remote Sensing, 3d ed., Vol. 2: Principles and Applications of Imaging Radar. Wiley, New York 1998.
- Jakowatz, C. V., et al., *Spotlight Mode Synthetic Aperture Radar*. Kluwer, Boston, 1996.
- Jakowatz, C. V., Jr., and D. E. Wahl, "An Eigenvector Method for Maximum Likelihood Estimation of Phase Errors in SAR Imagery," *Journal of Optical Society of America*, vol. 10, no. 12, pp. 2539–2546, Dec. 1993.
- Jakowatz, C. V., Jr., D. E. Wahl, and D. A. Yocky, "Beamforming as a Foundation for Spotlight-Mode SAR Image Formation by Backprojection," *Proceedings SPIE vol. 6970, Algorithms for Synthetic Aperture Radar Imagery XV*, Mar. 2008.
- Kennedy, T. A., "The Design of SAR Motion Compensation Systems Incorporating Strapdown Inertial Measurement Units," *Proceedings of the IEEE 1988 National Radar Conference*, pp. 74–78, Apr. 1988a.
- Kennedy, T. A., "Strapdown Inertial Measurement Units for Motion Compensation for Synthetic Aperture Radars," *IEEE AESS Magazine*, vol. 3, no. 10, pp. 32–35, Oct. 1988b.
- Kirk, J. C., Jr., "Discussion of Digital Processing in Synthetic Aperture Radar," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-11, no. 3, pp. 326–337, May 1975.
- Lacomme, P., J.-P. Hardange, J.-C. Marchais, and E. Normant, *Air and Spaceborne Radar Systems*. William Andrew Publishing, Norwich, NY, 2001.
- Munson, D. C., Jr., J. D. O'Brien, and W. K. Jenkins, "A Tomographic

Formulation of Spotlight-Mode Synthetic Aperture Radar," *Proceedings of the IEEE*, vol. 71, no. 8, pp. 917–925, Aug. 1983.

- Munson, D. C., Jr., et al., "A Comparison of Algorithms for Polar-to-Cartesian Interpolation in Spotlight Mode SAR," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 10, pp. 1364–1367, 1985.
- Munson, D. C., Jr., and R. L. Visentin, "A Signal Processing View of Strip-Mapping Synthetic Aperture Radar," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 12, pp. 2131–2147, Dec. 1989.
- Oliver, C., and S., Quegan, *Understanding Synthetic Aperture Radar Images*. SciTech Publishing, Raleigh, NC, 2004.
- Raney, R. K., "A New and Fundamental Fourier Transform Pair," Proceedings of IEEE 12th International Geoscience & Remote Sensing Symposium (IGARSS '92), pp. 106–107, 26–29 May 1992.
- Richards, J. A., *Remote Sensing with Imaging Radar*. Springer, Heidelberg 2009.
- Richards, M. A., "Nonlinear Effects in Fourier Transform Processing," Chap. 6 in J. A. Scheer and J. L. Kurtz (eds.), *Coherent Radar Performance Estimation*. Artech House, Norwood, MA, 1993.
- Richards, M. A., "Interferometric SAR and Coherent Exploitation," Chap. 8 in W. L. Melvin and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, 2013.
- Schleher, D. C., *MTI and Pulsed Doppler Radar*. Artech House, Norwood, MA, 1991.
- Sherwin, C. W., J. P. Ruina, and R. D. Rawcliffe, "Some Early Developments in Synthetic Aperture Radar Systems," *IRE Transactions on Military Electronics*, vol. MIL-6, no. 2, pp. 111–115, Apr. 1962.
- Showman, G. A., "An Overview of Radar Imaging," Chap. 21 in M. A. Richards, J. A. Scheer, and W. A. Holm (eds.), *Principles of Modern Radar: Basic Principles*. SciTech Publishing, Raleigh, NC, 2010.
- Showman, G. A., "Stripmap SAR," Chap. 7 in W. L. Melvin and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Edison, NJ, 2013.
- Stimson, G. W., *Introduction to Airborne Radar*, 2d ed. SciTech Publishing, Mendham, NJ, 1998.
- Soumekh, M., Synthetic Aperture Radar Signal Processing With Matlab Algorithms. Wiley, New York, 1999.
- Sullivan, R. J., *Microwave Radar: Imaging and Advanced Concepts*. Artech House, Norwood, MA, 2000.
- Wahl, D. E., P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, Jr., "Phase Gradient Autofocus—A Robust Tool for High Resolution SAR Phase Correction," *IEEE Transactions on Aerospace & Electronics Systems*,

vol. AES-30, no. 5, pp. 827-835, Jul. 1994.

- Wiley, C. A., "Pulsed Doppler Radar Methods and Apparatus," U. S. patent no. 3,196, 436, 1965 (originally filed 1954).
- Wiley, C. A., "Synthetic Aperture Radars—A Paradigm for Technology Evolution," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-21, pp. 440–443, 1985.
- Wu, C., K. Y. Liu, and M. Jin, "Modeling and a Correlation Algorithm for Spaceborne SAR Signals," *IEEE Transactions Aerospace & Electronics Systems*, vol. AES-18, no. 5, pp. 563–574, Sep. 1982.

Problems

- 1. Compute the sidelooking SAR synthetic aperture size D_{SAR} and aperture time T_a for cross-range resolutions of 50, 5, and 0.5 m for a 10-GHz airborne radar at a range of 10 km and a 5-GHz LEO satellite at a range of 770 km. Assume the aircraft velocity is 100 m/s and the satellite velocity is 7500 m/s.
- 2. Compute the integration angle γ in degrees needed to obtain cross-range resolutions of 100, 10, and 1 m at RFs of 1 and 10 GHz.
- 3. Compute the maximum swath length L_s for an elevation beamwidths of 10° at a grazing angle of $\delta = 30^\circ$ and ranges of 10 and 50 km.
- 4. Compute the elevation beamwidth in degrees required to obtain a swath length of 100 km from an altitude of 770 km. Assume a grazing angle of $\delta = 30^{\circ}$ and ignore earth curvature.
- 5. What is the area coverage rate of the radar in problem 4 in km²/h? Assume v = 7500 m/s.
- 6. For a spaceborne SAR having v = 7500 m/s and a grazing angle of $\delta = 30^{\circ}$, what is the finest cross-range resolution possible if the swath length required is 100 km? What is the longest swath length possible if the cross-range resolution is required to be 1 m? Assume sidelooking stripmap operation.
- 7. What is the upper bound on the number of images per hour obtainable by a spotlight SAR with a platform velocity of 50 m/s, an RF of 35 GHz, a standoff range of 10 km, and a cross-range resolution of 0.25 m? Assume that 10 percent of the time is used for re-steering the antenna between spots or other "overhead" operations.
- 8. The series expansion of the square root leading to Eq. (8.28) (and eventually to the DBS and range-Doppler algorithms) requires $|u x_p| \ll R_p$. Interpret "much less than" to mean $|u x_p| < R_p/100$ (a tight standard). Determine whether a radar on the space shuttle (250 km altitude) having a

resolution of $\Delta CR = 10$ m at an RF of 5 GHz meets this requirement. Repeat for an airborne radar having a resolution of $\Delta CR = 0.25$ m at an RF of 16 GHz and a range of 5 km.

- 9. Develop formulas for range walk and range curvature normalized to range resolution, $\Delta R_w / \Delta R$ and $\Delta R_c / \Delta R$. Assume "square pixels," i.e., $\Delta R = \Delta CR$. If the radar resolution is improved by a factors of 10×, by what factors do the normalized range walk and range curvature increase?
- 10. It is desired to use DBS to form an image 10 km on a side with resolution of 3 m. The RF is 10 GHz and the standoff range is 10 km. Considering the range walk and quadratic phase constraints, can these specifications be achieved with good focus quality? Assume "square pixels."

Problems 11 to 17 consider the same airborne sidelooking stripmap SAR system. For simplicity the problem is treated as two-dimensional rather than three-dimensional so that the slant plane and the ground plane are the same. The goal of the SAR is to image a swath extending from 10 to 20 km slant range from the aircraft; thus the swath length is $L_s = 10$ km. The resolution goal in both range and cross range is 1.0 m. The aircraft is flying straight and level at a constant velocity of 150 m/s. The radar operates at K_u band (16 GHz). The physical antenna beamwidth is $\theta_{az} = 2^\circ$.

- 11. Estimate the azimuth size of the physical antenna aperture, D_{az} , in meters. *Hint:* It is not determined by the desired resolution, which may or may not meet the stripmap constraint that $\Delta CR \ge D_{az}/2$; use other information provided to determine D_{az} . What pulse bandwidth β is required to obtain the desired range resolution? (Do not allow for any weighting for range sidelobe control.)
- 12. What is the maximum Doppler bandwidth of the slow-time data? What is the bound on the PRF due to the Doppler bandwidth? What is the bound on the PRF due to the swath length?
- 13. What is the aperture time T_a required to achieve the desired cross-range resolution ΔCR at the near edge of the swath? At the far edge? What are the corresponding synthetic aperture lengths D_{SAR} ?
- 14. Assume the waveform is an LFM pulse with a time-bandwidth product $\beta \tau = 50$; its bandwidth β was found in Prob. 11. How many range (fast time) samples are required to process the full swath length? How many pulses will be transmitted in an aperture time at the near and far edges? Assume a PRF equal to the Nyquist rate for the Doppler bandwidth of the data.
- 15. What is the range walk for a scatterer within the swath? Does it vary over the swath? Express the answer in both meters and in range resolution cells. What is the maximum range curvature for a scatterer within the swath?

Does it occur for scatterers at the near or far edge of the swath? Express the answer in both meters and in range resolution cells.

- 16. What is the depth of focus for this system? Can a single point spread response (PSR) be used to focus well everywhere in the swath? If not, what is the minimum number of subswaths needed to achieve good focus throughout?
- 17. Consider Doppler beam sharpening processing (probably not a good choice for this fine a resolution, but relatively easy for hand calculations). Suppose that Fourier analysis (an FFT) of the slow time data at range R = 15 km reveals a peak in the spectrum at $F_t = 106.7$ Hz. What is the cross-range position x_t of the scatterer that caused the peak?
- 18. A sidelooking imaging radar system achieves a cross-range resolution of $\Delta CR = 1$ m. The radar is an L-band (1-GHz) system with an azimuth antenna width $D_{az} = 3$ m. It is mounted on an aircraft that flies at 100 m/s. The center of the imaged area is at a range of 50 km. Must the radar be operated in stripmap mode, spotlight mode, or can either be used? Justify the answer.
- 19. Regardless of the answer to the previous problem, assume that spotlight mode is used. What is the line-of-sight rotation angle γ in degrees required to achieve the 1 m cross-range resolution? What are the resulting aperture size D_{SAR} in meters and aperture time T_a in seconds?
- 20. Starting with Eq. (8.32) for range curvature, derive Eq. (8.56). Note that R_P in Eq. (8.32) and R_0 in Eq. (8.56) represent the same quantity.
- 21. Equation (8.62) depends on the RVP term being negligible. Because this is a quadratic phase term in the integrand, it has negligible effect if $\pi(\beta/\tau)$ $(\delta t_i)^2 < \pi/4$ for the maximum absolute value of δt_i . For a scene of depth L_u $(\pm L_u/2 \text{ around the CRP})$, this is L_u/c . Derive an upper bound on L_u in terms of β and τ . *Note*: The result will not be the same as Eq. (8.72), which requires further assumptions, but will have some similarities.
- 22. Consider a spotlight SAR using the polar format algorithm. The system operates at K_u band (16 GHz) from a standoff range $R_0 = 30$ km and is designed for a range and cross-range resolution of 0.2 m. The LFM pulse sweep rate is $\alpha = \pi (7.5 \times 10^{12})$ Hz/s. Compute the limits on scene size for using the PFA algorithm due to isorange curvature and residual video phase.
- 23. Determine the ambiguous terrain height change Δh_{ua} for a K_u-band (16-GHz) IFSAR system at an altitude of Z = 3 km and a nominal grazing angle of $\psi = 20^{\circ}$. The IFSAR baseline is B = 1 m. Δh_{ua} is the change in terrain height that produces a change of 2π in the IPD $\Delta \phi_{fg}$.

- 24. Determine the SNR needed to obtain coherence values $y_{\text{noise}} = 0.5, 0.8$, and 0.95.
- 25. Compute the critical baseline $B_{c\perp}$ for a C-band (5 GHz) LEO radar. The range is R = 770 km, the range resolution is 30 m, and the grazing angle is $\psi = 15^{\circ}$. If the baseline coherence is to be no less than 0.9, what is the maximum spacing of two parallel orbits that can be used for IFSAR processing?
- 26. Table 8.1 states that a linear phase error term in slow-time results in a cross-range displacement of the scatterer image. Consider a sidelooking 10 GHz SAR and a scatterer directly abreast of the SAR at a range of 5 km. Assume the platform velocity is 100 m/s. Suppose that crosswinds cause the aircraft to drift such that there is a velocity component of +5 m/s along the LOS toward the imaged scatterer. The scatterer should be imaged as being at x = 0. Using Doppler beam sharpening as a simple model of the processing, at what x will the scatterer actually be imaged? Include the sign of x and state whether the shift is in the direction of the aircraft motion or its opposite.
- 27. Confirm that the slope of the linear phase component of $y'_2[m]$ of Eq. (8.93) is the negative of the slope of $y'_1[m]$, which can be obtained from Eq. (8.94).

¹ "False color" or "pseudocolor" SAR imagery is often produced by combining multiple SAR images of the same scene collected at different polarizations and/or radar frequencies.

²This image is a grayscale representation of a false color image generated by assigning red to the L band, horizontal polarization transmit/horizontal polarization receive ("HH") data; green to the L band horizontal transmit/vertical receive (HV) data; and blue to the C band HV data.

³Subsequently part of Veridian Corporation, then General Dynamics Corporation, and now Lockheed Martin Corporation.

 $^{4^{\}circ}$ Cross-range" is the direction orthogonal to the radar range direction, and thus to the radar antenna boresight. It differs from azimuth in that azimuth specifies an angular displacement relative to the boresight, while cross range specifies a displacement in an orthogonal Cartesian coordinate. If the radar is sidelooking and the platform is not crabbed, the cross-range direction is parallel to the platform velocity vector. Thus the cross-range dimension is sometimes referred to as the along-track dimension.

⁵Successive synthetic apertures are usually overlapped, with offsets ranging from as little as one pulse position to 50 percent or more of D_{SAR} . The two data subsets in Fig. 8.7 are shown nonoverlapping for clarity.

⁶This is equivalent to viewing the three-dimensional scenario in the *slant plane* formed by the velocity vector and the vector traced out by the motion of the antenna boresight along the ground plane.

 $[\]frac{7}{\tilde{\rho}_i}$ includes range weighting and other range equation factors; see Sec. 2.7.

 $[\]frac{8}{2}$ The results here are generalized to include use of a common transmit aperture and two independent receive apertures in Richards (2013). Both one- and two-transmitter versions of IFSAR are of practical

importance.

 $\frac{9}{2}$ The two apertures could also be displaced vertically, with similar results.

 $\frac{10}{10}$ SAR images are naturally formed in the slant plane, but are usually projected into the ground plane for display.

¹¹If operated in squint mode, the image will also have layover in the cross-range dimension, so pixels must be shifted in both dimensions. See Jakowatz et al. (1996) for details.

 $\frac{12}{2}$ Long term drift may be quite important, however, to image interpretation. Targets or land features detected in a SAR image may need to be precisely geolocated for targeting or for surveying purposes. It is then essential to have accurate information on the absolute position of the radar platform.

CHAPTER 9

Introduction to Beamforming and Space-Time Adaptive Processing

In <u>Chap. 3</u>, the concept of the radar datacube was introduced to describe the data collected in a coherent processing interval. Figure 3.8, repeated in part here as Fig. 9.1, illustrates the datacube y[l, m, n] with its independent axes of fast time, slow time, and antenna phase center. The radar signal processing described up to this point has dealt almost entirely with a fast-time/slow-time matrix for a single antenna phase center; little use has been made of the third datacube dimension other than to discuss the array factor of a phased array antenna pattern in <u>Chap. 1</u> and the accuracy of angle estimation in <u>Chap. 7</u>.



FIGURE 9.1 Radar datacube for a single CPI.

Just as the temporal sampling of the slow-time axis enables analysis and processing of signals in a given range bin based on their temporal Doppler frequency content, the phase center axis enables analysis and processing of signals within a range bin based on their spatial frequency content, which is equivalent to the *angle of arrival* (AOA). In this chapter, a basic introduction to *beamforming* and *space-time adaptive processing* (STAP) is presented. Beamforming refers to the coherent combination of data from multiple phase centers to provide selectivity in the AOA, i.e., to form and steer an antenna beam. STAP combines both spatial and temporal filtering on a moving radar platform to discriminate targets from both clutter and jamming. Excellent introductions to adaptive beamforming and STAP are Aalfs (2013) and Melvin

9.1 Spatial Filtering

9.1.1 Beamforming

Consider a monochromatic plane wave with temporal variation $A\exp(j\Omega t)$ impinging on a uniform linear one-dimensional array as shown in Fig. 9.2. If the wave's AOA relative to the array normal is θ radians, the signal observed at the *n*th array element is



FIGURE 9.2 Wavefront impinging on a uniform linear array.

$$\overline{y}_n(t) = A e^{j[\Omega(t - nd\sin\theta/c) + \phi_0]}$$
(9.1)

where the phase offset ϕ_0 accounts for the absolute phase at the n = 0 element. Now consider the set of samples y[n] formed from the individual array signals, sampled at a common time t_0 :

$$y[n] \equiv \overline{y}_n(t_0) = A e^{j[\Omega(t_0 - nd\sin\theta/c) + \phi_0]}$$

= $\hat{A} e^{-j\Omega nd\sin\theta/c} = \hat{A} e^{-j2\pi nd\sin\theta/\lambda}, \quad n = 0, ..., N-1$
(9.2)

Assembling the *N* element samples into column vector form gives a *snapshot* of the array at a fixed time:

$$\mathbf{y} = \begin{bmatrix} y[0] & y[1] & \cdots & y[N-1] \end{bmatrix}^{T}$$

$$= \hat{A} \begin{bmatrix} 1 & e^{-j2\pi d \sin \theta/\lambda} & \cdots & e^{-j2\pi (N-1)d \sin \theta/\lambda} \end{bmatrix}^{T}$$

$$= \hat{A} \begin{bmatrix} 1 & e^{-jk_{\theta}} & \cdots & e^{-j(N-1)k_{\theta}} \end{bmatrix}^{T}$$

$$\equiv \hat{A} \mathbf{a}_{s}(\theta)$$
(9.3)

where $k_{\theta} \equiv 2\pi d \sin \theta / \lambda$ is the normalized spatial frequency in radians per sample as projected into the plane of the array face and $\mathbf{a}_s(\theta)$ is the *spatial steering vector*. Thus, there is a one-to-one relationship between the AOA of a plane wave and the spatial frequency across the array face. The range of θ is $\pm \pi$, so the range of k_{θ} is $\pm 2\pi d / \lambda$. It is also useful to define the normalized spatial frequency in cycles per sample, $f_{\theta} \equiv k_{\theta}/2\pi$. If the phase centers are spaced by $d = \lambda/2$, common for individual elements, the range of k_{θ} is $\pm \pi$ and of f_{θ} is ± 0.5 . If phase centers correspond to subarrays their spacing will be larger and the range of k_{θ} and f_{θ} will be larger.

Conventional nonadaptive beamforming is implemented as a weighted sum of the element signals, $z = \mathbf{h}^T \mathbf{y}$, where **h** is a vector of complex weights

$$\mathbf{h} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{N-1} \end{bmatrix}^T$$
(9.4)

A special case of interest occurs when \mathbf{h} takes the form

$$\mathbf{h} = \begin{bmatrix} w_0 & w_1 e^{+jk_{\theta}} & \cdots & w_{N-1} e^{+j(N-1)k_{\theta}} \end{bmatrix}^T$$
$$= \begin{bmatrix} w_0 & w_1 & \cdots & w_{N-1} \end{bmatrix}^T \odot \mathbf{a}_s^*(\theta)$$
$$= \mathbf{w} \odot \mathbf{a}_s^*(\theta)$$
(9.5)

The symbol \odot represents the Hadamard (element-by-element) product of two vectors. If **a** and **b** are two *N*-element column vectors, then

$$\mathbf{a} \odot \mathbf{b} \equiv \begin{bmatrix} a_0 b_0 & a_1 b_1 & \cdots & a_{N-1} b_{N-1} \end{bmatrix}^T$$
(9.6)

Equation (9.5) represents **h** as the Hadamard product of two factors: a shape vector **w** that provides for any weighting for sidelobe control and the conjugate of a steering vector $\mathbf{a}_s(\theta)$ that provides maximum coherent integration for signals arriving from angle θ .

Suppose the weights are matched to an AOA of θ_0 ; the array is said to be "steered" to θ_0 . The response of a beamformer steered to θ_0 to an incoming

wavefront at angle θ is

$$z(\theta) = \mathbf{h}^T \mathbf{y} = \hat{A} \sum_{n=0}^{N-1} w_n e^{-j(k_\theta - k_{\theta_0})_n}$$
(9.7)

The scalar output $z(\theta)$ is just the discrete Fourier transform of the weight sequence $\{h_n\}$, shifted to a center spatial frequency of k_{θ_0} and multiplied by \hat{A} . When all of the weight amplitudes $\{w_n\} = 1$, this is a standard asinc pattern in the variable $\sin\theta$,

$$z(\theta) = e^{j(N-1)(\pi d/\lambda)(\sin\theta - \sin\theta_0)} \left\{ \frac{\sin[N(\pi d/\lambda)(\sin\theta - \sin\theta_0)]}{\sin[(\pi d/\lambda)(\sin\theta - \sin\theta_0)]} \right\}$$
(9.8)

More commonly, the weights are chosen to reduce antenna sidelobes at the cost of degraded resolution and gain in the form of a wider mainbeam. Figure 9.3 illustrates the antenna pattern $|z(\theta)|$ for an array steered to $\theta_0 = 30^\circ$ with N = 11 phase centers, both with and without Hamming weighting. In the latter case, w is simply a vector of all ones; in the former case, it is a Hamming window function. The windowed case exhibits the same effects seen in windowing for sidelobe control in range and Doppler: much lower sidelobes at the cost of reduced peak gain and loss of resolution due to a wider mainlobe.



FIGURE 9.3 Antenna pattern with and without Hamming weighting. N = 11

phase centers, steering angle $\theta_0 = 30^\circ$.

The beamformer output of Eq. (9.8) is a function of $\sin\theta$. This is a consequence of the spatial frequency k_{θ} being proportional to $\sin\theta$. A function of $u = \sin\theta$ is often said to be a function in *sine space*, a concept also mentioned briefly in the section on Kalman filtering in Chap. 7.² In this chapter, most figures involving antenna patterns are plotted against θ instead of u. The result is that the patterns are distorted near $\theta = \pm \pi$ by the nonlinear transformation from u to θ , as seen for instance in Fig. 9.3.

The previous results can be obtained as the matched filter solution to a spatial filtering problem using the same results developed for Doppler filtering in <u>Chap. 5</u>. Suppose it is desired to maximize the output $z(\theta)$ when the input is a snapshot of a plane wave having AOA θ , plus white noise. The mathematics are exactly the same as used for vector matched filtering in <u>Chap. 5</u> and the same results can be applied. Specifically,

$$\mathbf{h} = \kappa \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^* \tag{9.9}$$

where S_I = covariance matrix of the interference at the N phase center outputs

t = model of the desired target signal vector (i.e., steering vector)

k = arbitrary constant

Recall that $\mathbf{S}_{\mathbf{I}} = \mathbf{S}_{\mathbf{I}}^{H}$ for covariance matrices. If the signal of interest is a monochromatic plane wave with AOA θ , the target model **t** is (letting $\hat{A} = 1$ without loss of generality) exactly the spatial steering vector $\mathbf{a}_{s}(\theta)$ defined in Eq. (9.3). When the interference is independent identically distributed (i.i.d.) white noise with variance σ_{w}^{2} at each element, $\mathbf{S}_{\mathbf{I}} = \sigma_{w}^{2}\mathbf{I}$. Choosing $\kappa = \sigma_{w}^{2}$ gives

$$\mathbf{h} = \mathbf{a}_{s}^{*}(\theta) = \begin{bmatrix} 1 & e^{+jk_{\theta}} & \cdots & e^{+j(N-1)k_{\theta}} \end{bmatrix}^{T}$$
(9.10)

The filter output is then

$$z(\theta) = \mathbf{h}^{T} \mathbf{y} = \sum_{n=0}^{N-1} y[n] e^{+jk_{\theta}n} = \sum_{n=0}^{N-1} y[n] e^{-j\tilde{k}_{\theta}n}$$
(9.11)

which shows that the antenna pattern is the DTFT of the data snapshot evaluated at $\tilde{k}_{\theta} \equiv -k_{\theta}$. If the observed signal y[n] is in fact a snapshot of a plane wave at AOA θ_0 , **y** is of the form of Eq. (9.3) with $\theta = \theta_0$ and $z(\theta)$ is again given by Eq. (9.7). The peak output occurs when the filter steering vector matches the actual

steering vector of the input, i.e., at $z(\theta_0)$, and is just *N* times the amplitude of y[n]. A plot of $z(\theta)$ as a function of AOA θ gives the (unweighted) normalized antenna gain pattern of Fig. 9.3. The beamformer signal processing gain G_{sp} is just this coherent gain *N*.

While the FFT can be used to efficiently compute the antenna pattern for a given steering vector **t**, the FFT output will be in terms of \tilde{k}_{θ} and must be flipped to obtain *z* as a function of k_{θ} . The flipped pattern will be sampled at constant increments of spatial frequency k_{θ} corresponding to constant increments in $\sin\theta$. Thus, the samples are not spaced uniformly in angle θ . If pattern samples at constant increments in θ are required, $z(\theta)$ must be computed explicitly using $z(\theta) = \mathbf{h}^T \mathbf{y}$ and varying \mathbf{y} for each desired value of AOA.

Digital beamforming enables several advantageous capabilities for phased array antennas. One is the ability to form multiple simultaneous beams on receive. The same snapshot data \mathbf{y} can be weighted with several different weight vectors \mathbf{h} , effectively forming multiple receive beams steered in different directions. Another major capability is the opportunity to compute the weights adaptively based on the interference environment characteristics. This makes possible adaptive nulling of jammers, discussed next. Other enhanced capabilities are mentioned in Aalfs (2013).

9.1.2 Adaptive Beamforming

The vector matched filter approach leads directly to a means for designing array weight vectors **h** that can steer zeros of the antenna pattern, often called *nulls*, in specific directions to cancel interference sources. This capability is useful in combating *jammers*, which are hostile interfering signal sources that seek to degrade radar performance by any of a number of mechanisms, such as degrading the signal-to-interference ratio (SIR) by increasing the noise level, or creating false detections to overwhelm the radar with false targets. One of the most common forms of jamming is a simple noise jammer. This device radiates a relatively high-power waveform at the victim radar from a specific air-, space-, or ground-based platform. The jammer waveform is a random noise process that is white over the receiver bandwidth of the victim radar.³ From the radar's point of view, the jammer signal is therefore a white noise process arriving from some specific AOA. A major advantage enjoyed by the jammer in interfering with target detection is that it must only propagate in one direction, as opposed to the two-way propagation of the transmitted waveform and target echo. The jammer signal suffers only a R^{-2} spreading loss instead of the R^{-4} loss of the target signal, making it imperative to try to attenuate the jammer signal on reception. Forming a beampattern with a null at the jammer's AOA accomplishes this.

The antenna pattern that maximizes the output SIR in the presence of both white noise and jamming is still given by Eq. (9.9); however, the model for the

interference covariance matrix S_I must be extended to incorporate a model for the noise jammer. The temporal variation of the jammer signal $J_n(t)$ received from AOA θ at each array phase center can be expressed as

$$J_n(t) = \sigma_J w(t) e^{j[\Omega(t - nd\sin\theta/c) + \phi_0]}$$
(9.12)

where σ_l^2 is the jammer power and w(t) is a unit variance zero mean white random process. A snapshot of the array response to the jammer gives

$$J[n] \equiv J_n(t_0) = \sigma_J w(t_0) e^{j[\Omega(t_0 - nd \sin \theta/c) + \phi_0]}$$

= $\sigma_J \hat{w}(t_0) e^{-j\Omega nd \sin \theta/c} = \sigma_J \hat{w}(t_0) e^{-j2\pi nd \sin \theta/\lambda}$
= $\sigma_J \hat{w}(t_0) e^{-jk_{\theta}n}$
(9.13)

In vector form, this is

$$\mathbf{J} = \sigma_{J} \hat{w}(t_{0}) \begin{bmatrix} 1 & e^{-jk_{\theta}} & \cdots & e^{-j(N-1)k_{\theta}} \end{bmatrix}^{T}$$
$$= \sigma_{J} \hat{w}(t_{0}) \mathbf{a}_{s}(\theta)$$
(9.14)

The covariance matrix of the jammer signals is then

$$S_{J} = E\{J^{*}J^{T}\} = \sigma_{J}^{2}E\{|w(t_{0})|^{2}\}a_{s}^{*}(\theta)a_{s}^{T}(\theta)$$

$$= \sigma_{J}^{2}a_{s}^{*}(\theta)a_{s}^{T}(\theta)$$

$$= \sigma_{J}^{2}\begin{bmatrix}1 & e^{-jk_{\theta}} & \cdots & e^{-j(N-1)k_{\theta}}\\e^{+jk_{\theta}} & 1 & e^{-jk_{\theta}} & e^{+j(N-2)k_{\theta}}\\\vdots & \vdots & \ddots & \vdots\\e^{+j(N-1)k_{\theta}} & e^{-j(N-2)k_{\theta}} & \cdots & 1\end{bmatrix}$$

$$(9.15)$$

 S_J is a Hermitian matrix, $S_J = S_J^H$. It is also positive semidefinite, but has a rank of only one because the columns are all linearly dependent; in fact, they are all simple multiples of the first column.

The total interference covariance matrix for the sum of receiver noise and some number P of mutually uncorrelated jammers is

$$\mathbf{S}_{\mathbf{I}} = \sigma_w^2 \mathbf{I} + \sum_{p=0}^{p-1} \mathbf{S}_{\mathbf{J}_p}$$

The matched filter output can then be maximized by using the interference model of Eq. (9.16) in Eq. (9.9).

As an example, consider a case with N = 16 antenna phase centers. Assume that two jammers are present, one at an AOA of +18° with a *jammer-to-signal ratio* (JSR) of +50 dB, and another at -33° and a JSR of +30 dB. The SNR = 0 dB, resulting in a total SIR at the element level (before any beamforming) of – 50.04 dB. Figure 9.4 shows the resulting antenna patterns when the antenna is steered to 0° (normal to the array face). In part *a* of the figure, no adaptation was used; i.e., the beamformer weights were designed using $S_1 = \sigma_{\omega}^{2}I$. A standard asinc pattern results. The two dotted vertical lines indicate the angles of arrival of the two jammers. Although in the sidelobes, both are on or near sidelobe peaks and, given their high power, will result in significant jamming energy in the beamformer output. In part *b* of the figure, the beamformer weights were designed using a covariance matrix S_1 computed via Eq. (9.16). The resulting pattern places deep nulls at the location of both jammers, effectively canceling them.



FIGURE 9.4 Antenna pattern for N = 16 phase centers. Two jammers present at indicated angles: (*a*) without adaptation, (*b*) with adaptation.

