Cyclostationarity in Communications and Signal Processing

Edited by

William A. Gardner

IEEE PRESS

IEEE Communications Society, *Sponsor*

The Institute of Electrical and Electronics Engineers, Inc., New York

This book may be purchased at a discount from the publisher when ordered in bulk quantities. For more information contact:

IEEE PRESS Marketing
Attn: Special Sales
P.O. Box 1331
445 Hoes Lane
Piscataway, NJ 08855-1331

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-7803-1023-3
IEEE Order Number: PC0373-1

Library of Congress Cataloging-in-Publication Data

Cyclostationarity in communications and signal processing / edited by William A. Gardner.
p. cm.
Includes bibliographical references and index.
ISBN 0-7803-1023-3
1. Signal processing—Mathematics. 2. Cyclostationary signals--Mathematical models. 3. Spectral theory (Mathematics) I. Gardner, William A.
TK5102.9.C94 1994
621.382'2—dc20 93-21059
CIP

Contents

PART I

Editor's Introduction

Many conventional statistical signal-processing methods treat random signals as if they were statistically stationary, that is, as if the parameters of the underlying physical mechanisms that generate the signals do not vary with time. But for most man-made signals encountered in communication, telemetry, radar, and sonar systems, some parameters do vary periodically with time. In some cases even multiple incommensurate (not harmonically related) periodicities are involved. Examples include sinusoidal carriers in amplitude, phase, and frequency modulation systems, periodic keying of the amplitude, phase, or frequency in digital modulation systems, periodic scanning in television, facsimile, and some radar systems, and periodic motion in rotating machinery. Although in some cases these periodicities can be ignored by signal processors, such as receivers that must detect the presence of signals of interest, estimate their parameters, and/or extract their messages, in many cases there can be much to gain in terms of improvements in performance of these signal processors by recognizing and exploiting underlying periodicity. This typically requires that the random signal be modeled as *cyclostationary* or, for multiple periodicities, *polycyclostationary*, in which case the statistical parameters vary in time with single or multiple periods. Cyclostationarity also arises in signals of natural origins, because of the presence of rhythmic, seasonal, or other cyclic behavior. Examples include time-series data encountered in meteorology, climatology, atmospheric science, oceanology, astronomy, hydrology, biomedicine, and economics.

Important work on cyclostationary processes and time-series dates back over three decades, but only recently has the number of published papers in this area grown exponentially. Fueled by recent advances in applications to communications, signal processing, and time-series analysis that demonstrate substantial advantages of exploiting cyclostationarity in both design and analysis, the appetite for learning about cyclostationarity exhibited by research and development communities in areas such as wireless and cable communications, signals intelligence and covert communications, and modeling and prediction for natural systems (hydrology, climatology, meteorology, oceanology, biology/medicine, economics, etc.) has outgrown the available tutorial literature. This edited book is intended to help fill this void by presenting individual tutorial treatments of the major subtopics of cyclostationarity and by featuring selected articles that survey the latest developments in various specific areas.

The book is composed of two parts. Part I consists of six chapters, the first four of which are adapted from the four plenary lectures at the Workshop on Cyclostationary Signals, which was held August 16–18, 1992, at the Napa Valley Lodge in Yountville, California. Part II consists of seven articles. Part I is strongly tutorial and provides in-depth surveys of major areas of work. Similarly, Part II, which focuses on more specific topics, also has a tutorial survey flavor. Each of these two parts treats both theory and application.

Chapter 1 provides a historical perspective on cyclostationarity and discusses in detail both the practical and mathematical motives for studying cyclostationarity. It also treats the philosophy of aesthetics and utility that underlies alternative conceptual/mathematical frameworks within which theory and method can be developed. The latter half of the chapter surveys the theory and application of wide-sense cyclostationarity, touching on the problems of detection, modulation recognition, source-location, and extraction of highly corrupted signals, and the roles that the spectral-line generation and spectral-redundancy properties of cyclostationarity play in tackling these and other problems. This chapter provides an introduction to cyclostationary signals that serves as a foundation for the remainder of the book.

Chapter 2 provides an overview of the recently formulated theory of higher-order temporal and spectral moments and cumulants of cyclostationary time-series. It is shown that the nth-order polyperiodic cumulant of a polycyclostationary time-series is the solution to the problem of characterizing the strengths of all sine waves that are produced by multiplying n delayed versions of the time-series together, with the parts of those sine waves that result from products of sine waves that are present in lower-order factors of the nth-order product removed. Thus, the study of higher-order cumulants is motivated by a practical problem that arises in signal processing. The chapter also discusses other motivations for studying the moments and cumulants and provides a historical account of cumulants and their uses. The properties of these statistical functions that render them useful in signal processing are discussed and compared to the properties of similar statistical functions for stationary time-series. Applications of the unique signal-selectivity property of the polyperiodic cumulants to the tasks of weak-signal detection and source location are briefly described.

Chapter 3 provides an overview of sensor array processing for cyclostationary signals, focusing on adaptive spatial filtering and direction-of-arrival estimation, and presenting some new results on blind equalization and channel identification. It briefly describes many recently introduced methods and highlights their advantages

and disadvantages relative to each other and to more conventional techniques that ignore cyclostationarity. Applications of cyclostationarity-exploiting methods to existing problems in array processing and to the design of new wireless communication systems are suggested.

Chapter 4 supplements the material on cyclostationary processes by reviewing the basic theory of periodically and polyperiodically time-varying linear systems. Such systems are extensively employed as filters for processing and modeling cyclostationary signals. Various input-output and state-variable descriptions together with filter structures that are appropriate for implementing the desired response characteristics in both continuous- and discrete-time are discussed. The chapter concludes with a brief discussion and some examples of polyperiodic filtering for waveform extraction.

Chapter 5 provides an overview of the state-space theory of cyclostationary processes in discrete time. The three alternative descriptions, (i) jointly periodic autocovariance functions, (ii) state-space stochastic models (Markovian representations), and (iii) autoregressive moving average models with periodic coefficients, are investigated, and connections among them are explained. Innovations representations, linear prediction, spectral factorization, and model identification are all studied and the current state of knowledge on these topics is summarized.

Chapter 6 provides a review of the spectral theory of cyclostationary (periodically and almost periodically correlated) random processes and of existing results on the consistent estimation of the Fourier coefficients of the autocorrelation function and their Fourier transforms, the spectral correlation densities. The representation of these processes in terms of sets of jointly stationary processes and in terms of unitary operators also is reviewed.

Article 1 in Part II addresses the joint transmitter/receiver optimization problem for multiuser communications and presents a coherent view of system design approaches that include different but related multiinput/multioutput models on the basis of analytical optimization. The present state of knowledge in this area is summarized, and the potential for suppression of cochannel interference that is afforded by the cyclostationarity of the signals is emphasized. The results demonstrate analytically that greatly improved cross-talk rejection is achievable when the spectral correlation property of the cyclostationary signals is properly exploited.

In Article 2 the objective is to provide insight into the nature of the self-noise that is present in the timing wave produced by a square-law synchronizer acting on a cyclostationary pulse-amplitude modulated signal and to provide a quantitative analysis of the mean square phase jitter in the timing wave. The results obtained show explicitly how the design and performance analysis of the square-law synchronizer is characterized by the spectral correlation function and the fourth-order spectral-moment function of the signal.

Article 3 provides a tutorial review of recent methods for multipath channel identification using known test signals. By exploiting the signal-selectivity properties of the cyclic autocorrelation function or the associated spectral correlation function, these methods can perform well in severely corruptive noise and interference environments. Several such identification methods are compared in terms of their performance characteristics by analysis and simulation.

Article 4 provides a brief overview of the various approaches to blind channel equalization and identification that have been reported in the literature and then explains the potential advantages to be gained by exploiting the cyclostationarity of digital-quadrature-amplitude-modulated signals. The theoretical possibility of accomplishing blind identification with the use of only second-order statistics is explained, and a frequency-domain approach is described.

The fifth article presents a time-domain approach to the blind channel equalization and identification problem. The results of simulations presented therein suggest that exploitation of second-order cyclostationarity can be an effective alternative to methods that ignore it in favor of higher-than-second-order statistics. A connection between the frequency-domain and time-domain approaches also is explained.

Article 6 reviews the theory and implementation of digital spectral correlation analysis. The performance characteristics and computational requirements of various algorithms based on either time smoothing or frequency smoothing are compared analytically, and two specific implementation studies are briefly presented.

Article 7 briefly reviews recent developments in the theory of prediction for cyclostationary processes. The fundamental role in the theory played by multivariate stationary representations of univariate cyclostationary processes is explained, and both discrete-time and continuous-time processes are considered.

The chapters in Part I and articles in Part II collectively cover a wide range of topics in the theory and application of cyclostationarity. We hope that the tutorial style of these contributions coupled with the broad survey and comprehensive reference lists they provide will make this volume instrumental in furthering progress in understanding and using cyclostationarity not only in the fields of communications and signal processing, but in all fields where cyclostationary data arises.

ACKNOWLEDGMENTS

The editor gratefully acknowledges the sponsorship of the 1992 Workshop on Cyclostationary Signals by the National Science Foundation (Grant # MIP-91-12800), the Army Research Office (Grant # DAAL03-92-G-0297), the Air Force Office of Scientific Research (Grant # F49620-92-J-0303), and the Office of Naval Research (Grant # N0001492-J-1218). This workshop motivated this volume and most of the contributions contained herein grew out of presentations made at this workshop.

PART I

Chapter 1

An Introduction to Cyclostationary Signals

William A. Gardner
Statistical Signal Processing, Inc.
www.sspi-tech.com
1909 Jefferson Street
Napa, CA 94559

This introductory chapter has five objectives. The first is to give a broad but thorough view of the status of the development of the theory and application of cyclostationary signals. This entails discussing and answering the following questions: What is cyclostationarity? How is it useful? Why publish a book on cyclostationarity? What are some of the seminal contributions to the study of cyclostationarity? and What are some of the specific motivations—both practical and mathematical—for studying cyclostationarity?

The second objective is to explain the philosophies behind the two alternative approaches to the subject: the orthodox approach based on stochastic processes and ensemble averaging and the more recently developed approach based on nonstochastic time-series and time averaging. Since some controversy regarding these two approaches is said to exist (it is more misunderstanding than it is controversy), the

This chapter is adapted from the opening plenary lecture at the Workshop on Cyclostationary Signals, held August 16-18, 1992 in Yountville, CA. The reference style (author(s), date(s)) is used in this chapter to help the reader put the contributions surveyed into historical perspective. In the remainder of the book, references are identified by number according to the order listed at the end of each chapter and article.

discussion here is intended to be particularly thorough, including both pragmatic and mathematical arguments and illuminating both strengths and weaknesses of each approach. The goal is to provide a sound basis for choice for everyone interested in studying cyclostationarity.

The third objective is to provide a comprehensive introduction to the principles of second-order cyclostationarity, which involve only second-order statistical moments of signals in the time and frequency domains. This treatment considers primarily discrete-time signals, and in this way it complements previous treatments by this author, which focus on continuous-time signals.

The fourth objective of this chapter is to survey applications of second-order cyclostationarity in the areas of communications and signal processing. The focus here is on exploiting the spectral redundancy and sine-wave generation properties of cyclostationary signals to perform difficult signal-processing tasks.

The fifth objective is to provide a reasonably comprehensive bibliography of work on the theory and application of cyclostationarity (which is complemented by the more focused bibliographies in subsequent chapters and articles).

Altogether, this chapter provides a foundation for the rest of the book that will help the reader to put each individual contribution into perspective and to integrate the parts into a whole reference source that not only will chart the past, but also will serve as a primary vehicle for taking us into the future.

1 BACKGROUND

1.1 What Is Cyclostationarity?

Let us begin with the most obvious question: "What is a *cyclostationary* signal?"[1] One answer is that a signal is cyclostationary of order n (in the wide sense) if and only if we can find some nth-order nonlinear transformation of the signal that will generate finite-strength additive sine-wave components, which result in spectral lines. For example, for $n = 2$, a quadratic transformation (like the squared signal or the product of the signal with a delayed version of itself, or the weighted sum of such products) will generate spectral lines. For $n = 3$ or $n = 4$, cubic or quartic transformations (i.e., sums of weighted products of 3 or 4 delayed versions of the signal) will generate spectral lines. In contrast, for stationary signals, only a spectral line at frequency zero can be generated.

Another answer to this question, which is completely equivalent to the first answer but does not appear to be so upon first encounter, is that a signal is cyclostationary of order n (in the wide sense) if and only if the time fluctuations in n spectral bands with center frequencies that sum to certain discrete nonzero values are statistically dependent in the sense that their joint nth-order moment (the infinite time average of their product in which each factor is shifted in frequency to have a center frequency of zero) is nonzero. In contrast, for stationary signals, only those bands whose center frequencies sum to zero can exhibit statistical dependence.

[1] For the moment, it is not important to be specific about whether or not we conceive of a signal as a member of the ensemble of some stochastic process. This issue is addressed later. Similarly the modifier *wide sense* is also explained later, in footnote 12.

In fact, for a cyclostationary signal, each sum of center frequencies for which the nth-order spectral moment is nonzero is identical to the frequency of a sine wave that can be generated by putting the signal through an appropriate nth-order nonlinear transformation.