The SIR achieved by the optimum beamformer can be computed using the result of Eq. (5.7), namely

$$SIR = \frac{\mathbf{h}^{H} \mathbf{t}^{*} \mathbf{t}^{T} \mathbf{h}}{\mathbf{h}^{H} \mathbf{S}_{I} \mathbf{h}}$$
(9.17)

For the case where the interference is noise only, $\mathbf{S}_{I} = \sigma_{\omega}^{2}\mathbf{I}$, $\mathbf{h} = \mathbf{a}_{s}^{*}(\theta)$, $\mathbf{t} = \hat{A}\mathbf{a}_{s}(\theta)$, and since $\mathbf{a}_{s}^{T}(\theta)\mathbf{a}_{s}^{*}(\theta) = \sum_{n} |\mathbf{a}_{sn}(\theta)|^{2}$, Eq. (9.17) reduces to

$$SIR = \frac{\left(|\hat{A}_{s}|^{2} \sum_{n=0}^{N-1} |\mathbf{a}_{sn}(\theta)|^{2}\right)^{2}}{\sigma_{w}^{2} |\hat{A}_{s}|^{2} \sum_{n=0}^{N-1} |\mathbf{a}_{sn}(\theta)|^{2}} = \frac{|\hat{A}_{s}|^{2} N^{2}}{\sigma_{w}^{2} N} = N \frac{|\hat{A}_{s}|^{2}}{\sigma_{w}^{2}}$$
(9.18)

Equation (9.18) shows again that a coherent integration gain of $G_{sp} = N$ is obtained against white noise using the *N*-element snapshot, while for a more general interference environment the gain is the SIR of Eq. (9.17) divided by the pre-beamforming SIR. For the example of Fig. 9.4 without jammers the interference is just the noise, so the pre-beamforming SIR is the single-element SNR $|\hat{A}_s|^2 / \sigma_w^2 = 0$ dB. The post-beamforming SIR is 16 = 12.04 dB and the SIR gain is therefore also $G_{sp} = 12.04$ dB. With jammers present the pre-beamforming SIR is -50.04 dB and the post-beamforming SIR is 15.64 (11.94 dB), so the gain is 61.98 dB. The slight reduction in post-beamforming SIR (0.36 dB) when the jammers are present occurs because some of the *degrees of freedom* (DOF) are consumed canceling the jammers instead of providing coherent integration gain for the signal.

This variation in peak gain in the direction of the desired target signal **t** can be compensated by requiring that $\mathbf{h}^T \mathbf{t} = 1$, resulting in what is often called a *distortionless* beamformer. Using this condition in Eq. (9.9) gives

$$\mathbf{l} = \mathbf{h}^{T} \mathbf{t} = \left(\kappa \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^{*}\right)^{T} \mathbf{t} = \kappa \mathbf{t}^{H} \left(\mathbf{S}_{\mathbf{I}}^{-1}\right)^{T} \mathbf{t}$$
$$= \kappa \mathbf{t}^{H} \left(\mathbf{S}_{\mathbf{I}}^{-1}\right)^{*} \mathbf{t} \quad \Rightarrow$$
$$\kappa = \frac{1}{\mathbf{t}^{H} \left(\mathbf{S}_{\mathbf{I}}^{-1}\right)^{*} \mathbf{t}}, \quad \mathbf{h} = \frac{\mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^{*}}{\mathbf{t}^{H} \left(\mathbf{S}_{\mathbf{I}}^{-1}\right)^{*} \mathbf{t}}$$
(9.19)

Since κ is simply a scalar, this choice does not affect the shape of the antenna pattern $z(\theta)$ but merely scales it up or down in amplitude so that $z(\theta_0) = 1$, where θ_0 is the AOA of the target signal **t**. This design approach can be extended to constrain the pattern gain at multiple AOAs; details are given in Van Trees (2002).

If the interference is white, $\mathbf{S}_{\mathbf{I}} = \sigma_{\omega}^{2}\mathbf{I}$, the optimum weight vector is a scaled version of the steering vector, $\mathbf{h} = \kappa \mathbf{t}^{*}$. An interesting interpretation of the

optimum adaptive beamformer when the interference is not white starts by noting that any covariance matrix of interest can be factored in the form $S_I = V \cdot V$. That is, V is the "square root" of S_I . It is also the case that $S_I^{-1} = V^{-1}V^{-1}$ and that S_I , V, and their inverses are Hermitian. Using this decomposition and Eq. (9.9) with $\kappa = 1$ gives the beamformer output as

$$z(\theta) = \mathbf{h}^{T} \mathbf{y} = \mathbf{t}^{H} \left(\mathbf{S}_{\mathbf{I}}^{-1} \right)^{T} \mathbf{y}$$

= $\mathbf{t}^{H} [(\mathbf{V}\mathbf{V})^{-1}]^{T} \mathbf{y} = \mathbf{t}^{H} (\mathbf{V}^{-1})^{T} (\mathbf{V}^{-1})^{T} \mathbf{y}$
= $[\mathbf{t}^{T} (\mathbf{V}^{-1})^{H}] [(\mathbf{V}^{-1})^{*} \mathbf{y}] = [(\mathbf{V}^{-1})^{*} \mathbf{t}]^{T} [(\mathbf{V}^{-1})^{*} \mathbf{y}]$
= $\tilde{\mathbf{t}}^{T} \tilde{\mathbf{y}}$ (9.20)

The last line shows that the optimum filter output can be interpreted as a transformed steering vector $\tilde{\mathbf{t}} = (\mathbf{V}^{-1})^* \mathbf{t}$ applied to similarly transformed data $\tilde{\mathbf{y}} = (\mathbf{V}^{-1})^* \mathbf{y}$. The fact that the weight vector is just the transformed steering vector in this formulation suggests that the transformed interference in $\tilde{\mathbf{y}}$ should be white.

To see that this is indeed the case, consider the transformed data snapshot \tilde{y} when the data is interference only. The covariance of \tilde{y} is

$$\begin{split} \tilde{\mathbf{S}}_{\mathbf{I}} &= \mathbf{E}\{\tilde{\mathbf{y}}^* \, \tilde{\mathbf{y}}^T\} = \mathbf{V}^{-1} \mathbf{E}\{\mathbf{y}^* \, \mathbf{y}^T\} (\mathbf{V}^{-1})^H \\ &= \mathbf{V}^{-1} \mathbf{S}_{\mathbf{I}} (\mathbf{V}^{-1})^H = \mathbf{V}^{-1} (\mathbf{V} \mathbf{V}) (\mathbf{V}^{-1})^H \\ &= \mathbf{V}^{-1} (\mathbf{V} \mathbf{V}) (\mathbf{V}^{-1}) = \mathbf{I} \end{split}$$
(9.21)

Equation (9.21) states that transforming the data snapshot with the operator (V^- ¹)* *whitens* the interference. The appropriate weight vector is then just a scaled version of the similarly transformed steering vector as was discovered in Eq. (9.20). Applying the original optimal weight vector **h** to the original snapshot is therefore equivalent to solving a modified problem in which the data is transformed to whiten the interference and then filtered using the transformed steering vector.

Computing the adaptive weights requires finding $\mathbf{h} = \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^*$, which is equivalent to solving the linear system of equations $\mathbf{t}^* = \mathbf{S}_{\mathbf{I}} \mathbf{h}$. Various algorithms exist for doing this in a numerically stable and somewhat efficient matter. Generally, the computational load is $O(P^3)$, where the notation $O(\cdot)$ means "on the order of" and $P \equiv MN$ is the order of the system of equations (Arakawa and Bond, 2008).

How many jammers can be cancelled? For each distinct jammer signal \mathbf{J}_p , the vector matched filter chooses \mathbf{h} such that $\mathbf{h}^T \mathbf{J}_p = 0$ (Guerci, 2003). If *P* jammers are present this creates *P* such conditions. The distortionless constraint

 $\mathbf{h}^T \mathbf{t} = 1$ adds a (P + 1)st condition. There are therefore P + 1 equations in the N unknowns of the filter vector \mathbf{h} . A solution will exist so long as $P + 1 \le N$. Thus, an N-phase center array can cancel up to N - 1 jammers.

If a jammer is located in the mainbeam of the antenna pattern, the adaptive pattern is seriously degraded. In the previous example, the 3-dB beamwidth is approximately 6°. If the jammer at -33° is moved instead to -2° and the distortionless beamformer of Eq. (9.19) is applied the pattern of Fig. 9.5 results. The two jammers are nulled, but the antenna pattern peak has been shifted from the target AOA of 0°. The distortionless constraint still guarantees that the antenna pattern gain is unity at 0°, but the peak is now +5.15 dB and occurs at $\theta = 3.3^{\circ}$. In this example, the post-beamforming SIR is reduced to only 3.64 (5.61 dB).



FIGURE 9.5 Adaptive pattern with one jammer in the mainlobe of the adapted pattern.

The elements of the steering vector and thus of the optimum adaptive weights depend on the RF through the wavelength λ or the spatial frequency κ_{θ} . The weights that steer the beam to a particular AOA for one RF steer it to a different AOA for a different RF. Because of this dependence on λ , a single set of phase weights is incapable of cancelling a jammer over a wide bandwidth. A very common way to deal with this issue is to use *subbanding*. This technique, diagrammed in Fig. 9.6, divides the array spatial data as a function of fast time into *K* subbands, each relatively narrowband. A separate steering vector and

adaptive beamformer is computed in each subband according to the effective value of λ for that band and applied to that data. After cancellation, the subband data is recombined to restore the full bandwidth with the jammer cancelled. Because the temporal bandwidth in each subband is 1/K times the full temporal bandwidth, the fast-time sampling rate can be reduced by the factor K in each subband. The full sampling rate is restored in the single output channel. This technique can be implemented quite efficiently using *polyphase filter banks* (Harris, 2004). Another approach to wideband cancellation uses fast-time filtering in each spatial channel (Aalfs, 2013).



FIGURE 9.6 Subbanding approach for adaptive cancellation of interference in wideband signals.

9.1.3 Adaptive Beamforming with Preprocessing

The discussion so far has applied adaptive interference calculation directly to the individual array element signals, an approach called *element space* processing. Some systems perform fixed conventional beamforming on the element signals first, and then apply adaptive processing to the individual beam outputs. This technique is called *beamspace processing*. Figure 9.3 was an example of forming a single fixed beam; a bank of such beams would provide coverage of the entire angle space. Beamspace processing tends to concentrate each interferer into the output of a small number of beams and can therefore reduce the dimensionality of the adaptive processing problem. Because of its cubic dependence on the DOF, this reduction can lead to major savings in computational load.

Beams are formed as linear combinations of the element signals as shown in the example of Eq. (9.7). A set of *P* beams formed from an *N*-element array can thus be represented as a $P \times N$ transformation matrix **T** acting on the $N \times 1$ spatial snapshot **y** to form a new $P \times 1$ vector of beam outputs $\mathbf{\tilde{y}}$ (Ward, 1994; Melvin, 2013):
$$\tilde{\mathbf{y}} = \mathbf{T}\mathbf{y} \tag{9.22}$$

For instance, a set of conventional "DFT beams" at spatial frequencies k_{θ_0} , k_{θ_1} , ..., $k_{\theta_{P-1}}$ can be formed with the transformation matrix:

$$\mathbf{T} = \begin{bmatrix} 1 & e^{-jk_{\theta_0}} & e^{-j2k_{\theta_0}} & \cdots & e^{-j(N-1)k_{\theta_0}} \\ 1 & e^{-jk_{\theta_1}} & e^{-j2k_{\theta_1}} & \cdots & e^{-j(N-1)k_{\theta_1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-jk_{\theta_{p-1}}} & e^{-j2k_{\theta_{p-1}}} & \cdots & e^{-j(N-1)k_{\theta_{p-1}}} \end{bmatrix}$$
(9.23)

If the $[k_{\theta_p}]$ are evenly spaced over the interval $(-\pi, +\pi)$, **T** computes the *P*-point DFT of the element snapshot.

To see how to apply adaptive processing to the preprocessed beamspace data, note that the snapshot **y** is now replaced with $\tilde{\mathbf{y}}$. This applies to whatever signals are present at the array face, so the transformation is applied to the interference as well as the target signals. The new interference covariance matrix becomes

$$\tilde{\mathbf{S}}_{\mathbf{I}} = \mathbf{E}\left\{\tilde{\mathbf{y}}^* \tilde{\mathbf{y}}^T\right\}$$
$$= \mathbf{E}\left\{(\mathbf{T}\mathbf{y})^* (\mathbf{T}\mathbf{y})^T\right\} = \mathbf{T}^* \mathbf{E}\left\{\mathbf{y}^* \mathbf{y}^T\right\} \mathbf{T}^T$$
$$= \mathbf{T}^* \mathbf{S}_{\mathbf{I}} \mathbf{T}^T$$
(9.24)

The adaptive weight vector and the filtered output are

$$\tilde{\mathbf{h}} = \kappa \, \tilde{\mathbf{S}}_{\mathbf{I}}^{-1} \tilde{\mathbf{t}}^{*}$$
$$\tilde{z}(\theta) = \mathbf{h}^{T} \, \tilde{\mathbf{y}}$$
(9.25)

where $\tilde{t} = Tt$ is the transformed target steering vector.

To illustrate this process, a DFT matrix **T** with P = 16 beams was applied to the same N = 16 element example used previously. Figure 9.7 shows the magnitudes of the adaptive weights with and without the beamspace transformation. No window for sidelobe control was included. Without the transformation, the weights are all similar in magnitude, suggesting that all of the data is important to obtaining good cancellation results. With the transformation, most of the weights are close to zero, suggesting that many channels could be discarded with little impact on the adapted pattern. Figure 9.8 shows the adapted pattern with sidelobe jammers when the beamspace weights are thresholded at 5 percent of their maximum value, corresponding to the dashed line in Fig. 9.7. This discards three-quarters of the elements in $\tilde{\mathbf{h}}$, keeping only elements 0, 4, 13, and 14. Both in the absence of jammers and when the jammers are located at +18° and -33°, the resulting adapted antenna patterns are very close to those formed using the full 16 DOF in Fig. 9.4*a* and <u>*b*</u>.



FIGURE 9.7 Magnitude of adapted weights with sidelobe jammers for element space and DFT beamspace processing.



FIGURE 9.8 Adapted pattern using a 10-beam DFT fixed beamformer preprocessor. Compare to Fig. 9.4*b*.

The thresholded beamspace adaptive weight computation requires solving only a fourth-order system of equations, while the element space version requires the solution of a 16th-order system. The computational complexity of the weight vector computation will therefore decrease by a factor of about $(16/4)^3 = 64 \times$. The beamspace approach has the extra step of applying the beam formation preprocessor **T**, but in many cases the total computational load is still substantially reduced, especially if **T** has a structure that can be implemented efficiently, such as an FFT matrix.

The approach described here can be used with any linear transformation **T**. Additional examples include beams uniformly spaced in θ instead of κ_{θ} or approaches that form conventional beams and then combine sums of adjacent beams to achieve better cancellation. The technique can also be extended to include windowing for sidelobe control and gain constraints similar to the distortionless constraint discussed previously.

9.2 Space-Time Signal Environment

In a multi-phase center radar, filtering is possible in both Doppler shift and angle of arrival. It is therefore important to characterize the data in a given range bin in terms of Doppler and AOA. Figure 9.9 is a notional sketch of the general behavior of noise, jamming, clutter, and moving targets in the space-time environment. Receiver noise has no structure in time or frequency, and therefore appears as a uniform noise floor throughout the angle-Doppler space. As described in the previous section, broadband noise jammers are localized in AOA but spread across the entire Doppler spectrum. This is reflected in Fig. 9.9 as a ridge of energy localized in AOA but spread across all Doppler shifts. Because the jamming energy is the same at all values of Doppler, discrimination against jamming must be based primarily on spatial filtering as described in the previous section.



FIGURE 9.9 Space-time signal environment.

Clutter is more complicated. Assume a platform with velocity v and a sidelooking radar, $\frac{4}{2}$ as shown in Fig. 9.10. The height dimension is neglected for simplicity. In a given range bin, clutter scatterers anywhere on the isorange circle corresponding to the range of interest contribute to the total clutter return. (If the system is range ambiguous, there will be multiple isorange rings contributing to a given range bin.) A clutter scatterer directly on the radar boresight is at a squint angle of 90° with respect to the velocity vector; therefore the Doppler shift for that scatterer is zero. More generally, scatterers at an angle of θ radians with respect to the antenna boresight will have a Doppler shift of $(2v/\lambda)\sin\theta$ Hz. Note that there are two such patches that will produce the same Doppler shift, one in the radar look direction, and one behind it in the backlobes. The backlobes are often ignored due to low antenna gain in that direction, but in some systems they must be considered. If the backlobe return is ignored, there is a one-to-one relationship between AOA and Doppler shift for clutter:



FIGURE 9.10 Clutter contributing to the angle-Doppler spectrum.

$$F_D = \frac{2v}{\lambda} \sin\theta \text{ Hz}$$
(9.26)

or in normalized units

$$f_D = \left(\left(\frac{2vT}{\lambda} \sin \theta \right) \right)_{1,0} \quad \text{cycles per sample}$$
(9.27)

where *T* is the PRI and the notation $((\cdot))_{1,0}$ indicates arithmetic modulo 1.0 due to the use of discrete-time Fourier analysis. Echo from stationary ground clutter that has a Doppler shift of F_D Hz can therefore be presumed to be arriving at an angle of θ radians with respect to the sidelooking antenna boresight. Consequently, clutter echoes tend to fall on a diagonal ridge in a Doppler-sin θ space as diagrammed in Fig. 9.9. The amplitude of ground clutter from different AOAs is determined by the clutter reflectivity and antenna gain in that direction. In homogenous clutter it is therefore largest near boresight and lower in the sidelobes and backlobes of the antenna.

Equation (9.27) can be rewritten as

$$\omega_{D} = 2\pi f_{D} = \left(\left(\frac{2vT}{d} k_{\theta} \right) \right)_{2\pi}$$
$$= ((\beta k_{\theta}))_{2\pi} \text{ radians per sample}$$
(9.28)

where $\beta \equiv 2vT/d$. β is the slope of the clutter ridge when plotted in (k_0, ω_D) coordinates. It also represents the number of times the clutter ridge spans the range of $-\pi$ to $+\pi$ radians per sample (-0.5 to +0.5 cycles per sample) in normalized Doppler as the AOA varies from $-\pi$ to $+\pi$.

Because F_D is proportional to $\sin\theta$ rather than θ itself, the clutter ridge is a straight line in Doppler- $\sin\theta$ space; however, if plotted as a function of θ instead of $\sin\theta$, the ridge curves visibly as the AOA approaches ±180°. This effect is visible in Fig. 9.11*a* which is a simulation of the angle-Doppler spectrum for a medium PRF radar. Two jammers are present, at approximately -40° and +60°. The clutter ridge is diagonal through the center of the spectrum but curves noticeably at ±60°. Also, the discrete time Doppler spectrum is periodic with period equal to the PRF. Thus, if the magnitude of F_D of Eq. (9.26) exceeds *PRF*/2, the clutter ridge will alias. This effect is seen in Fig. 9.11*b* that simulates a low PRF angle-Doppler spectrum.



FIGURE 9.11 Simulation of angle-Doppler spectrum. Amplitude is in dB: (a) illustration of clutter ridge curvature and two jammers, (b) illustration of clutter ridge aliasing. (Figure courtesy of Dr. W. L. Melvin, GTRI.)

The angle-Doppler characteristics of the echo from a (possibly) moving point target depend on both the radar platform motion and the target motion. Assuming the target RCS is small enough that detection is likely only if it is in the radar mainbeam, the target can be presumed to be within a few degrees of the radar boresight. The Doppler shift will depend on the total radial velocity. If the target is stationary and directly on the boresight, the Doppler shift will be zero and it will fold in with the clutter. However, if the target is moving, it will separate from the clutter on the Doppler axis as shown in Fig. 9.9. This fact illustrates a key reason for the interest in space-time processing techniques, especially for the detection of relatively slow-moving ground targets ("slow movers"). As seen in Chap. 5, if only Doppler processing is used, the target Doppler shift must typically exceed the Doppler width of the clutter spectrum to achieve a signal-to-interference ratio adequate for detection.

If the platform velocity is high or the mainbeam relatively wide, ground clutter can fill most of the Doppler spectrum, making detection of slow movers very difficult. Figure 9.9 shows that introducing spatial processing gives a second dimension in which to separate the target from the clutter. Any Doppler shift of the target echo causes it to compete with clutter arriving from a different AOA; the added capability of filtering based on AOA then allows separation of target and clutter having the same Doppler shift.

9.3 Space-Time Signal Modeling

Space-time adaptive processing applies vector matched filtering to the combined slow-time/phase center data set in each range bin. It is usually assumed that pulse compression, if used, has been applied prior to STAP processing. The two-dimensional slice of the datacube in range bin l_0 , $y[l_0, m, n]$, is called a *space-time snapshot* (or just *snapshot*) of the data. To proceed, the $N \times M$ two-dimensional snapshot is converted to an $NM \times 1 = P \times 1$ column vector by stacking the columns:

$$\mathbf{y} = \begin{bmatrix} y[l_0, 0, 0] \\ y[l_0, 0, 1] \\ \vdots \\ y[l_0, 0, N-1] \\ y[l_0, 1, 0] \\ y[l_0, 1, 1] \\ \vdots \\ y[l_0, 1, N-1] \\ \vdots \\ y[l_0, M-1, 0] \\ y[l_0, M-1, 1] \\ \vdots \\ y[l_0, M-1, N-1] \end{bmatrix}$$

This process of converting the data from a given range bin to a one-dimensional vector is illustrated in Fig. 9.12.



FIGURE 9.12 Mapping of a datacube range bin to a two-dimensional space-time snapshot and then to a one-dimensional vector.

Next, the filter weight vector **h** must be designed using Eq. (9.9) or (9.19). The target model vector $\mathbf{t} = \mathbf{t}(f_D, \theta)$ must represent the expected signal from a target at some specific Doppler shift f_{Dt} and AOA θ_t of interest. Define a *temporal steering vector*

$$\mathbf{a}_{t}(f_{Dt}) = \begin{bmatrix} 1 & e^{-j2\pi f_{Dt}} & \cdots & e^{-j2\pi(M-1)f_{Dt}} \end{bmatrix}^{T}$$
(9.30)

This is simply the model for the slow-time data sequence corresponding to a target at normalized Doppler frequency f_{Dt} . The two-dimensional snapshot of the data from a target at Doppler f_{Dt} and AOA θ_t would have the temporal variation of $\mathbf{a}_t(f_{Dt})$ in each row and the spatial variation of $\mathbf{a}_s(\theta_t)$ in each column. The snapshot therefore has the form of the outer product of \mathbf{a}_s with \mathbf{a}_t

$$y[l_0, m, n] = \mathbf{a}_s(\theta_t) \mathbf{a}_t^T(\theta_t)$$

= $[a_{t0}(f_{Dt}) \mathbf{a}_s(\theta_t) \quad a_{t1}(f_{Dt}) \mathbf{a}_s(\theta_t) \quad \cdots \quad a_{t(M-1)}(f_{Dt}) \mathbf{a}_s(\theta_t)]$
(9.31)

where $a_{tm}(f_{Dt})$ is the *m*th element of $\mathbf{a}_t(f_{Dt})$. When this matrix is vectorized by stacking the columns, the result is the Kronecker product of the desired spatial and temporal steering vectors (Guerci, 2002; Melvin, 2013):

$$\mathbf{t} = \mathbf{a}_{t}(f_{Dt}) \otimes \mathbf{a}_{s}(\theta_{t}) = \begin{bmatrix} a_{t0}(f_{Dt})\mathbf{a}_{s}(\theta_{t}) \\ a_{t1}(f_{Dt})\mathbf{a}_{s}(\theta_{t}) \\ \vdots \\ a_{t(M-1)}(f_{Dt})\mathbf{a}_{s}(\theta_{t}) \end{bmatrix}$$
(9.32)

Next, a model for the $P \times P$ covariance matrix S_I of the interference is needed. The interference is the sum of receiver noise (**n**), jammer (**J**), and clutter (**c**) components. It is assumed that these are all uncorrelated with one another, with the result that the total interference covariance is the sum of the covariances of the three components, $S_I = S_n + S_J + S_c$.

Receiver noise, as usual, is assumed to be i.i.d. zero-mean complex Gaussian at each phase center and time sample so that $S_n = \sigma_w^2 I_p$, a scaled *P*th-order identity matrix. Now consider a single noise jammer signal. The spatial variation was given in Eq. (9.14). The temporal variation can be modeled as

$$\mathbf{a}_{t_{f}} = [a_{t_{f}0} \quad a_{t_{f}1} \quad \cdots \quad a_{t_{f}(M-1)}]^{T}$$
(9.33)

where the $\{a_{i_jm}\}$ are uncorrelated i.i.d. random variables with equal power σ_I^2 . The covariance of the jammer temporal variation is

$$\mathbf{E}\left\{\mathbf{a}_{t_{J}}^{*} \mathbf{a}_{t_{J}}^{T}\right\} = \sigma_{J}^{2} \mathbf{I}_{M}$$
(9.34)

If the AOA of the jammer is θ_{J0} , the space-time data vector for the jammer component is

$$\mathbf{J} = \sigma_J^2 \, \mathbf{a}_{t_J} \otimes \mathbf{a}_{s_J}(\theta_{J0}) \tag{9.35}$$

and its covariance matrix can be shown to be (Ward, 1994)

$$\mathbf{S}_{\mathbf{J}} = \mathbf{E}\{\mathbf{J}^*\mathbf{J}^T\} = \sigma_J^2 \mathbf{I}_M \otimes \mathbf{a}_{s_J}^*(\theta_{J0}) \mathbf{a}_{s_J}^T(\theta_{J0})$$

(9.36)

which is a block-diagonal matrix. If *R* multiple uncorrelated jammers are present, S_J is the sum of *R* terms of the form in Eq. (9.34), each with its own AOA θ_{JR} .

The clutter signal is the sum of contributions from all of the clutter scatterers within the isorange ring of interest (or multiple rings, if the system is ambiguous in range). Strictly speaking, this is an integral of the pertinent scatterers as with the angle-averaged reflectivity of <u>Chap. 2</u>. However, in STAP the integrated clutter is generally approximated as the sum of Q elemental clutter patches (see Fig. 9.10), each typically of an angular extent approximately equal to the radar beamwidth. For clutter patch q, the space-time data vector becomes

$$\mathbf{c}_{q} = \sigma_{c_{q}}^{2} \mathbf{a}_{t_{c}}(f_{D_{cq}}) \otimes \mathbf{a}_{s_{c}}(\theta_{cq})$$
(9.37)

where $\sigma_{c_q}^2$ represents the power of the θ th clutter patch as determined by the radar range equation. $\sigma_{c_q}^2$ is proportional to the antenna gain in the direction of that patch $G(q_q)$. The normalized Doppler shift and AOA of the clutter patch are related through Eq. (9.27). The total clutter vector is

$$\mathbf{c} = \sum_{q=0}^{Q-1} \mathbf{c}_q = \sum_{q=0}^{Q-1} \sigma_{c_q}^2 \, \mathbf{a}_{t_c}(f_{D_{cq}}) \otimes \mathbf{a}_{s_c}(\theta_{cq})$$
(9.38)

The covariance matrix of the clutter is

$$\mathbf{S}_{c} = \mathbf{E}\{\mathbf{c}^{*}\mathbf{c}^{T}\} = \sum_{q=0}^{Q-1} \sigma_{c_{q}}^{2} \mathbf{c}_{q}^{*} \mathbf{c}_{q}^{T}$$
$$= \sum_{q=0}^{Q-1} \sigma_{c_{q}}^{2} \left[\mathbf{a}_{t_{c}}^{*}(f_{D_{cq}}) \mathbf{a}_{t_{c}}^{T}(f_{D_{cq}}) \right] \otimes \left[\mathbf{a}_{s_{c}}^{*}(\theta_{cq}) \mathbf{a}_{s_{c}}^{T}(\theta_{cq}) \right]$$
(9.39)

This is an $M \times M$ block matrix, where each "element" of the block matrix is the $N \times N$ cross-covariance of the spatial snapshots from two different PRIs. S_c can be factored as (Ward, 1994)

$$S_{c} = C\Sigma_{c}C$$

$$C = [c_{0} \ c_{1} \ \cdots \ c_{Q-1}]$$

$$\Sigma_{c} = diag\left(\left[\sigma_{c_{q}}^{2} \ \sigma_{c_{q}}^{2} \ \cdots \ \sigma_{c_{q}}^{2}\right]\right)$$

The discussion of clutter so far assumes that it is uncorrelated in the spatial dimensions (range and cross-range) but perfectly correlated in slow time. It is indeed common to model the clutter as uncorrelated in space, assuming that clutter patches are separated by a distance on the order of a resolution cell. However, as discussed in <u>Chap. 2</u> and again in the analysis of Doppler processing in <u>Chap. 5</u>, the clutter echo from any given ground patch cannot be reasonably modeled as constant in slow time over the time scale of a coherent processing interval. Natural clutter exhibits reflectivity fluctuations in time due to *internal clutter motion* (ICM, also called *intrinsic clutter motion*). This is simply the physical movement of scatterers due to wind or waves. The radar system itself contributes sources of temporal modulation such as antenna scanning modulation or pulse-to-pulse instabilities. All of these temporal reflectivity fluctuations cause a broadening of the clutter power spectrum and cause decorrelation of the temporal snapshots.

ICM is easily incorporated into the model of Eq. (9.39) by replacing the temporal snapshot for patch q, $\sigma_{c_q}^2 a_{t_c}(f_{D_{c_q}})$, with a modified temporal snapshot that replaces the constant amplitude $\sigma_{c_q}^2$ with time-varying amplitudes $\sigma_{c_q}^2 \alpha_{q}$, where

$$\alpha_q = [\alpha_{0q} \ \alpha_{1q} \ \cdots \ \alpha_{(M-1)q}]^T$$
(9.41)

Let \mathbf{A}_q be the covariance matrix of the temporal fluctuations α_q ; \mathbf{A}_q is an $M \times M$ Toeplitz matrix. The modified space-time data vector for the *q*th clutter patch is then

$$\mathbf{c}_{q} = \sigma_{c_{q}}^{2} \left[\alpha \cdot \mathbf{a}_{t_{c}}(f_{D_{cq}}) \right] \otimes \mathbf{a}_{s_{c}}(\theta_{cq})$$
(9.42)

and the clutter covariance matrix becomes

$$\mathbf{S}_{c} = \sum_{q=0}^{Q-1} \sigma_{c_{q}}^{2} \left[\alpha \cdot \mathbf{a}_{t_{c}}^{*} \left(f_{D_{cq}} \right) \mathbf{a}_{t_{c}}^{T} \left(f_{D_{cq}} \right) \right] \otimes \left[\mathbf{a}_{s_{c}}^{*} \left(\theta_{cq} \right) \mathbf{a}_{s_{c}}^{T} \left(\theta_{cq} \right) \right]$$

$$(9.43)$$

An alternative approach for modeling ICM using covariance matrix tapers is described in <u>Sec. 9.7</u>.

Various models for the temporal correlation of the data are available in the literature. One that is popular in STAP simulations is the Billingsley model (Billingsley, 2001). This model assumes that the clutter temporal power spectrum is the sum of a two-sided exponential function and an impulse at the

origin in Doppler frequency space:

$$S_{c}(F) = \sigma_{c}^{2} \left[\frac{\alpha}{\alpha+1} \delta_{D}(F) + \frac{1}{\alpha+1} \left(\frac{\beta\lambda}{4} \right) \exp\left(-\frac{\beta\lambda}{2} |F| \right) \right]$$
(9.44)

where α is the ratio of the DC to AC components, and β is a parameter dependent primarily on wind conditions.

The corresponding autocorrelation function is

$$s_c(\tau) = \sigma_c^2 \left(\frac{\alpha}{\alpha+1} + \frac{1}{\alpha+1} \frac{(\beta\lambda)^2}{(\beta\lambda)^2 + (4\pi\tau)^2} \right)$$
(9.45)

Experimental data are available to choose the parameters α and β to fit various scenarios distinguished by the type of clutter, radar wavelength, weather conditions, and so forth. Simple autoregressive filters can be used to implement the model in simulations (Mountcastle, 2004).

9.4 Processing the Space-Time Signal

9.4.1 Optimum Matched Filtering

Optimal space-time processing of the two-dimensional data snapshot consists of the following steps. For a range bin of interest,

- 1. Form the interference covariance matrix S_I . In practice, S_I must be estimated from the radar data, but discussion of this issue is deferred to <u>Sec. 9.4.4</u>.
- 2. Select a Doppler shift and angle of arrival at which to test for the presence of a target signal, and form the appropriate target space-time steering vector \mathbf{t} using Eq. (9.32).
- 3. Compute the optimal filter weight **h** using either Eq. (9.9) or (9.19), depending on the normalization desired.
- 4. Form the space-time data vector **y** as illustrated in Fig. 9.12 and described in Eq. (9.29).
- 5. Apply the weight vector to the data to obtain the test statistic $z = z(f_D, \theta)$ = $\mathbf{h}^T \mathbf{y}$.