For the simplest nontrivial case, which is $n = 2$, this means that a signal $x(t)$ is cyclostationary with *cycle frequency* α if and only if at least some of its delay-product waveforms, $y(t) = x(t-\tau)x(t)$ or $z(t) = x(t-\tau)x^*(t)$ (where $(\cdot)^*$ denotes conjugation) for some delays τ, exhibit a spectral line at frequency α, and if and only if the time fluctuations in at least some pairs of spectral bands of $x(t)$, whose two center frequencies sum (for the case of $y(t)$) or difference (for the case of $z(t)$) to α, are correlated.

If not all cycle frequencies α for which a signal is cyclostationary are multiples of a single fundamental frequency (equal to the reciprocal of a fundamental period), then the signal is said to be *polycyclostationary* (although the term cyclostationary also can be used in this more general case when the distinction is not important). This means that there is more than one statistical periodicity present in the signal.

1.2 Is Cyclostationarity Useful?

Perhaps the second most obvious question an engineer would ask is, "Is the property of cyclostationarity useful?" The answer is emphatically "Yes!" Cyclostationarity can generally be exploited to enhance the accuracy and reliability of information gleaned from data sets such as measurements of corrupted signals. This enhancement is relative to the accuracy and reliability of information that can be gleaned from stationary data sets or from cyclostationary data sets that are treated as if they were stationary. Such information includes the following:

1. A decision as to the presence or absence of a random signal, or about the number of random signals present, with a particular modulation type in a data set that also contains background noise and other modulated signals,
2. A classification of multiple received signals present in a noisy data set according to their modulation types,
3. An estimate of a signal parameter, such as carrier phase, pulse timing, or direction of arrival, based on a noise-and-interference-corrupted data set,
4. An estimate of an analog or digital message being communicated by a signal over a channel corrupted by noise, interference, and distortion,
5. A prediction of a future value of a random signal,
6. An estimate of the input-output relation of a linear or nonlinear system based on measurements of the system's response to random excitation,
7. An estimate of the degree of causality between two data sets, and
8. An estimate of the parameters of a model for a data set.

1.3 Why Publish a Book on Cyclostationarity?

The next question we should consider is "Why publish a book on cyclostationarity?" Some of the primary reasons are

1. There is a growing awareness in signal processing and communications communities that the cyclostationarity inherent in many man-made random signals and some signals of natural origins (that were previously modeled as stationary) must be properly recognized and modeled if analyses of systems involving such signals are to properly reflect actual behavior;
2. There is a growing awareness of the potential for considerable enhancement of performance of signal-processing algorithms by recognizing and exploiting cyclostationarity in the design process rather than ignoring it by treating signals as if they were stationary;
3. There is a growing awareness by theoreticians that cyclostationary processes are, in many ways, much more than a trivial variation on stationary processes and do, therefore, merit their attention to further develop and refine the theory of these processes;
4. There is a perception by engineers and scientists that cyclostationary processes are much more than a trivial variation on stationary processes and do, therefore, merit their effort to retrain—to expand their theoretical background (their analytical/conceptual "tool boxes") from stationary to cyclostationary processes; and
5. Technological advances, which enable the implementation of increasingly sophisticated signal-processing algorithms, have made the exploitation of cyclostationarity more viable in practice.

We have important work on cyclostationary processes dating back twenty to thirty years (Bennett, 1958; Gladyshev, 1961; Brelsford, 1967; Franks, 1969; Hurd, 1969; Gardner, 1972) and the author's research group at the University of California, Davis, has contributed for the last twenty years. Also, there have been relatively isolated contributions from many others to the development of this subject over the last twenty years. However, the growth in the number of research papers has recently accelerated, and it is only in the last five years that research groups, journal editors, and program directors at funding agencies have shown real interest. The accelerated growth in research activity is illustrated by the histogram of the number of papers on cyclostationarity published per two-year period that is shown in Fig. 1.[2]

[2]The statistics in this graph were compiled by the author using a comprehensive bibliography that he has created over the last five years using his personal files, computerized literature searches, and the assistance of colleagues and students, most notably L. Paura, C. M. Spooner, and K. Vokurka.

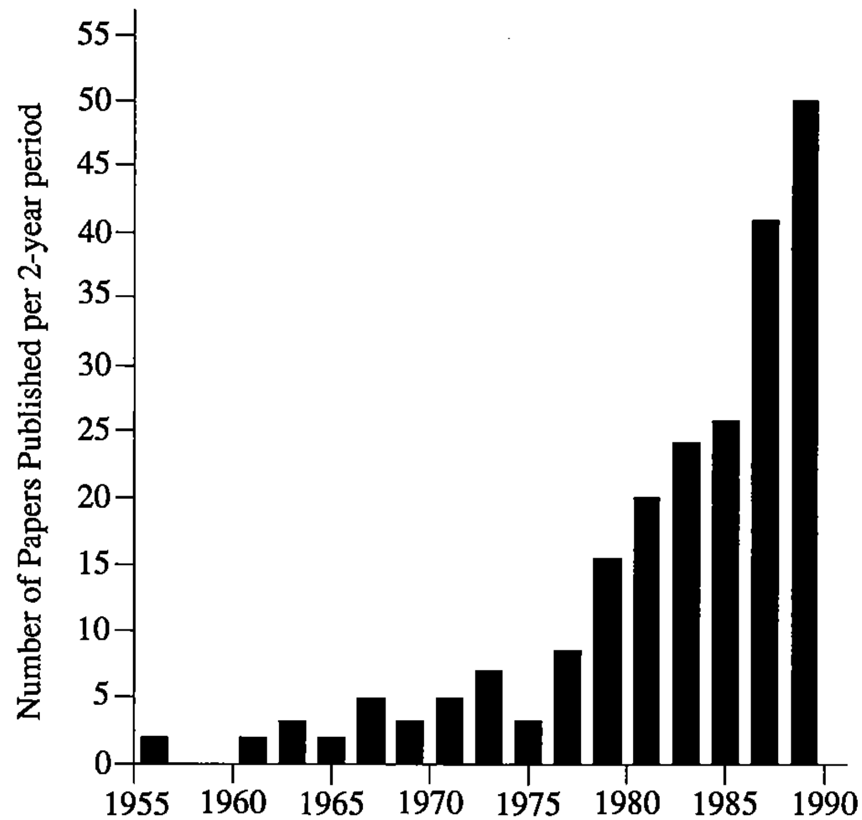

Figure 1: Histogram of papers on cyclostationarity.

Considering the following indicators, it appears that a "critical mass" of interest has been reached and, as a result, that research activity will undergo explosive growth:

1. Acceleration in production of research papers on cyclostationarity;
2. Interest of the National Science Foundation, Army Research Office, Air Force Office of Scientific Research, and Office of Naval Research in supporting the recent workshop on cyclostationarity;
3. Interest demonstrated by the participants of the recent workshop on cyclostationarity;
4. Recent increases in both industrial and government funding of research on cyclostationarity.

Thus, the time is right for publishing a book that provides comprehensive tutorial treatments of the major subtopics of cyclostationarity and surveys the latest developments in various specific areas.

1.4 What Are Some of the Seminal Contributions to the Study of Cyclostationarity?

To expand our perspective on this subject, let us consider the following brief historical survey of some of the seminal contributions to the theory and application of cyclostationarity:[3]

(*Bennett, 1958; Franks, 1969*): Establishment of cyclostationary processes as appropriate models for many communications signals.

(*Jacobs, 1958; Gladyshev, 1963; Gardner, 1978*): First studies of cyclostationary processes with multiple periods.

(*Gudzenko, 1959*): First study of consistency of nonparametric estimates of the Fourier coefficients of periodic autocorrelations.

(*Gladyshev, 1961 and 1963*): Discovery of equivalences among a cyclostationary process (with one period) and several vector-valued stationary processes. Initial work on spectral representation.

(*Brelsford, 1967*): Seminal work on periodic autoregressive modeling and periodic linear prediction.

(*Hurd, 1969, 1989a; Gardner, 1986c, 1987a; Brown 1987*): First studies of consistency of nonparametric estimates of spectral moments of cyclostationary processes with one period (Hurd) and with multiple periods (Gardner and Brown).

(*Gardner, 1972; Gardner and Franks, 1975*): First development and application of several series representations of continuous-time cyclostationary processes in terms of jointly stationary processes for optimum periodically time-variant linear filtering of cyclostationary processes. First characterization of Fourier coefficients of periodic autocorrelations and periodic spectra (the cyclic autocorrelations and cyclic spectra) as crosscorrelations and cross-spectra of frequency-shifted versions of the process.

(*Rootenberg and Ghozati, 1977, 1978; Bittanti 1987; Bittanti and DeNicolao, 1993*): First efforts to develop the Gauss-Markov theory of cyclostationary processes; formulation and partial solution of the cyclospectral factorization problem.

(*Pagano, 1978*): Development of equivalence between univariate periodic AR modeling and multivariate constant AR modeling.

(*Miamee and Salehi, 1980*): Extension—from stationary to cyclostationary processes—of the Wold-Cramér decomposition of a process (and its spectrum) into regular (continuous) and singular (discrete) components.

(*Nedoma, 1963; Boyles and Gardner, 1983*): First formulation and development of cycloergodicity for cyclostationary processes with single (Nedoma) and multiple (Boyles and Gardner) periods.

[3]Contributions from the untranslated Russian literature are not included here, but it is mentioned that several Russian authors, most notably Ya. P. Dragan, have published a substantial amount on cyclostationarity.

(*Gardner, 1985*): First general treatise on cyclostationary processes and their applications to signal processing and communications (1 book chapter).

(*Gardner, 1986b, 1987a*): First formulation and development of the nonstochastic statistical theory of cyclostationary time-series and its applications to signal processing and communications (6 book chapters).

(*Gardner, 1987a; Brown, 1987; Chen, 1989; Agee et al., 1990; Schell, 1990; Spooner, 1992*): First studies of the exploitability of the separability of individual-signal contributions to cyclic temporal and spectral moments (of second order) of multiple interfering signals for the problems of detection, modulation recognition, time-delay estimation, blind-adaptive spatial filtering, and high-resolution direction finding. Discovery that spectrally overlapping signals can be separated with linear temporal processing by exploiting spectral redundancy.

(*Gardner and Spooner, 1992b; Spooner and Gardner, 1992a, b; Spooner, 1992*): First formulation and development of the temporal and spectral moment and cumulant theory of cyclostationary time of order series $n > 2$.

1.5 What about Terminology?

A few words about terminology are in order. The first term given to this class of processes is the term *cyclostationary*, which was introduced by Bennett (1958), who also introduced the term *cycloergodic*. Other terms used include *periodically stationary*, *periodically nonstationary*, and *periodically correlated*. This last term is appropriate only for second-order (wide-sense) cyclostationarity, whereas the preceding three terms admit the modifiers wide-sense, nth-order, and strict-sense, and are, therefore, more general. The most commonly used term is *cyclostationary*. When multiple periodicities exist, this term is modified to *polycyclostationary*, although the terms *almost cyclostationary* and *almost periodically correlated* are used also.

1.6 What Are Some of the Specific Motivations for Studying Cyclostationarity?

There is a great deal of motivation for studying cyclostationarity. Let us consider first some of the practical motives and then some of the mathematical motives and, while we are at it, we can recognize many of the existing contributions to the study of cyclostationarity. The practical motives cited here are specified in terms of a series of facts.

Fact 1: Cyclostationary models, such as PAR (periodic autoregressive), PMA (periodic moving average), and PARMA (periodic autoregressive moving average), can be more parsimonious—better fit with fewer parameters—than stationary models (AR, MA, and ARMA) are. This has been illustrated with real data from

- climatology/meteorology (Brelsford, 1967; Hasselmann and Barnett, 1981; Barnett, 1983; Barnett et al., 1984; Johnson et al., 1985)
- hydrology (Salas, 1972; Salas and Smith, 1980; Vecchia, 1983, 1985; Thompstone et al., 1985; Obeysekera and Salas, 1986; McLeod et al., 1987; Bartolini et al., 1988)

- medicine/biology (Newton, 1982)
- oceanology (Dragan and Yavorskii, 1982; Dragan et al., 1984, 1987)
- economics (Parzen and Pagano, 1979).

Fact 2: Periodic prediction of cyclostationary processes can be done (and periodic causality between cyclostationary processes can be found) when time-invariant prediction is not possible or is inferior (and time-invariant causality is not found or is weaker). Examples are given in Section 2.

Fact 3: Spectrally overlapping cyclostationary signals can never be separated using time-invariant linear filters (e.g., optimum filters of the Wiener and Kalman type for stationary models of the cyclostationary signals). But they can possibly be separated using periodic filters that exploit spectral redundancy. This has been demonstrated for PAM (pulse-amplitude modulation), digital QAM (quadrature-amplitude modulation), AM (amplitude modulation), ASK (amplitude-shift-keying), and PSK (phase-shift-keying) signals (Brown, 1987; Gardner, 1987a; Gardner and Brown, 1989; Gardner and Venkataraman, 1990; Reed and Hsia, 1990; Petersen, 1992; Gardner, 1993).