The test statistic can then be used for detection or for other purposes such as angle estimation. If used for detection, typically |z| or $|z|^2$ is computed and compared to an appropriate threshold computed using the techniques of <u>Chap. 6</u>.

The previous procedure computes the optimum (maximum SIR) test statistic for a single Doppler frequency and AOA and in a single range bin. It therefore must be repeated for each Doppler and AOA of interest, and then for each range bin. Within a given range bin the covariance matrix will remain constant and can be reused, but steps 2 through 5 must be repeated for each (f_D , θ) combination. **S**_I should ideally be recomputed for each range bin. Because the product P = MN of the number of pulses in the CPI and the number of phase centers in the antenna can easily be in the hundreds, the procedure above implies solving a system of hundreds of linear equations for each Doppler-AOA point of interest and each range bin.

While the optimum weight vector is given by $S_I^{-1}t^*$, in practice it is often desirable to include windowing of the data to reduce sidelobes as was done in the spatial beamforming case in Fig. 9.3 and for Doppler processing in <u>Chap. 5</u>. Combined angle-Doppler weighting can be included by computing the weight vector as

$$\mathbf{h} = \kappa \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}_{w}^{*} \tag{9.46}$$

where the windowed target steering vector is

$$\mathbf{t}_w = (\mathbf{w}_f \otimes \mathbf{w}_\theta) \cdot \mathbf{t} \tag{9.47}$$

and \mathbf{w}_f and \mathbf{w}_{θ} are the temporal and spatial weight vectors. That is, the conventional steering vector **t** is multiplied by a space-time window vector that is the Kronecker product of the desired Doppler and angle weighting functions.

9.4.2 STAP Metrics

A common metric used to visualize STAP filtering performance is the *adapted* pattern (Ward, 1994). This is simply a two-dimensional plot of $|z(f_D, \theta)|^2 = |\mathbf{h}^T \mathbf{t}|^2 = \mathbf{h}^H \mathbf{t}^* \mathbf{t}^T \mathbf{h}$ as f_D and θ are stepped over a regular grid. If the array has uniformly spaced phase centers and a constant PRI was used, the adapted pattern can be computed as the two-dimensional DFT of the weight vector \mathbf{h} after it is remapped to a two-dimensional snapshot form, allowing the use of the FFT to efficiently compute the pattern.⁵ However, the angle samples will be evenly spaced in k_{θ} rather than in θ itself. If it is preferred to display the pattern in terms of f_D and θ , one-dimensional FFTs can be applied to the temporal dimension of \mathbf{h} but the angle dimension samples must be computed individually using one-dimensional DFTs at the non-uniformly spaced values of k_{θ} that correspond to uniformly spaced samples of θ .

Figure 9.13 illustrates the adapted pattern for an idealized example.⁶ The case considered corresponds to a sidelooking radar at an RF of 675 MHz. The

platform velocity is 50 m/s and the PRF is 200 pulses per second. There are eight phase centers spaced by $\lambda/2$ meters and eight pulses in the CPI so that M =N = 8. Thus the clutter ridge slope $\beta = 1.5$ in this example. Jammers are present at $+30^{\circ}$ and -44.43° , corresponding to normalized spatial frequencies of +0.25and -0.35 cycles/sample. The *clutter-to-noise ratio* (CNR) is +40 dB, as is the *jammer-to-noise ratio* (JNR) for both jammers. Part *a* of the figure shows the loci of the jammer and clutter energy in angle-Doppler space. Part b of the figure shows the resulting adapted pattern when a target is present at an AOA of 0° and a normalized Doppler shift of 0.2 cycles/sample. The adapted pattern clearly shows vertical nulls at the two jammer AOAs and a three-part diagonal null corresponding to the clutter ridge. The adapted pattern has a strong peak at the target location of 0° and 0.2 cycles/sample. Because no windowing was used in this example, high Doppler and spatial sidelobes are evident as horizontal and vertical ridges extending from the peak. Figure 9.14 plots Doppler and spatial cuts through the adapted pattern taken at the target coordinates, showing the nulls in the spatial pattern at the jammer and clutter locations and in the Doppler pattern at the clutter ridge location.



FIGURE 9.13 (*a*) Clutter and jamming loci for idealized example: (*b*) adapted pattern. See text for details.



FIGURE 9.14 Pattern cuts from the adapted pattern of Fig. 9.13: (*a*) Doppler pattern, (*b*) spatial pattern. Both cuts pass through the actual target location; cut lines are shown in Fig. 9.13*b*.

Another important class of metrics involves SIR when the interference is noise only and when it is noise plus clutter and/or jamming. Metrics in this class include SIR and *SIR loss*, which is the reduction in SIR for the clutter-plusjamming case compared to the noise-only case. Each of these is a function of target angle and Doppler; typically they are plotted as a function of Doppler for a fixed AOA.

Consider a target that produces an SNR of χ_t for a single pulse and phase center, i.e., for a single sample of the two-dimensional snapshot. The optimum SNR is then

$$\chi_0 = MN\chi_t \tag{9.48}$$

The factor MN represents the coherent integration gain achievable by combining the MN samples of the snapshot. The SIR is, from Eq. (9.17),

$$SIR = \frac{\mathbf{h}^{H} \mathbf{t}^{*} \mathbf{t}^{T} \mathbf{h}}{\mathbf{h}^{H} \mathbf{S}_{I} \mathbf{h}} = \frac{|z|^{2}}{\mathbf{h}^{H} \mathbf{S}_{I} \mathbf{h}}$$
(9.49)

From <u>Chap. 5</u>, when using the optimum weight vector this becomes just

$$SIR_{\max} = \mathbf{t}^T \mathbf{S}_{\mathbf{I}}^{-1} \mathbf{t}^*$$
(9.50)

The SIR loss is then defined as

$$L_{SIR} = \frac{SIR}{\chi_0}$$
(9.51)

If the steering vector used in the STAP filter design exactly matches the target Doppler and angle, SIR_{max} is used in Eq. (9.51); however, this definition of L_{SIR} can be used with suboptimal filter designs as well, in which case Eq. (9.49) is used as the numerator of L_{SIR} . Note that L_{SIR} is defined here as a number less than one (negative dB), in keeping with common practice in the STAP literature.

 L_{SIR} is a function of Doppler and angle of arrival. It is typically plotted as a function of Doppler with the array steered in the correct target direction. Figure 9.15 is a plot of L_{SIR} versus Doppler for the interference scenario of Fig. 9.13. Each point on this curve is obtained by positing a target at the corresponding Doppler shift and then evaluating Eq. (9.51) using the matched steering vector for that target Doppler. When the target velocity is far from the clutter ridge, for example $|f_D| > 0.3$, the loss is minimal, about 0.7 dB at the plot edges. As the target velocity nears the clutter at zero Doppler the losses increase, reaching over 50 dB at $f_D = 0$.



FIGURE 9.15 SIR loss versus Doppler for the scenario of Fig. 9.13.

This plot serves as the basis for two additional metrics (Ward, 1994). *Minimum detectable velocity* (MDV) is the velocity closest to the clutter notch at which an acceptable SIR loss is achieved. *Minimum detectable Doppler* (MDD) is the corresponding Doppler shift in either absolute or normalized units. Clearly, MDV is $\lambda/2$ times the MDD in hertz or $\lambda/2T$ times the MDD in normalized frequency units of cycles/sample. The frequency at which the SIR loss becomes acceptable is not necessarily symmetric about the clutter notch. Using normalized frequency units, define the upper and lower MDD as

$$MDD_{+} = \min_{f_{D}} \{ f_{D} \text{ such that } L_{SIR} \ge L_{0}, f_{D} > 0 \}$$

$$MDD_{-} = \max_{f_{D}} \{ f_{D} \text{ such that } L_{SIR} \ge L_{0}, f_{D} < 0 \}$$
(9.52)

where L_0 is the maximum acceptable loss threshold. This definition assumes that the clutter notch is at $f_D = 0$; it is straightforward to generalize it to clutter notches at other frequencies. The MDD is then defined as the average of these two offsets, taking into account the fact that MDD_1 is negative:

$$MDD = \frac{1}{2}(MDD_{+} - MDD_{-})$$
(9.53)

The choice of L_0 is a system design decision. Based on radar range

equation considerations, $L_0 = -12$ dB would correspond to a 50 percent reduction in detection range, while $L_0 = -5$ dB would correspond to a 25 percent reduction. In Fig. 9.15, a value of $L_0 = -3$ dB has been selected. In this example, this choice results in $MDD_+ = 0.115$ and $MDD_- = -0.112$, giving MDD= 0.1135.

Another metric related to MDD is the *usable Doppler space fraction* (UDSF). This is the fraction of the Doppler space over which the SIR loss is acceptable, that is, $L_{SIR} > L_0$. UDSF is simply expressed in terms of the MDD in normalized frequency units:

$$UDSF = 1 - (MDD_{+} - MDD_{-})$$
$$= 1 - 2MDD$$
(9.54)

In the example of Fig. 9.15, UDSF = 0.773, that is, the SIR loss is considered acceptable over 77.3 percent of the Doppler spectrum.

As a final example, Fig. 9.16 shows the adapted patterns obtained with the optimum filter for the two interference environments of Fig. 9.11. In the medium PRF case, there are now two vertical nulls in the Doppler dimension where the jammer energy was located. In addition, the STAP processing has implemented an S-shaped null² to follow the clutter ridge and attenuate the clutter energy. The large response at approximately $F_D = 400$ Hz and $\theta = 0^\circ$ is a target that was not visible in the original data. The two vertical ridges of energy in Doppler and angle are the sidelobes of the target response. Similarly, the low PRF case shows the ability of STAP to implement a three-part null to remove aliased clutter, again revealing a hidden target.



FIGURE 9.16 Result of optimal STAP processing of data in Fig. 9.11: (*a*) medium PRF case with clutter and two jammers, (*b*) low PRF case with aliased clutter. (Figure courtesy of Dr. W. L. Melvin, GTRI.)

9.4.3 Relation to Displaced Phase Center Antenna Processing

<u>Section 5.7</u> introduced displaced phased center antenna processing in the context of slow-moving target indication from a moving radar platform. DPCA

processing combines data from multiple pulses and multiple antenna phase centers to form a clutter-cancelled output; as such, it appears to be related to STAP processing. To identify this connection, consider nonadaptive DPCA processing using two phase centers separated by a distance d_{pc} so that N = 2. Assume that the DPCA condition is met with a time slip of M_s pulses; that is

$$M_{\rm s} = \frac{d_{pc}}{2vT} \implies \beta = \frac{2vT}{d_{pc}} = \frac{1}{M_{\rm s}}$$

$$(9.55)$$

The space-time snapshot in two-dimensional form is therefore a $2 \times M$ array, with spatial index n = 0 corresponding to the fore phase center and n = 1 corresponding to the aft phase center as shown in the upper portion of Fig. 9.17.



FIGURE 9.17 Schematic diagram of two-dimensional space-time snapshot for a DPCA processor, and its decomposition into sub-CPIs.

A nonadaptive DPCA processor is constrained to be of the form

$$z[m] = y_f[m] - y_a[m + M_s]$$
(9.56)

where $y_f[m]$ and $y_a[m]$ are the fore and aft phase center outputs, respectively. It is sufficient to work with a single "sub-CPI", i.e., an interval of M_s pulses, that spans the two pulses combined in the DPCA processor. The complete CPI snapshot is then constructed of a series of $M - M_s$ overlapping sub-CPIs as

shown in the bottom portion of Fig. 9.17 for the case $M_s = 2$. The weight vector \mathbf{h}_1 that implements Eq. (9.56) for a single sub-CPI and the corresponding weight vector \mathbf{h} for the entire CPI are shown in Fig. 9.18 in two-dimensional form.



FIGURE 9.18 Full-CPI weight vector \mathbf{h} and sub-CPI weight vector \mathbf{h}_1 .

Viewed as a two-dimensional discrete function, the sub-CPI weight vector \mathbf{h}_1 can be written as

$$\mathbf{h}_1 \Longrightarrow h_1[m, n] = \delta[m, n] - \delta[m + M_s, n - 1]$$
(9.57)

The adapted pattern is the two-dimensional DTFT of this function:

$$H_{1}(\omega_{D}, \tilde{k}_{\theta}) = \sum_{m=0}^{M_{g}-1} \sum_{n=0}^{1} h_{1}[m, n] e^{-j(m\omega_{D}+n\tilde{k}_{\theta})}$$

$$= 1 - \exp[-j(-M_{s}\omega_{D} - \tilde{k}_{\theta})]$$

$$= 1 - \exp[j(M_{s}\omega_{D} + \tilde{k}_{\theta})]$$

$$= 1 - \exp[j(M_{s}\omega_{D} - k_{\theta})]$$

(9.58)

Note that this function has a zero at $\omega_D = k_{\theta}/M_s = \beta k_{\theta}$, which is exactly the expression for the clutter ridge in Eq. (9.28). Figure 9.19 shows $|H_1(f_D, k_{\theta})|$ for the case $M_s = 2$. Clearly, the DPCA processor implements a null along the clutter ridge. Repeating the single sub-CPI DPCA process for the additional sub-CPIs within the total CPI simply provides coherent integration by a factor of $M - M_s$. On the other hand, the constrained form of the weight vector [Eq. (9.57)] prevents any combining of samples from different phase centers on the same pulse so that the DPCA processor cannot provide any spatial beamforming capability. As a result, DPCA processing can cancel the clutter but not jammers.



FIGURE 9.19 Adapted pattern for nonadaptive DPCA processor with $M_s = 2$.

The model of Eq. (9.58) implicitly assumes an omnidirectional antenna pattern for each of the antennas or subarrays that form the DPCA phase centers. A more realistic model that accounts for the subarray patterns can be obtained by repeating the analysis with the output from each phase center filtered in angle by the antenna pattern of the antenna or subarray that forms that phase center. Denoting the fore and aft subarray antenna patterns as $E_f(k_\theta)$ and $E_a(k_\theta)$, respectively, Eq. (9.58) can be generalized to

$$H_1(\omega_D, k_\theta) = E_f(k_\theta) - E_a(k_\theta) \exp[-j(M_s\omega_D - k_\theta)]$$

= $E_f(k_\theta) \{1 - \exp[-j(M_s\omega_D - k_\theta)]\}$
(9.59)

where the last step holds only if $E_f(k_\theta) = E_a(k_\theta)$. If the two patterns differ, the first line of Eq. (9.59) shows that there will not be a null at the clutter ridge location as desired. This emphasizes the need for carefully matched subarray patterns in DPCA processing.

Adaptive DPCA improves on these results by applying the vector matched filtering framework as described in <u>Sec. 5.7.2</u>. The adaptive DPCA processor is still constrained to a weight vector structure that prohibits combining of phase

center outputs from the same pulse and therefore, like nonadaptive DPCA, provides no spatial beamforming. In addition, the use of the target model $\mathbf{t} = [1 \ 0]^T$ provides a result that is optimized "on average" over all target Doppler frequencies, rather than for any specific target Doppler.

9.4.4 Adaptive Matched Filtering

In <u>Chap. 6</u> it was seen that the knowledge of the interference power was required to set the detection threshold. CFAR techniques were introduced to estimate the noise level from the radar data. Exactly the same issue exists with STAP processing, which requires knowledge of the interference covariance matrix S_I to compute the optimal weight vector **h** and perform the matched filtering. Again, it is not realistic to assume that S_I can be known a priori, so it must be estimated from the radar data.

The most common approach to STAP when S_I is unknown is called the *sample matrix inverse* (SMI) method (Ward, 1994). This technique is exactly analogous to cell-averaging CFAR. Figure 9.20 illustrates a datacube with the cell under test (CUT) indicated, as well as a number of adjacent range bins. The data in these reference cells adjacent to the CUT are assumed to consist of distributed i.i.d. interference having the same statistics as the interference in the CUT. It is also assumed that the reference cells do not contain any target signals.⁸ Consequently, they can be used to compute a sample mean estimate of S_I . Specifically, the sample covariance of the data from a single reference cell is



FIGURE 9.20 Datacube showing selection of reference cells for estimating the interference covariance matrix.

$$\hat{\mathbf{S}}_l = \mathbf{y}_l^* \mathbf{y}^T \tag{9.60}$$

If L_S reference cells are used, the estimate of S_I is simply

$$\hat{\mathbf{S}}_{\mathbf{I}} = \frac{1}{L_{s}} \sum_{l=1}^{L_{s}} \hat{\mathbf{S}}_{l}$$
(9.61)

Note that filter weight vectors computed using \hat{S}_{I} instead of S_{I} will be suboptimum because of the imperfect interference estimate.

The reduction in SIR due to the use of $\hat{\mathbf{S}}_{\mathbf{I}}$ instead of $\mathbf{S}_{\mathbf{I}}$ is denoted by L_{CFAR} . It has been shown to be a beta-distributed random variable with expected value (Reed et al., 1974; Nitzberg, 1984)

$$E(L_{CFAR}) = \frac{L_{s} + 2 - P}{L_{s} + 1}$$
(9.62)

Note that, as defined, $E(L_{CFAR})$ will be less than one (negative dB). It is plotted as a function of L_s/P , the size of the reference window relative to the number of DOF (snapshot size), in Fig. 9.21. This figure shows that the number of reference cells must be twice the number of DOF in the processor to limit the loss due to covariance estimation to 3 dB. To limit the loss to 1 dB requires L_s > 5P. The conclusion that the reference window size should usually be two to five times the snapshot size is known as the "Reed-Mallet-Brennan" or *RMB* rule. The shape of the curve is a very weak function of P; the example shown is for P = 256, corresponding for example to a system with N = 8 phase centers and M = 32 pulses in the CPI. These results apply only when there is no mismatch between the actual target steering vector and the model vector t used in the filter design; this also implies no use of windowing for reduced sidelobes. Generalizations of Eq. (9.62) for mismatched and windowed target model vectors are discussed in Boroson (1980) and Kelly (1989).



FIGURE 9.21 RMB estimate of SIR loss due to estimation of covariance matrix S_{I} for P = 256.

Once the covariance matrix has been estimated, the weight vector is computed in the usual fashion. One particular choice of the scale factor κ of interest when using a square-law detector is the *adaptive matched filter* (AMF) (Kelly, 1986; Chen and Reed, 1991; Robey et al., 1992)

$$\kappa = \frac{1}{\sqrt{\mathbf{t}^{H} \left(\hat{\mathbf{S}}_{\mathbf{I}}^{-1}\right)^{*} \mathbf{t}}}$$
(9.63)

Notice the similarity to Eq. (9.19). With this choice, the filter output becomes

$$|z|^{2} = \frac{\left|\mathbf{t}^{H}(\hat{\mathbf{S}}_{\mathbf{I}}^{-1})^{*}\mathbf{y}\right|^{2}}{\mathbf{t}^{H}(\hat{\mathbf{S}}_{\mathbf{I}}^{-1})^{*}\mathbf{t}}$$
(9.64)

It can be shown that a threshold test applied to this test statistic exhibits CFAR behavior. It is also claimed that the AMF is more robust to targets in the training data than alternative tests such as the generalized likelihood ratio test (Steinhardt and Guerci, 2004).

The adapted pattern when using the SMI technique is subject to pattern degradations such as elevated sidelobes and distorted mainbeams, especially when L_s is relatively small. In addition, the pattern may vary from update to

update, a condition called *weight jitter*. If $L_s < P$, then \hat{S}_I may also become nonsingular. A common extension of the basic SMI technique that addresses these issues is *diagonal loading*, in which a bias term is added to the diagonal elements of the estimated covariance matrix (Carlson, 1988):

$$\hat{\mathbf{S}}_{\mathbf{I}} \Rightarrow \hat{\mathbf{S}}_{\mathbf{I}} + \varepsilon \mathbf{I}$$
(9.65)

Since diagonal loading adds a factor that has the same structure as the covariance matrix of white noise, its effect is to increase the apparent noise floor of the data. The loading factor ε is typically set 10 to 30 dB above the actual noise level σ_{ω}^2 (Kim et al., 1998). Diagonal loading tends to ensure nonsingularity of \hat{S}_I and reduce distortion of the adapted pattern, but also reduces the depth of the nulls (Guerci, 2003).

9.5 Reduced-Dimension STAP

It was shown in Sec. 9.1.3 that preprocessing of the element data could be used to reduce the dimensionality of the adaptive processing equations. The same technique can be applied to STAP processing. However, reduced-dimension processing is especially important in STAP because the snapshot dimensionality can be quite large, possibly in the hundreds. This aggravates two problems. First, the RMB rule states that the number of reference range bins required to maintain acceptable losses due to covariance estimation is on the order of 2P to 5P. Consequently, complex systems (many phase centers, long CPIs) can have long reference windows, with the result that the training data are very unlikely to be statistically homogeneous as assumed in the SMI algorithm due to terrain variations. Second, the computational load is of order P^3 . Reducing the dimensionality by only a factor of two produces nearly an order of magnitude reduction in computational load in solving the SMI equations. Increasing it by a factor of two for better performance (for instance, by reducing straddle losses or covariance matrix estimation losses) would similarly increase the load by nearly an order of magnitude.

Because STAP operates on two data dimensions, preprocessing can be applied in the slow-time dimension, phase center dimension, or both. Figure 9.22 illustrates a taxonomy of reduced dimension STAP techniques based on the choice of preprocessing options (Ward, 1994). There are many variants of each general class. As an example, Fig. 9.23 illustrates a particular variant of the beamspace post-Doppler class of STAP processors. DFTs in both dimensions are used to form a grid of fixed angle-Doppler bins. For a given target model vector \mathbf{t} , adaptation is performed using only a small number of angle and Doppler bins around the target AOA and Doppler frequency, analogous to the

earlier example of adaptive beamforming using DFT beamspace preprocessing. Beamspace post-Doppler architectures can isolate the interference in both angle and Doppler and substantially reduce the order of the SMI equations. Typical implementations use three to five bins in each of the Doppler and angle dimensions, giving nine to 25 total degrees of freedom.



FIGURE 9.22 Taxonomy of reduced-dimension STAP algorithms. (After Ward, 1994.)



FIGURE 9.23 Structure of a particular beamspace post-Doppler STAP processor.

9.6 Advanced STAP Algorithms and Analysis

Only the most basic STAP algorithm, the SMI approach based on the vector matched filter, has been introduced in this chapter, and only metrics related to the adapted pattern and SIR have been used to evaluate its performance. A deeper understanding of the character of the interference and the behavior of STAP algorithms requires analysis of the eigenstructure of the signal environment, a topic beyond the scope of this book but thoroughly covered in texts devoted to STAP such as (Klemm, 2002; Van Trees, 2002; and Guerci, 2003). The latter includes (p. 112) a taxonomy of modern STAP algorithms proposed for radar data.

A number of ideas in advanced STAP are based on the realization that the covariance matrix of neither the clutter or the jammers is full rank. For instance, the rank of an ideal covariance matrix for J independent jammers is MJ (Guerci, 2003). The rank of the clutter-only covariance matrix under ideal conditions (no crab, constant velocity) can be estimated as (Brennan and Staudaher, 1992)

$$\operatorname{rank}\{\mathbf{S}_{c}\}\approx N+(M-1)\boldsymbol{\beta}$$

(9.66)

where β is the clutter ridge slope of Eq. (9.28). Because β is usually in the range of zero to three, this is well below the full rank P = NM. Thus, both the clutter and jamming signals can be represented with a relatively few basis vectors in the *P*-dimensional signal space. Eigenanalysis provides a convenient means for decomposing the signal components, defining a reduced-dimension representation, and analyzing the performance of the resulting algorithms.

This approach leads to algorithms that are fundamentally different from the SMI approach. Suppose that the combined clutter and jamming covariance matrix has Q dominant eigenvalues, with the remainder at or near the noise floor eigenvalue level. The *principal components* (PC) method forms the adaptive filter weight vector as a linear combination of the Q eigenvectors corresponding to the dominant eigenvalues, with weights related to the eigenvalue associated with each eigenvector used. The PC method and similar *subspace projection* techniques can construct high quality adapted patterns without the degradation of sidelobes that often occurs with SMI techniques when estimated covariance matrices are used. Furthermore, because the filter vector is derived from only Q eigenvectors, these algorithms provide another means of dimensionality reduction. Algorithms based on projecting data into lower-dimensional spaces are generally called *reduced rank* STAP techniques (as opposed to the reduced dimension techniques of Sec. 9.5).

Even with effective rank reduction, the nonstationarity and heterogeneity of the data due to error effects to be discussed in <u>Sec. 9.7</u>, and more fundamentally to variations in the physical clutter scene over the reference windows, remain major limiting factors in STAP performance. Significant research effort has been focused on knowledge-aided (KA) STAP (Weiner et al., 1998; Guerci, 2002). KA STAP attempts to use auxiliary sources of information to improve the interference covariance estimate for the cell under test. For example, map data can be used to identify changes in terrain type, roadways, and other variations in the characteristics of the clutter in the reference cells. Preprocessing algorithms can then excise some cells from the covariance matrix estimation process in an attempt to provide a more homogeneous set of training data and an estimate of $\hat{\mathbf{S}}_{\mathbf{I}}$ more consistent with the actual covariance matrix $\mathbf{S}_{\mathbf{I}}$ in the cell under test. The high level architecture of one KA STAP system is shown in Fig. 9.24. The knowledge-aided preprocessing edits or modifies the data-cube to provide a modified datacube with more homogeneous statistics. Any of the conventional STAP algorithms can then be applied to cancel the now better-behaved clutter. In another example, digital terrain elevation data (DTED) maps can be used along with terrain type maps to predict clutter characteristics. In-band electromagnetic interference (EMI) can be identified using data on known emitters such as television stations and wireless services

that may corrupt portions of the radar frequency band. The EMI can then be removed by prefiltering of the data in the receiver.



FIGURE 9.24 Basic architecture of a knowledge-aided STAP processor. (After Guerci, 2002.)

9.7 Limitations to STAP

In <u>Chap. 5</u> it was seen that the performance of MTI is degraded by several factors, some internal to the radar system and some external to it. Internal factors include pulse-to-pulse variability in pulse amplitude, oscillator phase drifts, I/Q channel errors, and antenna scanning modulation. External factors are principally internal clutter motion and clutter heterogeneity.

All of these factors also degrade STAP performance, but additional complications exist due to the multi-phase center radar and the moving platform. For instance, mismatches will exist among the *N* channels of the receive array. These can be classified further as angle-independent mismatches, which are differences in the frequency responses $H_n(\Omega)$ of the channels, and angle-dependent mismatches. The latter arise from a variety of sources, including element placement errors, wideband dispersion, and mutual coupling of elements. Platform motion effects such as misalignment of the array face with the platform velocity vector due to platform crab angle create additional degradations. Because adaptive weights are not generally updated every CPI, changing geometry between the moving platform and stationary or moving jammers can cause a jammer to move out of the adapted filter null until the weights are updated again, a phenomenon referred to as the "stale weights"

problem.

All of these effects tend to increase the rank of the interference covariance matrix, a phenomenon called *interference subspace leakage* (ISL) because the expanded rank implies that the size of the subspace required to represent the interference is increased. The rank increase implies an increase in reference window size requirements. Many ISL effects can be modeled as *covariance matrix tapers* (CMTs) applied to the true covariance matrix. A CMT is a $P \times P$ matrix **G** that combines with the ideal interference covariance matrix in a Hadamard product to give a modified and higher-rank covariance matrix:

$$\tilde{\mathbf{S}}_{\mathbf{I}} = \mathbf{S}_{\mathbf{I}} \odot \mathbf{G} \tag{9.67}$$

G, in turn, can often be modeled as the Hadamard product of several component CMTs representing different ISL effects (Guerci, 2003),

$$\mathbf{G} = \mathbf{G}_0 \odot \mathbf{G}_1 \odot \cdots \odot \mathbf{G}_{\mathbf{G}-1} \tag{9.68}$$

for some number G of component effects. A one-dimensional example of this in adaptive beamforming defines an N-element CMT vector as samples of a sinc function (Mailloux, 1995; Zatman, 1995). Applying this CMT to the estimated covariance matrix widens the interference notch in the adapted pattern, providing increased immunity to the stale weight problems caused by changing jammer AOAs. As another example, in the space-time case it can be shown that an appropriate CMT for angle-independent channel mismatch is the weighted sum of a matrix of all ones and an identity matrix (Guerci, 2003). CMTs can also be used to represent the effects of internal clutter motion. A CMT based on the Billingsley model, discussed earlier, is obtained by sampling Eq. (9.45):

$$G(m, n) = s_c(|m-n|T)$$
(9.69)

where *T* is the PRI.

However defined, CMTs can be used in at least two ways. The first uses the CMT in Eq. (9.67) to improve the model of the interference covariance in standard STAP algorithms. The second develops new algorithms that take account of the covariance structure, including the CMT component, to provide improved performance. Examples are given in Guerci (2003).

As noted earlier, clutter heterogeneity is an especially significant concern in STAP because of the large reference window size often required. Losses due to clutter heterogeneity can range from an insignificant 0.1 dB to as much as 16 dB for realistic scenarios (Melvin, 2000). Approaches to combating these losses include reduced dimension and reduced rank techniques described previously to limit reference window size, and knowledge-assisted algorithms to improve the homogeneity of the data used to estimate the clutter covariance in the cell under test.

References

- Aalfs, D.,"Adaptive Digital Beamforming," Chap. 9 in W. L. Melvin and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Edison, NJ, 2013.
- Arakawa, M., and R. A. Bond, "Computational Characteristics of High Performance Embedded Algorithms," Chap. 5 in D. R. Martinez, R. A. Bond, and M. M. Vai (eds.), *High Performance Embedded Computing Handbook: A Systems Perspective*. CRC Press, Boca Raton, FL, 2008.
- Billingsley, J. B., Radar Clutter. Artech House, Norwood, MA, 2001.
- Boroson, D. M., "Sample Size Consideration in Adaptive Arrays," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-16, no. 4, pp. 446–451, Jul. 1980.
- Brennan, L. E., and F. M. Staudaher, "Subclutter Visibility Demonstration," Technical Report RL-TR-92-21, Adaptive Sensors, 1992.
- Carlson, B. D., "Covariance Matrix Estimation Errors and Diagonal Loading in Adaptive Arrays," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-24, no. 3, pp. 397–401, Jul. 1988.
- Chen, W. S., and I. S. Reed, "A New CFAR Detection Test for Radar," *Digital Signal Processing*, vol. 1, pp. 198–214. Academic Press, New York, 1991.
- Guerci, J. R., "Knowledge Aided Sensor Signal Processing and Expert Reasoning," *Proceedings of Knowledge Aided Sensor Signal Processing and Expert Reasoning* (KASSPER) Workshop, Apr. 2002.
- Guerci, J. R., *Space-Time Adaptive Processing for Radar*. Artech House, Norwood, MA, 2003.
- Harris, F. J., *Multirate Signal Processing for Communication Systems*. Prentice-Hall, New York, 2004.
- Kelly, E. J., "An Adaptive Detection Algorithm," *IEEE Aerospace & Electronic Systems Magazine*, vol. 28, no. 1, pp. 115–127, March 1986.
- Kelly, E. J., "Performance of an Adaptive Detection Algorithm: Rejection of Unwanted Signals," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-25, no. 2, pp. 122–133, Mar. 1989.
- Kim, Y. L., Su. U. Pillai, and J. R. Guerci, "Optimal Loading Factor for Minimal Sample Support Space-Time Radar," *Proceedings* of 1998 *IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), vol. 4, pp. 2505–2508, May 1998.
- Klemm, R., Principles of Space-Time Adaptive Processing. Institution of

Electrical Engineers (IEE), London, 2002.