Fact 4: The biases and variances of parameter estimators (e.g., for TDOA (time-difference-of-arrival), FDOA (frequency-difference-of-arrival), and AOA (angle-of-arrival) of propagating waves) can be much lower, especially for multiple interfering signals, when algorithms that exploit the signal selectivity associated with cyclostationarity (rather than ignore it by treating the signals as if they were stationary) are used. This has been demonstrated for various types of communications signals (Gardner, 1987a, 1988a, 1990a; Gardner and Chen, 1988, 1992; Chen 1989; Chen and Gardner, 1992; Schell and Gardner 1989, 1990a,b,c, 1991, 1992, 1993a; Schell, 1990; Gardner and Spooner, 1993; Izzo et al., 1989, 1990, 1992; Xu and Kailath, 1992).

Fact 5: For the design and analysis of systems that synchronize local digital clocks and sine-wave generators to the frequencies and phases of periodicities embedded in received communications and telemetry signals, the property of cyclostationarity is crucial (Franks, 1980; Franks and Bubrouski, 1974; Moeneclaey, 1982, 1983, 1984; Gardner 1986a).

Fact 6: For the design of algorithms that blindly adapt sensor arrays to perform spatial filtering (for beam/null steering and/or mitigation of multipath fading effects), exploitation of signal selectivity associated with cyclostationarity has proven to be extremely powerful (Agee et al., 1987, 1988, 1990; Schell and Gardner, 1990a; Gardner 1990a) and application to multiuser wireless communications appears to be promising (Gardner et al., 1992; Schell et al., 1993).

Fact 7: For the design of algorithms that adapt channel equalizers to remove intersymbol interference in digital communication systems, exploitation of the phase information contained in second-order cyclostationary statistics of the channel output enables blind adaptation without the use of higher-order statistics (cf. Chapter 3 and Articles 4 and 5 in this volume).

Fact 8: For radio-signal analysis, including detection, classification, modulation recognition, source location, etc., the cyclic spectrum analyzer and related algorithms that exploit cyclostationarity have proven to be ideally suited (Gardner, 1985, 1986b,c, 1987a,b, 1988b,c, 1990a,c, 1991a; Gardner et al., 1987; Brown, 1987; Roberts, 1989; Roberts et al., 1991; Brown and Loomis, 1992; Spooner and Gardner, 1991, 1992a,b; Gardner and Spooner, 1990, 1992a; Spooner, 1992).

Fact 9: For the design and analysis of communications systems that accommodate unintentional nonlinearities that inadvertently generate spectral lines from modulated message signals, the property of cyclostationarity is crucial (Campbell et al., 1983; Albuquerque et al., 1984).

Fact 10: For acoustic-noise analysis for rotating machinery, the cyclic spectrum analyzer holds promise for improved diagnosis of machine wear (e.g., in ground, air, and water vehicles, and hydroelectric plants) and for detection, classification, and location of cyclostationary noise sources (e.g., submarines) (Sherman, 1992).

Fact 11: Many statistical inference and decision problems involving multiple interfering cyclostationary signals in noise can exploit the cyclostationarity to great advantage because of the inherent noise-tolerance and separability of the cyclic features in the signals (Gardner, 1987a, 1990a, 1991a, 1992).

Let us now consider some of the mathematical motives for studying cyclostationarity. Cyclostationary processes (including one or more periods), as a subclass of nonstationary processes, have more in common with stationary processes than do other subclasses of nonstationary processes. The common structure shared by cyclostationary processes suggests (and in some ways this has already been proven) that important theorems and special theories for stationary processes can be extended and/or generalized, and that important theorems for generally nonstationary processes can be specialized, to cyclostationary processes. This potential for mathematical progress, coupled with the increasingly recognized importance of cyclostationarity to practical problems, provides strong motivation for mathematicians to study these processes.

A few examples of important theorems/theories for stationary (or nonstationary) processes that should be—or have been—extended/generalized (or specialized) are given here (consult the key given below[4]).

Topic 1: †† *Wiener-Khinchin and Shiryaev-Kolmogorov theorems relating temporal and spectral moments and cumulants* (Gardner 1986b, 1987a, 1990c; Gardner and Spooner, 1990; Spooner and Gardner, 1992a; Spooner, 1992; Chapter 2 in this volume)

Topic 2: † *Spectral representation theory* (e.g., for harmonizable processes) (Gladyshev, 1963; Hurd, 1974a, 1989b, 1991; Honda, 1982; Rao and Chang, 1988; Chapter 6 in this volume)

[4]* Some progress has been made for cyclostationary processes with one period.
† Substantial progress has been made for cyclostationary processes with one period.
** Some progress has been made for polycyclostationary processes with multiple periods.
†† Substantial progress has been made for polycyclostationary processes with multiple periods.

Topic 3: † *Wold-Cramér theorem on decomposition of a process into singular and regular components and decomposition of its spectrum into discrete and continuous components* (Miamee and Salehi, 1980; article 7 in this volume)

Topic 4: ** *Wiener and Kalman smoothing, filtering, and prediction theory* (Gardner, 1972; Gardner and Franks, 1975; Gardner 1985, 1987a, 1993; Brown, 1987; Gardner and Brown, 1989; Chapter 5 and Article 1 in this volume)

Topic 5: * *Theory of AR, MA, and ARMA models, linear prediction, and parametric spectral estimation* (Brelsford, 1967; Pagano, 1978; Miamee and Salehi, 1980; Tiao and Grupe, 1980; Sakai, 1982, 1983, 1990, 1991; Pourahmadi and Salehi, 1983; Vecchia, 1985; Obeysekera and Salas, 1986; Li and Hui, 1988; Anderson and Vecchia, 1992; Chapter 5 and Article 7 in this volume)

Topic 6: * *Theory of fast algorithms for linear prediction and filtering* (Sakai, 1982, 1983)

Topic 7: * *Markov theory of state-space representations* (Rootenberg and Ghozati, 1977, 1978; Bittanti, 1987; Bittanti and DeNicolao, 1993; Chapter 5 in this volume)

Topic 8: * *Birkhoff Ergodic Theorem and associated ergodic theory* (Nedoma, 1963; Blum and Hansen, 1966; Boyles and Gardner, 1983; Honda, 1990)

Topic 9: ** *Theory of consistent nonparametric estimation of temporal and spectral moments and cumulants* (Gudzenko, 1959; Hurd, 1969, 1989a; Alekseev, 1988, 1991; Gardner, 1985, 1986c, 1987a, 1991b; Dehay, 1991; Spooner, 1992; Spooner and Gardner, 1991, 1992c; Genossar et al., 1993; Hurd and Leskow, 1992a, 1992b; Chapter 2 in this volume)

Topic 10: †† *Theory of higher-order statistics (temporal and spectral moments and cumulants)* (Gardner, 1990c; Gardner and Spooner, 1990, 1992b; Spooner and Gardner, 1992a,b; Spooner, 1992; Chapter 2 in this volume)

2 FUNDAMENTAL CONCEPTS, PHILOSOPHY, AND DEFINITIONS

What do we need to accomplish here? We need a general description of the types of signals that motivate the work being done under the name of cyclostationarity; we need a generally useful definition of the signal property called cyclostationarity, and we need to understand what mathematical/conceptual frameworks are particularly useful for formulating and solving practical problems involving cyclostationary signals, particularly those arising in communication system design and analysis and more general signal processing.

We shall see that an empirically motivated approach to accomplishing these things leads naturally to a probabilistic conceptual framework. However, this framework is distinct from that of stochastic processes in that it does not involve the concept of an ensemble of random samples.

2.1 Signal Types

The types of signals of primary interest here are those normally encountered in communication systems. These signals are typically unpredictable and occur over long periods of time. That is, they are in some sense random (this does not necessarily mean stochastic) and they are persistent rather than transient. These signals also typically originate from physical sources with parameters that are either time-invariant, periodic, or polyperiodic. Thus, the characteristics of the physical signal-generating mechanism vary polyperiodically with time (this includes as special cases periodic variation and time invariance). In some cases the signal-generating mechanism can be decomposed into more elementary signal generators whose outputs are mixed together to form the signal of interest. Some of these more elementary signal generators can have characteristics all of which are time-invariant, thereby giving rise to stationary random signals. Other elementary signals can be simply periodic or polyperiodic functions of time. Thus, the signals of interest often consist of combinations—additive, multiplicative, and other types—of stationary and polyperiodic signals, and are called *polycyclostationary* signals.

2.2 Operational Definition of Polycyclostationarity

What physical evidence in a signal reveals that there is polyperiodic time variation present in its generating mechanism? Fortunately, there is a unique unambiguous answer to this question that appears to be adequate for the general purpose of designing and analyzing signal-processing algorithms that exploit or in some way involve the underlying polyperiodic time variation: We shall say that polyperiodic time variation exists in the generating mechanism of a signal if and only[5] if it is possible to generate finite-amplitude additive polyperiodic components from the signal by passing it through some appropriate nonlinear transformation that is time-invariant and stable. We can take this as an operational definition of polycyclostationarity.

2.3 Operational Origin of Probabilistic Models

There are two particularly interesting ways to characterize an adequately large class of all nonlinear transformations that could potentially generate polyperiodic components from polycyclostationary signals. One way is to require that all transformations of interest be representable in a generalized[6] Volterra series (which is a multivariate Taylor series with a continuum of variables indexed by time). Thus, for a signal x

---

[5]One can conceive of polyperiodic variation in a signal generator that is unobservable in the signal. This is analogous to the concept of unobservability in system theory. Since the focus here is on modeling signals in which there is physical evidence of underlying polyperiodic variation, we are not interested in unobservable polyperiodic variation.

[6]"Generalized" in the sense that the transformations are not constrained to be causal.

and a transformation $g(\cdot)$, we have

$$g(x) = \sum_n \int_{\tau_n} k_n(\tau_n) L_x(t, \tau_n)_n \, d\tau,$$

where $L_x(t, \tau_n)_n$ is the nth-order delay product

$$L_x(t, \tau_n)_n = \prod_{j=1}^{n} x(t + \tau_j),$$

and $k_n(\tau_n)$ is the nth-order Volterra kernel. For example, all transformations $g(\cdot)$ that are continuous and have finite memory admit a convergent Volterra series representation.

Another way is to require that all transformations be representable as a convolution with a finite product of Dirac deltas:

$$\begin{aligned} g(x) &= \int g(y_1, y_2, \ldots, y_n) \prod_{j=1}^{n} \delta\left[y_j - x(t + \tau_j)\right] dy_j \\ &= g[x(t + \tau_1), x(t + \tau_2), \ldots, x(t + \tau_n)], \end{aligned}$$

which can be reexpressed as a Riemann-Stieltjes integral

$$g(x) = \int g(y_n) d^n I_x(t, \tau_n, y_n)_n$$

where

$$I_x(t, \tau_n, y_n)_n \triangleq \prod_{j=1}^{n} I\left[y_j - x_j(t + \tau_j)\right]$$

and $I(\cdot)$ is the indicator function

$$I(z) = \begin{cases} 1, & z > 0 \\ 0, & z \leq 0, \end{cases}$$

for which $dI(z) = \delta(z) dz$. For example, all continuous transformations with finite discrete memory admit this representation. Although this representation of $g(\cdot)$ in terms of itself appears to accomplish nothing, we shall see that it is very useful for our purpose here.

The operation, denoted by $P\{\cdot\}$, for extracting the additive polyperiodic component of a signal is linear (as explained subsequently). Therefore, the polyperiodic component of a weighted sum of signals is the weighted sum of polyperiodic components. Consequently, the polyperiodic component of the first type of transformed signal is given by

$$P\{g(x)\} = \sum_n \int_{\tau_n} k_n(\tau_n) P\{L_x(t, \tau_n)_n\} \, d\tau_n$$

and that of the second type of transformed signal is given by

$$P\{g(x)\} = \int g(y_n) d^n P\{I_x(t, \tau_n, y_n)_n\}.$$

As explained later on, the function $P\{L_x(t, \tau_n)_n\}$ is mathematically equivalent to the joint nth-order moment of n random variables $X_j = x(t + \tau_j)$, $j = 1, 2, \ldots, n$, and the function $P\{I_x(t, \tau_n, y_n)_n\}$ is mathematically equivalent to the joint nth-order probability distribution for the same N random variables, and these equivalences reveal that the polyperiodic component extraction operation $P\{\cdot\}$ is mathematically equivalent to the probabilistic expectation operation. In fact, choosing

$$g(x) = L_x(t, \tau_n)_n$$

in the preceding equation yields

$$P\{L_x(t, \tau_n)_n\} = \int \left[\prod_{j=1}^{n} y_j\right] d^n P\{I_x(t, \tau_n, y_n)_n\}$$

which is the standard formula from probability theory for the nth-order moment in terms of the nth-order probability distribution.

We see, then, that the physical evidence in a signal of polyperiodic time-variation in the generating mechanism of the signal is completely characterized by the signal's temporal moment functions or its temporal probability distribution functions. That is, the polyperiodic component of the delay product of the signal is a temporal moment function and the polyperiodic component of the indicator product is a temporal probability distribution. Hence, we are led naturally by a practically motivated inquiry into the problem of mathematically characterizing physical evidence of polyperiodic time-variation in an unpredictable signal, to a probabilistic description of the signal. Moreover, as explained later on, these moments and distributions are identical to those corresponding to a polycyclostationary stochastic process with appropriate ergodic properties (called *cycloergodicity*), in which case the signal $x(t)$ can be *interpreted* as a sample path (one ensemble member) of the stochastic process. However, in spite of this equivalence between the mathematical model of polyperiodic time variation underlying a signal and a corresponding stochastic process model, the conceptual framework of a stochastic process and its associated ensemble is fundamentally different from the conceptual framework of a single signal that is characterized by all the polyperiodic components that can be generated from it using nonlinear transformations. It is the latter conceptual framework, not the former, that is motivated by the desire to design and analyze signal processors that exploit the generatable polyperiodic components.

2.4 Stochastic vs. Nonstochastic Operational Models

We need to understand the similarities and differences between the stochastic-process approach and the nonstochastic signal (or time-series) approach to conceptualizing, defining, and modeling stationary (S), cyclostationary (CS), and polycyclostationary (PCS) signals, and to developing theory—like the classical theory of statistical inference and decision—to guide the practice of designing and analyzing signal-processing algorithms.

The nonstochastic time-series approach to this subject has not gained the wide level of acceptance that the stochastic-process approach enjoys, particularly for stationary processes. This is believed to be primarily a result of the limited exposure that the time-series approach has received. The aim in recent work (Gardner, 1987a) on developing the time-series approach has been to bring the aesthetics of mathematics and the utility of engineering pragmatism together to produce elegant problem solving.[7] The treatment presented here aims at the same target: the focusing of attention on important concepts for mathematicians who care about the applicability of the mathematics of polycyclostationary signals and for engineers who seek more than a superficial understanding of not only the "how" but also the "why" of polycyclostationary signal processing.

However, before embarking on a discussion of specific mathematical definitions and properties, questions of mathematical existence, and unsolved mathematical problems, a brief summary of the essence of the differences and similarities, from an operational standpoint, of the two alternative approaches is presented.

When properly restricted to appropriate domains of definition (i.e., requiring stochastic process models to exhibit certain ergodic properties and requiring time-series models to exhibit certain regularity properties that guarantee the existence of infinitely long time averages), either approach can be used to obtain the same results in deriving signal-processing algorithms and analyzing their performances (Gardner, 1990a). However, it is not guaranteed that any particular user will in fact obtain the same results regardless of the approach used, because each approach has its own unique conceptual attributes. Thus, it is argued here that the most proficient problem solvers need to understand how to use both approaches.[8] Some problems may naturally fit one approach or the other, and some other problems may benefit from application of both approaches. For example, sometimes it is easier to see how to carry out a particular mathematical calculation using one or the other approach (the stochastic-process approach seems to be favored here), and sometimes it is easier

[7]The title of the book (Gardner, 1987a), which seems to have sparked some controversy, *Statistical Spectral Analysis: A Nonprobabilistic Theory*, can be misleading since it is shown in this book that an empirically motivated inquiry into the problem of quantifying the average behavior of spectral measurements leads naturally to a probabilistic theory. Since this probabilistic theory is nonstochastic (it involves only time averages, not ensemble averages), the title could have been *Statistical Spectral Analysis: A Nonstochastic Theory*. Nevertheless, the majority of the concepts and methods developed in the book are not only nonstochastic, they are indeed nonprobabilistic, and a primary goal of the book is to show that in an empirically motivated development of the fundamental concepts and methods of statistical spectral analysis, probability does not play a seminal role. It does play an important role in the mechanics of quantifying average behavior, but it plays no role in conceptualizing the objectives and methods (parametric and nonparametric) of statistical spectral analysis of single time-series.

[8]It is curious that some followers of the stochastic-process approach insist that the alternative approach is of no value or, worse yet, has negative value. Perhaps the stochastic process faith should be formally recognized as a religion.

to relate the mathematics to the real-world problem at hand using one or the other approach (the nonstochastic time-series approach seems to be favored here with regard to many of the applications discussed in this book[9]).

Mathematicians have, for the most part, chosen the stochastic-process framework for their work because it is apparently more amenable to deep mathematical treatment. Statisticians have, for the most part, chosen the approach to statistical inference and decision that is based on stochastic processes because it does naturally fit the problem of making inferences about a total population on the basis of limited "random" samples from the population, which is the statistician's classical problem. The concept of a population, or ensemble, also naturally fits a number of situations in communications engineering and signal processing or time-series analysis for engineering purposes; however, there are many other engineering (and science) problems involving time-series data where the ensemble concept is fictitious, irrelevant, or otherwise inappropriate. In these cases, users often force an application of the theory of stochastic processes onto their real-world problem because they have not learned that there is a viable alternative for statistical inference and decision. This can lead to substantial confusion and less effective engineering.

2.5 Nonstochastic Statistical Inference and Decision

Let us briefly consider how a theory of statistical inference and decision can be based on the concept of a single time-series without reference to an ensemble. Many—but by no means all—real-world problems in engineering and science involve time-series data for which no population exists; that is, for which replication of the "experiment" is impossible or impractical. However, many of these time-series arise from physical phenomena that can be considered to be unchanging in their basic nature for a very long time. In such cases, conceptually idealizing this time-invariance by extending the length of time without bound enables us to conceive of a model that is derivable from the data in the limit as the amount of data used for measuring the parameters of the model approaches infinity. This leads us to the concept of a fraction-of-time (FOT) probability model that is free from the abstract concept of a population. For example, the FOT probability that a time-series exceeds some specified level is defined to be the fraction of time that this event occurs over the life of the time-series.

Once we have accepted the idea of an infinitely long time-series with an FOT probability model, we can develop a theory of statistical inference and decision that is isomorphic to the theory for stationary stochastic processes. This was briefly pointed out in (Wold, 1948), developed in (Hofstetter, 1964), and extended from stationary to cyclostationary and polycyclostationary time-series in (Gardner, 1987a), (Gardner and Brown, 1991). But one might ask what it is that motivates the development of such a FOT probability theory. One of the answers to this question is analogous to that which motivates the theory that is based on the concept of a population: We want to make inferences about the physical phenomenon that gave rise to the observed

[9]A compelling example of this is the novel derivation of the cumulant in the study of higher-order cyclostationarity presented in Chapter 2.

time-series. To the extent that this phenomenon is characterized by the FOT model for the time-series $x(t)$, (i.e., the set of joint FOT probability distributions for all finite sets of time translates $\{x(t+t_i) : i = 1, \ldots, n\}$ for all n translations τ_i and all natural numbers n), we can interpret our objective as that of making inferences about the infinitely long time-series or its generating mechanism on the basis of finite-length observations. We can use the FOT probability model to calculate bias, variance, and confidence intervals for parameter estimates, and we can calculate probabilities of correct and incorrect decisions. We also can formulate and solve optimization problems.

2.6 A Historical Perspective

The stochastic-process approach (to the exclusion of the nonstochastic time-series approach) is currently the orthodox approach because this is the approach that dominated for sixty years in mathematics and statistics and it is, therefore, the approach in terms of which the theory of statistical inference and decision has been formulated and is taught. It does not follow that the stochastic-process approach is orthodox because it is always the superior approach. This last point can be illustrated with a brief history of statistical inference and decision in communications engineering.

Why have communications engineers focused on using theoretical measures of performance that average over an ensemble of signals and/or noises? Because they have wanted to design systems that would perform well on the average over the ensemble, and because mathematicians and statisticians had developed a powerful theory of statistical inference and decision that was based on ensemble averages, and because probability theory itself is an immensely powerful conceptual tool, and few engineers have realized that probability theory can be based entirely on time-averages. But why then have communications engineers focused almost exclusively on measuring system performance in practice by averaging over time for a single system? Because of economics (the high cost of making measurements on many systems) and because they also want each system to perform well on the average over time.

In order to match the theory based on ensembles of data to the practice based on a single record of data, they invoked the concept of ergodicity. That is, they agreed to use stationary stochastic-process models that were ergodic so that the mathematically calculated expected values (ensemble averages) would equal the measured time averages (in the limit as averaging time approaches infinity).

Unfortunately, however, the logic seems to have stopped at this point. It apparently was not recognized (except by too few to make a difference) that once consideration was restricted to ergodic stationary models, the stochastic process and its associated ensemble could be dispensed with because a completely equivalent theory of statistical inference and decision that was based entirely on time-averages over a single record of data could be used (Hofstetter, 1964). Any calculations made using a model based on the time-average theory could be applied to any one member of an ensemble if one so desired because the arguments that justify the ergodic stochastic-

process model also guarantee that the time-average for one ensemble member will be the same (with probability one) as the time-average for any other ensemble member.

Because the time-average framework is more conceptually straightforward for application to problems where time-average performance is of primary concern, it is a more natural choice; but because of history and inertia, it may never gain its rightful place in engineering. This is even more likely to be the case when the utility of nonergodic stochastic processes is taken into account. For example, whenever transient behavior is of interest, ergodic models are ruled out, because all transient behavior is lost in an infinitely long time-average. Thus, to counter the conceptual simplicity and realism offered by the time-average approach, the stochastic-process approach offers the advantage of more general applicability.

Nevertheless, there is a special class of signals that includes more than just those that can be modeled as stationary ergodic processes, for which there is a compelling argument, which has only recently surfaced, to adopt an alternative nonstochastic approach. And these are the signals that are appropriately modeled as polycyclostationary time-series. As explained earlier here, the use of time-averages to extract additive polyperiodic components from nonlinear transformations of these signals leads naturally to a probability theory based entirely on time averaging. Let us now consider in some detail these two alternative approaches to conceptualizing, defining, and modeling signals.

2.7 Dual Theoretical Frameworks

The concepts and definitions presented here apply equally well to continuous-time and discrete-time signals. We need only choose either the continuous-time-averaging operation

$$\langle\cdot\rangle \triangleq \lim_{Z\to\infty} \frac{1}{2Z} \int_{-Z}^{Z} (\cdot)\, dt$$

or its discrete-time counterpart

$$\langle\cdot\rangle \triangleq \lim_{Z\to\infty} \frac{1}{2Z+1} \sum_{t=-Z}^{Z} (\cdot).$$

We consider first stochastic processes, and then we consider nonstochastic time-series. Let $X(t)$ be a real-valued stochastic process on the real line $-\infty < t < \infty$, with measure μ on the probability space Ω. Consider the event indicator

$$I[x - X(t)] \triangleq \begin{cases} 1, & X(t) < x \\ 0, & X(t) \geq x. \end{cases}$$

The expected value of this event indicator is the probability distribution (PD) function for the random variable $X(t)$,

$$F_{X(t)}(x) \triangleq Prob\{X(t) < x\} = E\{I[x - X(t)]\}$$

where the expectation operation $E\{\cdot\}$ is defined by

$$E\{H\} \triangleq \int_{\Omega} H(\omega)\, d\mu(\omega)$$

for any random variable H defined on Ω. Therefore, the joint PD function for the set of random variables

$$\boldsymbol{X}(t) \triangleq \{X(t+t_1), \ldots, X(t+t_n)\},$$

where t is the time-translation parameter, is given by the expectation

$$F_{\boldsymbol{X}(t)}(\boldsymbol{x}) = E\left\{\prod_{j=1}^{n} I\left[x - X(t+t_j)\right]\right\},$$

and the joint probability density (Pd) function is

$$f_{\boldsymbol{X}(t)}(\boldsymbol{x}) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{\boldsymbol{X}(t)}(\boldsymbol{x}),$$

which contains Dirac deltas when the PD contains step discontinuities. We have the following theorem from probability theory.

Fundamental Theorem of Expectation

For any nonrandom function $g(\cdot)$ for which $E\{g[\boldsymbol{X}(t)]\}$ exists, we have

$$E\{g[\boldsymbol{X}(t)]\} \triangleq \int y f_{g[\boldsymbol{X}(t)]}(y)\, dy$$
$$= \int g(\boldsymbol{x}) F_{\boldsymbol{X}(t)}(\boldsymbol{x})\, d\boldsymbol{x}$$

That is, the Pd for $g[\boldsymbol{X}(t)]$ need not be found from the Pd of $\boldsymbol{X}(t)$ in order to evaluate the expected value of this random variable. This theorem can be used to verify that the PD is indeed equal to the expected value of the event indicator by letting

$$g[\boldsymbol{X}(t)] = \prod_{j=1}^{n} I\left[x_j - X(t+t_j)\right]$$

to obtain

$$E\{g[\boldsymbol{X}(t)]\} = \int \prod_{j=1}^{n} I\left[x_j - z_j\right] f_{\boldsymbol{X}(t)}(\boldsymbol{z})\, d\boldsymbol{z}$$
$$= \int_{-\infty}^{\boldsymbol{x}} f_{\boldsymbol{X}(t)}(\boldsymbol{z})\, d\boldsymbol{z}$$
$$= F_{\boldsymbol{X}(t)}(\boldsymbol{x}).$$

Let us now consider time-series. Let $x(t)$ be a well-behaved[10] real-valued time-series (a nonstochastic real-valued function) on the real line $-\infty < t < \infty$. Consider the event indicator

$$I[x - x(t)] \triangleq \begin{cases} 1, & x(t) < x \\ 0, & x(t) \geq x. \end{cases}$$

The time-average of this event indicator is the *fraction-of-time* (FOT) PD

$$\hat{F}^0_{x(t)}(x) \triangleq \hat{P}rob\{x(t) < x\} = \hat{E}^0\{I[x - x(t)]\}$$

where the time-average operation $\hat{E}^0\{\cdot\}$ is defined by

$$\hat{E}^0\{h(t)\} \triangleq \lim_{Z\to\infty} \frac{1}{2Z} \int_{-Z}^{Z} h(t+t')\, dt'$$

for any time function h. (The superscript 0 will be explained subsequently.) Therefore the joint FOT PD for the set of variables $\boldsymbol{x}(t) \triangleq \{x(t+t_1), \ldots, x(t+t_n)\}$ is given by

$$\hat{F}^0_{\boldsymbol{x}(t)}(\boldsymbol{x}) = \hat{E}^0\left\{\prod_{j=1}^{n} I\left[x_j - x(t+t_j)\right]\right\}$$

and the joint FOT Pd is

$$\hat{f}^0_{\boldsymbol{x}(t)}(\boldsymbol{x}) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} \hat{F}^0_{\boldsymbol{x}(t)}(\boldsymbol{x}).$$

We have the following theorem.