- Mailloux, R. J., "Covariance Matrix Augmentation to Produce Adaptive Array Pattern Troughs," *Electronics Letters*, vol. 31, no. 10, pp. 771–772, 1995.
- Melvin, W. L., "Space-Time Adaptive Radar Performance in Heterogeneous Clutter," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-36, no. 2, pp. 621–633, Apr. 2000.
- Melvin, W. L., "Clutter Suppression using Space-Time Adaptive Processing," Chap. 10 in W. L. Melvin and J. A. Scheer (eds.), *Principles of Modern Radar: Advanced Techniques*. SciTech Publishing, Edison, NJ, 2013.
- Mountcastle, P. D., "New Implementation of the Billingsley Clutter Model for GMTI Data Cube Generation," *Proceedings of IEEE 2004 Radar Conference*, pp. 398–401, Apr. 2004.
- Nitzberg, R., "Detection Loss of the Sample Matrix Inversion Technique," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-26, no. 6, pp. 824–827, Nov. 1984.
- Reed, I. S., J. D. Mallet, and L. E. Brennan, "Rapid Convergence Rate in Adaptive Arrays," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-10, no. 16, pp. 853–863, Nov. 1974.
- Robey, F. C., et al,. "A CFAR Adaptive Matched Filter Detector," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-28, no. 1, pp. 208–216, Jan. 1992.
- Steinhardt, A., and J. Guerci, "STAP for RADAR: What Works, What Doesn't, and What's in Store," *Proceedings of IEEE Radar Conference*, pp. 469–473, April 2004.
- Van Trees, H. L., *Optimum Array Processing: Part IV of Detection*, *Estimation, and Modulation Theory*. Wiley, New York, 2002.
- Ward, J., "Space-Time Adaptive Processing for Airborne Radar," Technical Report 1015, Massachusetts Institute of Technology, Lincoln Laboratory, Dec. 13, 1994.
- Weiner, D. D., G. T. Capraro, C. T. Capraro, G. B. Berdan, and M. C. Wicks, "An Approach for Using Known Terrain and Land Feature Data in Estimation of the Clutter Covariance Matrix," *Proceedings of* IEEE 1998 National Radar Conference, pp. 381–386, Dallas, TX, May 1998.
- Zatman, M., "Production of Adaptive Array Troughs through Dispersion Synthesis," *Electronics Letters*, vol. 31, no. 25, p. 2141, Dec. 1995.

Problems

1. Equation (9.8) gave the beamformer response as a function of AOA θ when the array is steered to angle θ_0 . Let $d = \alpha \lambda/2$ for some integer α . Show that the magnitude of the pattern $z(\theta)$ will have peaks at α different AOAs θ_0 in the interval $(-\pi/2, \pi/2]$.

- 2. Verify that the pre-beamforming SIR in the example of Fig. 9.4 is -50.04 dB.
- 3. Consider an adaptive beamformer with N = 2 phase centers in an interference environment consisting of noise and one jammer. The noise and jammer powers are σ_n^2 and σ_l^2 . The jammer AOA is θ_J radians. Define the noise-to-jammer ratio as $NJR = \sigma_n^2/\sigma_l^2$. By factoring out the jammer power, write the total interference covariance matrix in terms of σ_l^2 , NJR, and θ_J .
- 4. For the beamformer and interference environment in the previous problem, compute the optimum matched filter vector **h** as a function of the target angle θ_t . All overall scale factors, for instance those resulting from matrix inversion, can be ignored. What is the form of **h** as $NJR \rightarrow 0$ (i.e., the jammer dominates the interference)? Now let the target angle $\theta_t \rightarrow \theta_J$. What is the form of **h** in this case? Interpret the result.
- 5. Now assume that $NJR \rightarrow \infty$ (noise limited environment) in Problem 4. Determine the limiting form of **h** for this case. Interpret the result.
- 6. Suppose the interference covariance matrix for a second-order (N=2) adaptive beamformer is

$$\mathbf{S}_{\mathbf{I}} = \begin{bmatrix} 17 & j8 \\ -j8 & 17 \end{bmatrix}$$

Find the matrix square root V such that $S_I = V \cdot V$. *Hint*: At least one valid V is of the form:

v	а	e ^{jθ}
v =	$e^{-j\theta}$	а

Use this information to find the values of *a* and θ . (Systematic ways to find **V** without assuming a particular structure for **V** use the eigenvalues and eigenvectors of **S**_I or its Cholesky decomposition and are not covered here.)

- 7. Compute \mathbf{S}_{I} using the first form of the second line of Eq. (9.21) and show that it is an identity matrix to verify that applying the transformation V whitens the interference covariance matrix \mathbf{S}_{I} of the previous problem. MATLAB® or a similar computational tool can be used for the calculations.
- 8. Sketch the STAP clutter ridge in (k_{θ}, ω_D) coordinates assuming the following system parameters: v = 150 m/s, T = 1 ms, d = 0.3 m, and $F_0 = 1$
GHz. What is the value of the clutter ridge slope β in this case? Repeat for d = 0.15 m and T = 0.25 ms. Plot the ridge for the range of k_{θ} corresponding to AOAs of $-\pi$ to $+\pi$.

- 9. Write out explicitly the target steering vector **t** of Eq. (9.32) for a small case with M = 3 pulses and N = 2 phase centers. The target vector should be steered to an angle of $\theta_t = 30^\circ$ and $f_D = 0.25$ cycles per sample. Assume $d = \lambda/2$.
- 10. Consider a small STAP radar system operating at X band (10 GHz) with M = 4 slowtime samples in the CPI and N = 2 phase centers. The goal is to compute the optimum matched filter coefficients for detecting a target directly on boresight of the antenna ($\theta_t = 0^\circ$) and having a normalized Doppler shift of $f_{Dt} = 0.25$ cycles per sample.
 - a. What is the temporal steering vector for the target, \mathbf{a}_t ? Simplify the resulting expression (e.g., $e^{j\pi} = -1$).
 - b. What is the spatial steering vector for the target, \mathbf{a}_s ? Simplify.
 - c. What is the complete target model vector **t**?
- 11. For the STAP system in the previous problem, assume the interference is white noise with power σ_w^2 only (no clutter, no jammers). What will be the optimum beamforming filter coefficient vector **h**? Do not simplify or combine constants.
- 12. According to the RMB rule, how many range bins should be used in the STAP system of Prob. 10 to estimate the interference covariance matrix S_{I}^{-1} so that the mismatch loss is less than 1 dB? Use the exact formula of Eq. (9.62). How does the exact result compare to the rule-of-thumb estimate?
- 13. Verify that for a uniform linear array and constant PRI, computing the adapted pattern of a STAP processor over a uniform grid in f_D and k_θ is equivalent to computing the magnitude squared of the two-dimensional DFT of the outer product of the weight vector when it is remapped to a two-dimensional function.
- 14. Using MATLAB® or another computational tool, compute the adapted pattern for the case of noise interference only and the target parameters from Prob. 10. Let $\kappa = 1$ and assume no windows are used for sidelobe control in either Doppler or angle. Use a two-dimensional DFT of size 30 \times 30 or larger to get reasonable detail in the pattern.
- 15. Consider the optimum weight vector for the jammer cancellation example of Prob. 4 for the general case (*NJR* neither zero nor infinite) as a function of *NJR* when $\theta_t = \theta_j$, i.e. at the null. What is the expected value of the output power $|z|^2$ when the input is a jammer signal from angle θ_j ? The jammer signal can be modeled using Eq. (9.13) or (9.14).

16. Recompute $E\{|z|^2\}$ for the conditions of Prob. 15, but with a diagonal loading matrix of $\varepsilon \cdot \sigma_n^{21}$ added to the covariance matrix. Is the output power from the jammer input increased, unchanged, or decreased?

 $\frac{4}{2}$ These results generalize readily to non-side-looking cases, but for simplicity only the sidelooking case is considered in this chapter.

 $\frac{5}{2}$ Note that the DFT will compute the pattern in terms of \hat{k}_{θ} , not k_{θ} .

 $\frac{6}{5}$ Figures 9.13 through 9.15 were generated using LL_STAP[©], a MATLAB[®] program for demonstrating basic STAP patterns developed by Massachusetts Institute of Technology Lincoln laboratory (MIT/LL)

² The null would follow a straight line in $(f_D, \sin\theta)$ space.

 $\frac{8}{2}$ If desired, guard cells can be included immediately adjacent to the CUT to prevent target contamination of the interference estimate.

¹ Most STAP literature defines the filter according to $z = \mathbf{h}^H \mathbf{y}$. The convention $z = \mathbf{h}^T \mathbf{y}$ is retained here for consistency with the discussion of vector matched filtering in <u>Chap. 5</u> and the usual definition of convolution in DSP literature. A consequence is that the forms for **h** obtained here are the conjugate of the results in common STAP literature.

 $[\]frac{2}{2}$ Not to be confused with the variable *u* used for cross-range platform position in SAR in <u>Chap. 8</u>.

 $[\]frac{3}{2}$ Effective jamming therefore requires knowledge of the victim radar frequency and bandwidth. Unless the jammer has a priori knowledge of the precise portion of the band used by the radar or can estimate it by detecting and analyzing the radar signal, it must spread its energy over a bandwidth wider than that actually used by the radar.

APPENDIX A

Selected Topics in Probability and Random Processes

A-1.A Probability Density Functions and Likelihood Functions

The reader is assumed to be familiar with the concept of the probability density function (PDF) $p_x(x)$ for a continuous random variable x. In brief, the PDF describes the range of values x may possibly take on and the probability that x falls in a particular interval. Specifically, the probability that x falls in the interval x_1 to x_2 is computed as

$$\Pr\{x_1 \le x \le x_2\} = \int_{x_1}^{x_2} p_x(x) dx$$
(A.1)

Setting $x_1 = x_0 - \Delta x/2$ and $x_2 = x_0 + \Delta x/2$ and taking the limit as $\Delta x \rightarrow 0$ shows that the probability that *x* falls in a narrow interval around the value x_0 is

$$\Pr_{\Delta x \to 0} \left\{ x_0 - \frac{\Delta x}{2} \le x \le x_0 + \frac{\Delta x}{2} \right\} \approx p_x(x_0) \cdot \Delta x$$
(A.2)

This equation shows that the probability that $x \approx x_0$ is proportional to the PDF evaluated at x_0 . Thus the PDF tells us not only the range of x but also the relative probability of observing a measurement of the random variable near a given value of x. For example, a uniform PDF over the interval [0,1] dictates that the random variable described by that PDF will never take on a value less than zero or greater than one, but that a value in any small interval in between those limits is equally probable. On the other hand, a zero-mean, unit-variance Gaussian PDF indicates that the random variable so described could take on any real value whatsoever. However, values near zero are much more probable than values greater than +3 or less than -3. The symmetry of the PDF also implies that positive and negative values are equally probable.

Another specific example of Eq. (A.1) of importance in radar is the probability that *x* exceeds some threshold value *T*, a calculation that arises frequently in detection theory. This is the probability that *x* lies between *T* and $+\infty$,

$$\Pr\{x > T\} = \int_{T}^{\infty} p_x(x) dx$$
(A.3)

and is often described as a "right-tail probability" because it is the area under the right-hand tail of the PDF, from *T* to $+\infty$.

A possibly less familiar use of the PDF is as a *likelihood function*, an interpretation important in estimation theory. Consider a Gaussian random variable x with mean A and variance σ^2 , $x \sim N(x; A, \sigma^2)$.¹ The PDF is

$$p_{x}(x) = \frac{1}{\sqrt{2\pi\sigma^{2}}} \exp[-(x-A)^{2}/2\sigma^{2}]$$
(A.4)

x might model a noisy measurement of a fixed value *A*. Suppose that Eq. (A.4) is known to be an appropriate model for the PDF of *x*, and that the parameter σ^2 is known but *A* is not. Suppose it is desired to estimate *A* from a measurement of *x*. Consider a plot of the PDF *as a function of A* for a given known value of σ^2 and an observed value of *x*, e.g., x = 3. Denoting this "likelihood function" as $\ell(A|\sigma^2, x = 3)$ gives

$$\ell(A \mid \sigma^2, x = 3) = p_x(3) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(3-A)^2/2\sigma^2]$$
(A.5)

Figure A.1*a* plots $\ell(A|\sigma^2, x = 3)$ for $\sigma^2 = 1$. Again, note this is a function of *A*, not of *x*; the value of *x* is fixed by the measurement that has been made. This plot indicates that the most likely value of *A*, given that x = 3, is A = 3. Figure A.1*b* shows why this is the case. It plots $p_x(x)$, now as a function of *x*, for two different values of *A*. When A = 3, the probability that the measured value of *x* will be in a small neighborhood of width Δx around the value 3 is, from Eq. (A.2), approximately $p_x(3) \cdot \Delta x = 0.399 \cdot \Delta x$. For any other value of *A* that probability that *x* is close to 3 will be $p_x(3) \cdot \Delta x = 0.004432 \cdot \Delta x$, nearly two orders of magnitude lower. Clearly, the estimate of *A* that will produce the highest probability of the observation x = 3 is, for this simple problem, A = 3, the peak of the likelihood function. The estimate of *A* obtained by choosing the value that maximizes the likelihood function for a given observed *x* is called the *maximum likelihood estimate*.



FIGURE A.1 (*a*) The likelihood function of Eq. (A.5) when the observed x = 3 and $\sigma^2 = 1$. (*b*) The PDF of *x* when A = 3 and A = 5.

Any monotonically increasing transformation can be applied to the likelihood function without changing the value of *A* which maximizes it. In many problems of interest, the log-likelihood function $\ln[\ell(\cdot|\cdot)]$ is convenient to work with. Continuing the above example, the log-likelihood function is

$$\ln\left[\ell\left(A\left|\sigma^{2}, x=3\right)\right] = \frac{-(3-A)^{2}}{2\sigma^{2}} - \frac{1}{2}\ln(2\pi\sigma^{2})$$
(A.6)

Differentiating Eq. (A.6) with respect to A and setting the result equal to zero quickly gives A = 3 as the maximum likelihood estimate again.

A-2.A Important Probability Distributions in Radar

Many probability distributions have been used in modeling radar target and interference signals, and their number and sophistication is constantly increasing in attempts to better model observed phenomena. This section provides basic information on some of the most commonly encountered distributions in radar signal processing. The focus is PDFs most commonly used to represent the signal power (magnitude-squared), voltage or amplitude (magnitude), and phase of complex-valued radar receiver output signals. The corresponding characteristic functions, useful in calculating the distributions of sums of random variables, are given as well in several cases. The use of these distributions in radar signal modeling and the many extensions and modifications made for that purpose are discussed in the appropriate chapters. Most of the results in this section can be found in Papoulis and Pillai (2001) and Omura and Kailath (1965).

A.2.1 Power Distributions

Chi-Squared, Exponential, Erlang, and Gamma

Chi-squared distributions have a long history in modeling the power of radar target and interference signals. For example, versions of the chi-squared PDF are used for the Swerling target models and the standard model of white noise. The Erlang PDF is a generalization of the chi-squared PDF, and the gamma PDF is a generalization of both. The exponential PDF is a special case of the chi-squared and thus of the gamma and Erlang as well. This section describes the relationship between these four PDFs.

Consider a random variable x with mean and variance \bar{x} and σ_x^2 . The common definition of the chi-squared PDF with N degrees of freedom (DOF) (also called the chi-squared of duo-degree N) is

$$p_x(x) = \chi^2(x; N) = \begin{cases} \frac{x^{N/2 - 1}}{2^{N/2} \Gamma\left(\frac{N}{2}\right)} \exp(-x/2), & x \ge 0\\ 0, & x < 0 \end{cases}$$

(A.7)

where the notation $\chi^2(x; N)$ has been defined to indicate that the random variable *x* has a χ^2 PDF with parameter *N*. The mean and variance of *x* are

$$\overline{x} = N, \quad \sigma_x^2 = 2N \tag{A.8}$$

The *N*-DOF chi-squared PDF arises in radar as the PDF of a random variable x that is the sum of the squares of N independent zero-mean, unit-variance Gaussian random variables:

$$x = \sum_{i=1}^{N} x_i^2 \tag{A.9}$$

where the x_i are i.i.d. and $x_i \sim N(x; 0,1)$. Of particular interest are the cases N = 2 and N = 4 because of their use in the Swerling target RCS fluctuation models (see <u>Chap. 2</u>).

It is more useful to consider a slight generalization that allows for an arbitrary mean. Recall that for a general random variable x, the PDF, mean, and variance of the new random variable y obtained by the linear transformation y = ax + b are related to those of x according to

$$p_{y}(y) = \frac{1}{|a|} p_{x} \left(\frac{y-b}{a} \right)$$

$$\overline{y} = a\overline{x} + b$$

$$\sigma_{y}^{2} = |a|^{2} \sigma_{x}^{2}$$

(A.10)

If all the x_i in Eq. (A.9) have variance σ^2 instead of 1, the new sum-of-squares variable $x' = \sigma^2 x$ so that the PDF of x' is (dropping the "prime" in the notation)

$$p_{x}(x) = \chi^{2}(x; N, \sigma^{2}) = \begin{cases} \frac{x^{N/2 - 1}}{(2\sigma^{2})^{N/2} \Gamma\left(\frac{N}{2}\right)} \exp(-x/2\sigma^{2}), & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.11)

with mean and variance

$$\overline{x} = N\sigma^2, \quad \sigma_x^2 = 2N\sigma^4 \tag{A.12}$$

This generalized chi-squared PDF in Eq. (A.11) is a function of two parameters (*N* and σ^2) rather than just one (*N*).

Figure A.2*a* illustrates the chi-squared PDF for various values of *N*. In this plot, the variance of the underlying Gaussian random variables has been set to $\sigma^2 = 1/N$ so that all the PDFs have a common mean $\bar{x} = 1$. Notice that as *N* increases, the PDF becomes more Gaussian in shape, as would be expected for the sum of a large number of random variables of any distribution.



FIGURE A.2 The central and noncentral chi-squared families of PDFs: (*a*) central chi-squared PDFs with unit mean and varying N, (*b*) noncentral chi-squared PDFs with $\sigma^2 = 1$, N = 10, and varying noncentrality parameters λ .

The characteristic function (CF) of a PDF $p_x(x)$ is defined as

$$C(q) = \int_{-\infty}^{\infty} p_x(x) e^{jqx} dx$$
(A.13)

The CF is essentially the Fourier transform of the PDF, though with the sign of the complex exponential chosen opposite from the definition commonly used in electrical engineering texts. The CF corresponding to the generalized chisquared PDF is

$$C_{\chi^2}(q) = \frac{1}{(1 - j2\sigma^2 q)^{N/2}}$$
(A.14)

The exponential PDF with mean γ , Exp($x; \gamma$), is

$$p_x(x) = \operatorname{Exp}(x;\gamma) = \begin{cases} \frac{1}{\gamma} e^{-x/\gamma}, & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.15)

The mean and variance of x are

$$\overline{x} = \gamma, \quad \sigma_x^2 = \gamma^2 \tag{A.16}$$

Its characteristic function is

$$C_{\rm Exp}(q) = \frac{1}{1 - j\gamma q} \tag{A.17}$$

The exponential PDF is a special case of the generalized chi-squared with N = 2. This choice reduces it to a single-parameter distribution. In order that the mean of the exponential variate be γ when N = 2, it is necessary that $\gamma = 2\sigma^2$. Thus, $\text{Exp}(x;\gamma) = \chi^2(x; 2, \gamma/2)$.

In radar, the exponential PDF most often arises in connection with the standard model of noise in a coherent radar receiver. If the I and Q channel noise processes are assumed to be i.i.d. N(x; 0, σ^2), the magnitude-squared of the complex noise (noise power) is the sum of the I and Q voltages squared and so will be distributed as $\chi^2(x; 2, \sigma^2)$. The Swerling 1 and 2 target fluctuation models also use the exponential PDF to describe the RCS fluctuations of a target having many scatterers with none dominant, as discussed in <u>Chap. 2</u>.

The gamma PDF $\Gamma(x; \alpha, \beta)$ is a more general distribution that includes the

chi-squared and exponential as special cases, but can fit a wider range of experimental data. It is

$$p_{x}(x) = \Gamma(x;\alpha,\beta) = \begin{cases} \frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \exp(-x/\beta), & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.18)

where $\Gamma(\cdot)$ is the usual gamma function. The mean and variance of x are

$$\overline{x} = \alpha \beta, \quad \sigma_x^2 = \alpha \beta^2$$
 (A.19)

The CF corresponding to the gamma PDF is

$$C_{\Gamma}(q) = \frac{1}{\left(1 - j\beta q\right)^{\alpha}} \tag{A.20}$$

The generalized chi-squared PDF is related to the gamma PDF as $\chi^2(x; N, \sigma^2) = \Gamma(x; N/2, 2\sigma^2)$. Since $\text{Exp}(x; \gamma) = \chi^2(x; 2, \gamma/2)$, it follows also that $\text{Exp}(x; \gamma) = \Gamma(x; 1, \gamma)$.

The Erlang or Erlang-*k* PDF is sometimes mentioned in radar signal modeling. It lies between the gamma and generalized chi-squared models in that it restricts the parameter α of $\Gamma(x; \alpha, \beta)$ to an integer value *k*. The parameter β is arbitrary but is usually expressed in the form $1/\lambda k$ for integer *k* and some λ . With this choice, the Erlang PDF E($x; k, \lambda$) = $\Gamma(x; k, 1/\lambda k)$ is

$$p_{x}(x) = \mathbf{E}(x;k,\lambda) = \begin{cases} \frac{(\lambda k)^{k} x^{k-1}}{\Gamma(k)} \exp(-\lambda kx), & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.21)

Recall that $\Gamma(k) = (k-1)!$ for integer k. The mean and variance of x are

$$\overline{x} = \frac{1}{\lambda}, \quad \sigma_x^2 = \frac{1}{k\lambda^2}$$
(A.22)

and the characteristic function is

$$C_{\rm E}(q) = \frac{1}{\left(1 - jq/\lambda k\right)^k}$$

(A.23)

Noncentral Chi-Squared

The $\chi^2(x; N, \sigma^2)$ distribution is also called the central chi-squared with *N* degrees of freedom. The traditional *non*central chi-squared PDF with *N* degrees of freedom, $\chi^2_{nc}(x;N,\lambda)$, describes the sum of squares of normal random variables with different, nonzero means but identical unit variances:

$$x = \sum_{i=1}^{N} x_i^2$$
 (A.24)

with the $x_i \sim N(x; \mu_i, 1)$. The resulting PDF is

$$p_{x}(x) = \chi_{\mathrm{nc}}^{2}(x; N, \lambda) = \begin{cases} \frac{1}{2} \left(\frac{x}{\lambda}\right)^{\frac{N-2}{4}} e^{-(x+\lambda)/2} \cdot I_{\frac{N}{2}-1}(\sqrt{\lambda x}), & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.25)

where the noncentrality parameter λ is defined as

$$\lambda = \sum_{i=1}^{N} \mu_i^2 \tag{A.26}$$

and is the magnitude-squared of the vector of means $\mu = [\mu_1 - \mu_N]^T$. $I_a(\cdot)$ is the modified Bessel function of the first kind and order α .

Again, it is useful to slightly generalize this to the case where the x_i all have the same non-unit variance σ^2 . This is done by scaling the x_i by the factor σ so that now $x_i \sim N(x; \sigma \mu_i, \sigma^2)$. The new sum variable and noncentrality parameter are

$$x' = \sum_{i=1}^{N} x_i'^2 = \sigma^2 \sum_{i=1}^{N} x_i^2 = \sigma^2 x, \quad \lambda' = \sum_{i=1}^{N} (\sigma \mu_i)^2 = \sigma^2 \sum_{i=1}^{N} \mu_i^2 = \sigma^2 \lambda$$
(A.27)

Applying these scalings to Eq. (A.25) and again dropping the primes from the notation, the generalized noncentral chi-squared PDF is

$$p_{x}(x) = \chi_{\mathrm{nc}}^{2}(x; N, \lambda, \sigma^{2}) = \begin{cases} \frac{1}{2\sigma^{2}} \left(\frac{x}{\lambda}\right)^{\frac{N-2}{4}} e^{-(x+\lambda)/2\sigma^{2}} \cdot I_{\frac{N}{2}-1} \left(\frac{\sqrt{\lambda x}}{\sigma^{2}}\right), & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.28)

with mean and variance

$$\overline{x} = \sigma^2 N + \lambda, \quad \sigma_x^2 = 2N\sigma^4 + 4\sigma^2\lambda \tag{A.29}$$

and characteristic function

$$C_{\chi^{2}_{R}}(q) = \frac{1}{(1 - j2\sigma^{2}q)^{N/2}} \exp\left(\frac{j2\lambda\sigma^{2}q}{1 - j2\sigma^{2}q}\right)$$
(A.30)

This PDF reduces to the (central) chi-squared when $\lambda = 0$ and approaches a Gaussian as $N \rightarrow \infty$. Figure A.2b illustrates the noncentral chi-squared PDF for fixed values N = 10 and $\sigma^2 = 1$ and various values of λ . The nonzero mean of the noncentral chi-squared PDF allows modeling of signal statistics when the underlying complex data includes a persistent component.

Weibull and Log-Normal

The Weibull and log-normal distributions are used when longer-"tailed" PDFs are needed, implying that the phenomenon being modeled has a higher rate of occurrence of large values than predicted using the distributions above. Such data is often referred to as being "spiky." They have become popular for modeling land and sea clutter returns at shallow grazing angles, especially at finer resolutions and higher radar frequencies. They are sometimes used to model fine resolution, high frequency target echoes as well.

One common form of the Weibull PDF is

$$p_{x}(x) = W(x;\alpha,\beta) = \begin{cases} \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}}, & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.31)

The mean, median x_m , and variance of x are

$$\overline{x} = \beta \Gamma \left(1 + \frac{1}{\alpha} \right), \quad x_m = \beta (\ln 2)^{1/\alpha}, \quad \sigma_x^2 = \beta^2 \left\{ \Gamma \left(1 + \frac{2}{\alpha} \right) - \Gamma^2 \left(1 + \frac{1}{\alpha} \right) \right\}$$

The Weibull characteristic function is

$$C_{\rm W}(q) = \sum_{n=0}^{\infty} \frac{(j\beta q)^n}{n!} \Gamma\left(1 + \frac{n}{\alpha}\right)$$
(A.33)

The parameter α is called the "shape parameter," while β is the "scale parameter." As α varies from 1 to 2, the Weibull varies from an exponential PDF to a Rayleigh PDF. However, α is not restricted to that range; it can take on any non-negative value. Figure A.3*a* illustrates the Weibull PDF with $x_m = 1$ and several values of the shape parameter.



FIGURE A.3 The "long-tailed" Weibull and log-normal families of PDFs: (*a*) Weibull PDFs with a unit median and varying shape parameters, (*b*) log-normal PDFs with unit median and varying values of β^2 .

The log-normal PDF describes a random variable whose logarithm (in any base) is normally distributed:

$$p_x(x) = \text{LN}(x;\alpha,\beta) = \frac{1}{\sqrt{2\pi} \cdot \beta x} \exp\left[-\frac{(\ln x - \alpha)^2}{2\beta^2}\right]$$
(A.34)

Its mean, median, and variance are

$$\overline{x} = e^{\alpha + \beta^2/2}, \quad x_m = e^{\alpha}, \quad \sigma_x^2 = (e^{\beta^2} - 1)e^{2\alpha + \beta^2}$$
(A.35)

<u>Figure A.3*b*</u> illustrates the log-normal PDF with $x_m = 1$ and several values of β^2 .

Power Distribution for K-Distributed Amplitude

The K distribution is also used for modeling spiky clutter and targets. It is generally discussed in terms of amplitude (voltage) rather than power in radar literature, and so the main definition and discussion are deferred to the next section. Since power x is the square of amplitude y, and y is nonnegative, applying the transformation $p_x(x) = p_y(\sqrt{x})/\sqrt{x}$ to Eq. (A.49) gives the following PDF for the power of a random variable with a K-distributed amplitude:

$$p_{x}(x) = \begin{cases} \frac{2c^{a+1}}{\Gamma(a)} x^{(a-1)/2} K_{a-1}(2c\sqrt{x}), & x \ge 0\\ 0, & x < 0 \end{cases}$$
(A.36)

The shape and scale parameters a and c are described with the K distribution in the next section.

A.2.2 Voltage Distributions

Given a random variable x representing the magnitude-squared and thus the power of a complex-valued voltage, the quantity $y = \sqrt{x}$ represents the amplitude (magnitude) of the complex voltage. The actual complex voltage cannot be obtained from the power without also knowing the signal phase. Nonetheless, the amplitude alone is often of interest, for instance when using a linear detector instead of a square-law detector in a radar receiver. Using standard results from random variable theory and noting that the power x is always non-negative, the PDF of y can be obtained from the PDF of x with the transformation

$$p_y(y) = 2y \cdot p_x(y^2) \tag{A.37}$$

Rayleigh

Applying this result, a signal whose power is described by the exponential PDF has an amplitude described by the Rayleigh PDF:

$$p_{y}(y) = \mathbf{R}(y;\gamma) = 2y \cdot \mathbf{Exp}(y^{2};\gamma) = \begin{cases} \frac{2y}{\gamma} e^{-y^{2}/\gamma}, & y \ge 0\\ 0, & y < 0 \end{cases}$$
(A.38)

The mean and variance of *y* are

$$\overline{y} = \frac{1}{2}\sqrt{\pi\gamma}, \quad \sigma_y^2 = \left(1 - \frac{\pi}{4}\right)\gamma$$
(A.39)

Its characteristic function is

$$C_{\rm R}(q) = \left(1 + j \frac{\sqrt{\pi\gamma}}{2} q\right) e^{-\gamma q^2/4} \tag{A.40}$$

Central Chi

As with the exponential and chi-squared distributions, the generalizations of the Rayleigh amplitude distribution are the central and noncentral chi distributions. Using the transformation Eq. (A.37) in Eq. (A.11) gives the central chi distribution:

$$p_{y}(y) = \chi(y; N, \sigma^{2}) = \begin{cases} \frac{2y^{N-1}}{(2\sigma^{2})^{N/2}} \Gamma\left(\frac{N}{2}\right)^{\exp(-y^{2}/2\sigma^{2})}, & y \ge 0\\ 0, & y < 0 \end{cases}$$
(A.41)

The mean, mean-square, and variance are

$$\overline{y} = \sqrt{2\sigma^2} \cdot \frac{\Gamma[(N+1)/2]}{\Gamma(N/2)}$$
$$\overline{y^2} = 2\sigma^2 \frac{\Gamma(N/2+1)}{\Gamma(N/2)} = N\sigma^2$$
$$\sigma_y^2 = \overline{y^2} - \overline{y}^2$$
(A.42)

Equation (A.41) retains the parameters N and σ^2 that emphasize the connection to the amplitude of the sum of N squared $N(x; 0, \sigma^2)$ variates. As will be seen in the discussion of the K distribution, the chi PDF is also used to empirically model variations of the local mean of a nonstationary Rayleigh process. In this case, there is no particular connection to the sum of squared Gaussians so the PDF is better expressed in the more neutral form

$$p_{y}(y) = \chi(y; a, b) = \begin{cases} \frac{2b^{2a}}{\Gamma(a)} y^{2a-1} \exp(-b^{2}y^{2}), & y \ge 0\\ 0, & y < 0 \end{cases}$$
(A.43)

where the substitutions $b = \sqrt{1/2\sigma^2}$ and a = N/2 have been made. In this form, *a* is the shape parameter and *b* is the scale parameter. The first and second moments are

$$\overline{y} = \frac{1}{b} \cdot \frac{\Gamma(a+1/2)}{\Gamma(a)}, \quad \overline{y^2} = \frac{1}{b^2} \frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{b^2}, \quad \sigma_y^2 = \overline{y^2} - \overline{y}^2$$
(A.44)

Noncentral Chi

The noncentral chi distribution is the result of applying Eq. (A.37) to the noncentral chisquared PDF, and thus represents the PDF of the magnitude (as opposed to the magnitude-squared) of the sum of *N* random variables distributed as $N(x;\sigma\mu_i,\sigma^2)$. Defining the magnitude of the vector of means as $v = \sqrt{\lambda}$, the PDF is

$$p_{y}(y) = \chi_{nc}(y; N, v, \sigma^{2}) = \begin{cases} \frac{v}{\sigma^{2}} \left(\frac{y}{v}\right)^{\frac{N}{2}} e^{-(y^{2}+v^{2})/2\sigma^{2}} \cdot I_{\frac{N}{2}-1} \left(\frac{vy}{\sigma^{2}}\right), & y \ge 0\\ 0, & y < 0 \end{cases}$$
(A.45)

The moments of the noncentral chi PDF do not take a simple form. The mean, mean-square, and variance are

$$\begin{split} \overline{y} &= \sqrt{2\sigma^2} \cdot e^{-v^2/2\sigma^2} \frac{\Gamma[(N+1)/2]}{\Gamma(N/2)} \, _1F_1\!\left(\frac{N+1}{2}, \frac{N}{2}, \frac{v^2}{2\sigma^2}\right) \\ \overline{y^2} &= 2\sigma^2 e^{-v^2/2\sigma^2} \frac{\Gamma(N/2+1)}{\Gamma(N/2)} \, _1F_1\!\left(\frac{N}{2}\!+\!1, \frac{N}{2}, \frac{v^2}{2\sigma^2}\right) = N\sigma^2 e^{-v^2/2\sigma^2} \, _1F_1\!\left(\frac{N}{2}\!+\!1, \frac{N}{2}, \frac{v^2}{2\sigma^2}\right) \\ \sigma_y^2 &= \overline{y^2} - \overline{y}^2 \end{split}$$

(A.46)

where $_1F_1(\cdot,\cdot,\cdot)$ is the confluent hypergeometric function² and the gamma function recurrence relation $\Gamma(x + 1) = x\Gamma(x)$ was used to get the second form of $\overline{y^2}$. The noncentral chi PDF reduces to the Rayleigh PDF when v = 0, $\gamma = 2\sigma^2$, and N = 2.