Fundamental Theorem of Time-Averaging

For every time-invariant function $g(\cdot)$ for which $\hat{E}^0\{g[\boldsymbol{x}(t)]\}$ exists, we have

$$\hat{E}^0\{g[\boldsymbol{x}(t)]\} \triangleq \lim_{Z\to\infty} \frac{1}{2Z} \int_{-Z}^{Z} g[\boldsymbol{x}(t+t')]\, dt'$$
$$= \int g(\boldsymbol{x}) \hat{f}^0_{\boldsymbol{x}(t)}(\boldsymbol{x})\, d\boldsymbol{x}.$$

This theorem motivates us to call the time-average operation $\hat{E}^0\{\cdot\}$ the *temporal-expectation operation*. To illustrate the validity of this theorem, we can substitute the definition of the FOT PD into the definition of the FOT Pd, which can be substituted into the result of this theorem to obtain

[10]We mean "well-behaved" in the sense that $x(t)$ exhibits the regularity required for all time averages of interest to exist.

$$\int g(x) f_{x(t)}(x)\, dx = \int g(x) \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{x(t)}(x)\, dx$$

$$= \int g(x) \frac{\partial^n}{\partial x_1 \ldots \partial x_n} \hat{E}^0 \left\{ \prod_{j=1}^{n} I\left[x_j - x(t+t_j)\right] \right\} dx$$

$$= \int g(x) \hat{E}^0 \left\{ \prod_{j=1}^{n} \delta\left[x_j - x(t+t_j)\right] \right\} dx$$

$$= \hat{E}^0 \left\{ \int g(x) \prod_{j=1}^{n} \delta\left[x_j - x(t+t_j)\right] \right\} dx$$

$$= \hat{E}^0\{g[x(t)]\},$$

where we have used the sampling property of the Dirac delta δ, which is the derivative of the unit-step function $I(\cdot)$.

For any function $h(t)$ for which $\hat{E}^0\{h(t)\}$ exists, we have

$$h(t) = c + r(t)$$

where $c = \hat{E}^0\{h(t)\}$ is a constant (independent of t) and $r(t) = h(t) - c$ is the residual for which $\hat{E}^0\{r(t)\} = 0$. Consequently, the temporal expectation operation can also be called the *constant-component extractor*.

We can see from these two theorems that there is a *duality* between the *probability-space* theory of stochastic processes based on the operation $E\{\cdot\}$ and what we shall call the *time-space* theory of time-series based on $\hat{E}^0\{\cdot\}$. Wold (Wold, 1948) tried to formalize this duality in terms of an isomorphism based on the mapping

$$x(t+\sigma) \to X(t, \omega(\sigma))$$

where $X(t, \omega)$ is a sample path of the stochastic process $X(t)$ corresponding to the sample point $\omega = \omega(\sigma)$, indexed by σ, in Ω. That is, the ensemble members of $X(t)$ correspond to translates of $x(t)$ in this isomorphism. While this isomorphism is conceptually useful, a mathematically rigorous study of it has not (to my knowledge) been performed. (For example, does such a stochastic process corresponding to a given time-series actually exist?)

We can justifiably ask, "Just how viable is the time-space theory?—do 'well-behaved' time-series models exist?" The answer to the latter question is "Yes"; examples are provided by typical sample paths of ergodic stochastic processes. But, "Can we construct useful time-series models?" The answer, again, is "Yes": We can construct models in the same way we do for stochastic processes, except we specify $\hat{F}_{x(t)}(x)$ instead of $F_{X(t)}(x)$.

Regarding the answer to the first question, we might ask "Does this apparent reliance, of existence of time-series, on stochastic processes detract from the conceptual simplicity of working with time-series rather than stochastic processes?" The answer, in my opinion, is "No."

Let us trace the conceptual paths for both stochastic processes and time-series so that we can see specifically where they are parallel and where they diverge. As before, we begin with stochastic processes by giving the definitions of the classes of processes of interest in the study of cyclostationarity, namely, processes that are S, CS, or PCS of order n (in the strict sense).

2.8 Stochastic-Process Definitions

Definition 1: $X(t)$ is a S process if and only if $F_{X(t)}(x)$ is independent of the time-translation parameter t.

Definition 2: $X(t)$ is a CS process with period T if and only if $F_{X(t)}(x)$ is periodic in t with period T.

Definition 3: $X(t)$ is a PCS process with periods $\{T\} = T_1, T_2, T_3, \ldots$ if and only if $F_{X(t)}(x)$ is polyperiodic in t with periods $\{T\}$ (which is a sum of periodic functions with single periods $T_1, T_2, T_3, \ldots$).

The relationships among the class of generally nonstationary (NS) processes and the three classes S, CS, and PCS can be described with the Venn diagram shown in Fig. 2.

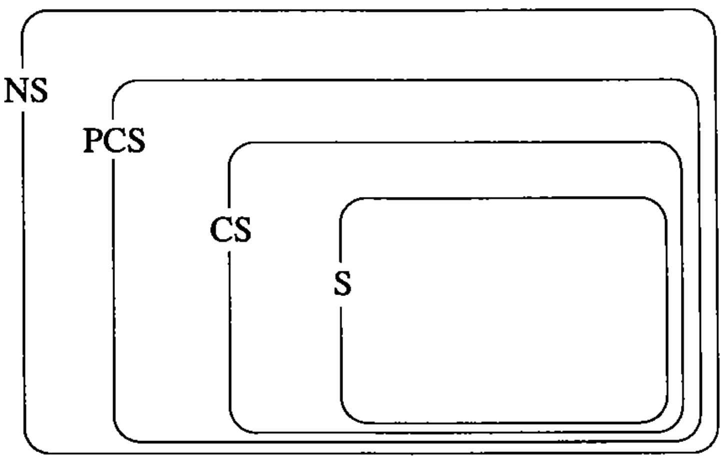

Figure 2: Venn diagram of classes of stochastic processes. With the class NS omitted, also the Venn diagram of classes of time-series.

It is useful to expand the polyperiodic PD function in a Fourier series:

$$F_{X(t)}(x) = \sum_{\alpha} F^{\alpha}_{X(0)}(x)\, e^{i2\pi\alpha t} = \sum_{\alpha} F^{\alpha}_{X(t)}(x),$$

where $F^{\alpha}_{X(0)}(x)$ are the Fourier-coefficients, and where the sinusoidal-component functions are given by

$$F^{\alpha}_{X(t)}(x) \triangleq \lim_{Z\to\infty} \frac{1}{2Z} \int_{-Z}^{Z} F_{X(t+t')}(x)\, e^{-i2\pi\alpha t'}\, dt'$$
$$= \hat{E}^0 \left\{ F_{X(t)}(x)\, e^{-i2\pi\alpha t} \right\} e^{i2\pi\alpha t}$$
$$\triangleq \hat{E}^\alpha \left\{ F_{X(t)}(x) \right\}.$$

For any function $h(t)$ for which $\hat{E}^\alpha\{h(t)\}$ exists, we have

$$h(t) = c\, e^{i2\pi\alpha t} + r(t)$$

where c is a constant, $ce^{i2\pi\alpha t} = \hat{E}^\alpha\{h(t)\}$, and $r(t) = h(t) - ce^{i2\pi\alpha t}$ is the residual for which $\hat{E}^\alpha\{r(t)\} = 0$. Consequently, we call the operation

$$\hat{E}^\alpha\{\cdot\} \triangleq \hat{E}^0 \left\{ (\cdot) e^{-i2\pi\alpha t} \right\} e^{i2\pi\alpha t}$$

the *sine-wave-component extractor*. It can be thought of as the limit, as bandwidth goes to zero, of a bandpass filter with center frequency α and unity gain at α. For $\alpha = 0$, it reduces to the constant-component extractor $\hat{E}^0\{\cdot\}$.

2.9 Time-Series Definitions

Now let us turn from stochastic processes to time-series. Before we can give the dual time-space definitions of S, CS, and PCS time-series, we need to generalize the temporal expectation operation $\hat{E}^0\{\cdot\}$. The appropriate generalization is simply the sum of sine-wave-component extractors

$$\hat{E}^{\{\alpha\}}\{\cdot\} \triangleq \sum_{\alpha\in\{\alpha\}} \hat{E}^\alpha\{\cdot\}$$

for all sine-wave frequencies in some set $\{\alpha\}$ of interest. Thus, $\hat{E}^{\{\alpha\}}\{\cdot\}$ is called the *multiple-sine-wave-component extractor* or, equivalently, the *polyperiodic-component extractor*. (It is identical to the operator $P\{\cdot\}$ discussed in Section 2.3.) The sine-wave frequencies α are the harmonics of the reciprocals of the periods $1/T_1, 1/T_2, 1/T_3, \ldots$ of interest.

In terms of the generalized temporal expectation operation, we can define the *polyperiodic FOT PD*:

$$\hat{F}^{\{\alpha\}}_{x(t)}(x) = \hat{E}^{\{\alpha\}} \left\{ \prod_{j=1}^{n} I\left[x_j - x(t+t_j)\right] \right\}$$

and the *polyperiodic FOT Pd*:

$$\hat{f}^{\{\alpha\}}_{x(t)}(x) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} \hat{F}^{\{\alpha\}}_{x(t)}(x).$$

It is not obvious that $\hat{F}^{\{\alpha\}}_{x(t)}(x)$ is indeed a valid probability distribution function, but this proposition is proved in (Gardner and Brown, 1991). We have the following theorem.

Fundamental Theorem of Polyperiodic-Component Extraction

For every time-invariant function $g(\cdot)$ for which $\hat{E}^{\{\alpha\}}\{g[x(t)]\}$ exists, we have

$$\hat{E}^{\{\alpha\}}\{g[x(t)]\} \triangleq \sum_{\alpha} \lim_{Z\to\infty} \frac{1}{2Z} \int_{-Z}^{Z} g[x(t+t')]\, e^{-i2\pi\alpha t'}\, dt'$$
$$= \int g(x) \hat{f}^{\{\alpha\}}_{x(t)}(x)\, dx$$

The validity of this theorem can be illustrated in the same way the validity of the fundamental theorem of constant-component extraction is illustrated. Also, this theorem is valid more generally if $g(\cdot) = g(t, \cdot)$ is polyperiodic in time t.

We are now in a position to define the classes of S, CS, and PCS of order n (in the strict-sense) time-series. Let $\{\alpha\}$ be the set of all α for which $\hat{F}^{\{\alpha\}}_{x(t)} \not\equiv 0$.

Definition 4: $x(t)$ is a S time-series if and only if $\hat{F}^{\{\alpha\}}_{x(t)}(x)$ exists and $\not\equiv 0$ and is independent of the time-translation parameter t (that is, $\{\alpha\} = \{0\}$).

Definition 5: $x(t)$ is a CS time-series with period T if and only if $\hat{F}^{\{\alpha\}}_{x(t)}(x)$ exists and $\not\equiv 0$ and is periodic in t with period T (that is, $\{\alpha\} = \{$harmonics of $1/T\}$).

Definition 6: $x(t)$ is a PCS time-series with periods $\{T\} = T_1, T_2, T_3, \ldots$ if and only if $\hat{F}^{\{\alpha\}}_{x(t)}(x)$ exists and $\not\equiv 0$ and is polyperiodic in t with periods $\{T\}$).

The relationships among these three classes of time-series can be described with the Venn diagram shown in Fig. 2 except that for time-series, unlike stochastic processes, the superclass NS does not exist. Generally nonstationary FOT PDs cannot be defined (although locally S FOT PDs, which are NS, can be defined by limiting the time-averaging interval used in $\hat{E}^0\{\cdot\}$ to one of finite length Z).

These definitions of S, CS, and PCS time-series represent a modification of previous terminology. Wold (Wold, 1948) defined a stationary (S′) time-series to be one for which $\hat{F}^{0}_{x(t)}(x)$ exists and $\not\equiv 0$. To refine this definition, (Gardner, 1987a) defined a purely S′ time-series to be a S′ time-series for which $\hat{F}^{\alpha}_{x(t)}(x) \equiv 0$ for all $\alpha \neq 0$. Following Wold, (Gardner, 1987a) also defined a cyclostationary (CS′) time-series with period T to be one for which $\hat{F}^{\{\alpha\}}_{x(t)}(x)$ exists and $\not\equiv 0$ for some $\{\alpha\} \in \{$harmonics of $1/T\}$, and to refine this a purely CS′ time-series was defined to be a CS′ time-series for which $\hat{F}^{\{\alpha\}}_{x(t)}(x) \equiv 0$ for all $\alpha \notin \{$harmonics of $1/T\}$.

The relationships among these previously defined classes of time-series can be described with the Venn diagram shown in Fig. 3. Observe that the nesting of the classes S′, CS′, and PCS′ $\equiv$ PCS is inverted from that in Fig. 2 for the classes S, CS, and PCS. The definitions of S, CS, and PCS form the basis for a theory of time-series that in some ways has a stronger duality with the theory of stochastic processes than does the theory that could be based on the previous definitions of S′, CS′, and PCS′.