Rice

The Rice (also called "Rician") distribution is the amplitude distribution that corresponds to the generalized noncentral chi-squared power distribution with N = 2, i.e., it is the noncentral chi with N = 2. The PDF becomes

$$p_{y}(y) = \operatorname{Ri}(y; N, v, \sigma^{2}) = \chi_{\operatorname{nc}}(y; 2, v, \sigma^{2}) = \begin{cases} \frac{y}{\sigma^{2}} e^{-(y^{2} + v^{2})/2\sigma^{2}} \cdot I_{0}\left(\frac{vy}{\sigma^{2}}\right), & y \ge 0\\ 0, & y < 0 \end{cases}$$
(A.47)

The mean, mean-square, and variance are given by Eq. (A.46) with N = 2:

$$\overline{y} = \sqrt{2\sigma^2} \cdot e^{-v^2/2\sigma^2} \Gamma\left(\frac{3}{2}\right) {}_1F_1\left(\frac{3}{2}, 1, \frac{v^2}{2\sigma^2}\right) = \sqrt{\frac{\pi\sigma^2}{2}} \cdot e^{-v^2/2\sigma^2} {}_1F_1\left(\frac{3}{2}, 1, \frac{v^2}{2\sigma^2}\right)$$

$$\overline{y^2} = 2\sigma^2 e^{-v^2/2\sigma^2} {}_1F_1\left(2, 1, \frac{v^2}{2\sigma^2}\right)$$

$$\sigma_y^2 = \overline{y^2} - \overline{y}^2$$
(A.48)

The relations $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(x + 1) = x\Gamma(x)$ have been used to get the second form of \overline{y} in Eq. (A.48). The Rice PDF reduces to the Rayleigh PDF when v = 0.

Weibull and Log-Normal

Applying the transformation of Eq. (A.37) to the Weibull PDF of Eq. (A.31) shows that the amplitude PDF that corresponds to a Weibull power PDF is also Weibull but with the parameters modified to $\alpha' = 2\alpha$ and $\beta' = \sqrt{\beta}$. The earlier formulas for the mean, median, variance, and characteristic function can be used with these new parameters.

Similarly, applying Eq. (A.37) to Eq. (A.34) yields a log-normal distribution with new parameters $\alpha' = \alpha/2$ and $\beta' = \beta/2$ for the amplitude corresponding to a log-normal power variate. The mean, median, and variance of Eq. (A.35) may be used for the amplitude variate with these new values.

K Distribution

The K distribution is a relative newcomer to radar signal modeling. Like the log-normal and Weibull PDFs, it is most often used to model "spiky" ground or sea clutter returns, particularly for finer resolution radars. The rationale is that with these systems, the assumption underlying other PDFs of relatively homogeneous and stationary clutter scenes containing "many" scatterers per radar resolution cell becomes invalid; see Watts (1985).

The K distribution is a "compound PDF" composed from two more basic PDFs. Specifically, the signal amplitude is assumed to follow a Rayleigh PDF with mean z. However, z is itself modeled as a random variable described by a central chi distribution with shape and scale parameters a and b. The resulting PDF for the amplitude y is

$$p_{y}(y) = \int_{0}^{\infty} p_{y}(y|z) p_{z}(z) dz = \begin{cases} \frac{4c}{\Gamma(a)} (cy)^{a} K_{a-1}(2cy), & y \ge 0\\ 0, & y < 0 \end{cases}$$
(A.49)

where $K_{a-1}(\cdot)$ is the modified Bessel function of the second kind and order a - 1 and $c = b\sqrt{\pi/4}$. In this equation, $p_y(y|z)$ is a Rayleigh PDF with mean z and $p_z(z)$ is a central chi PDF. The first and second moments of the K distribution are

$$\overline{y} = \frac{\sqrt{\pi}}{2c} \cdot \frac{\Gamma(a+1/2)}{\Gamma(a)}, \quad \overline{y^2} = \frac{a}{c^2}, \quad \sigma_y^2 = \overline{y^2} - \overline{y}^2$$
(A.50)

A.2.3 The Unfortunate Tendency in Radar to Call Power Distributions by the Name of the Voltage Distribution

It is an unfortunate fact that it is common in the radar community to apply the correct name for a voltage PDF to the PDF for power that results when the voltage variable is squared. For example, a random variable representing radar cross section (RCS) and described by the exponential PDF may be referred to as a Rayleigh RCS. The reader must be careful to realize that such a reference probably means Rayleigh voltage, but exponential power. Similarly, a variable described by a noncentral chi-squared PDF with N = 2 may be referred to as a Rice variate.

Although not an error, another source of confusion arises when the PDF of both a voltage variable and its corresponding power variable have the same general form (though with different parameters). As seen above, this occurs with Weibull and log-normal variates. When discussing these PDFs, extra caution is needed to determine whether the variate in question is modeling voltage or power.

A.2.4 Phase Distributions

Uniform

The uniform PDF U(x; x_1 , x_2) describes a random variable that takes on any value in an ordered interval (x_1 , x_2) with equal likelihood. It is

$$p_{x}(x) = \mathbf{U}(x; x_{1}, x_{2}) = \begin{cases} \frac{1}{x_{2} - x_{1}}, & x_{1} \le x \le x_{2} \\ 0, & \text{otherwise} \end{cases}$$
(A.51)

The mean and variance of x are

$$\overline{x} = \frac{1}{2}(x_2 - x_1), \quad \sigma_x^2 = \frac{(x_2 - x_1)^2}{12}$$
(A.52)

Its characteristic function is

$$C_{\rm U}(q) = \frac{e^{jx_1q} - e^{-jx_2q}}{j(x_2 - x_1)q} = e^{j[(x_2 + x_1)/2]q} \operatorname{sinc}\left[\left(\frac{x_2 - x_1}{2}\right)q\right]$$
(A.53)

The most common use of the uniform PDF in radar is to describe a completely random phase that is uniformly distributed over $[0, 2\pi]$ radians (equivalently, $[-\pi, \pi]$ radians). Another common usage is to describe quantization error (see <u>App. B</u>).

Tikhonov

The Tikhonov PDF, also called the Von Mises PDF, is given in Van Trees (1968) as

$$p_{x}(x;\alpha) = \mathbf{T}(x;\alpha) = \begin{cases} \frac{1}{2\pi I_{0}(\alpha)} e^{\alpha \cos(x-\mu)}, & -\pi \le x \le \pi \\ 0, & \text{otherwise} \end{cases}$$
(A.54)

Most often, the moments of the complex variable $z = e^{jx}$ are considered rather than the moments of x itself. The moments of z are called the *circular moments* of x. They are

$$\overline{z} = \frac{I_1(\alpha)}{I_0(\alpha)} e^{j\mu}, \quad \sigma_z^2 = 1 - \frac{I_1(\alpha)}{I_0(\alpha)}$$

 $I_0(\cdot)$ and $I_1(\cdot)$ are the modified Bessel functions of the first kind and orders zero and one, respectively. The mean of *x* itself is the argument of \overline{z} :

$$\overline{x} = \mu \tag{A.56}$$

The Tikhonov distribution provides a family of PDFs suitable for modeling phase and phase errors that vary from the uniform PDF when $\alpha = 0$ to, in the limit, a nonrandom phase of μ as $\alpha \to \infty$. The PDF is defined over the interval $[-\pi, \pi]$ radians, and its integral over that interval is unity, so that it is a strictly valid PDF for phase. However, the same formula can be used over any interval of length 2π (e.g., 0 to 2π) because of the periodicity of the cosine function. Figure A.4 illustrates the Tikhonov PDF for $\mu = 0$ and several different values of the parameter α .



FIGURE A.4 The Tikhonov PDF for phase with $\mu = 0$.

A-3.A Estimators and the Cramèr-Rao Lower Bound

Consider an observed signal y(t) that is the sum of a target component s(t) and a noise component w(t):

$$y(t) = s(t) + w(t) \tag{A.57}$$

y(t) is a function of one or more parameters Θ_i . These might be, for example, the time delay, amplitude, Doppler shift, or angle of arrival of the target component. Once a target is detected, estimation of these parameters will usually be the next goal of the radar. In many cases, the estimated values may serve as inputs to tracking algorithms. Thus, the quality of these estimates is of great interest.

Suppose y(t) is sampled multiple times (intrapulse and/or over multiple pulses) to give a vector of N observations,

$$\mathbf{y} = \{y_1, y_2, \dots y_N\}$$
(A.58)

Now define an *estimator* f of a parameter Θ_i as some function or procedure that produces an estimated value $\hat{\Theta}_i$ from the data **y**,

$$\Theta_i = f(\mathbf{y}) \tag{A.59}$$

Because y is random, the estimate $\hat{\Theta}_i$ is itself a random variable and therefore has a probability density function with a mean and variance. The *accuracy* of the estimator is the error in the mean, $\mathbb{E}[\hat{\Theta}_i - \Theta_i]$, also known as the bias. The *precision* of the estimator is the standard deviation $\sigma_{\hat{\Theta}_i}$.

Two desirable properties of an estimator are that it be *unbiased* and *consistent*. These mean that

$$E\{\hat{\Theta}_i\} = \Theta_i \quad \text{(unbiased)}$$
$$\lim_{N \to \infty} \{\sigma^2_{\hat{\Theta}_i}\} \to 0 \quad \text{(consistent)}$$
(A.60)

In this text, only unbiased estimators will be considered, and so the issue becomes how large or small is the variance of the estimate and how it behaves versus the amount of data N. The derivation of the real, single-parameter CRLB in this section closely follows (Peebles, 1998). Most of the other results follow Kay (1993).

A.3.1 The Cramèr-Rao Lower Bound on Estimator Variance

The *Cramèr-Rao Lower Bound* (CRLB) is a famous and very important result that establishes the minimum variance of an unbiased estimator. Any particular unbiased estimator must then have a variance (square of precision) equal to or greater than the CRLB, and the quality of an unbiased estimator can be judged by how close its actual variance comes to achieving the CRLB. An estimator that is unbiased and achieves the CRLB is said to be *efficient*.

Consider a signal dependent on a single scalar parameter Θ . Denote the *N*-dimensional joint PDF of the observations y, given the actual value of Θ , as

 $p_{\mathbf{v}}(\mathbf{y}|\Theta)$. The assumption of an unbiased estimator requires

$$\mathbf{E}\{\hat{\Theta} - \Theta\} = \int \cdots \int (\hat{\Theta} - \Theta) p_{\mathbf{y}}(\mathbf{y}|\Theta) d\mathbf{y} = 0$$
(A.61)

Using Leibniz' integral rule to differentiate this equation with respect to Θ gives

$$\frac{\partial \mathbf{E}\{\hat{\Theta} - \Theta\}}{\partial \Theta} = \int \cdots \int (\hat{\Theta} - \Theta) \frac{\partial p(\mathbf{y}|\Theta)}{\partial \Theta} d\mathbf{y} - \underbrace{\int \cdots \int p_{\mathbf{y}}(\mathbf{y}|\Theta) d\mathbf{y}}_{=1} = 0$$
(A.62)

The second multiple integral equals one because it has just the PDF $p_y(\mathbf{y}|\Theta)$ as the integrand, and the integral of any valid PDF is one. Note also that for any function $g(\mathbf{y},\Theta)$

$$\frac{\partial}{\partial \Theta} \ln\{g(\mathbf{y}, \Theta)\} = \frac{1}{g(\mathbf{y}, \Theta)} \frac{\partial g(\mathbf{y}, \Theta)}{\partial \Theta} \implies \frac{\partial g(\mathbf{y}, \Theta)}{\partial \Theta} = \left[\frac{\partial}{\partial \Theta} \ln\{g(\mathbf{y}, \Theta)\}\right] g(\mathbf{y}, \Theta)$$
(A.63)

Using Eq. (A.63) in Eq. (A.62) gives

$$\int \cdots \int (\hat{\Theta} - \Theta) \frac{\partial \ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}}{\partial \Theta} p_{\mathbf{y}}(\mathbf{y}|\Theta) d\mathbf{y} = 1$$
(A.64)

This relationship is the consequence of the unbiased estimator assumption.

The Schwarz inequality in integral form can be applied to Eq. (A.64). The scalar version was used in discussing the matched filter in <u>Chap. 4</u>. The multivariable version is

$$\left\{\int \cdots \int A(\mathbf{y})B(\mathbf{y})d\mathbf{y}\right\}^2 \leq \int \cdots \int A^2(\mathbf{y})d\mathbf{y} \cdot \int \cdots \int B^2(\mathbf{y})d\mathbf{y}$$
(A.65)

with equality if and only if $A(\mathbf{y}) = \alpha B(\mathbf{y})$ for some scalar α . Choose

$$A(\mathbf{y}) = (\hat{\Theta} - \Theta) \sqrt{p_{\mathbf{y}}(\mathbf{y}|\Theta)}, \quad B(\mathbf{y}) = \frac{\partial \ln\{p(\mathbf{y}|\Theta)\}}{\partial \theta} \sqrt{p_{\mathbf{y}}(\mathbf{y}|\Theta)}$$
(A.66)

to get

$$\begin{split} \left\{ \int \cdots \int (\hat{\Theta} - \Theta) \frac{\partial \ln \{ p_{\mathbf{y}}(\mathbf{y}|\Theta) \}}{\partial \Theta} p_{\mathbf{y}}(\mathbf{y}|\Theta) d\mathbf{y} \right\}^{2} \leq \\ \int \cdots \int (\hat{\Theta} - \Theta)^{2} p_{\mathbf{y}}(\mathbf{y}|\Theta) d\mathbf{y} \cdot \int \cdots \int \left[\frac{\partial \ln \{ p_{\mathbf{y}}(\mathbf{y}|\Theta) \}}{\partial \Theta} \right]^{2} p_{\mathbf{y}}(\mathbf{y}|\Theta) d\mathbf{y} \end{split}$$
(A.67)

From Eq. (A.64), the left-hand side of Eq. (A.67) equals one. The first multiple integral on the right hand side is the estimator variance σ_{Θ}^2 , while the second multiple integral is, by definition, $\mathbf{E} \{\partial \ln \{p_y(\mathbf{y}|\Theta)\}/\partial\Theta\}^2$. Using these relationships and rearranging Eq. (A.67) then gives the CRLB:

$$\sigma_{\hat{\Theta}}^{2} \ge \frac{1}{\mathbb{E}[\{\partial \ln \{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}/\partial\Theta\}^{2}]}$$
(A.68)

An alternate form of the CRLB is also common. If $p_y(\mathbf{y}|\Theta)$ is twice differentiable and obeys some other mild regularity conditions, it can be shown that

$$\mathbf{E}\left[\left\{\frac{\partial \ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}}{\partial\Theta}\right\}^{2}\right] = -\mathbf{E}\left[\frac{\partial^{2}\ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}}{\partial\Theta^{2}}\right]$$
(A.69)

which gives another version of the CRLB:

$$\sigma_{\hat{\Theta}}^{2} \ge \frac{1}{-\mathbf{E}\left[\partial^{2}\ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}/\partial\Theta^{2}\right]}$$
(A.70)

The choice between Eq. (A.68) and Eq. (A.70) is a matter of convenience, depending on the functional form of $\ln\{p_v(\mathbf{y}|\Theta)\}$.

The denominator of Eq. (A.70) is called the *Fisher information I*(Θ) for the data **x**. The CRLB is therefore the inverse of *I*(Θ). The Fisher information shows that signals whose log-likelihood function has a high curvature with respect to the parameter Θ will have small values of σ_{Θ}^2 .

Consideration of $I(\Theta)$ leads to an interesting observation about i.i.d. samples. Suppose N i.i.d. observations y_i of an RV dependent on a parameter Θ are available. For a single observation $I(\Theta) = -\mathbf{E}[\partial^2 \ln\{p_y(y|\Theta)\}/\partial\Theta^2]$. Because the N observations are independent and identically distributed, $p_y(y|\Theta)=[p_y(y|\Theta)]^N$ so that $I(\Theta)$ is N times larger, $-\mathbf{E}[\partial^2 \ln\{[p_y(y|\Theta)]^N\}/\partial\Theta^2] = -N\mathbf{E}[\partial^2 \ln\{p_y(y|\Theta)\}/\partial\Theta^2]$. Consequently, the CRLB for N i.i.d. observations is 1/N times the CRLB for a single observation.

A.3.2 The CRLB for Transformed Parameters

If the CRLB for a particular parameter Θ in some estimation problem is known, it is easy to find the CRLB for a related parameter $\Phi = g(\Theta)$ for some function g. The result is

$$\sigma_{\hat{\Phi}}^2 = \left(\frac{\partial g(\Theta)}{\partial \Theta}\right)^2 \sigma_{\hat{\Theta}}^2 \tag{A.71}$$

As a trivial example, if the CRLB for estimating time delay t_0 is known, the CRLB for the corresponding range $R_0 = c \cdot t_0/2$ is $\sigma_{\hat{R}_0}^2 = (c/2)^2 \sigma_{\hat{t}_0}^2$.

It is easily shown that if the estimator for Θ is efficient, it is not the case in general that the estimator for Φ is also efficient. However, if the transformation *g* is affine (a linear transformation followed by a translation), efficiency of the estimator is maintained. Even in the case of a nonlinear transformation, the transformed estimator will be *asymptotically efficient*, i.e., it will become efficient as $N \rightarrow \infty$.

A.3.3 Signals in Additive White Gaussian Noise

The CRLB takes on a special form for the very important case of a signal in additive white Gaussian noise (AWGN). Assume the measurement vector \mathbf{y} is composed of the *N* real-valued signal + noise samples

$$y[n] = s[n;\Theta] + w[n], \quad n = 0, ..., N-1$$
 (A.72)

where Θ is the real-valued parameter to be estimated and the variance of w[n] is σ_w^2 . Because the noise is white and Gaussian, the PDF $p_y(\mathbf{y}|\Theta)$ is a multidimensional real-valued Gaussian function

$$p_{\mathbf{y}}(\mathbf{y}|\Theta) = \frac{1}{\left(2\pi\sigma_{w}^{2}\right)^{N/2}} \exp\left\{\frac{-1}{2\sigma_{w}^{2}}\sum_{n=0}^{N-1} (y[n] - s[n;\Theta])^{2}\right\}$$
(A.73)

so that

$$\ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\} = -\frac{N}{2} \left(2\pi\sigma_{w}^{2}\right) - \frac{1}{2\sigma_{w}^{2}} \sum_{n=0}^{N-1} (y[n] - s[n;\Theta])^{2}$$
(A.74)

The first and second partial derivatives of $\ln\{p_y(\mathbf{y};\Theta)\}\$ with respect to the unknown parameter Θ are

$$\frac{\partial [\ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}]}{\partial \Theta} = \frac{1}{\sigma_{w}^{2}} \sum_{n=0}^{N-1} (y[n] - s[n;\Theta]) \frac{\partial s[n;\Theta]}{\partial \Theta}$$
$$\frac{\partial^{2} [\ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}]}{\partial \Theta^{2}} = \frac{1}{\sigma_{w}^{2}} \left\{ \sum_{n=0}^{N-1} (y[n] - s[n;\Theta]) \frac{\partial^{2} s[n;\Theta]}{\partial \Theta^{2}} - \left(\frac{\partial s[n;\Theta]}{\partial \Theta}\right)^{2} \right\}$$
(A.75)

The expected value of the second partial derivative is

$$\mathbf{E}\left\{\frac{\partial^{2}[\ln\{p_{\mathbf{y}}(\mathbf{y}|\Theta)\}]}{\partial\Theta^{2}}\right\} = \frac{-1}{\sigma_{w}^{2}}\sum_{n=0}^{N-1} \left(\frac{\partial s[n;\Theta]}{\partial\Theta}\right)^{2}$$
(A.76)

This result follows because the term $(y[n] - s[n;\Theta])$ is just the noise w[n], which is zero mean. Finally, using Eq. (A.76) in Eq. (A.70) gives the CRLB for a real-valued signal in real white Gaussian noise:

$$\sigma_{\Theta}^{2} \ge \frac{\sigma_{w}^{2}}{\sum_{n=0}^{N-1} \left(\frac{\partial s[n;\Theta]}{\partial \Theta}\right)^{2}}$$
 (real-valued signal in real AWGN)
(A.77)

The denominator of Eq. (A.77) shows that signals that are more sensitive to changes in the value of the parameter Θ (large magnitudes of the derivative) will have smaller values of σ_{Θ}^2 , suggesting that systems having finer resolution in a parameter of interest will also exhibit finer precision than coarser-resolution systems. This point is illustrated more clearly in <u>Chap. 9</u>.

A.3.4 Signals with Multiple Parameters in AWGN

The CRLB can be generalized to describe estimators of multiple simultaneous scalar parameters by defining a vector parameter Θ . For example, it might be desirable to simultaneously estimate the amplitude, frequency, and initial phase of a sinusoid in additive WGN. The Fisher information for a signal **x** dependent on a real-valued parameter vector Θ is now the $N \times N$ matrix:

$$[\mathbf{I}(\boldsymbol{\Theta})]_{ij} = -\mathbf{E}\left[\frac{\partial^2 \ln p_{\mathbf{x}}(\mathbf{x} |\boldsymbol{\Theta}|)}{\partial \Theta_i \, \partial \Theta_j}\right]$$
(A.78)

The vector parameter CRLB then states that the covariance matrix of any estimate $\hat{\Theta}$ of Θ must satisfy

$$C_{\hat{\Theta}} - I^{-1}(\Theta) \ge 0$$

where the notation " ≥ 0 " means that the left-hand side is positive semidefinite. In particular, consideration of the diagonal elements gives

$$\sigma_{\hat{\Theta}_i}^2 \ge [\mathbf{I}^{-1}(\boldsymbol{\Theta})]_{ii} \tag{A.80}$$

If the signal x has a Gaussian PDF with mean vector μ and general covariance matrix C_x (not necessarily white noise), both possibly dependent on a real-valued parameter vector Θ , the Fisher information matrix becomes (Kay, 1993, <u>Chap. 3</u>)

$$[\mathbf{I}(\Theta)]_{ij} = \frac{1}{2} \operatorname{tr} \left[\mathbf{C}_{\mathbf{x}}^{-1}(\Theta) \frac{\partial \mathbf{C}_{\mathbf{x}}(\Theta)}{\partial \Theta_{i}} \mathbf{C}_{\mathbf{x}}^{-1}(\Theta) \frac{\partial \mathbf{C}_{\mathbf{x}}(\Theta)}{\partial \Theta_{j}} \right] + \left[\frac{\partial \mu(\Theta)}{\partial \Theta_{i}} \right]^{T} \mathbf{C}_{\mathbf{x}}^{-1}(\Theta) \left[\frac{\partial \mu(\Theta)}{\partial \Theta_{j}} \right]$$
(A.81)

where

$$\begin{bmatrix} \mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta}) \end{bmatrix}_{ij} = \mathbf{E}[(x_{i} - x_{i})(x_{j} - x_{j})]$$

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\Theta})}{\partial \Theta_{i}} = \begin{bmatrix} \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\Theta})]_{0}}{\partial \Theta_{i}} & \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\Theta})]_{1}}{\partial \Theta_{i}} & \cdots & \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\Theta})]_{N-1}}{\partial \Theta_{i}} \end{bmatrix}$$

$$\frac{\partial \mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})}{\partial \Theta_{i}} = \begin{bmatrix} \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{0,0}}{\partial \Theta_{i}} & \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{0,1}}{\partial \Theta_{i}} & \cdots & \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{0,N-1}}{\partial \Theta_{i}} \\ \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{1,0}}{\partial \Theta_{i}} & \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{1,1}}{\partial \Theta_{i}} & \cdots & \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{1,N-1}}{\partial \Theta_{i}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{N-1,0}}{\partial \Theta_{i}} & \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{N-1,1}}{\partial \Theta_{i}} & \cdots & \frac{\partial [\mathbf{C}_{\mathbf{x}}(\boldsymbol{\Theta})]_{N-1,N-1}}{\partial \Theta_{i}} \end{bmatrix}$$
(A.82)

Now consider the common additional restriction that the signal is of the form $\mathbf{x} = \mathbf{s}(\mathbf{\Theta}) + \mathbf{w}$ with $\mathbf{s} = [s[0;\mathbf{\Theta}] \ s[1;\mathbf{\Theta}] \ \dots \ s[N-1;\mathbf{\Theta}]]^T$ being real and deterministic, and \mathbf{w} being i.i.d. real-valued Gaussian noise with zero mean and covariance matrix $\mathbf{C}_{\mathbf{w}} = \sigma_{\mathbf{w}}^2 \mathbf{I}$. Then \mathbf{x} is Gaussian with covariance $\mathbf{C}_{\mathbf{x}} = \mathbf{C}_{\mathbf{w}}$ and mean $\mu = \mathbf{s}$. Furthermore, for most problems of interest $\mathbf{C}_{\mathbf{w}}$ and therefore $\mathbf{C}_{\mathbf{x}}$ does not depend on $\hat{\mathbf{\Theta}}$, so that Eq. (A.92) reduces to³

$$\begin{split} \left[\mathbf{I}(\Theta)\right]_{ij} &= \frac{1}{\sigma_w^2} \left[\frac{\partial s(\Theta)}{\partial \Theta_i} \right]^T \cdot \left[\frac{\partial s(\Theta)}{\partial \Theta_j} \right] \\ &= \frac{1}{\sigma_w^2} \sum_{n=0}^{N-1} \frac{\partial s[n;\Theta]}{\partial \Theta_i} \frac{\partial s[n;\Theta]}{\partial \Theta_j} \quad \text{(real-valued signal in real AWGN, multiple parameters)} \end{split}$$

$$(A.83)$$

A.3.5 Complex Signals and Parameters in AWGN

Similar results can be derived for the case where the signal and noise are complex-valued and the parameters of interest are real or complex. One way to derive the needed results defines the derivative of a complex function with respect to a complex variable. This can be done in more than one way, but all produce some surprising results. For example, if Θ is a complex variable, the definition used in Kay (1993) results in $\partial(|\Theta|^2)/\partial\Theta = \Theta^*$ and $\partial\Theta^*/\partial\Theta = 0$. Another method is to treat complex parameters as two real parameters, $\Theta_R = \text{Re}\{\Theta\}$ and $\Theta_I = \text{Im}\{\Theta\}$. This allows the mixture of both real and complex parameters in Θ and the use of conventional calculus. Either method, used correctly, produces the same results. The latter approach is used here.

The general CRLB in terms of PDFs given in Eq. (A.68) or (A.70) for the scalar case and Eqs. (A.78) and (A.80) in the vector case still apply. For a complex signal dependent on a real parameter vector $\boldsymbol{\Theta}$ in complex AWGN the Fisher information matrix can be shown to be (Kay, 1993, Chap. 15)

$$\left[\mathbf{I}(\boldsymbol{\Theta})\right]_{ij} = \frac{2}{\sigma_w^2} \operatorname{Re}\left\{\left[\frac{\partial \mathbf{s}(\boldsymbol{\Theta})}{\partial \Theta_i}\right]^H \cdot \left[\frac{\partial \mathbf{s}(\boldsymbol{\Theta})}{\partial \Theta_j}\right]\right\} = \frac{2}{\sigma_w^2} \operatorname{Re}\left\{\sum_{n=0}^{N-1} \left[\frac{\partial s[n;\boldsymbol{\Theta}]}{\partial \Theta_i}\right]^* \left[\frac{\partial s[n;\boldsymbol{\Theta}]}{\partial \Theta_j}\right]\right\}$$
(A.84)

Because Θ is real, $\partial s^*[n; \Theta] / \partial \Theta_i = [\partial s[n; \Theta] / \partial \Theta_i]^*$. The diagonal elements, which can be compared to the single parameter real case of Eq. (A.77), are

$$[I(\Theta)]_{ii} = \frac{2}{\sigma_w^2} \sum_{n=0}^{N-1} \left| \frac{\partial s[n;\Theta]}{\partial \Theta_i} \right|^2 \quad \text{(complex signals in complex AWGN, real parameters)}$$
(A.85)

The CRLBs for the individual parameters are still given by Eq. (A.80).

Comparing Eqs. (A.84) and (A.85) to (A.83) shows that the CRLBs for signals in AWGN have the same general form in the complex case as in the real case, but are smaller by a factor of two. Thus, the real CRLB is *not* a special case of the complex CRLB.

Because a complex parameter Θ is treated as two real parameters Θ_R and Θ_I , the results above produce the CRLBs for Θ_R and Θ_I . However, the CRLB for Θ is usually the quantity of interest. It is easy to show that $\sigma_{\Theta}^2 = \sigma_{\Theta_R}^2 + \sigma_{\Theta_I}^2$, so the

CRLB for Θ is simply the sum of the CRLBs for Θ_R and Θ_I .

A.3.6 Finding Minimum Variance Estimators

Equations (A.68) or (A.70) give the minimum variance of an unbiased estimator. Clearly, it would be good to know how to construct an estimator that achieves this minimum. While this is not always possible, it is possible to see what form it must take if it exists. The minimum variance will be achieved when the condition for equality in the Schwarz inequality is met. From Eq. (A.66), this will occur when

$$A(\mathbf{x}) = (\hat{\Theta} - \Theta)\sqrt{p_{\mathbf{x}}(\mathbf{x}|\Theta)} = \alpha \frac{\partial \ln\{p(\mathbf{x}|\Theta)\}}{\partial \Theta} \sqrt{p_{\mathbf{x}}(\mathbf{x}|\Theta)} = B(\mathbf{x})$$
(A.86)

so that the estimator takes the form

$$\hat{\Theta} = \Theta + \alpha \frac{\partial \ln\{p_{\mathbf{x}}(\mathbf{x}|\Theta)\}}{\partial \Theta}$$
(A.87)

for some α .

This estimator has a serious problem: the estimate $\hat{\Theta}$ depends on knowing the actual value Θ of the parameter being estimated! Obviously, if Θ is known, there is no need to estimate it. In many, but not all cases, this problem can be solved by judicious choice of the scalar α .

To clarify this issue, consider a minimum variance estimate \hat{A} of the mean A in the constant-plus AWGN example of <u>Sec. 7.1</u>. Using the intermediate result in Eq. (7.13) so that the dependence on the data x_i is explicit in Eq. (A.87) gives

$$\hat{A} = A + \alpha \left\{ \frac{1}{\sigma_w^2} \sum_{i=0}^{N-1} (x_i - A) \right\} = \left(1 - \frac{\alpha N}{\sigma_w^2} \right) A + \frac{\alpha}{\sigma_w^2} \sum_{i=0}^{N-1} x_i$$
(A.88)

Note in passing that the estimate is indeed unbiased for any choice of α :

$$\mathbf{E}\{\hat{A}\} = A + \frac{\alpha}{\sigma_w^2} \left\{ \sum_{i=0}^{N-1} \mathbf{E}\{x_i - A\} \right\} = A$$
(A.89)

The dependence of \hat{A} on A can be removed by choosing $\alpha = \sigma_w^2/N$, giving the unbiased minimum variance estimator

$$\hat{A} = \frac{1}{N} \sum_{i=0}^{N-1} x_i$$

(A.90)

Note that the required choice of α in fact was the CRLB for this problem. It is true in general that $\alpha = I^{-1}(\Theta)$, the CRLB, so that Eq. (A.87) is more specifically

$$\hat{\Theta} = \Theta + \frac{1}{I(\Theta)} \frac{\partial \ln\{p_{\mathbf{x}}(\mathbf{x}|\Theta)\}}{\partial \Theta}$$
(A.91)

Equation (A.91) can be generalized for the vector parameter case. The minimum variance estimator must be of the form

$$\hat{\Theta} = \Theta + \mathbf{I}^{-1}(\Theta) \frac{\partial \ln\{p_{\mathbf{x}}(\mathbf{x}|\Theta)\}}{\partial \Theta}$$
(A.92)

Equations (A.91) and (A.92) are both necessary and sufficient for the existence of a minimum variance estimator. Thus, an estimator that satisfies these equations, however it was found, is efficient.