On the other hand, the previous concepts of pure stationarity and pure cyclostationarity arise also within the framework of stochastic processes. Since these concepts depend on notions of ergodicity, let us now consider the relevant types of ergodicity.

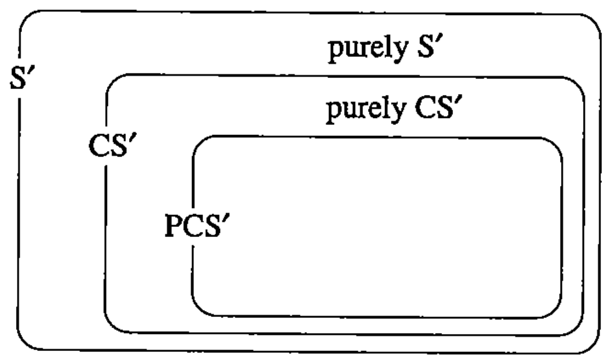

Figure 3: Venn diagram for previously defined classes of time-series.

2.10 Cycloergodicity and Refined Stochastic-Process Definitions

Definition 7: $X(t)$ is an *ergodic* (E) process if and only if for every natural number n and every nonrandom function $g(\cdot)$ of n variables for which $E\{g[X(t)]\}$ exists, we have

$$\hat{E}^0\{E\{g[X(t)]\}\} = \hat{E}^0\{g[X(t)]\} \ w.p.\ 1$$

(where w.p. 1 means with probability equal to 1). For a S process, $\hat{E}^0\{E\{\cdot\}\} = E\{\cdot\}$ (since the constant component of a constant is that constant) and the outer operation on the left side of this defining equation can be deleted.

Definition 8: $X(t)$ is a *cycloergodic* (CE) process with period T if and only if for every natural number n and every nonrandom function $g(\cdot)$ of n variables for which $E\{g[X(t)]\}$ exists, we have (with $\{\alpha\} \in \{\text{harmonics of } 1/T\}$)

$$\hat{E}^{\{\alpha\}}\{E\{g[X(t)]\}\} = \hat{E}^{\{\alpha\}}\{g[X(t)]\} \ w.p.\ 1.$$

For a CS process, $\hat{E}^{\{\alpha\}}\{E\{\cdot\}\} = E\{\cdot\}$ (since the periodic component of a periodic function is that periodic function) and the outer operation on the left side of this defining equation can be deleted.

Definition 9: $X(t)$ is a *polycycloergodic* (PCE) process with periods $\{T\}$ if and only if it is cycloergodic with period T_k for $k = 1, 2, 3, \ldots$.

Stochastic processes that are not CE or PCE can exhibit hidden cyclostationarity. For example, if $X(t)$ is S and PCE with all periods, then its sample paths are stationary time-series (w.p. 1); however, if $X(t)$ is S (and possibly E) but not CE, its sample paths can be CS (w.p. 1). Similarly, if $X(t)$ is CS and PCE with all periods, then its sample paths are CS time-series (w.p. 1); however, if $X(t)$ is CS (and possibly CE), but not PCE, its sample paths can be PCS (w.p. 1). Such non-CE and non-PCE models typically result from the (explicit or implicit) inclusion of random-phase variables in

the stochastic-process model. This hidden cyclostationarity motivates the following refined probability-space definitions.

Definition 10: If $X(t)$ is S and PCE with all periods, then it is defined to be *purely stationary*, and its sample paths are purely stationary (S or purely S′) time-series (w.p. 1): there is no hidden CS.

Definition 11: If $X(t)$ is CS and PCE with all periods, then it is defined to be *purely CS*, and its sample paths are purely cyclostationarity (CS or purely CS′) time-series (w.p. 1): there is no hidden PCS.

Definition 12: If $X(t)$ is PCS and PCE with all periods, then it is defined to be *purely PCS*: there is no hidden PCS.

The relationships among all the classes of stochastic processes defined so far are illustrated with the Venn diagram shown in Fig. 4.

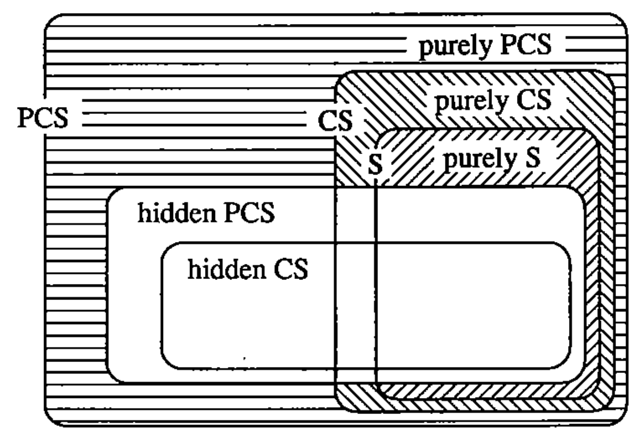

Figure 4: Venn diagram of classes of stochastic processes.

Let us now consider an example that illustrates the various classes of stochastic processes and time-series that have been introduced. Let the *stochastic process* $X(t)$ be specified by

$$X(t) = A(t) + B(t)\cos(\omega_1 t + \theta_1) + C(t)\cos(\omega_2 t + \theta_2),$$

where $A(t)$, $B(t)$, and $C(t)$ are purely stationary ergodic processes. If θ_1 and θ_2 are nonrandom, then the *stochastic process* $X(t)$ is PCS and PCE. On the other hand, if θ_1 and/or θ_2 is random, then (depending on their PDs) the *stochastic process* $X(t)$ can be PCS (with periods T_1 and T_2), or it can be CS (with period T_1 or T_2), or it can be S, and $X(t)$ is not PCE. Furthermore, with probability one, the sample paths of $X(t)$ are PCS *time-series*, the sample paths of $A(t)$, $B(t)$, and $C(t)$ are S *time-series*, and the sample paths of the components $B(t)\cos(\omega_1 t + \theta_1)$ and $C(t)\cos(\omega_2 t + \theta_2)$ are CS *time-series*.

2.11 Phase Randomization

To pursue the concept of phase randomization a little further, it is noted that even if θ_1 and θ_2 are both nonrandom, we can introduce a random phase Θ into $X(t)$ to obtain

$$Y(t) = X(t + \Theta)$$

which can be changed from PCS to CS or to S by choosing the distribution for Θ (Gardner, 1978; Hurd, 1974b). Thus, we see that there is a nonuniqueness of models for stochastic processes. We can change the stochastic process from PCS to CS to S by phase-randomizing with a single phase variable: $X(t) \to X(t+\Theta)$. Or, equivalently, we can change the PD function from polyperiodic to periodic to constant by time-averaging; e.g.,

$$\hat{E}^0 \left\{ F_{X(t)}^{\{\alpha\}}(x) \right\} = F_{X(t)}^0(x).$$

That is, phase-randomizing a CS or PCS process or time-averaging its PD function can result in hidden cyclostationarity. Similarly, we can change the PD function for a time-series from polyperiodic to periodic to constant by time-averaging; e.g.,

$$\hat{E}^0 \left\{ \hat{F}_{x(t)}^{\{\alpha\}}(x) \right\} = \hat{F}_{x(t)}^0(x).$$

2.12 Pitfalls of the Stochastic-Process Framework

There are some significant pitfalls associated with nonunique models. One such pitfall is "hidden statistical dependence." Let SI denote statistical independence (e.g., of two variables). We can show that SI in a CS model does not necessarily imply SI in the corresponding S model, and that SI in an S model does not necessarily imply SI in the associated CS model. To prove the first statement we simply observe that the equality

$$f_{X_1(t),X_2(t)} = f_{X_1(t)} f_{X_2(t)}$$

that results from the SI of jointly CS processes $X_1(t)$ and $X_2(t)$ does not necessarily imply the equality

$$\hat{E}^0 \left\{ f_{X_1(t),X_2(t)} \right\} = \hat{E}^0 \left\{ f_{X_1(t)} \right\} \hat{E}^0 \left\{ f_{X_2(t)} \right\},$$

which would hold if, in the associated S model, $X_1(t)$ and $X_2(t)$ were SI. To prove the second statement, we consider the example of discrete-time processes

$$X_1(t) = Z(t) = i.i.d. \pm 1$$

$$X_2(t) = Z(t)\cos(t).$$

We can easily show that

$$\hat{E}^0 E\left\{X_1^n(t) X_2^m(t)\right\} = \hat{E}^0 E\left\{X_1^n(t)\right\} \hat{E}^0 E\left\{X_2^m(t)\right\}$$

for all n and m and, therefore, $X_1(t)$ and $X_2(t)$ are SI in the S model. However,

$$E\left\{X_1^n(t) X_2^m(t)\right\} \neq E\left\{X_1^n(t)\right\} E\left\{X_2^m(t)\right\}$$

for n and m odd and, therefore, $X_1(t)$ and $X_2(t)$ are not SI in the CS model.

We are now ready to take stock. One conclusion we can draw is that when a process is not PCE, the hidden CS or hidden PCS can result in single-sample-path behavior (w.p. 1) that cannot be predicted from probabilistic analysis (unless the hidden CS can be revealed by conditioning on certain random phase variables).

An important fact concerning this conclusion is that the theory of PCE is mostly nonexistent and appears to require nontrivial extensions/generalizations of the theory of E and the incomplete theory of CE.

Another important fact is that the commonplace approach to deriving ad hoc signal-processing algorithms, of replacing expectation operations $E\{\cdot\}$ in analytical expressions with time-average operations $\hat{E}^0\{\cdot\}$, or (when a composition of both operations are present in the analytical expression) of deleting the expectation operation, cannot be justified (and will often fail to produce the desired results) when the stochastic-process model used is not PCE.

A related important fact is that the "optimum" solutions to inference and decision problems (e.g., for signal estimation and detection) that are based on S and E, but not CE (or based on CS and CE, but not PCE), process models can be highly inferior to inference and decision rules that exploit the hidden CS (or hidden PCS).

Let us consider some examples that illustrate the ramifications of the preceding conclusion and associated facts.

Example 1: Let $Y(t)$ be the output of a time-invariant nonlinear transformation with input $X(t)$. Let $X(t)$ be S (for all n) and E, but not CE, with no spectral lines. Then $Y(t)$ is S (for all n) and E but, because of the hidden CS in $X(t)$, contains spectral lines. The presence of these spectral lines cannot be explained except by virtue of the hidden CS in $X(t)$. As a specific example, let $X(t) = A(t)\cos(\omega_1 t + \Theta_1)$, where $A(t)$ is S and E, and Θ_1 is independent of $A(t)$ and uniformly distributed on $[0, 2\pi]$, and let $Y(t) = X^2(t)$. Then $Y(t)$ has spectral lines at frequencies 0 and $\pm\omega_1/\pi$ Hz.

Example 2: Let $X(t)$ be S (for $n = 2$) and E, but not CE, and let $X(t)$ admit an exact AR model (with white residuals). The sample paths of the white residuals can be partially predictable (w.p. 1) using linear periodic predictors derived from the sample-path statistics. For example, let $U(t)$ be any nonwhite CS process and let $V(t)$ be an independent S process with complementary spectrum; that is, the sum of the spectra of $V(t)$ and the stationarized version $U(t+\Theta)$ of $U(t)$ equals a constant over all frequency. Then the S process $W(t+\Theta) = U(t+\Theta) + V(t+\Theta)$ is white (the spectra of $V(t)$ and $V(t+\Theta)$ are identical) but the CS process $W(t)$ is nonwhite. That is, the autocorrelation of $W(t+\Theta)$ is proportional to a delta function in the lag variable but the autocorrelation of $W(t)$ is not. Thus, $W(t)$ is predictable using linear periodic predictors and so too are its sample paths. Since the sample paths of $W(t+\Theta)$ and $W(t)$ are the same except for a time-shift, the sample paths of $W(t+\Theta)$

are also predictable with a periodic predictor. Furthermore, the sample paths of a S and E, but not CE, process $X(t)$ can admit periodic AR (PAR) models (w.p. 1) with finite periodic order even though the stochastic process $X(t)$ admits no finite-order AR model. For example, let the CS and CE process $Z(t)$ be a first-order PAR process. The stationary process $X(t) = Z(t + \Theta)$ will not, in general, admit any finite-order AR model. But its sample paths, being the same as the sample paths of $Z(t)$ except for a time shift, admit first-order PAR models that can be identified from the sample path statistics.