A-4.A Random Signals in Linear Systems

A rigorous definition of a *random signal* is subtle; see Papoulis and Pillai (2001) for a discussion. For this text, it is adequate to define a random signal as a continuous or discrete signal x(t) or x[n] whose value at each instant of t or n is a random variable. If the PDF of that random variable is the same for each value of t or n, the random signal is *stationary*. The principal concern here is the effect of linear shift-invariant (LSI) systems such as filters on the properties of a random signal. Discrete signal notation will be used, but all the results translate readily to the continuous case.

A.4.1 Correlation Functions

The mean of a stationary random process x is denoted $\mathbf{E}\{x\}$, m_x , or \overline{x} , as convenient. The (stochastic) *cross-correlation function* of two possibly complex random signals x[n] and y[n] is defined as

$$s_{xy}[k] = \mathbf{E}\{x[n]y^*[n+k]\}$$
(A.93)

The variable k is called the correlation *lag*. This definition is obviously very similar to the deterministic autocorrelation function described in <u>App. B</u>, with a stochastic expectation instead of a deterministic averaging.

The *autocorrelation function* (ACF) results when y = x, i.e. x is correlated with itself:

$$s_x[k] = \mathbb{E}\{x[n]x^*[n+k]\}$$
(A.94)

The normalized ACF is defined as

$$\rho_{xx}[k] = \frac{1}{s_x[0]} s_x[k]$$
(A.95)

The ACF has several useful properties. The average power of x is given by the zero lag:

$$s_{x}[0] = \mathbf{E}\{x[n]x^{*}[n]\} = \mathbf{E}\{|x[n]|^{2}\}$$
(A.96)

 s_x has a symmetry property,

$$s_x[-k] = s_x^*[k]$$
 (A.97)

a "shape property,"

$$\left|s_{x}[k]\right| \le s_{x}[0] \tag{A.98}$$

and in many cases of interest an asymptotic limit property:

$$\lim_{k \to \infty} \{s_x[k]\} = |m_x|^2 \tag{A.99}$$

The shape property states the ACF always has a maximum at the zero lag; it can reach the same amplitude at other lags but can never exceed it. It also implies that the normalized ACF has a maximum value of one at zero lag and is bounded between ± 1 at other lags. An infinite-duration sinusoid is an example of a signal that matches the maximum autocorrelation amplitude at lags other than zero, specifically at lags equal to integer multiples of the signal period. The asymptotic limit property applies only to signals whose samples tend to become uncorrelated as the time between them (the lag) becomes large. This is true for most signals of interest, but the infinite-duration sinusoid is an example of a signal which does not have this behavior. If a signal exhibits this decorrelation for large lags, the property states that the ACF approaches the square of the process mean in the limit. For zero mean random processes, this means that the ACF must decay to zero.

A.4.2 Correlation and Linear Estimation

A frequent problem is estimating a random variable y based on measurements of another random variable x. An example would be estimation of a random process at sample n_2 based on its value at n_1 . A common approach is to find the minimum mean-squared error (MMSE) estimator, which is the estimate \hat{y} that minimizes the mean-squared error (MSE)

$$\xi = \mathbf{E}\{(y - \hat{y})^2\}$$
(A.100)

Only real-valued variables are considered for now. It is frequently convenient to restrict the estimator to be linear so that \hat{y} is of the form

$$\hat{y} = ax + b \tag{A.101}$$

Specifying the estimator is then a matter of specifying the constant a and b. Linear estimators are convenient because a and b can be found by solving linear equations and because they will turn out to depend only on the second-order moments of x and y, whereas more general estimators require knowledge of the joint PDFs of x and y, information that is less likely to be known.

The MMSE linear estimator can be found by using Eq. (A.101) in Eq. (A.100) and then setting the derivatives of ξ with respect to *a* and *b* equal to zero so as to find the minimum value of ξ . Since the expectation and partial derivative operators commute,

$$\frac{\partial\xi}{\partial a} = -2E\{xy\} + 2aE\{x^2\} + 2bE\{x\} = 0$$
$$\frac{\partial\xi}{\partial b} = -2E\{y\} + 2aE\{x\} + 2b = 0$$
(A.102)

It is convenient to first solve the second equation for b and then use that result in the first equation to get a. The result is (using the overbar notation for expected values)

$$a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x^2}, \quad b = \overline{y} - a\overline{x}$$
(A.103)

This solution for *a* is reminiscent of the definition of the normalized *correlation coefficient* between two random variables *x* and *y*, which is

$$\rho_{xy} = \frac{\mathbf{E}\{(x - \overline{x})(y - \overline{y})\}}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y}$$
(A.104)

Consequently, $a = (\sigma_y / \sigma_x) \rho_{xy}$. Combining these results gives the MMSE linear estimator as

$$\hat{y} = \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - \overline{x}) + \overline{y}$$
(A.105)

It is useful to consider two limiting cases. It can be shown that $|\rho_{xy}|^2 \le 1$ (Hayes, 1996). If x and y are uncorrelated, then $\rho_{xy} = 0$ and $\hat{y} = \bar{y}$. The lack of correlation implies that knowing x gives no information regarding the value of y, so the best linear estimate is just to guess the mean value of y. At the opposite extreme, suppose $\rho_{xy} = 1$, meaning that x is highly correlated with y. In this event, the best linear estimate of y is obtained by adding the deviation of x from its mean, scaled by the ratio of standard deviations of x and y, to the mean of y.

An important example occurs when the variables x and y are samples of the same random process at different times, i.e., x = w[n] and y = w[n + k]. If the random process is stationary, then $\sigma_x = \sigma_y$ and ρ_{xy} is the normalized autocorrelation function evaluated at lag m, $\rho_w[k]$. Equation (A.105) then becomes

$$\overline{w[n+k]} = \rho_w[k](w[n] - \overline{w}) + \overline{w}$$
$$= \rho_w[k]w[n] + (1 - \rho_w[k])\overline{w}$$
(A.106)

If ρ_w evaluated at a lag of *m* samples is zero, the MMSE estimate of w[n + k] is just the mean of *w*. If $\rho_w[k] = 1$, the MMSE estimate of w[n + k] is w[n]. Thus, a high value of the normalized ACF predicts that the value of the future sample will be close to that of the current sample. A small value of *r* does not indicate that the future value will necessarily be different from the current value, but rather that it cannot be predicted based on the current value.

A.4.3 Power Spectrum

The *power spectral density* (PSD) or *power spectrum* of a random process is the Fourier transform (DTFT in the discrete case) of the autocorrelation function:

$$S_x(\omega) = F\{s_x[k]\}$$
(A.107)

Its properties include

$$E\{|x|^2\} = s_x[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\omega) d\omega$$

$$S_x^*(\omega) = S_x(\omega)$$

$$S_x(-\omega) = S_x(\omega) \text{ if } x[n] \text{ is real}$$

$$S_x(\omega) \ge 0$$
(A.108)

The first property follows from Parseval's theorem and states that average power of a random process equals the integral of the PSD, which tends to explain the term PSD. The second property states that the PSD is real-valued, even if x is complex. The third states that when x is real, the PSD is symmetric about the origin. The fourth states that it is nonnegative, again consistent with the idea of a measure of power versus frequency. These last three properties are also seen in the magnitude-squared of the Fourier transform of a deterministic signal.

A.4.4 White Noise

White noise is a random signal having the specific ACF (discrete case)

$$s_x[k] = \sigma_x^2 \delta[k] \tag{A.109}$$

where $\delta[k]$ is the discrete impulse function. It follows from Eq. (A.107) that the PSD of white noise is

$$S_x(\omega) = \sigma_x^2 \tag{A.110}$$

Thus the PSD of white noise is a constant for all frequencies. This fact accounts for the term "white noise," in analogy to the idea that white light contains all visible colors (wavelengths or frequencies) in equal proportion.

White noise is necessarily zero mean. To see this, suppose x has a nonzero mean m_x . Then x could be written as the sum of a zero-mean term \bar{x} and m_x , $x = \tilde{x} + m_x$. The autocorrelation would take the form

$$s_{x}[k] = s_{\tilde{x}}[k] + |m_{x}|^{2}$$
(A.111)

The constant term in s_x will result in an impulse $2\pi |m_x|^2 \delta_D(\omega)$ in $S_x(\omega)$, where $\delta_D(\omega)$ is the continuous-variable Dirac impulse function. The resulting PSD

cannot be a constant for all ω and x cannot be white noise.

A.4.5 The Effect of LSI Systems on Random Signals

Suppose a random signal x[n] is passed through an LSI system characterized by its impulse response h[n] and frequency response $H(\omega)$. How are characteristics such as the mean or the power spectrum altered?

The output y[n] is given by the convolution sum:

$$y[n] = \sum_{n = -\infty}^{\infty} h[l]x[n-l]$$
(A.112)

The mean of y, m_y , is

$$m_{y} = \mathbf{E}\left\{\sum_{n=-\infty}^{\infty} h[l]x[n-l]\right\} = \sum_{n=-\infty}^{\infty} h[l]\mathbf{E}\left\{x[n-l]\right\}$$
$$= m_{x}\sum_{n=-\infty}^{\infty} h[l] = m_{x}H(0)$$
(A.113)

If m_x is thought of as the "DC value" of x, then the DC value of y is just m_x multiplied by the frequency response at DC.

The ACF of *y* is

$$s_{y}[k] = \mathbf{E}\{y[n]y^{*}[n+k]\}$$

$$= \mathbf{E}\left\{\left(\sum_{l=-\infty}^{\infty} h[l]x[n-l]\right)\left(\sum_{m=-\infty}^{\infty} h^{*}[m]x^{*}[n+k-m]\right)\right\}$$

$$= \sum_{l=-\infty}^{\infty} h[l]\sum_{m=-\infty}^{\infty} h^{*}[m]\mathbf{E}\{x[n-l]x^{*}[n+k-m]\}$$

$$= \sum_{l=-\infty}^{\infty} h[l]\sum_{m=-\infty}^{\infty} h^{*}[m]s_{x}[k-m+l]$$
(A.114)

Making the change of variables p = m - l gives

$$s_{y}[k] = \sum_{l=-\infty}^{\infty} h[l] \sum_{p=-\infty}^{\infty} h^{*}[l+p]s_{x}[k-p] = \sum_{p=-\infty}^{\infty} \left(\sum_{l=-\infty}^{\infty} h[l]h^{*}[l+p]\right)s_{x}[k-p]$$
$$= \sum_{p=-\infty}^{\infty} s_{h}[p]s_{x}[k-p] = s_{h}[k] * s_{x}[k]$$
(A.115)
The output autocorrelation is the stochastic autocorrelation of the input random signal with the deterministic autocorrelation of the LSI system impulse response. The relationship between power spectra immediately follows:

$$S_{y}(\omega) = S_{h}(\omega) \cdot S_{x}(\omega) = |H(\omega)|^{2} S_{x}(\omega)$$
(A.116)

The power in the output random signal can be found from Eqs. (A.96), (A.108), and (A.116):

$$\mathbf{E}\{|y[n]|^{2}\} = s_{y}[0] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{y}(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega)|^{2} S_{x}(\omega) d\omega$$
(A.117)

The white noise input case is of special interest. If x[n] is a white noise signal with power σ_x^2 and ACF $\sigma_x^2 \delta[k]$, the autocorrelation of y[n] will be

$$s_{y}[k] = s_{h}[k] * s_{x}[k] = s_{h}[k] * \sigma_{x}^{2} \delta[k] = \sigma_{x}^{2} s_{h}[k]$$
(A.118)

and the output power will simplify to

$$\sigma_y^2 = s_y[0] = \sigma_x^2 s_h[0] = \sigma_x^2 \left(\sum_{n = -\infty}^{\infty} |h[n]|^2 \right)$$
(A.119)

An important consequence is that if the input to an LSI system is a white random signal, the output will no longer be white (assuming the system is nontrivial, $h[n] \neq \delta[n]$). This is illustrated in the following example.

Suppose h[n] is the impulse response of a four-point averager so that h[n] = 1 for n = 0, 1, 2, 3 and zero otherwise. Let x[n] be 100 samples of WGN with $\sigma_x^2 = 1$, and pass it through the averaging filter to get y[n]. The normalized autocorrelation and PSD of the random data are computed at the input and output of the filter. Since these functions are themselves random because of the random input, the experiment is repeated 100 times and the various sample ACFs and PSDs are averaged to get the results shown in Fig. A.5. Part (*a*) shows the average input normalized ACF, which well-approximates the ideal unit-amplitude impulse. Shown in part (*b*) is the average sample PSD, which well-approximates the theoretical white-noise spectrum also shown. Part (*c*) shows that the average output normalized ACF closely matches the theoretical result, which is just $s_h[k]$. Finally, part (*d*) shows that the average output PSD closely matches the expected result $S_h(\omega) = |H(\omega)|^2$.





FIGURE A.5 Effect of four-point averaging filter on 100 samples of white noise. Averages of 100 sample realizations shown: (*a*) normalized input autocorrelation $s_x[k]$. (*b*) Input PSD $S_x(f)$ and ideal white-noise PSD, (*c*) normalized output autocorrelation $s_y[k]$ and theoretical result $s_h[k]$, (*d*) output PSD $S_y(f)$ and theoretical result $S_h(f) = |H(f)|^2$.

As another example, continuous-time white noise with a power spectral density of σ_x^2 passed through a unity gain ideal lowpass filter with a total bandwidth of *B* Hz (cutoff frequency of $\pm B/2$ Hz) will have a PSD that is white only over $\pm B/2$ Hz. The corresponding ACF is

$$s_y(t) = \sigma_x^2 \frac{\sin(\pi Bt)}{\pi t} = \sigma_x^2 B \operatorname{sinc}(Bt)$$
(A.120)

where the sinc function is defined as $\operatorname{sinc}(z) = \frac{\sin(\pi z)}{\pi z}$. The filtered white noise is still zero mean but is no longer white. However, because it is bandlimited to *B* Hz the ACF of the output noise is zero at lags of integer multiples of 1/B seconds. Consequently, the samples of a discrete-time sequence obtained by sampling the filtered noise at the Nyquist rate of *B* samples per second will be uncorrelated with one another, so that the discrete-time ACF will be $\sigma_y^2 = \sigma_x^2 B \delta[k]$. The sampled noise is therefore white. Oversampling will result in noise having a sampled sinc function ACF and therefore a PSD that is bandlimited to some degree on the normalized frequency scale.

References

- Hayes, M. H., *Statistical Digital Signal Processing and Modeling*. Wiley, New York, 1996.
- Kay, S. M., Fundamentals of Statistical Signal Processing, Vol. 1: Estimation Theory. Prentice-Hall, Upper Saddle River, New Jersey, 1993.
- Olver, F. W. J., D. W. Lozier, R. F. Boisvert, and C. W. Clard (eds.), *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, 2010.
- Omura, J., and T. Kailath, "Some Useful Probability Distributions," Technical Report 7050-6, Stanford Electronics Laboratories, Sep. 1965. Available as free downloadable PDF from the U.S. Defense Technical Information Center at <u>www.dtic.mil</u> as report AD747128.
- Papoulis, A., and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, 2001.
- Peebles, P. Z., Jr., Radar Principles. Wiley, New York, 1998.
- Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part 1.* Wiley, New York, 1968, p. 338.
- Watts, S., "Radar Detection Prediction in Sea Clutter Using the Compound *K*-Distribution Model," *IEE Proceedings*, vol. 132, pt. F, no. 7, pp. 613–620, Dec. 1985.

¹ The symbol ~ means "is distributed as" and the notation N(x; a, b) denotes a normal or Gaussian distribution with mean *a* and variance *b*.

² The confluent hypergeometric function $1F_1(\alpha, \beta, x)$ is also called Kummer's function and denoted $M(\alpha, \beta, x)$. See Olver et al. (2010) for more information.

³ An exception would be if the noise variance σ_{w}^{2} was the parameter to be estimated.

Selected Topics in Digital Signal Processing

A few basic concepts and signal processing operations appear over and over again, even if sometimes in disguise, in radar signal processing. In this section, several frequently encountered concepts and operations are reviewed with special emphasis on a few aspects not often emphasized in some other application fields.

cB-1.B Fourier Transforms

The Fourier transform is as ubiquitous in radar signal processing as in most other signal processing fields. Frequency domain representations are often used to separate desired signals from interference; the Doppler shift is a frequency domain phenomenon of critical importance; and it will be seen that in some radar systems, especially imaging systems, the collected data is related to the desired end product by a Fourier transform.

Both continuous and discrete signals are of interest, and therefore Fourier transforms are required for both. Consider a signal x(u) that is a function of a continuous variable in one dimension called the *signal domain*.¹ Its Fourier transform, denoted $X(\Omega)$, is given by

$$X(\Omega) = \int_{-\infty}^{\infty} x(u) e^{-j\Omega u} du, \quad \Omega \in (-\infty, \infty)$$
(B.1)

and is said to be a function in the transform domain. The inverse transform is

$$x(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) e^{+j\Omega u} d\Omega, \quad u \in (-\infty, \infty)$$
(B.2)

In Eqs. (B.1) and (B.2), the frequency variable Ω is in radians per unit of u. For example, if u = t, i.e., u is in units of seconds, then Ω is the usual radian frequency in units of radians per second; if u is a spatial variable in meters, then Ω is spatial frequency in units of radians per meter.

An equivalent transform pair using a cyclical frequency variable $F = \Omega/2\pi$ is

$$X(F) = \int_{-\infty}^{\infty} x(u) \, e^{-j2\pi F u} \, du, \quad F \in (-\infty, \infty)$$
(B.3)

$$x(u) = \int_{-\infty}^{\infty} X(F) e^{+j2\pi F u} dF, \quad u \in (-\infty, \infty)$$
(B.4)

If the signal domain is time (u = t), then *F* is in cycles per second, or hertz. Ω and *F* will sometimes be referred to as *analog frequencies*, representing their correspondence to continuous-variable signals.

There are many excellent textbooks on Fourier transforms and their properties. Two classics are Papoulis (1987) and Bracewell (1999). Papoulis uses primarily the radian frequency notation, while Bracewell uses cyclical frequency.

A useful example in radar is the Fourier transform of a constant-frequency complex exponential pulse. Define

$$x(t) = \begin{cases} A \exp(j\Omega_0 t), & -\tau/2 \le t \le \tau/2 \\ 0, & \text{otherwise} \end{cases}$$
(B.5)

Inserting Eq. (B.5) into the definition Eq. (B.1) gives

$$X(\Omega) = \int_{-\tau/2}^{+\tau/2} A e^{+j\Omega_0 t} e^{-j\Omega t} dt = A \int_{-\tau/2}^{+\tau/2} e^{-j(\Omega - \Omega_0)t} dt$$
$$= \frac{A}{j(\Omega - \Omega_0)} \left\{ e^{+j(\Omega - \Omega_0)\frac{\tau}{2}} - e^{-j(\Omega - \Omega_0)\frac{\tau}{2}} \right\}$$
(B.6)

Applying Euler's formula gives the result

$$X(\Omega) = \frac{2A}{(\Omega - \Omega_0)} \sin\left[(\Omega - \Omega_0)\frac{\tau}{2}\right]$$
$$= A\tau \frac{\sin[\pi\tau(F - F_0)]}{\pi\tau(F - F_0)} \equiv A\tau \operatorname{sinc}[\tau(F - F_0)]$$
(B.7)

where $\operatorname{sinc}(z) \equiv \operatorname{sin}(\pi z)/\pi z$ is consistent with the MATLAB[®] definition of the sinc function. The main portion of this function is shown in Fig. B.1 for $F_0 = 5$ MHz and $\tau = 1 \,\mu s$. The peak occurs at $F = F_0$ and has a value of $A\tau$. It is easy to see from Eq. (B.7) that the Rayleigh width (peak to first null) is $1/\tau$ Hz. It can be shown numerically that the 3 dB width is $0.89/\tau$ Hz and that the magnitude of the largest sidelobe (first negative peak) relative to the mainlobe magnitude is $-13.26 \,\mathrm{dB}$.



FIGURE B.1 The Fourier transform of a 1 μ s long, 5 MHz complex exponential pulse.

Conditions on a signal for the existence of its Fourier transform are discussed in Bracewell (1999). For current purposes, it can be assumed that the Fourier transform of any signal of interest exists, and that the signal and its Fourier transform form a unique one-to-one pair.

The Fourier transform for continuous signals is important for some analyses, particularly those relating to establishing sampling requirements, but most actual processing will be performed with discrete-variable signals. There are two classes of Fourier transforms for discrete-variable signals. Directly analogous to the continuous-variable case is the following transform pair for a discrete-variable signal x[n]:²

$$X(\omega) = \sum_{n = -\infty}^{\infty} x[n] e^{-j\omega n}, \quad \omega \in (-\infty, \infty)$$
(B.8)

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{+j\omega n} d\omega, \quad n \in (-\infty, \infty)$$
(B.9)

In this pair, *n* is the signal index in units of samples and ω is a *continuous* frequency variable with units of radians per sample (not radians per second or

radians per meter).³ The frequencies ω and f will sometimes be referred to as *digital* frequencies because the corresponding signals are discrete. They are also called *normalized* frequencies for reason to be explained in <u>Sec. B.2</u>.

A related Fourier transform pair using a normalized cyclical frequency $f = \omega/2\pi$ in cycles per sample is given by

$$X(f) = \sum_{n = -\infty}^{\infty} x[n] e^{-j2\pi fn}, \quad f \in (-\infty, \infty)$$
(B.10)

$$x[n] = \int_{-0.5}^{0.5} X(f) e^{+j2\pi f n} df, \quad n \in (-\infty, \infty)$$
(B.11)

The function $X(\omega)$ or X(f) is called the *discrete-time Fourier transform* (DTFT) of x[n]. It is readily seen from Eq. (B.8) that $X(\omega)$ is continuous in the frequency variable ω with a period of 2π radians per sample; that is, the DTFT repeats itself every 2π radians per sample. Consequently, though it is defined for all ω , normally only the principal period $-\pi \le \omega < \pi$ is discussed and illustrated. Similarly, X(f) has a period of one cycle per sample and the principal period is $-0.5 \le f < 0.5$ cycles per sample. The properties of DTFTs are described in most modern digital signal processing textbooks, for example Oppenheim and Schafer (2010).

Similar to the continuous-time case, the DTFT of a sampled constantfrequency complex exponential is of interest. Defining the waveform as

$$x[n] = Ae^{j\omega_0 n}, \quad 0 \le n \le N - 1$$
 (B.12)

a derivation very similar to that leading to Eq. (B.7) gives

$$X(\omega) = A \frac{\sin[(\omega - \omega_0)N/2]}{\sin[(\omega - \omega_0)/2]} e^{-j(\omega - \omega_0)(N-1)/2}$$
$$X(f) = A \frac{\sin[\pi(f - f_0)N]}{\sin[(f - f_0)]} e^{-j\pi(f - f_0)(N-1)}$$
(B.13)

The result is shown for $f_0 = 0.285$, N = 20, and A = 1 as the solid curve in Fig. B.2*a* or *b*. Examination of Eq. (B.13) and Fig. B.2 shows that this DTFT is very similar to the sinc function of Eq. (B.7) in appearance, especially around the mainlobe. The sidelobes do not decay indefinitely because of the requirement that the DTFT be periodic in frequency. The phase term in Eq. (B.13) arises because the pulse was not centered on the origin; if N were odd and it was centered, the phase term would not be present.

While the DTFT is well-suited for mathematical analysis of discrete-time signals, $X(\omega)$ or X(f) are not suitable for computation and manipulation in a digital processor because they cannot be computed for all of the uncountably infinite values of the continuous frequency variable ω or f. A finite, discrete set of frequency values is needed. The *discrete Fourier transform* (DFT, not to be confused with the DTFT) is a computable Fourier transform defined for *finite-length* discrete-variable signals. The *K*-point DFT and its inverse for an *N*-point signal x[n] are given by⁴

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi nk/K}, \quad k \in [0, K-1]$$
(B.14)

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{+j2\pi nk/K}, \quad n \in [0, N-1]$$
(B.15)

Inspection shows that for x[n] of finite duration N samples, X[k] is simply a sampled version of $X(\omega)$ or X(f), with K frequency samples distributed uniformly across the period $[0, 2\pi)$ in ω or [0, 1) in f:

$$X[k] = X(\omega) \Big|_{\omega = \frac{2\pi}{K}k} = X(f) \Big|_{f = \frac{1}{K}k}, \quad k \in [0, K-1]$$
(B.16)

Because the DTFT is periodic, so is the DFT, with period K samples.

It is usually preferable to plot the spectrum of a signal on a symmetric scale that places zero frequency in the center. Because of the periodicity of the DTFT, the DFT samples corresponding to frequencies in the interval $f \in (0.5, 1]$ cycles per sample are identical to those in the interval $f \in (-0.5, 0]$. The upper half of the DFT samples $(k > \lceil K/2 \rceil)$ are therefore often plotted first, followed by the lower half of the samples.⁵

Figure B.2 illustrates the difference between the DTFT and the DFT for a signal consisting of a 20-sample complex exponential pulse with a frequency f = 0.285 cycles per sample. The solid curves are the DTFT of the data over the interval $f \in (-0.5, 0.5]$. The small circles represent the DFT values, which have been re-arranged as discussed above. In part (*a*) of the figure, a K = 20-point DFT was used, resulting in a sparsely sampled representation of the DTFT that misses the true peak frequency and amplitude by a significant amount, and provides only a crude representation of the sidelobe structure. In part (*b*), a K = 48-point DFT was used. The underlying DTFT is unchanged; it is determined only by the data x[n]. The larger DFT simply samples it on a denser frequency



grid. A still larger DFT would trace out the details of the DTFT quite accurately.

FIGURE B.2 Sampling of the DTFT of a 20-sample complex sinusoid by the DFT: (*a*) K = 20-point DFT, (*b*) K = 48-point DFT.

So far three versions of the Fourier transform have been discussed: the original "analog" Fourier transform, the DTFT, and the DFT. Most readers will be aware of the *fast Fourier transform* (FFT). This is not a fourth variation, but an efficient algorithm for computing the DFT (Oppenheim and Schafer, 2010). Table B.1 summarizes the relationship between the various forms of the Fourier transform.

Signal Variable	Transform Variable	Transform
Continuous	Continuous	Fourier Transform (FT)
	Sample in Time	
Discrete	Continuous	Discrete-Time Fourier Transform (DTFT)
	Sample in Frequency	
Discrete	Discrete	Discrete Fourier Transform (DFT)
Fast Algorithm	for Computing the DFT	
Discrete	Discrete	Fast Fourier Transform (FFT)

TABLE B.1 Relationship between Fourier Transform Versions

cB-2.B Sampling, Quantization, and A/D Converters

The subject of this text is *digital* processing of radar signals. A digital signal is one that is discretized in two independent ways. Both are necessary to represent the signal in a digital processor. The first discretization represents the signal as a function of a discrete, rather than continuous, variable: one discrete variable for one-dimensional signals, two variables for two-dimensional signals, and so forth. Most discrete-variable signals of interest here will be obtained by *sampling* a continuous-variable physical quantity at selected, usually equispaced, values of the independent variable. An example is a discrete-time sampled version of the continuous-time output voltage of an antenna and receiver, but sampled functions of frequency and spatial position are of concern as well. A discrete representation is necessary to allow a finite-length signal to be represented in the processor by a finite number of values.

The second discretization is *quantization* of the signal's values. Each sample of a continuous-varying signal can take on an infinity of possible values. Quantization maps the continuous amplitude of a signal to one of a finite set of values so that each sample can be represented in the processor with a finite number of bits. The number of permissible values is determined by the number

of bits available in the quantized signal representation and the encoding used. Figure B.3 illustrates the distinction between continuous, sampled, quantized, and digital (sampled *and* quantized) signals.



FIGURE B.3 Relationship between continuous signals, sampling, quantization, and discrete signals.

B.2.1 Sampling

The most fundamental question in developing a sampled-variable representation is "how many samples are enough"? That is, how should the sampling interval be chosen? The Nyquist sampling theorem provides the answer. It states that if the Fourier transform X(F) of a signal x(u) is bandlimited to an interval in the Fourier domain of total width β_F cyclical frequency units (equivalently, β_{Ω} radian frequency units), then the signal can be recovered from a set of samples taken at a sampling interval

$$T_{\rm s} < \frac{1}{\beta_{\rm F}} = \frac{2\pi}{\beta_{\Omega}} \tag{B.17}$$

by an appropriate interpolation operation. That is, the sampling rate $F_s = 1/T_s$ must satisfy the simple relation

$$F_{\rm s} > \beta_{\rm F} \tag{B.18}$$

This formula will be used in this text to establish sampling rates for pulse echoes, Doppler processing, and phased array antenna element spacing, among other things.

The Nyquist theorem is easy to derive. Details supporting the following outline are provided in Oppenheim and Schafer (2010). Start by modeling the process of sampling x(u) as multiplication by an infinite Dirac impulse train:

$$x_{s}(u) = x(u) \left(\sum_{n=-\infty}^{+\infty} \delta_{D}(u - nT_{s}) \right)$$
(B.19)

 T_s is the sampling interval in units of u. For instance, if the independent variable is time, u = t, T_s is in seconds. The sampled signal $x_s(u)$ then has the Fourier transform, in cyclical frequency units, given by

$$X_{s}(F) = \frac{1}{T_{s}} \sum_{k=-\infty}^{+\infty} X\left(F - \frac{k}{T_{s}}\right) = \frac{1}{T_{s}} \sum_{n=-\infty}^{+\infty} X(F - kF_{s})$$
(B.20)

where $F_s = 1/T_s$ is the sampling rate. Equation (B.20) illustrates a very important effect of sampling: infinite *replication* of the original spectrum, at an interval of F_s units in cyclical frequency ($2\pi F_s$ in radian units). This Fourier domain replication occurs any time a signal is subsampled at regular intervals. Because the forward and inverse Fourier transforms are dual operations, this also occurs in reverse: if one samples the Fourier domain signal, then the signal is replicated in the original signal domain.

The fact that sampling in one domain results in replication in the complementary domain is an important property of Fourier analysis in its own right, independent of the Nyquist theorem that will follow. It provides the basis for an alternative to the usual modulation technique of time-domain multiplication by sinusoids for shifting a signal in the frequency domain and is used, for instance, for one method of efficient formation of digital in-phase and quadrature signals described in <u>Chap. 3</u>.

Replication also occurs when "sampling" a discrete-variable signal by decimation. Consider a signal x[n] with DTFT X(f). Define a *decimated* signal by subsampling x[n]:

$$y[n] = x[nM], -\infty < n < +\infty, M$$
 integer (B.21)

The signal x[n] is said to be decimated by the factor M to produce y[n]. The DTFT of y[n] is related to that of x[n] according to Oppenheim and Schafer (2010):

$$Y(f) = \frac{1}{M} \sum_{k=0}^{M-1} X\left(\frac{f-k}{M}\right)$$
(B.22)

The summation is finite because the DTFT already repeats periodically. This is exactly the same type of scaling and replication of the spectrum seen also in the sampling of a continuous-variable signal.

Resuming the derivation of the Nyquist theorem, the next step is to relate the analog and digital signals and their spectra to one another. A discrete-time signal is formed from the sampled data by the simple assignment

$$x[n] = x(nT_s), \quad n \in (-\infty, +\infty)$$
(B.23)

The DTFT of x[n], computed using Eq. (B.8) or Eq. (B.10), will be simply $X_s(F)$ expressed on a "digital" frequency scale:

$$X(f) = \frac{1}{T_{\rm s}} \sum_{k=-\infty}^{+\infty} X((f+k)F_{\rm s})$$
(B.24)

in cyclical units. Comparing Eq. (B.20) with Eq. (B.24) shows that the analog and digital frequency scales are related according to $F = fF_s$, so that

$$f = \frac{F}{F_s} = FT_s$$
$$\omega = 2\pi \frac{F}{F_s} = \Omega T_s$$
(B.25)

Equation (B.25) provides the basis for converting units between analog and digital frequencies. The first form, $f = F/F_s$, is the reason that the digital frequencies ω and f are often called *normalized* frequencies: f is the frequency in hertz normalized to the sampling frequency, while ω is its radian counterpart. Also, combining Eqs. (B.16) and (B.25) shows that in continuous-variable units, the *K*-point DFT frequency sample locations are equivalent to frequencies of $k/KT_s = (k/K)F_s$ cycles per unit (hertz if T_s is in seconds) or $2\pi k/KT_s = 2\pi (k/K)$ F_s radians per unit, where k is the DFT sample index number and ranges from zero to K - 1.

Figure B.4 illustrates the spectrum replication effect of sampling and the relation of the resulting DTFT X(f) to the original spectrum X(F). A notional Fourier transform X(F) is shown in Fig. B.4*a*. X(F) is not assumed to be bandlimited or to be centered at F = 0. Because x(u) may be complex-valued,

X(F) also is not assumed to exhibit Hermitian symmetry. Part (*b*) of the figure shows the replicated analog spectrum $X_s(F)$, while part (*c*) is the corresponding DTFT X(f) or $X(\omega)$. Note that X(f) is periodic with a period of 1 and $X(\omega)$ with period 2π as required for spectra of discrete-variable signals.



FIGURE B.4 Spectrum replication effect of sampling, and relation between spectra: (*a*) original non-bandlimited spectrum X(F), (*b*) spectrum of sampled signal, $X_s(F)$, (*c*) DTFT of sampled data, X(F) or $X(\omega)$.

The Nyquist theorem relies on the fact that a signal and its Fourier transform form a unique one-to-one pair. Consequently, reconstruction of the original spectrum X(F) is equivalent to reconstructing x(u). Consideration of Fig. B.4 makes clear the condition needed to enable recovery of the original spectrum: the replicas must not overlap. If each spectrum replica is distinct, then the sampled signal can be passed through a linear shift-invariant system (a filter) with a frequency response that selects only the original copy of the spectrum. The output of that system must then be the original signal.

This strategy leads to two conditions, one on the signal and one on the sampling rate. First, x(u) must be strictly bandlimited to some bandwidth β_F in cyclical frequency units:

$$X(F) \equiv 0, \quad F < F_c - \frac{\beta_F}{2} \text{ and } F > F_c + \frac{\beta_F}{2}$$
(B.26)

where F_c is the center of the region of support of X(F). If this is not the case, no sampling rate will separate the spectrum replicas; they will always overlap. Assuming x(u) is bandlimited, then the sampling rate must be chosen large enough to ensure that the spectrum replicas do not overlap. Figure B.5 shows a notional bandlimited spectrum and the corresponding sampled spectrum $X_s(F)$. By inspection, the condition needed is $F_c + \beta_F/2 < F_c + F_s - \beta_F/2$, which gives the simple result $F_s > \beta_F$ already quoted in Eq. (B.18).



FIGURE B.5 Illustration of Nyquist sampling requirement for bandlimited signals: (*a*) original bandlimited spectrum X(F), (*b*) spectrum of sampled signal, $X_s(F)$.

If the original x(u) is a *baseband* signal, meaning its spectrum is centered at $F_c = 0$, then the region of support of X(F) is $(-\beta_F/2, +\beta_F/2)$. In this event, Eq. (B.18) expresses the conventional wisdom that the sampling frequency should be at least twice the highest frequency component in the signal. However, a more direct and versatile interpretation of Eq. (B.18) is that the sampling frequency should be greater than the total spectral width of the signal. This rule is more easily applied to non-baseband signals.

In outlining the derivation of the Nyquist theorem, no assumption has been made that x(u) is real-valued. The theorem applies equally well to real or complex signals. In the case of a complex signal of total spectral width β_F Hz, the Nyquist criterion implies collecting at least β_F complex samples per second, equivalent to $2\beta_F$ real samples per second. It was also not assumed that the original signal spectrum was baseband, i.e., centered at F = 0. The spectrum can be offset to any location on the frequency axis without changing the required sampling rate; only the bandwidth matters. However, if the original spectrum is not centered at zero, none of the replicas will necessarily be centered at zero in the discrete-variable signal's spectrum unless appropriate relationships between the sampling frequency, spectrum bandwidth, and spectrum offset are maintained. In the end, it is usually desirable to have the information-bearing portion of the spectrum (the gray shaded part of Fig. B.5*a*) end up centered at the origin for ease of processing.

The Nyquist theorem is not peculiar to sampling of time domain signals. It is a consequence of the replication effect of sampling on the Fourier transform of the sampled signal, and can equally well be applied to sampling of frequency spectra or spatial signals. For example, when one samples a frequency spectrum, the corresponding time domain signal is replicated. So long as those time domain replicas are of finite duration and do not overlap, the original spectrum can be recovered, meaning that the frequency spectrum is adequately sampled. This fact was be used in <u>Chap. 3</u> to establish the minimum DFT size for computing the Doppler spectrum.

B.2.2 Quantization

Binary representations of numeric data can generally be categorized as either fixed point or floating point. In fixed-point representations, the *b* bits of a binary word are used to represent 2^b distinct and evenly spaced numerical values.⁶ In a floating-point representation, some number *e* of the *b* total bits is used to represent an exponent, and the remaining m = b - e bits are used to represent the mantissa. Here, fixed-point representations are of primary interest, as this is the representation output by most analog-to-digital (A/D) converters.

The numeric value corresponding to each of the 2^b possible binary words in a fixed-point encoding is determined by the arithmetic coding scheme that is used and the quantization step size Δ . The step size is the change in input value in volts for each increment or decrement of the binary word value. Thus, the value of the quantized data sample is simply Δ times the binary number output by the quantizer.

The two most common encodings are called *sign-magnitude* encoding and *two's complement* encoding. In sign-magnitude encoding, the most significant bit of the binary word represents the sign of the data sample; usually a value of zero represents a positive number, while a value of one represents a negative number. The remaining b - 1 bits encode the magnitude of the sample. Thus, sign-magnitude encoding can represent numbers from $(-2^{b-1}-1)\Delta$ to $(+2^{b-1}-1)\Delta$. Note that there are two codes for the value zero, corresponding to +0 and -0.

Two's complement is a somewhat more complex encoding that has advantages for the design of digital arithmetic logic. In two's complement encoding there is only one code for zero. The extra code value allows the representation of one more negative number, so the range of values becomes $-2^{b-1} \Delta$ to $(+2^{b-1} - 1)\Delta$. For any reasonable number of bits *b*, the difference in range is not significant. Details are given in Ercegovac and Lang (2003).

The choice of Δ and *b* is determined by the desired values of *signal-to-quantization noise* (SQNR) and dynamic range, and is also constrained by A/D converter technology. Consider the step size Δ first. Quantization entails rounding or truncating the analog sample value to one of the allowed quantized values. The difference between the unquantized and quantized samples is the *quantization error*. Although it is a deterministic function of the input data and A/D converter parameters, the behavior of the quantization error signal is usually complex enough that it is treated as a random variable uncorrelated from one sample to the next and independent of the original unquantized signal. Assuming rounding, the quantization error for each sample can vary between $\pm \Delta/2$, and it is commonly assumed that errors anywhere in that range are equally likely. Thus, the quantization noise process is modeled as a uniform random process over this range so that the quantization noise power $\sigma_q^2 = \Delta^2/12$.

To choose Δ , consider an A/D converter with a white noise input of power σ_n^2 . Ideally the noise power at the output of the A/D would be nearly the same as at the input so that quantization noise could be considered negligible. Let the step size $\Delta = \alpha \cdot f_n$ for some α . The increase in noise power at the A/D output due to quantization noise will be a factor of

$$\frac{\sigma_n^2 + \sigma_q^2}{\sigma_n^2} = 1 + \frac{\Delta^2 / 12}{\sigma_n^2} = 1 + \frac{\alpha^2}{12}$$
(B.27)

Figure B.6 plots the increase in decibels versus α . Values of α less than 1.76 produce a noise power increase of 1 dB or less. In practice, α is often chosen to be in the range of 0.25 to 0.5 to ensure that the receiver noise toggles one or two least significant bits. This choice ensures that very small signals at or below the noise level that should be reinforced by coherent integration are not suppressed instead.



FIGURE B.6 Effect of the quantization step size on the ratio of output power to input power for an A/D converter with a noise signal input.

This style of analysis is extended considerably in McClellan and Purdy (1978) to consider the saturation effects that result when Δ is made too small and the underflow that results when Δ is too large. It remains the case that Δ should normally be chosen to be around one-quarter to one-half the input noise standard deviation for good results.

Now consider the choice of b. This is a question of the *dynamic range* of the A/D converter. One definition of dynamic range is the ratio of the largest representable magnitude to the smallest nonzero magnitude. For the two's complement case this is, in both linear and decibel units,

$$DR = \frac{2^{b-1}\Delta}{\Delta} = 2^{b-1}$$

DR (dB) = 20log₁₀(2^{b-1}) = (b-1)20log₁₀(2) = 6.02b - 6.02
(B.28)

For the sign-magnitude case $DR = 2^{b-1} - 1 \approx 2^{b-1}$ for *b* more than just a few bits so that Eq. (B.28) still applies. This result shows that the dynamic range of the numbers that can be represented at the A/D converter output increases by 6 dB per bit. A 60-dB dynamic range would require 11 bits.

An alternative view of dynamic range comes from treating the input signal as a random process with variance σ_s^2 . Express the maximum input signal required to be represented without saturation (overflow) as $k[_s$. For a Gaussian-

distributed input signal *k* would normally be chosen in the range of 3 to 4 so that there is only a small probability of saturation. Equating this to the maximum representable number and recalling that $\Delta = \alpha [n]_n$ gives

$$k\sigma_{\rm s} = 2^{b-1}\Delta = 2^{b-1}\alpha\sigma_n \tag{B.29}$$

Now define the input signal-to-noise ratio $SNR_{IN} = \sigma_s^2 / \sigma_n^2$, substitute in Eq. (B.29), and convert to decibels:

$$SNR_{\rm IN} = \frac{2^{2b-2}\alpha^2}{k^2} \implies SNR_{\rm IN} (\rm{dB}) = 6.02b - 20\log_{10}\left(\frac{2k}{\alpha}\right) \tag{B.30}$$
(B.30)

Solving for the number of bits gives

$$b = \left[\frac{SNR_{\rm IN} \ (\rm{dB}) + 20\log_{10}(2k/\alpha)}{6.02}\right]$$
(B.31)

Equation (B.31) expresses the number of bits needed to quantize a signal with a given SNR at the quantizer input without significant saturation. For a typical choice of $\alpha = 0.5$ and k = 4, the second term in the numerator is 24.08 dB. For $\alpha = 1$ and k = 3, it is 15.56 dB. For the first case, representing signals with an SNR at the input of zero dB requires 4 bits. If the input SNR can be as high as 10 dB, this rises to 6 bits.

It is important to realize that this analysis gives the number of bits required to represent the signal present at the A/D converter input. It does not take into account increases in dynamic range due to processing gains achieved by coherent integration, matched filtering, and DFTs. If a post-A/D conversion operation achieves a coherent integration gain factor of N, an additional $\log_2 N$ bits will be required in fixed-point digital arithmetic to represent the operation result without overflow.

Another metric frequently cited is the SQNR. This is the ratio of the power in the A/D input signal to that of the quantization noise. For the Gaussian input signal, the signal power is σ_s^2 . Using the first relation in Eq. (B.29) and $\sigma_q^2 = \Delta^2/12$, and converting to dB gives

$$\frac{\sigma_s^2}{\sigma_q^2} = \frac{(2^{b-1}\Delta/k)^2}{\Delta^2/12} = \frac{12 \cdot 2^{2b-2}}{k^2} \implies$$

SQNR (dB) = 6.02b - 20log₁₀(k) - 4.77 (B.32)

Like the dynamic range, the SQNR also increases by 6 dB per bit.

Once the data are quantized, a choice must be made between implementing the processing in fixed- or floating-point arithmetic. Generally, floating-point arithmetic requires more digital logic to implement and is therefore slower and consumes more power. However, mathematical algorithms are easier to develop for floating-point arithmetic because numerical overflow and underflow are much less likely. Early radar digital processors relied mostly on fixed-point arithmetic, at least in the early processing stages that tend to be more computationally intensive, because it is faster. Increasing numbers of modern systems are using floating-point processing because the dramatic increases in processor power have made it possible to implement the desired algorithms in real time. When portions of the signal processing require greater speed or power efficiency than is available in floating point processors, field programmable gate array (FPGA) technology may sometimes provide a good alternative.

B.2.3 A/D Conversion Technology

As has been seen, the required sampling rate is determined through the Nyquist criterion by the instantaneous bandwidth β_F of the receiver output as discussed earlier, perhaps increased by a safety margin of 10 to 20 percent to allow for the imperfect bandlimiting of real waveforms and antialiasing filters, while the number of bits *b* in the digital sample is determined by the maximum dynamic range and quantization noise requirements of the system. Except in a few specialized situations, the minimum number of bits required for a signal sampled at its Nyquist rate is at least 6 and preferably 8, while 12 bits or more are desired in many applications (Merkel and Wilson, 2003).

In high-resolution radars the resulting sampling rates can be very high, often tens to hundreds of megasamples per second and in some cases even reaching rates of gigasamples per second. The use of digital I/Q or digital IF techniques for non-baseband sampling can increase the required rate relative to the waveform bandwidth by a factor of $2.5 \times$ to $4 \times$ (see <u>Chap. 3</u>), further exacerbating the problem. Many very wideband systems use a specialized linear FM waveform processing method called *stretch processing* (see <u>Chap. 4</u>) to reduce the analog signal bandwidth prior to A/D conversion, often by an order of magnitude or more.

In general, the higher the required sampling rates, the fewer the number of bits available in current A/D converter technology. Figure B.7 summarizes the state of the art in A/D converters in 2006 (Walden, 2008). The quantity plotted is the effective number of bits (ENOB) versus sampling rate. ENOB is inferred from measured SQNRs as described in Walden (1999), which also discusses many other A/D metrics and limiting factors. This data suggests that A/D converter wordlengths tend to drop about two to three bits per decade of sampling rate. However, an ENOB of 8 bits is achievable at rates up to about 1

gigasample per second, and an ENOB of 12 bits is achievable at rates up to approximately 100 megasamples per second. These are quite adequate for most radar applications.



FIGURE B.7 Analog-to-digital converter performance. (Data courtesy of Dr. Robert Walden and The Aerospace Corporation. Used with permission.)

B-3.B Spatial Frequency

The idea of spatial frequency is an important concept in any study involving propagating waves, and it will be needed to analyze spatial sampling and spacetime adaptive processing. A simplified intuitive introduction to the concept is given here. For a more complete discussion, see Johnson and Dudgeon (1993).

Consider a sinusoidal electromagnetic wave propagating in the +x direction with wavelength λ and velocity c as shown in Fig. B.8. An observer at a fixed spatial position x_0 will see successive crests of the electric field at a time interval (period) of $T = \lambda/c$ seconds; thus the temporal frequency of the wave is the usual $F = 1/T = c/\lambda$ Hz or $2\pi c/\lambda$ radians per second.



FIGURE B.8 Propagating electromagnetic wave.

A spatial period can also be defined; it is simply the interval between successive crests in space for a fixed observation time. The spatial period of the pulse is obviously λ meters. The spatial frequency is therefore $1/\lambda$ cycles per meter or $2\pi/\lambda$ radians per meter. It is common to call the latter quantity the *wavenumber* of the pulse and to denote it with the symbol *K*.²

Because position in space and velocity are three-dimensional vector quantities in general, so is the wavenumber. For simplicity of illustration, consider the two-dimensional version of Fig. B.8 shown in Fig. B.9. The pulse, now propagating at an angle in an *x*-*y* plane, still has a wavenumber $K = 2\pi/\lambda$ in the direction of propagation. However, measured in the +*x* direction, the wavenumber is $K_x = (2\pi/\lambda)\sin\theta$, where θ is the angle of incidence of the pulse measured relative to the +*y* axis. Similarly, the same signal has a wavenumber in the *y* direction of $K_y = (2\pi/\lambda)\cos\theta$. Note that as $\theta \to 0$, the wavelength in the *x* dimension tends to ∞ so that $K_x \to 0$.



FIGURE B.9 Propagating electromagnetic wave.

The extension to three dimensions of space is straightforward. The total wavenumber is related to the components in the obvious way

$$K = \sqrt{K_x^2 + K_y^2 + K_z^2}$$
(B.33)

and always equals $2\pi/\lambda$. Note that the temporal frequency remains c/λ Hz regardless of the direction of propagation.

B-4.B Correlation

Correlation is an operation that compares one signal against a reference signal to determine their degree of similarity. *Cross-correlation* is the correlation of two different signals; *autocorrelation* is the correlation of a signal with itself. Correlation is frequently defined in both a probabilistic sense, as a descriptive property of random signals, and as a deterministic operation performed on actual digital signals. If a random process is *ergodic*, the two interpretations are closely linked; see Oppenheim and Schafer (2010) or any text on random signals and processes for an introduction to these concepts. The probabilistic meaning is discussed in <u>App. A</u>. The deterministic processing operation is of concern

here.

Consider two signals x[n] and y[n] with DTFTs X(f) and Y(f). Their deterministic cross-correlation is defined as

$$s_{xy}[k] = \sum_{n=-\infty}^{+\infty} x[n]y^*[n+k], \quad -\infty < k < +\infty$$
(B.34)

and is denoted $s_{xy}[k] = x \otimes y$. If x[n] = y[n] this is the autocorrelation of x[n], denoted $s_x[k]$. The value $s_x[k]$ is called the *k*th correlation *lag*. It is straightforward to show that the deterministic cross-correlation of x[n] and y[n] is identical to the convolution of x[n] and $y^*[-n]$. The Fourier transform of the cross-correlation function is called the *cross-power spectrum*, and is just the product

$$S_{xy}(f) = \mathbf{F}\{s_{xy}[k]\} = X(f)Y^*(f)$$
(B.35)

The Fourier transform of the autocorrelation function (ACF) is usually called simply the *power spectrum* or the *power spectral density* (PSD). Notice that the power spectrum S_x is the squared-magnitude of the Fourier transform of the underlying signal x:

$$S_{x}(f) = \mathbf{F}\{s_{x}[m]\} = X(f)X^{*}(f) = |X(f)|^{2}$$
(B.36)

Thus, the power spectrum is not dependent on the phase of the signal spectrum.

The extension to two-dimensional versions is obvious:

$$s_{xy}[l,k] = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} x[m,n]y^*[m+l,n+k], \quad -\infty < l,k < +\infty$$
(B.37)

$$S_{xy}(f,g) = \mathbf{F}\{s_{xy}[l,m]\} = X(f,g)Y^*(f,g)$$
(B.38)

Graphically, correlation corresponds to overlaying the two constituent signals; multiplying them sample-by-sample; and adding the results to get a single value. One of the two signals is then shifted and the process repeated, creating a series of output values which form the correlation sequence $s_{xy}[k]$. This is illustrated notionally in Fig. B.10, which shows the cross-correlation of two functions: x[n], which is nonzero for $0 \le n \le M - 1$, and y[n], nonzero for $0 \le n \le N - 1$. Note that $s_{xy}[k]$ will be nonzero only for $1 - N \le m \le M - 1$.



FIGURE B.10 Computation of the deterministic cross-correlation function $s_{xv}[k]$.

It is sometimes convenient to define the *normalized correlation* function:

$$\rho_x[k] \equiv \frac{s_x[k]}{s_x[0]} \tag{B.39}$$

Normalized versions of the cross-correlation function, and of the twodimensional auto- and cross-correlation functions, are defined similarly.

The properties of correlation functions are described in many standard texts. Here only two properties of particular importance are presented, without proof:

$$s_x[k] \le s_x[0] \tag{B.40}$$

$$s_{xy}[k] = s_{yx}^*[-k]$$
 (B.41)

The first property states that the zero lag (k = 0) of an ACF is always the peak value. In terms of Fig. B.10, this corresponds to the case where the signal is being correlated with itself and the two replicas are completely overlapping. In this case Eq. (B.34) specializes to

$$s_{x}[0] = \sum_{n = -\infty}^{+\infty} x[n] x^{*}[n] = \sum_{n = -\infty}^{+\infty} |x[n]|^{2}$$
(B.42)

Note that $s_x[0]$ is necessarily real, even if x[n] is complex. Furthermore, $s_x[0]$ is

the total energy in x[n].

The second property above establishes the Hermitian symmetry of any auto- or cross-correlation function. From properties of the Fourier transform, it follows that the power spectrum will have even symmetry if the correlation function is also real-valued.

Computing the zero autocorrelation lag is a weighted integration of the signal samples. In coherent integration it is necessary to align the phases of the signal samples to maximize the coherent sum. This is exactly what happens in Eq. (B.42). If x[n] is represented in the magnitude-phase form $Ae^{i\phi}$, then multiplication by $x^*[n] = Ae^{-j\phi}$ cancels the phase component so that the products are all real positive values and therefore add in phase. Thus when performing an autocorrelation, the zero lag term is the weighted coherent integration of the signal samples. This observation will be useful in understanding matched filtering and synthetic aperture imaging.

Figure B.11 illustrates one of the important uses of correlation in radar signal processing: detecting and locating desired signals in the presence of noise. The sequence in part (*a*) of the figure is the sum of a zero mean, unit variance Gaussian random noise signal with the finite duration pulse $y[n] = 2, 9 \le n \le 38$ and zero otherwise. The signal-to-noise ratio is 6 dB. While there is some evidence of the presence of y[n] in the interval $9 \le n \le 38$ in Fig. B.11*a*, it does not stand out well above the noise and cannot be located precisely, nor its amplitude estimated well. However, the cross-correlation of the signal with y[n] shown in Fig. B.11*b* displays a clear peak at lag k = 9, indicating the presence of the signal y[n] in the noisy signal of part (*a*) and correctly locating it nine samples from the origin. Thus, correlation can be used to aid in identification and location of known signals in noise, i.e., as a way to "look for" a signal of interest in the presence of interference.



FIGURE B.11 Identifying signals in noise using cross-correlation: (a) signal consisting of rectangular pulse in noise, (b) result of correlating with reference rectangular pulse. See text for details.

B-5.B Vector-Matrix Representations and Eigenanalysis

B.5.1 Basic Definitions and Operations

It will sometimes be convenient to represent a finite length signal as a vector, rather than in indexed sequence notation. That is, if the signal x[n] is defined for $0 \le n \le N-1$, it can be denoted in column vector form as

$$\mathbf{x} = [x[0] \quad x[1] \quad \cdots \quad x[N-2] \quad x[N-1]]^T$$
(B.43)

where the superscript T denotes the matrix transpose. Signal vectors will usually be defined as column vectors, and boldface notation will be used to indicate vectors and matrices.

Many important signal processing operations can be expressed in matrixvector form. One of particular importance is calculation of a single output sample of a finite impulse response (FIR) linear filter. Suppose the filter impulse response is denoted h[n], $0 \le n \le L - 1$. The filter output is given by the convolution sum

$$y[n] = \sum_{l=0}^{L-1} h[l] x[n-l]$$
(B.44)

Vector notation can be used to represent the coefficients h[l] by the *L*-element column vector **h**:

$$\mathbf{h} = [h[0] \quad h[1] \quad \cdots \quad h[L-1]]^T$$
(B.45)

Now define the *L*-element vector \mathbf{x}_n of the *L* most recent signal samples as

$$\mathbf{x}_n = [x[n] \quad x[n-1] \quad \cdots \quad x[n-L+1]]^T$$
(B.46)

Equation (B.44) can then be written as the vector inner product

$$y[n] = \mathbf{h}^T \mathbf{x}_n \tag{B.47}$$

or simply as $y_n = \mathbf{h}^T \mathbf{x}_n$. This notation will be convenient in discussing matched filters and array processing in <u>Chaps. 4</u> and <u>9</u>.

Two special product operations on vectors are common in some radar signal processing. The symbol \odot represents the *Hadamard product*, which is the element-by-element product of two vectors. Specifically, if **a** and **b** are two *N*-element column vectors, then **a** \odot **b** is the *N*-element column vector

$$\mathbf{a} \odot \mathbf{b} \equiv \begin{bmatrix} a_0 b_0 & a_1 b_1 & \cdots & a_{N-1} b_{N-1} \end{bmatrix}^T$$
(B.48)

The *Kronecker product*, denoted by the symbol \otimes , is the replication of one vector weighted by the elements of the other. If **a** is an *M*-element column vector and **b** is an *N*-element column vector, then **a** \otimes **b** is the *MN*-element column vector vector

$$\mathbf{a} \otimes \mathbf{b} \equiv \begin{bmatrix} a_0 \cdot \mathbf{b} \\ a_1 \cdot \mathbf{b} \\ \vdots \\ a_{M-1} \cdot \mathbf{b} \end{bmatrix}$$
(B.49)

It is frequently necessary to compute the power or energy of various signals in vector notation. The energy E_x and power P_x in a vector **x** are given by

$$E_{\mathbf{x}} = \mathbf{x}^{H}\mathbf{x} = \sum_{n=0}^{N-1} |x_{n}|^{2}, \quad P_{\mathbf{x}} = \frac{1}{N}E_{\mathbf{x}} = \frac{1}{N}\mathbf{x}^{H}\mathbf{x} = \frac{1}{N}\sum_{n=0}^{N-1} |x_{n}|^{2}$$
(B.50)

where the superscript *H* denotes the Hermitian (conjugate) transpose. The energy or power in the result of an inner product (filtering) operation $y = \mathbf{h}^T \mathbf{x}$ is

$$E_y = P_y = |y|^2 = y^* \cdot y^T = \mathbf{h}^T \mathbf{x}^* \mathbf{x}^T \mathbf{h}$$
(B.51)

 E_{y} and P_{y} are the same because y is scalar (N = 1).

For a random vector \mathbf{w} (noise samples, for example), the expected value is required to get meaningful expressions for energy or power:

$$E_{\mathbf{w}} = \mathbf{E}\{\mathbf{w}^{H}\mathbf{w}\}, \quad P_{\mathbf{x}} = \frac{1}{N}E_{\mathbf{x}} = \frac{1}{N}\mathbf{E}\{\mathbf{w}^{H}\mathbf{w}\}$$
(B.52)

The power or energy in a filtered noise vector $y = \mathbf{h}^T \mathbf{w}$ then becomes

$$E_y = P_y = \mathbf{E}\{\mathbf{h}^T \mathbf{w}^* \mathbf{w}^T \mathbf{h}\}$$
$$= \mathbf{h}^T \mathbf{E}\{\mathbf{w}^* \mathbf{w}^T\} \mathbf{h} = \mathbf{h}^T \mathbf{S}_{\mathbf{w}} \mathbf{h}$$
(B.53)

where the *covariance matrix* S_w has been defined as the outer product²

$$\mathbf{S}_{\mathbf{w}} \equiv \mathbf{E}\{\mathbf{w}^* \, \mathbf{w}^T\} \tag{B.54}$$

Covariance matrices have numerous useful properties. For example, S_w is Hermitian, meaning that $S_w = S_w^H$ [where *H* is the Hermitian (conjugate transpose) operator], and Toeplitz (all the elements on a diagonal are identical). The inverse is also Hermitian, $S_w^{-1} = (S_w^{-1})^H$. Some more important properties are deferred until after a brief discussion of eigenanalysis.

The last form of Eq. (B.53) is an example of a *quadratic form* of a matrix. The quadratic form of a Hermitian matrix \mathbf{A} with respect to a vector \mathbf{x} is the real-valued scalar quantity:

$$Q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^{H} \mathbf{A} \mathbf{x} = \sum_{i=0}^{N-1} \sum_{i=0}^{N-1} x_{i}^{*} a_{ij} x_{j}$$
(B.55)

The name arises because $Q_A(\mathbf{x})$ is a second-order polynomial in the elements of \mathbf{x} .¹⁰ If $Q_A(\mathbf{x}) > 0$ for any \mathbf{x} , \mathbf{A} is said to be *positive definite*. If $Q_A(\mathbf{x}) \ge 0$, \mathbf{A} is said to be *positive semidefinite*. Matrices can also be negative definite or negative semidefinite. A matrix that is none of these is called indefinite.

The signal-to-interference (SIR) ratio for a signal **x** and interference **w** is just $P_x/P_w = E_x/E_w$. If the input to a filter **h** is the sum of **x** and **w**, the SIR of the output *y* will be the ratio of the powers in the filtered signal and noise from Eqs. (B.51) and (B.53), which takes on the often-seen form

$$SIR = \frac{\mathbf{h}^T \mathbf{x}^* \mathbf{x}^T \mathbf{h}}{\mathbf{h}^T \mathbf{S}_{\mathbf{w}} \mathbf{h}}$$
(B.56)

B.5.2 Basic Eigenanalysis

Consider the following system of linear equations described by an $N \times N$ matrix **A**:

$$Ae = \lambda e \tag{B.57}$$

Any vector **e** which satisfies this equation is called an *eigenvector* of **A** and the scalar λ is called the corresponding *eigenvalue*. An eigenvector is thus a vector that is unmodified (to within a scale factor) when operated on by **A**. It is frequently the case that eigenvectors are normalized to unity norm, $\|\mathbf{e}\| = 1$.

Rearranging Eq. (B.57) gives

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{e} = 0$$

For Eq. (B.58) to hold for a nonzero vector **e**, it must be the case that $\mathbf{A} - \lambda \mathbf{I} = 0$ so that in turn

$$\det(\mathbf{A} - \lambda \mathbf{I}) = p(\lambda) = 0 \tag{B.59}$$

The *characteristic polynomial* $p(\lambda)$ will be an *N*th-order polynomial in λ . Its roots are the eigenvalues of **A**.

Following are some useful properties of eigenvalues and eigenvectors, including a number which apply to Hermitian matrixes and therefore to covariance matrices. Proofs of these properties can be found in Hayes (1996).

For any matrix A:

- Nonzero eigenvectors e_i corresponding to distinct eigenvalues λ_i are linearly independent.
- If the rank of A is M, there will be M nonzero eigenvalues and N M eigenvalues with a value of zero.

For a Hermitian matrix A (includes covariance matrices):

- The eigenvalues of A are real.
- If A is the covariance matrix of a wide-sense stationary process (normally a reasonable assumption), the eigenvalues are also nonnegative. Consequently, A is a nonnegative definite matrix.
- Eigenvectors corresponding to distinct eigenvalues are orthogonal, $\mathbf{e}_i^T \mathbf{e}_j = 0$.
- (Spectral theorem) A can be decomposed in terms of its eigenvectors and eigenvalues as

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{H} = \sum_{i=0}^{N-1} \lambda_{i} \mathbf{e}_{i} \mathbf{e}_{i}^{H}$$
(B.60)

where V is a matrix whose columns are the normalized eigenvectors and Λ is matrix having the eigenvalues on the diagonal and zero elsewhere.

If B = A + αI, then A and B have the same eigenvectors. The eigenvalues of B are λ_i + α, where the λ_i are the eigenvalues of A. This property is useful when considering covariances matrices of the sum of a signal with white noise.

B.5.3 Eigenstructure of Sinusoids in White Noise

Many important problems in radar model the signal as a sum of sinusoids in random interference. Examples include Doppler processing, adaptive beamforming, and space-time adaptive processing. Certain important classes of processing algorithms rely on the structure of the covariance matrix of such signals. A brief summary of the main results is given here. A much more complete but still concise discussion is given in Hayes (1996).