Example 3: Let $X(t)$ and $Y(t)$ be jointly S (for all n) and E, but not CE, and, according to the usual definition of *causality*, let there be no causal relationship of $X(t)$ to $Y(t)$. That is, no linear or nonlinear *time-invariant* operation on $X(t)$ and its past has any prediction capability for $Y(t)$ and its future. Yet, each sample path of $Y(t)$ can possibly be perfectly *cyclically caused* by the corresponding sample path of $X(t)$. That is, a periodic operation on $X(t)$ can possibly perfectly predict $Y(t)$. As a specific example, consider the two continuous-time processes

$$X(t) = Z(t) = \pm 1$$

$$Y(t) = Z(t - \tau)\cos(t + \Theta)$$

where the transition times between +1 and −1 are arbitrary, and Θ is independent of $Z(t)$ and uniformly distributed on $[0, 2\pi]$. It can easily be shown that

$$E\left\{X^n(t-v)Y^m(t)\right\} = E\left\{X^n(t-v)\right\} E\left\{Y^m(t)\right\}$$

for all n and m, and all v; that is, $Y(t)$ is statistically independent of the past of $X(t)$. Nevertheless, each sample path of $Y(t)$ can be perfectly predicted from the past of the corresponding sample path of $X(t)$:

$$Y(t) = X(t - \tau)\cos(t + \Theta).$$

Example 4: Let $Z(t) = X(t) + Y(t)$, where $X(t)$ and $Y(t)$ are statistically independent, S (for $n = 2$), and E, but not CE, processes that have identical spectral densities. The Wiener filter for extracting $X(t)$ from $Z(t)$ (separating $X(t)$ and $Y(t)$) is essentially useless. Its transfer function is a constant. Yet the sample paths of $X(t)$ and $Y(t)$ can possibly be perfectly separated with a periodic filter. Examples include communication signals such as digital QAM, AM, PSK, ASK, and PAM. As a specific example, it can be shown (Gardner, 1993) that up to N spectrally coincident digital QAM signals with excess bandwidth $\geq (N-1)100\%$ can be perfectly separated.

Example 5: Let $\boldsymbol{X}(t) = \{X_1(t), X_2(t)\}$ be purely S (for all n) with a probabilistic model that is very similar to that of $\boldsymbol{Y}(t) = \{Y_1(t), Y_2(t)\}$, which is CS (e.g., $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are both Gaussian processes and the PSDs of $\boldsymbol{X}(t)$ equal those of the stationarized $\boldsymbol{Y}(t)$). The Cramér-Rao bounds of the same parameters in each of $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ (e.g., the relative time delay (TD) of an additive signal component common to both $X_1(t)$ and $X_2(t)$) can be drastically different. This has been demonstrated for TD at two reception platforms and for angle-of-arrival at a sensor array (Chen and

Gardner, 1992; Schell and Gardner, 1992b). Moreover, even the Cramér-Rao bound of the stationarized $\boldsymbol{Y}(t)$ can be drastically different from that of the purely S $\boldsymbol{X}(t)$.

Example 6: Let $X(t)$ and $Y(t)$ be independent, S (for all n), and E, but not CE, and let $Z(t)$ be specified under two hypotheses—

under hypothesis 1: $Z(t) = X(t) + Y(t)$

under hypothesis 2: $Z(t) = Y(t)$

The "optimum" (e.g., maximum-posterior-probability) detector for the presence of $X(t)$ in $Z(t)$ can be greatly outperformed by detectors that exploit the hidden CS in $X(t)$ and/or $Y(t)$, e.g., the joint maximum-posterior-probability detector and phase estimator (Gardner, 1988b; Gardner and Spooner, 1992a).

2.13 Two Paths into the Future

One approach to this unsettling situation, which is illustrated by the preceding examples, that should appeal to mathematicians is to take what shall be called *Path 1:* Develop the needed theory of PCE. The current status of the theory of CE and PCE is that substantial progress has been made for (1) CE w.p. 1 for discrete-time CS processes and Gaussian continuous-time CS processes, and (2) PCE in mean-square for finite-order moments of discrete- and continuous-time PCS processes. Little or no progress has been made for (1) PCE w.p. 1 for discrete-time PCS processes, (2) CE w.p. 1 for non-Gaussian continuous-time CS and PCS processes, and (3) PCE w.p. 1 for continuous-time CS and PCS processes.

The only paper to address PCE w.p. 1 (Boyles and Gardner, 1983) suggests that a substantial breakthrough will be required (even for the much less technical case of discrete time): conventional approaches and ideas apparently lead to dead ends. This suggests a challenge not unlike that Birkhoff faced around 1930 when he formulated and proved the fundamental ergodic theorem to replace the very unsatisfying "ergodic hypothesis." We need a fundamental polycycloergodic theorem that elegantly formalizes our informal notion of a PCE process in terms of a necessary and sufficient condition on the associated probability measure.

The most useful concept regarding PCE that we have for applications is the following unproved proposition.

Proposition *PCS processes constructed from stable (decaying-memory) nonrandom polyperiodic transformations of purely stationary ergodic processes are PCE.*

Stochastic-process models for many, if not most, communications signals can be constructed in this way.

In view of the difficulties before us, we should ask what some of the advantages of the stochastic-process approach are. The most apparent advantages are listed here:

1. It is the orthodox approach to modeling and studying evolutionary random phenomena and it is, therefore, attractive to those already familiar with it.

2. Mathematicians do know how, in principle, to construct stochastic-process models from elementary mathematical constructs (Borel fields, sigma algebras, probability measures, etc.). Therefore, there is a greater likelihood of success (compared with time-series) in constructing a mathematical theory of PCS and PCE processes from a few basic axioms.
3. It is possible, in principle, to exploit the hidden CS (or PCS) in a non-CE (or non-PCE) process within the conventional framework of stochastic processes. But, this requires that one have a model of the hidden CS (or PCS) that is *explicitly dependent* on one or more random phase variables Θ that are responsible for the lack of CE so that one can calculate probability densities and expectations *conditioned on* Θ.
4. Development of the theory of PCE will help clear the way for making the time-space theory of time-series mathematically rigorous.

In spite of these advantages, there is an alternative approach that should appeal to pragmatic engineers and scientists. Let us begin with the following perspective: The probability-space approach based on expectation introduces abstractions that, in many applications (e.g., many problems for which single-sample-path signal processing is of interest), have no redeeming *practical* value. Some of these abstractions can be properly dealt with only with a theory of PCE that is presently nonexistent. Regressing back to pre-1930 and adopting a "PCE hypothesis" is very unappealing (because the hypothesis can be false).

So, let us consider taking what shall be called *Path 2*: Adopt the time-space approach whose theory is in many ways dual to that of the probability-space approach, but without the practical drawbacks associated with cycloergodicity and the distracting abstraction associated with expectation over ensembles.

The essence of cyclostationarity from an operational standpoint is the fact that sine waves making up additive polyperiodic components can be generated from random data by applying certain nonlinear transformations. And, the time-space theory of cyclostationarity arises *naturally* out of the fundamental theorem of polyperiodic-component extraction using the generalized temporal expectation operation $\hat{E}^{\{\alpha\}}$; whereas the expectation E that gives rise to the probability-space theory has little to do with the essence of cyclostationarity.

But, can we construct time-series models? The answer is yes. Time-series models for many, if not most, communication signals can be constructed by subjecting one or more elementary time-series (e.g., purely stationary and white) to elementary transformations such as filters, periodic modulators, multiplexors, etc. (Gardner, 1987a). The defining properties of a discrete-time purely stationary white time-series are:

1. Whiteness: $\hat{f}^0_{x(t)}(x) = \hat{f}^0_{x(t+t_1)}(x_1)\hat{f}^0_{x(t+t_2)}(x_2)\dots\hat{f}^0_{x(t+t_n)}(x_n)$ for unequal $t_1, t_2, \dots, t_n$.
2. Pure stationarity: $\hat{f}^{\{\alpha\}}_{x(t)}(x) \equiv \hat{f}^0_{x(t)}(x)$ for all $\{\alpha\}$.
3. Existence: any sample path of any i.i.d. stochastic process will do (w.p. 1).

Okay. But, can we do probabilistic analysis using time-space theory? The answer, again, is yes. Performance measures such as bias, variance, Cramér-Rao bounds, confidence intervals, probabilities of decision-errors, etc., can be calculated using time-space theory just as well as they can using probability-space theory (Gardner, 1987a). But can we use the theory of statistical inference and decision? Yes, indeed.

The author's current assessment of progress along Path 2 can be summarized as follows: The considerable progress in the development and application of the time-space (or *temporal-probability*, or *fraction-of-time probability*) theory of CS and PCS time-series that has been made since its adoption by the UCD and SSPI groups in 1985, cf. (Gardner, 1987a, 1991a) includes:

1. Temporal and spectral second-order-moment theory (cyclic autocorrelation and cyclic spectra, or spectral correlation functions) (Gardner, 1987a; Section 3 in this chapter).
2. Temporal and spectral higher-order-moment and cumulant theory (cyclic cumulants and cyclic polyspectra, or spectral cumulants) (Gardner and Spooner, 1992b; Spooner and Gardner, 1992a,b; Spooner, 1992; Chapter 2 in this volume).
3. The rudiments of fraction-of-time probability distribution theory (Gardner, 1987a; Gardner and Brown, 1991).
4. A wide variety of applications of the theory to signal-processing and communications problems involving signal detection, signal classification, signal-parameter estimation, and signal-waveform estimation (Agee, et al., 1987, 1988, 1990; Brown, 1987; Chen, 1989; Chen and Gardner, 1992; Gardner, 1987a,b, 1988a,b,c, 1990a,b, 1991a,c, 1992, 1993; Gardner and Archer, 1993; Gardner and Brown, 1989; Gardner and Chen, 1988, 1992; Gardner and Paura, 1992; Gardner and Spooner, 1992a, 1993; Gardner and Venkataraman, 1990; Gardner et al., 1987, 1992; Schell, 1990; Schell and Agee, 1988; Schell and Gardner, 1989, 1990a,b,c, 1991, 1992, 1993a,b; Schell et al., 1989, 1993; Spooner, 1992; Spooner and Gardner, 1992b; Section 4 in this chapter; Chapter 2; Chapter 3).

Also, the conceptual gap between the existing time-space theory and its application to many signal-processing problems in communications is perceived by its current users to be much narrower than it is for the dual probability-space theory.

Further support for taking Path 2 includes the fact that the temporal-probability approach, which is centered on the concrete sine-wave extraction operation, has led naturally to a *derivation* of the cumulant as the solution to a fundamental problem in characterizing higher-order CS and PCS. It is doubtful that this derivation would have been discovered within the stochastic-process framework, which is centered on the abstract expectation operation. This derivation is discussed in Chapter 2 in this volume.

But, in the final analysis, the duality between the time-space and probability-space theories will likely result in either path taking the sufficiently persistent practical problem solver to the same places, although not necessarily in the same elapsed time or with the same energy. This duality can be formalized with the following loosely stated conjecture.

Conjecture *For every theorem that can be proved for a PCE PCS process, a dual theorem can be proved for a PCS time-series—and vice versa.*

This can be viewed as a generalization of Wold's isomorphism from S to PCS processes.

Nevertheless, there does remain a fundamental question that is not yet always answerable: Given a self-consistent set of probability distributions $\hat{F}^{\{\alpha\}}_{x(t)}$ for all orders n, does there exist a corresponding time-series $x(t)$? We have sufficient conditions on $\hat{F}^{\{\alpha\}}_{x(t)}$ that guarantee existence of $x(t)$: They are identical to the conditions that guarantee that $F_{X(t)}$ is PCE and they are called *mixing conditions* in the theory of stochastic processes. But we do not yet have a *necessary* and sufficient condition. (This presents another challenge for mathematicians.)

With regard to the taking of Paths 1 and 2, we can draw three conclusions:

1. The more abstract theory of PCS stochastic processes will undoubtedly be found to be of considerable value as it is developed, and those who are sufficiently mathematically inclined are encouraged to pursue this approach.
2. The less abstract theory of cyclostationary time-series is more accessible to engineers and scientists interested in theory as a *conceptual aid for solving practical problems.* It should be the preferred approach for the practically oriented whenever ensembles are not, in and of themselves, of primary concern.[11]
3. Both theories present important challenges to mathematicians.

In the remainder of this chapter, Path 2 is taken, and the theory and application of second-order (wide-sense) cyclostationarity is pursued in some detail.

3 INTRODUCTION TO THE PRINCIPLES OF SECOND-ORDER (WIDE-SENSE) CYCLOSTATIONARITY

The second-order (wide-sense[12]) theory of discrete-time stochastic processes deals with the probability-space autocorrelation function

$$R_X(t, t-\tau) = E\{X(t)X^*(t-\tau)\}.$$

[11] The practical value of this approach is amply demonstrated for parametric and nonparametric spectral analysis of S as well as CS and PCS time-series in (Gardner, 1987a).

[12] Wide-sense theory deals with moments, whereas strict-sense theory deals with probability distributions.

For a PCS process $X(t)$, this function is polyperiodic in t for each τ. The associated Fourier series for this function is

$$R_X(t, t-\tau) = \sum_{\{\alpha\}} R_X^\alpha(\tau)\, e^{i2\pi\alpha(t-\tau/2)}$$

where $\{\alpha\}$ includes all values of α in the principal domain $(-\frac{1}{2}, \frac{1}{2}]$ for which the corresponding Fourier coefficient is not identically zero as a function of τ:

$$R_X^\alpha(\tau) \triangleq \left\langle R_X(t, t-\tau)\, e^{-i2\pi\alpha(t-\tau/2)} \right\rangle \not\equiv 0.$$

If this PCS process is PCE, then (with probability equal to one)

$$R_X^\alpha(\tau) = \hat{R}_X^\alpha(\tau) \triangleq \left\langle X(t)X^*(t-\tau)\, e^{-i2\pi\alpha(t-\tau/2)} \right\rangle.$$

The sine waves $\exp[i2\pi\alpha(t-\tau/2)]$ in the Fourier series introduced here contain the time shift $-\tau/2$ so that the discrete-time theory presented here will match the continuous-time theory (cf. Gardner, 1985) in which the function $R_X(t+\tau/2, t-\tau/2)$ is expanded in a Fourier series with unshifted sine waves $\exp(i2\pi\alpha t)$.