Assume a length-*N* signal x[n] is the sum of a signal component s[n] consisting of *K* complex sinusoids, and white noise w[n] of variance σ_w^2 :

$$x[n] = s[n] + w[n] = \sum_{i=0}^{K-1} A_i e^{j\omega_i n} + w[n]$$
(B.61)

The amplitudes A_i are complex with uncorrelated phases randomly distributed on $[-\pi, \pi]$. The frequencies ω_i and magnitudes $|A_i|$ are nonrandom but unknown. Given a vector **x** consisting of *N* consecutive samples of x[n], the covariance matrix of **x** will be the sum of the covariance matrices of **s** and **w**,

$$\mathbf{S}_{\mathbf{x}} = \mathbf{S}_{\mathbf{s}} + \mathbf{S}_{\mathbf{w}} = \sum_{i=0}^{K-1} P_{i} \boldsymbol{\omega}_{i} \boldsymbol{\omega}_{i}^{H} + \sigma_{w}^{2} \mathbf{I} = \Omega \mathbf{P} \Omega + \sigma_{w}^{2} \mathbf{I}$$
(B.62)

In this equation, **I** is the $N \times N$ identity matrix, $\omega_i = [1 e^{j\omega_i} e^{j2\omega_i} \dots e^{j(N-1)\omega_i}]^T$, $\Omega = [\omega_0 \dots \omega_{K-1}]$ is a rank K matrix, and $\mathbf{P} = \text{diag}\{P_0, P_1, \dots, P_K-1\}$ is a diagonal matrix of the sinusoid powers $P_i = |A_i|^2$. While the ω_i vectors will not generally equal the eigenvectors of \mathbf{S}_x , they will lie in the space spanned by those eigenvectors, which is called the *signal subspace*. The subspace spanned by the remaining eigenvectors is called the *noise subspace*. Because of the last property of covariance matrices enumerated above, the N - K noise subspace eigenvalues will equal σ_w^2 , while the K signal subspace eigenvalues will equal the eigenvalues of $\Omega P\Omega$ plus σ_w^2 .

Figure B.12 illustrates these properties for a signal that is the sum of 15 samples of each of five unit-amplitude complex sinusoids with random normalized frequencies and phases and unit-variance, zero mean complex Gaussian noise. Thus N = 15 and K = 5. The solid line with round markers indicates the eigenvalues of the theoretical covariance matrix of Eq. (B.62). The square markers are the eigenvalues of a covariance matrix estimated by averaging 1000 sample covariance matrices formed from simulated signals in noise. The eigenvalues of the simulated data match the theoretical eigenvalues well. Both show that there are 10 noise eigenvalues with values near one. The remaining five signal eigenvalues are larger as expected. The associated eigenvectors define the signal and noise subspaces.



FIGURE B.12 Example of theoretical (round markers) and simulated (square markers) signal and noise eigenvalues for five sinusoids plus white noise.

B-6.B Instantaneous Frequency

Consider a real or complex sinusoidal function with an arbitrary phase function $\psi(t)$, $x(t) = A\cos\psi(t)$ or $x(t) = A\exp[j\psi(t)]$. The *instantaneous frequency* of this waveform is defined as

$$\Omega_i(t) = \frac{d\psi(t)}{dt} \text{ radians per second, } \quad F_i(t) = \frac{1}{2\pi} \frac{d\psi(t)}{dt} \text{ Hz}$$
(B.63)

As an example, a conventional constant-frequency sinusoid $A\cos(2\pi F_0 t)$ has $F_i(t) = F_0$ Hz, as should be expected. Note that the phase function of a constant-frequency sinusoid is linear in time.

Instantaneous frequency is sometimes useful in understanding the behavior of more complex signals. Consider a signal having a quadratic phase function $\psi(t) = \alpha t^2 + \beta t + \chi$. The instantaneous frequency in radians per second is $\Omega_i(t) = \alpha t + \beta$. Thus, a quadratic phase sinusoid has a frequency that varies linearly with time, as in the linear FM "chirp" waveforms of <u>Chap. 4</u>. As a final example, consider the following sinusoid and its instantaneous frequency:

$$x(t) = A \sin[2\pi F_0 t + \alpha \sin(2\pi\beta t)] \implies \psi(t) = 2\pi F_0 t + \alpha \sin(2\pi\beta t)$$

$$F_i(t) = F_0 + \alpha\beta \cos(2\pi\beta t)$$
(B.64)

This waveform is a sinusoid whose frequency varies by $\pm \alpha\beta$ Hz around a nominal frequency of F_0 Hz, with that variation occurring at a rate of β cycles per second.

B-7.B Decibels

Many radar parameters are commonly expressed in decibel (dB) units. Examples include antenna gains, radar cross sections, signal-to-interference ratios, and sidelobe levels. The advantage of using decibel units is that it reduces the numerical range of parameters that have very large variations. RCS is a good example. Average RCS values of interest in radar can range from 10^{-5} m² for insects to over 10^6 m² for a large ship such as an aircraft carrier viewed at certain aspect angles. This range of 11 orders of magnitude becomes a range of only 110 dB, from -50 dB to +60 dB.

Decibels in radar are a measure of *relative* power of an electrical signal. Consider a signal having a power of P watts. The power in decibels is defined as

$$P (dB) = 10\log_{10}\left(\frac{P}{P_{\text{ref}}}\right)$$
(B.65)

where P_{ref} is the reference power level. A value P_{dB} in decibels is easily converted back to a value *P* in non-decibel units (often called "linear units") by inverting Eq. (B.65):

$$P = P_{\rm ref} \cdot 10^{P_{\rm dB}/10}$$
(B.66)

The units of dB are often modified to reflect the reference value used. For instance, if $P_{ref} = 1$ mW and P = 100 mW, then P might be expressed as 20 dBm, with the abbreviation dBm interpreted as "decibels relative to 1 milliwatt." Another example is dBW for decibels relative to 1 watt. It is often the case, especially in signal processing, that the absolute units of a signal are not important, just its numerical value. For this reason, the reference value P_{ref} is often taken equal to one and the units of the decibel value are left as just dB with no modifier.

In electrical engineering, signals of interest are often voltages. In accordance with Ohm's law, the power of a quantity V in volts is proportional to V^2 . A quantity expressed in volts can be converted between decibel and linear scales using the relations
$$P (dB) = 10 \log_{10} \left(\frac{|V|^2}{|V_{ref}|^2} \right) = 20 \log_{10} \left(\frac{|V|}{|V_{ref}|} \right)$$
$$|V| = |V_{ref}| \cdot 10^{P_{dB}/20}$$
(B.67)

The use of $|V|^2$ instead of just V^2 allows for the possibility that the voltage V is complex. Note that converting from decibels back to the linear scale can recover only the magnitude of V. The phase of a complex voltage is lost in the conversion to decibels, as is the sign of a real-valued voltage. An arbitrary signal x(t) or x[n], for instance the output of a radar receiver, is usually considered to be a voltage unless otherwise specified. There are exceptions. The output of a square-law detector (Chap. 6) would be considered to be in power units, while the units of the output of a logarithmic detector would be the logarithm of power.

In addition to signal powers, decibel scales are commonly used in radar to describe antenna pattern features, filter time- and frequency-domain response features, and radar cross sections. For example, the power gain at the peak of the first sidelobe, relative to the peak mainlobe power gain, of an antenna with an ideal uniform illumination is about 0.047, corresponding to -13.26 dB. This same value surfaces frequently, for instance as the peak sidelobe level of the slow-time DTFT of a constant-Doppler target or the peak range sidelobe level of a matched filter for a linear FM waveform with a high time-bandwidth product. RCS is defined in terms of the ratio of two values of power density (see Chap. 2), and so becomes a factor in power calculations in radar range equations. Because RCS has units of m², on a decibel scale RCS is often given units of decibels relative to 1 m², written dBm. The distinction between the usage of dBm for dB relative to 1 square meter and dB relative to 1 milliwatt is usually clear from context.

Expressing parameters in decibels simplifies multiplication, division, and exponentiation calculations, as shown here:

$$z = x \cdot y \iff z_{dB} = x_{dB} + y_{dB}$$

$$z = x/y \iff z_{dB} = x_{dB} - y_{dB}$$

$$z = x^{\alpha} \iff z_{dB} = \alpha \cdot x_{dB}$$
(B.68)

These transformations were of great utility before the age of handheld scientific calculators and high-speed computers, but are of less importance today. Nonetheless, it is useful to be familiar with a few key correspondences between linear and decibel values given in <u>Table B.2</u>.

Decibel Value	Linear Factor
0	1
1	≈1.25 (1.2589)
3	≈2 (1.9953)
10	10
10α (e.g., 30 if $\alpha = 3$)	10^{α} (e.g., 1000 if $\alpha = 3$)
-10α (-30 if α = 3)	$10^{-\alpha}$ (e.g., 0.001 if $\alpha = 3$)

 TABLE B.2
 Linear Equivalences to Selected Decibel Values

With this table and the arithmetic properties of decibels one can approximate the linear value of any parameter given in dB without resorting to a scientific calculator. For example, the linear equivalent of 7 dB can be determined as

$$\alpha_{dB} = 7 \text{ dB} = 3 \text{ dB} + 3 \text{ dB} + 1 \text{ dB} \implies \alpha = 2 \cdot 2 \cdot 1.25 = 5$$
(B.69)

which is quite close to the actual linear value corresponding to 7 dB of 5.0119.

References

- Bracewell, R. N., *The Fourier Transform and Its Applications*, 3d ed. McGraw-Hill, New York, 1999.
- Ercegovac, M., and T. Lang, *Digital Arithmetic*. Morgan-Kauffman, San Francisco, CA, 2003.
- Hayes, M. H., *Statistical Digital Signal Processing and Modeling*. Wiley, New York, 1996.
- Johnson, D. H., and D. E. Dudgeon, Array Signal Processing: Concepts and Techniques. Prentice-Hall, Englewood Cliffs, NJ, 1993.
- McClellan, J. H., and R. J., Purdy, "Applications of Digital Signal Processing to Radar," Chap. 5 in *Applications of Digital Signal Processing*, A.V. Oppenheim (ed.). Prentice-Hall, Englewood Cliffs, NJ, 1978.
- Merkel, K. G., and A. L., Wilson, "A Survey of High Performance Analog-to-Digital Converters for Defense Space Applications," *Proceedings 2003 IEEE Aerospace Conference*, vol. 5, pp. 2415–2427, Mar. 8–15, 2003.
- Oppenheim, A. V., and R. W. Schafer, *Discrete-Time Signal Processing*, 3d ed. Pearson, Englewood Cliffs, NJ, 2010.
- Papoulis, A., *The Fourier Integral and Its Applications*, 2d ed. McGraw-Hill, New York, 1987.
- Walden, R. H., "Analog-to-Digital Converter Survey and Analysis," IEEE

Journal on Selected Areas in Communications, vol. 17, no. 4, pp. 539–550, Apr. 1999.

Walden, R. H., "Analog-to-Digital Conversion in the Early 21st Century," in Encyclopedia of Computer Science and Engineering, B. Wah (ed.). Wiley, New York, 2008.

³ Another fairly common DSP convention observed here is to use upper case variables to denote analog frequencies (F or Ω , in hertz or radians per second if the independent variable represents time) and lower case variables to denote "digital" frequency variables (f or ω). Many DSP texts consider the signal domain index n to be unitless and therefore units of f or ω to be simply cycles or radians. The preference here is to consider n to have units of samples so that f and ω have units of cycles per sample or radians per sample.

⁴ The definitions of Eqs. (B.14) and (B.15) are unusual in that most DSP texts define the DFT with K = N and introduce the idea of "zero padding" to account for the case where there are more DFT frequency samples than there are signal samples. The preference here is to distinguish these two numbers in the definition.

 $\frac{5}{2}$ The MATLAB® function fftshift exists precisely to perform this rearrangement of the output of the fft function for plotting convenience.

 $\frac{6}{2}$ "Nonlinear quantization" using unevenly spaced values of output voltage is common in some applications. Linear quantization (constant level spacing) is the norm in radar.

^{$\frac{7}{2}$} Most radar and array processing literature uses a lower case *k* for spatial frequency. Here upper case *K* is used in keeping with the convention to use uppercase letters for analog quantities and lower case for quantities normalized by a sampling interval.

⁸ Incidence angle will often be measured with respect to the normal to the y axis, i.e., the x axis, because this is convenient and conventional in analyzing antenna patterns. If the antenna aperture lies in the y dimension, then an incidence angle of $\theta = 0$ indicates a wave propagating normal to the aperture, i.e., in the boresight (x) direction.

 $\frac{9}{2}$ Strictly speaking, this is the *autocorrelation matrix*. The covariance matrix is obtained as the autocorrelation of the interference after the mean is subtracted out; see Hayes (1996) for details. The interference may almost always be assumed to have zero mean, in which case they are the same. The covariance terminology seems to be more prevalent in the radar community.

¹⁰ Since Q must be real, the second-order terms in the polynomial all appear in magnitude-squared form, $|x_i|^2$.

¹ To unify discussion of sampling in time, frequency, and space and to maintain generality, in this discussion the signal to be sampled will be referred to as a function in the *signal domain*, and its Fourier transform as a function in the *Fourier domain*.

² The fairly common convention in digital signal processing texts of enclosing the independent variable in square brackets when it is discrete, and parentheses when it is continuous, will be followed. Thus x(u) is a function of a continuous variable, while x[n] is a function of a discrete variable.

Index

— A — α - β filter correction stage prediction stage velocity estimates A/D conversion. See Analog-to-digital conversion Accuracy Adaptive displaced phase center antenna. See Displaced phase center antenna Additive white Gaussian noise (AWGN) Albersheim's equation Aliased sinc function. See Asinc function Ambiguity: Doppler. See also Doppler spectrum function. See Ambiguity function range resolution. See Ambiguity resolution velocity. See also Ambiguity, Doppler Ambiguity function complex ambiguity function of Costas waveform definition ideal of linear FM pulse properties of pulse burst of simple pulse of stepped chirp waveform thumbtack Ambiguity resolution Chinese remainder theorem coincidence algorithm congruences ghosts AN/FPS-108 radar Analog-to-digital conversion Angle estimation 2D angle lobing-based lobing bound

maximum likelihood phase-based Angle of arrival (AOA) Antenna: area constraint in SAR array array element spacing array factor beamwidth boresight effective aperture fan beam gain grating lobes monopulse one-way pattern pencil beam phase center phase front phased array power gain power pattern side lobes spatial spectra tapering and shading two-way pattern voltage pattern Asinc function ASR-9 radar ASR-12 radar Atmospheric attenuation Auction methods, for cost measurement Autocorrelation matrix Autofocus map drift algorithm phase gradient algorithm Automatic detection. See Constant false alarm rat detection Autoregressive (AR) model Azimuth angles

Bandwidth:

3-dB Doppler. See Doppler bandwidth expansion null-to-null Rayleigh root-mean-square (RMS) of a simple pulse Bandwidth-time product. See Time-bandwidt product Barker coded waveform. See Biphase-code waveform Barker codes **Baseband** signal **Baseline** decorrelation Batch processing **Bayesian** estimation Beam limited resolution Beamforming adaptive beamspace processing degrees of freedom distortionless element space jammers matched filter number of jammers cancelled nulls preprocessing signal-to-interference ratio (SIR) **Billingsley model** Biphase-coded waveform ambiguity function of Barker code Barker codes combined Barker code matched filter minimum peak sidelobe codes (MPS) nested Barker code pseudorandom sequences Bistatic radar Blind Doppler shift Blind velocity Blind zone maps

Cell-averaging CFAR: analysis of CFAR loss concept effect of false alarms extensions of greatest-of cell-averaging. See Greatest-of cell-averaging CFAR least-of cell-averaging. See Smallest-of cell-averaging CFAR limitations masking masking loss performance self-masking smallest-of cell averaging. See Smallest-of cell-averaging CFAR target masking threshold multiplier, linear detector threshold multiplier, square law detector Central chi distribution Central reference point Centroid interpolation Chaff Channel mismatch. See I/Q errors Characteristic function (CF) Chinese remainder theorem Chip Chirp. See Linear FM waveform Circular random process Circulator Clutter altitude line area Billingsley model compound RCS models constant gamma model Doppler spectrum edges GTRI model internal motion k distribution log-normal main lobe matched filters for suppression number flucutations

power spectrum sensitivity to threshold side lobe signal-to-clutter ratio temporal and spatial correlation volume Weibull Clutter fill pulses Clutter map **CFAR** in moving target detector threshold Clutter mapping Coherent change detection Coherent processing interval (CPI) Coherent scattering Coherent signals Conical scan Constant false alarm rate detection (CFAR) adaptive CFAR adaptive matched filter cell-averaging. See Cell-averaging CFAR cell under test (CUT) censored CFAR CFAR loss CFAR window for clutter map distribution-free guard cells lagging window leading window log CFAR non-parametric order statistic. See Order statistic CFAR power on false alarm probability rank-based. See Order statistic CFAR reference cells reference window switching trimmed mean two-parameter Continuous wave Correlation:

autocorrelation cross-correlation and linear estimation normalized Correlator performance estimate Cosine function Covariance matrix of clutter ofjammers of preprocessed data rank of space-time snapshot taper Cramèr-Rao Lower Bound (CRLB) complex signals in white Gaussian noise definition on estimator variance for transformed parameters signals in white Gaussian noise signal-to-noise ratio for sinusoid parameters time delay for transformed parameters Cross-power spectrum Cross-range resolution Cumulative distribution function Cumulative probability

____ D ____

Data association: auction methods gating global nearest neighbor nearest neighbor probabilistic data association (PDA) strongest neighbor Data processing Data turning Datacube Dechirp processing. *See* Stretch processing Decibels Decimated signal Decimation Deramp processing. See Stretch processing Detection Bayesian approach in coherent receivers coincidence detection complex constant-in-Gaussian example constant false alarm rate (CFAR). See Constant false alarm rate detection detector loss fluctuating targets likelihood ratio test. See Likelihood ratio test linear detector *M*-of-*N* detection Marcum's Q function Neyman-Pearson criterion nonfluctuating targets null hypothesis threshold real constant-in-Gaussian example **SNR** loss square law detector sufficient statistic. See Sufficient statistic Swerling 1 example Swerling 2 example Swerling cases threshold uniformly most powerful test (UMP) unknown parameters unknown phase example Difference signal in angle tracking Digital I/Q Lincoln Laboratory method Rader's method Digital IF. See Digital I/Q Digital signal processing Digital sinc function. See Asinc function Digital Terrain Elevation Data (DTED) Dirichlet function. See Asinc function Discrete cosine transform Displaced phase center antenna processing (DPCA) adaptive relation to STAP time slip Doppler:

ambiguity blind minimum detectable mismatch spatial tolerance unambiguous Doppler bandwidth due to intrinsic clutter motion due to platform motion Doppler beam sharpening (DBS) azimuth dechirp beam sharpening ratio cross-range/Doppler mapping geometric distortion quadratic phase errors range curvature range migration correction range walk limit secondary range compression Doppler bins Doppler estimation a priori bound CRLB discrete Fourier transform interpolation maximum likelihood sampling bound Doppler interpolation quadratic via DFT Doppler processing Doppler resolution Doppler shift cone angle estimation. See Doppler estimation spatial Doppler stop-and-hop approximation Doppler signal estimator angle estimators CRLB for sinusoid parameters Doppler spectrum: autoregressive model

Billingsley model characteristics clear region clutter region cubic model effect of platform motion Gaussian model micro moving platform effects sampling skirt region spectral width unambiguous width Downchirp Duplexer Dwell Dynamic range

— E —

Early-late technique Edge threshold detector Efficient estimator Eigenanalysis basic of sinusoids in white noise Eigenvector Electromagnetic interference (EMI) Elevation angle Error function complementary Estimator accuracy angle. See Angle estimation consistent Cramèr-Rao Lower Bound (CRLB) Doppler signal. See Doppler estimation efficient frequency. See Doppler estimation maximum likelihood minimum variance precision properties

range. *See* Range estimation sequential least squares time delay. *See* Range estimation unbiased Extended Kalman filter (EKF)

— F —

False alarm Fast Fourier transform (FFT) Fast time Feature-aided tracking FFT. See Fast Fourier transform Filter splitting interpolation method Fine Doppler estimation Finite impulse response FIR filter. See Finite impulse response Fisher information matrix Flat earth phase difference Fourier transform continuous derivative property in digital signal processing discrete discrete time fast. See Fast Fourier transform reciprocal spreading Frequency agility Frequency estimation error Frequency models Frequency-modulated continuous wave (FMCW)

— G —

Gain: antenna coherent integration noncoherent integration signal processing. *See* Signal processing gain Gain scheduling Gating Gaussian model Ghosts Glint Global nearest neighbor, cost measurement Global positioning system Greatest-of cell-averaging CFAR CFAR loss clutter edge performance target masking performance threshold multiplier Ground moving target indication (GMTI) Ground plane

— H —

Hadamard product Hamming window Hermitian High resolution range profile History of radar. *See* Radar history Hypothesis testing

I channel. See In-phase channel I/O errors in A-DPCA amplitude only correcting image component phase only I/Q imbalance. See I/Q errors IF sampling. See Digital I/Q Infinite impulse response IIR filter. See Infinite impulse response Imaging. See also Synthetic aperture radar Improvement factor autocorrelation method frequency domain method vector method In-phase channel Inertial measurement unit (IMU) Inertial navigation system (INS) Instantaneous frequency Integration binary coherent

noncoherent Integration gain coherent noncoherent Interacting multiple models (IMM) Interferometric phase difference (IPD) Interferometric SAR absolute height estimation baseline coherent change detection concept effect of height layover one-pass orthorectification phase estimation phase unwrapping pixel phase processing steps registration terrain motion mapping two-pass Internal clutter motion. See Clutter Interpolation: bandlimited kernel Doppler. See Doppler interpolation range spatial frequency two-point

_____J_____

Jamming

— K —

K distribution Kalman filter α - β filter as extended properties Kronecker product

Lagrange multipliers

Least-of cell-averaging CFAR. See Smallest-of cell-averaging CFAR Least-squares estimate (LSE) Likelihood ratio test (LRT) generalized likelihood ratio test log likelihood ratio test Linear detector Linear FM waveform ambiguity function definition range-Doppler coupling Rayleigh resolution Linear time-invariant system effect on random signals Lobe switching Lobing conical scan monopulse sequential Lobing bound Loss: amplitude antenna array steering atmospheric. See Atmospheric attenuation brightness CFAR. See Constant false alarm rate detection conversion detector mismatch. See also Doppler, mismatch processing. See Processing loss in processing gain. See Loss in processing gain propagation. See Atmospheric attenuation SNR straddle Loss in processing gain (LPG) Low Earth orbit

— M —

M-of-*N* method Mahalanobis distance Major-minor method Maneuverability index Marcum's Q function. See Detection Masking Matched filter all-range in beamforming clutter suppression impulse response impulse response shaping for moving target indication (MTI) for moving targets output peak SNR for pulse burst pulse-by-pulse processing. See Pulse-by-pulse processing in range estimation range resolution received pulse for simple pulse in stripmap SAR vector formulation whitening Maximum likelihood estimate (MLE) of clutter edge location of interference power of sinusoid parameters of phase gradient in probability density functions Mean-square error (MSE) bound Measurement precision Minimum detectable Doppler. See Space-time adaptive processing Minimum detectable velocity. See Space-time adaptive processing Minimum norm methods, phase unwrapping Minimum variance unbiased (MVU) estimator Monopulse tracking Monostatic radar Moore's law Motion compensation autofocus. See Autofocus global positioning system (GPS) inertial measurement unit (IMU) inertial navigation system (INS) Kalman filter phase errors PRI control

to a point Moving target detector (MTD) Moving target indication (MTI) blind speeds clutter attenuation combined with pulse Doppler concept figures of merit ground MTI improvement factor. See Improvement factor interference covariance matrix limitations. See Moving target indication limitations in moving target detector optimum filter pulse cancellers signal model as slow time highpass filter subclutter visibility surface MTI transient effects visibility factor Moving target indication limitations amplitude instability intrinsic clutter motion phase instability platform motion MTI. See Moving target indication Multipath Multiple hypothesis tracking Munson, David

N

Nearest neighbor, cost measurement NEXRAD Neyman-Pearson criterion. *See* Detection Noise DFT of effective temperature figure power spectrum of in quadrature receivers types Noise equivalent bandwidth Noise power Noise subspace Noisy measurement Noncoherent scattering Null hypothesis. *See* Detection Number fluctuations model Nyquist rate in 3-dB beamwidths in angle of arrival in Doppler in fast time in range in sine of angle of arrival Nyquist theorem

____0 ____

Order statistic Order statistic CFAR CFAR loss performance threshold multiplier in two-parameter clutter Oscillator local stable local

— P —

Path following method, phase unwrapping Phase history Phase unwrapping Phenomenology Pilot signal Points scatterer response (PSR) Polar format algorithm angular-range interpolation keystone grid polar-to-rectangular interpolation radial-keystone interpolation rectangular format algorithm Polarization scattering matrix Polyphase coded waveform

ambiguity function Frank codes P3 codes P4 codes precompression bandlimiting Power spectral density (PSD) Power spectrum Precision Principle of stationary phase Probabilistic data association Probability density functions central chi chi-squared circular Gaussian Erlang exponential Gamma Gaussian k distribution log-normal noncentral chi noncentral chi-squared for phase for power Rayleigh Rice Rician Tikhonov uniform for voltage Weibull Probability distributions in radar central chi K distribution log-normal noncentral chi-squared rice voltage Weibull Probability of detection Probability of false alarm Probability of miss Process noise

Processing gain Processing loss (PL) Projection-slice theorem Pulse burst waveform ambiguity function Doppler response relation of ambiguity function to slow time spectrum Pulse-by-pulse processing Pulse cancellers. See Moving target indication Pulse compression Pulse Doppler processing as bandpass filter bank combined with MTI **CPI-to-CPI stagger** DFT of Noise DTFT of moving target DTFT sampling fine Doppler estimation. See Doppler interpolation high PRF operation low PRF operation as matched filter medium PRF operation modern spectral estimation in in moving target detector **PRF** regimes signal processing gain steady-state operation transient effects Pulse pair processing autocorrelation method frequency domain method spectral subtraction spectral width Pulse repetition frequency Pulse repetition interval

Q channel. *See* Quadrature channel Q function. *See* Detection Quadratic interpolation Quadrature channel Quantization in digital signal processing Quantization error

— R —

Radar: **AN/FPS-108** antenna. See Antenna applications ASR-9 ASR-12 basic functions bistatic Chain Home continuous wave frequency bands imaging. See Synthetic aperture radar literature monostatic **NEXRAD** probability distributions in range equation. See Radar range equation receiver. See Receiver SCR-270 signal components SIR-C radar synthetic aperture. See Synthetic aperture radar systems and components time scales transmitter Radar cross section (RCS) of clutter. See Clutter common PDFs for RCS for voltage compound models convolutional models correlation in angle in frequency pulse-to-pulse scan-to-scan definition

meteorological spatial models spectral model statistical description statistical models. See Target models Swerling models. See Swerling models typical values Radar history Chain Home Hülsmeyer, Christian Page, Robert **Radiation Laboratory SCR-270** Taylor, Albert Watson-Watt, Sir Robert Young, Leo Radar range equation for area clutter for distributed targets for point scatterers for volume clutter Radial velocity Radio frequency interference (RFI) Range: ambiguity. See Ambiguity bins cells gates swath unambiguous window Range-Doppler algorithm narrowband approximation Range-Doppler coupling. See Linear FM waveform Range estimation a priori bound CRLB interpolation leading edge MLE MSE bound sampling bound signal model

split-gate Range migration range curvature range walk Range profile Range resolution Range side lobe control: for linear FM waveform by matched filter design minimum peak sidelobe codes mismatched phase codes by waveform spectrum shaping Range skew Range swaths Range window Rank sum double quantizer Real aperture radar Real-beam imaging Receiver effective temperature superheterodyne Receiver operating characteristic (ROC) Receiver output Reed-Mallett-Brennan (RMB) rule Reflectivity angle-averaged effective area clutter baseband complex effective factor projections range-averaged effective volume clutter Reflectivity factor Region of interest (ROI) Replication Residual video phase. See Spotlight SAR Resolution: 3-dB cell cross-range Doppler range

Rayleigh velocity Rice distribution Root-mean-square (RMS) bandwidth

____ S ____

Sampling: in angle array element spacing in digital signal processing in Doppler frequency in fast time Nyquist rate in slow time spatial Sampling bound (SB) ScanSAR Schwarz inequality Sequential least squares Sequential lobing Shnidman's equation Shuttle imaging radar-C. See SIR-C radar Sign-magnitude encoding Signal processing gain in coherent integration for coherent pulse burst for linear FM waveform in moving target indication in noncoherent integration for phase coded waveforms in pulse Doppler processing in synthetic aperture radar Signal subspace Signal-to-clutter ratio (SCR) Signal-to-interference ratio (SIR) Signal-to-noise ratio (SNR) CRLB peak precision Signal-to-quantization noise ratio (SQNR) Sinc-squared two-way voltage Sinusoid frequency

SIR-C radar Slant plane Slant range Slow time Smallest-of cell-averaging CFAR CFAR loss clutter edge performance target masking performance threshold multiplier Snapshot: space-time spatial temporal Space-time adaptive processing (STAP) adapted pattern **Billingsley model** clutter model clutter ridge slope column stacking covariance matrix. See Covariance matrix covariance matrix taper (CMT) diagonal loading interference subspace leakage knowledge-aided limitations metrics minimum detectable Doppler (MDD) minimum detectable velocity (MDV) principle components method rank of covariance matrix reduced dimension reduced rank Reed-Mallet-Brennan rule relation to DPCA sample matrix inverse signal model signal processing gain SIR loss subspace projection temporal correlation test statistic usable Doppler space fraction (UDSF) **Spatial Doppler**

Spatial frequency Spatial snapshot Split-gate technique Spotlight SAR central reference point convolution-backprojection cross-range resolution data characteristics limitations to polar format model mapping rate polar format algorithm. See Polar format algorithm polar format data projection rectangular format algorithm residual video phase sampling requirements Squint Staggered PRF: **CPI-to-CPI** pulse canceller frequency response pulse-to-pulse stagger ratio staggers Stationary point Steady-state operation Steering vector: space-time spatial temporal Stepped chirp waveform, ambiguity function Stepped frequency waveform ambiguity function array steering definition linear pulse-by-pulse processing Stop-and-hop assumption. See Doppler shift Straddle loss definition in Doppler sampling rate for specified limit Stretch processing Stripmap SAR

 ω - κ algorithm chirp scaling algorithm cross-range chirp data characteristics data set size depth of focus geometry lower bound on cross-range resolution mapping rate point scatterer response (PSR) range migration algorithm space-bandwidth product subswaths Strongest neighbor, cost measurement Sufficient statistic Sum signal, in angle tracking Surface moving target indication (SMTI) Swath length Swerling models Synthetic aperture radar antenna area constraint aperture time autofocus. See Autofocus beamwidth of synthetic array coverage and sampling cross-range resolution. See Resolution Doppler beam sharpening. See Doppler beam sharpening Doppler viewpoint interferometric SAR. See Interferometric SAR layover looks mapping rate motion compensation. See Motion compensation point scatterer response (PSR). See Stripmap SAR quadratic phase errors range curvature. See Range migration range-Doppler algorithm range migration. See Range migration range walk. See Range migration scanSAR sidelooking speckle speckle reduction

spotlight. See Spotlight SAR squinted SAR stripmap. See Stripmap SAR swath constraint swath length synthetic aperture size Synthetic array radar. See Synthetic aperture radar System loss factor

— T —

Target masking Target models: fluctuating effects on Doppler spectrum nonfluctuating Swerling. See Swerling models Target visibility. See Moving target indication, visibility factor Terrain motion mapping Threshold detection. See Detection Time-bandwidth product Time delay estimation Time delay measurement Tracking α - β filter association. See Data association cost matrix measurement cvcle extended Kalman filter (EKF) filters gain scheduling initiation initiation process Kalman filter least-squares estimate in spherical coordinates state Tracking index Transmit/receive switch Two-beam lobing Two's complement encoding Two-sample estimator

_____ U ____

Unambiguous range. See Range Upchirp

_____ V _____

Vector representation

Walker, Jack Waveform: biphase-coded. See Biphase-coded waveform Costas frequency codes linear FM (LFM). See Linear FM waveform nonlinear FM (NLFM) phase coded polyphase coded. See Polyphase coded waveform pulse burst. See Pulse burst waveform simple pulse stepped frequency. See Stepped frequency waveform Wavefront. See Antenna, phase front Weiss-Weinstein analysis White noise Whitening Wiley, Carl

____ Z ____

Zero padding Zero velocity filter Ziv-Zakai bound (ZZB)