The second-order theory of PCS discrete time-series $x(t)$ deals with the time-space autocorrelation function

$$\hat{E}^{\{\alpha\}}\left\{x(t)x^*(t-\tau)\right\} = \sum_{\{\alpha\}} \hat{E}^0\left\{x(t)x^*(t-\tau)\, e^{-i2\pi\alpha t}\right\} e^{i2\pi\alpha t}$$
$$= \sum_{\{\alpha\}} \hat{R}_x^\alpha(\tau)\, e^{i2\pi\alpha(t-\tau/2)},$$

where

$$\hat{R}_x^\alpha(\tau) \triangleq \hat{E}^0\{x(t)x^*(t-\tau)\, e^{-i2\pi\alpha(t-\tau/2)}\}.$$

That is, this theory deals with the sine-wave components in the delay product $x(t)x^*(t-\tau)$, whereas in the stochastic-process framework, we deal with an ensemble average that happens to be made up entirely of a sum of sine waves.

When our primary concern is the sine-wave components generated from $x(t)$ by the quadratic transformation $x(t)x^*(t-\tau)$, then the expectation operation $E\{\cdot\}$ and the associated ensemble are irrelevant. This being the case here, we proceed with the time-space theory. However, it is mentioned that the time-space theory presented can be translated to a probability-space theory (cf. Section 2) simply by following the rule:

> For all sinusoidally-weighted time averages $\langle z(t)e^{-i2\pi\alpha t}\rangle$ of time-series $z(t)$, replace $z(t)$ by the expected value $E\{Z(t)\}$ of the corresponding stochastic process $Z(t)$ to obtain $\langle E\{Z(t)\}e^{-i2\pi\alpha t}\rangle$ (when $\alpha = 0$, the operation $\langle\cdot\rangle$ can be omitted to obtain $E\{Z(t)\}$ *only* if $Z(t)$ is purely stationary).
>
> Common examples of $z(t)$ appearing in this presentation include delay products $z(t) = x(t)x^*(t-\tau)$ and cross products $z(t) = u(t)v^*(t)$.

In the remainder of this chapter, the circumflex notation (that was introduced in Section 2) on all time-average quantities is omitted for simplicity.

In the first part of this section, the possibility of generating spectral lines by simply squaring the signal is illustrated for two types of signals: the random-amplitude modulated sine wave and the random-amplitude modulated periodic pulse train. Then in the second part, it is explained that the property that enables spectral-line generation with some type of quadratic time-invariant transformation is called *cyclostationarity of order 2* (in the wide sense) and is characterized by the *cyclic autocorrelation function*, which is a generalization of the conventional autocorrelation function. Following this, it is shown that a signal exhibits cyclostationarity if and only if the signal is correlated with certain frequency-shifted versions of itself.

In the third and last part of this section, the correlation of frequency-shifted versions of a signal is localized in the frequency domain and this leads to the definition of a *spectral correlation density function*. It is then explained that this function is the Fourier transform of the cyclic autocorrelation function. This Fourier-transform relation between these two functions includes as a special case the well-known *Wiener relation* between the power spectral density function and the autocorrelation function. A normalization of the spectral correlation density function that converts it into a spectral correlation coefficient, whose magnitude is between zero and unity, is then introduced as a convenient measure of the degree of spectral redundancy in a signal.

Continuing in the final part of this section, the effects on the spectral correlation density function of several signal-processing operations are described. These include filtering and waveform multiplication, which in turn include the special cases of time delay and multipath propagation, bandlimiting, frequency conversion, and time sampling. These results are used to derive the spectral correlation density function for the random-amplitude modulated sine wave, the random-amplitude modulated pulse train, and the binary phase-shift keyed sine wave. The spectral correlation density functions for some other types of phase-shift keyed signals are also described graphically.

To conclude this section, the measurement of the (estimation of the ideal) spectral correlation density function is discussed and a particular algorithm for this purpose is illustrated with a simulation of a phase-shift keyed signal.

To complement similar treatments of this material (Gardner, 1987a; Gardner, 1991a), attention is focused in this section primarily on discrete-time signals rather than continuous-time signals.[13]

3.1 Spectral Line Generation

A discrete-time signal $x(t)$, for $t = 0, \pm 1, \pm 2, \pm 3, \ldots$, contains a *finite-strength additive sine-wave component* (an ac component) with frequency α, say

$$a\cos(2\pi\alpha t + \theta) \quad \text{with } \alpha \neq 0 \tag{1}$$

[13]For convenience, the notation herein is modified from that in (Gardner, 1987a; Gardner 1991a): here, $\tilde{R}_x^\alpha$ and $\tilde{S}_x^\alpha$ are used for continuous time and R_x^α and S_x^α are used for discrete time.

if the Fourier coefficient

$$M_x^\alpha = \langle x(t)\, e^{-i2\pi\alpha t}\rangle \tag{2}$$

is not zero, in which case (1) gives

$$M_x^\alpha = \frac{1}{2} a\, e^{i\theta}.$$

In (2), the operation $\langle\cdot\rangle$ is the time-averaging operation

$$\langle\cdot\rangle \triangleq \lim_{Z\to\infty} \frac{1}{2Z+1} \sum_{t=-Z}^{Z} (\cdot).$$

In this case, the power spectral density (PSD) of $x(t)$ includes a spectral line at $f = \alpha$ and its image $f = -\alpha$. (The PSD is defined later in this section.) That is, the PSD in the principal domain $(-1/2, 1/2]$ contains the additive term[14]

$$|M_x^\alpha|^2 [\delta(f-\alpha) + \delta(f+\alpha)], \tag{3}$$

where $\delta(\cdot)$ is the Dirac delta, or impulse, function. For convenience in the sequel, it is said that such a signal exhibits *first-order periodicity*, with frequency α.

Let $x(t)$ be decomposed into the sum of its finite-strength sine-wave component, with frequency α, and its residual, say $n(t)$,

$$x(t) = a\cos(2\pi\alpha t + \theta) + n(t), \tag{4}$$

where $n(t)$ is defined to be that which is left after subtraction of (1) from $x(t)$. It is assumed that $n(t)$ is random. Here, the term *random* is used to denote nothing more than the vague notion of erratic or unpredictable behavior. If the sine wave is weak relative to the random residual, it might not be evident from visual inspection of $x(t)$ that it contains a periodic component. Hence, it is said to contain *hidden periodicity*. However, because of the associated spectral lines, hidden periodicity can be detected and in some applications exploited through techniques of spectral analysis.

This presentation is concerned with signals that contain more subtle types of hidden periodicity that, unlike first-order periodicity, do not give rise to spectral lines in the PSD, but that can be converted into first-order periodicity by a nonlinear time-invariant transformation of the signal. In particular, we shall focus on the type of hidden periodicity that can be converted by a quadratic transformation to yield spectral lines in the PSD.

The discussion begins with two motivating examples. In the convention used here, the PSD for $x(t)$ is denoted by $S_x(f)$ and is periodic with unity period. $\tilde{S}_x(f)$

[14]The strength of the spectral line is $|M_x^\alpha|^2$ as indicated in (3) if and only if the limit (2) exists in the temporal mean square sense with respect to the time parameter u obtained by replacing t with $t+u$ in (2) (Gardner, 1987a, Chapter 15, exc. 6).

denotes the PSD restricted to the principal domain $(-1/2, 1/2]$; therefore,

$$S_x(f) = \sum_{n=-\infty}^{\infty} \tilde{S}_x(f+n).$$

On occasion, continuous-time signals also are discussed herein. In such cases it is assumed that the signal is time-scaled and bandlimited so that the PSD is restricted to the band $(-1/2, 1/2]$. Consequently, the PSD of the discrete-time sampled version, restricted to the principal domain, will be identical to the PSD of the continuous-time signal. Consequently, the same notation, $\tilde{S}_x(f)$, is used for both.

Example 1: AM Let $a(t)$ be a real random lowpass signal (say lowpass filtered thermal noise) with the PSD $S_a(f)$ shown in Fig. 5a, which contains no spectral lines. If $a(t)$ is used to modulate the amplitude of a sine wave, we obtain the amplitude-modulated (AM) signal

$$x(t) = a(t)\cos(2\pi f_o t), \tag{5}$$

whose PSD $S_x(f)$ is given by (Gardner, 1987a, Chapter 3, Sec. D)

$$S_x(f) = \frac{1}{4}S_a(f+f_o) + \frac{1}{4}S_a(f-f_o) \tag{6}$$

as shown in Fig. 5b.

Although the PSD is centered about $f = f_o$ and $f = -f_o$, there is no spectral line at f_o or $-f_o$. The reason for this is that, as shown in Fig. 5a, there is no spectral line in $S_a(f)$ at $f = 0$. This means that the dc component

$$M_a^0 \triangleq \langle a(t) \rangle \tag{7}$$

is zero, since the strength of any spectral line at $f = 0$ is $|M_a^0|^2$.

Let us now square $x(t)$ to obtain

$$\begin{aligned} y(t) &= x^2(t) = a^2(t)\cos^2(2\pi f_o t) \\ &= \tfrac{1}{2}\left[b(t) + b(t)\cos(4\pi f_o t)\right] \end{aligned} \tag{8}$$

where

$$b(t) = a^2(t). \tag{9}$$

Since $b(t)$ is nonnegative, its dc value must be positive: $M_b^0 > 0$. Consequently, the PSD of $b(t)$ contains a spectral line at $f = 0$, as shown in Fig. 5c. The PSD for $y(t)$ is given by

$$S_y(f) = \frac{1}{4}\left[S_b(f) + \frac{1}{4}S_b(f+2f_o) + \frac{1}{4}S_b(f-2f_o)\right] \tag{10}$$

and, as shown in Fig. 5d, it contains spectral lines at $f = \pm 2f_o$ as well as at $f = 0$. Thus, by putting $x(t)$ through a quadratic transformation (a squarer in this case) we

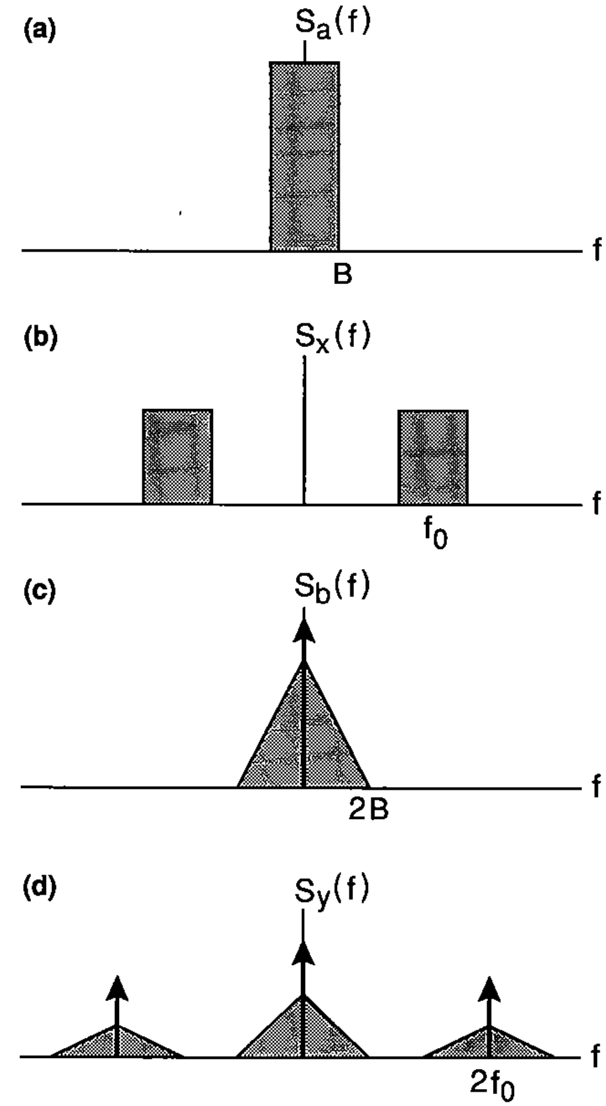

Figure 5: (a) Power spectral density (PSD) of a lowpass signal. (b) PSD of an amplitude-modulated (AM) signal. (c) PSD of a squared lowpass signal. (d) PSD of a squared AM signal.

have converted the hidden periodicity resulting from the sine-wave factor $\cos(2\pi f_o t)$ in (5) into first-order periodicity with associated spectral lines. This is particularly easy to see if $a(t)$ is a random binary sequence that switches back and forth between

+1 and −1 because then $b(t) \equiv 1$ and $y(t)$ in (8) is therefore a periodic signal:

$$y(t) = \frac{1}{2} + \frac{1}{2}\cos(4\pi f_o t).$$

Example 2: PAM As another example, we consider the real pulse-amplitude-modulated (PAM) signal

$$x(t) = \sum_{n=-\infty}^{\infty} a(nT_o)\, p(t - nT_o), \tag{11}$$

where the pulse $p(t)$ is confined within the interval $(-T_o/2, T_o/2)$ so that the pulse translates do not overlap, as shown in Fig. 6. For simplicity, we consider a continuous-time signal in this example (to avoid aliasing). The PSD of $x(t)$ is given by (Gardner, 1987a, Chapter 3, Sec. D)

$$\tilde{S}_x(f) = \frac{1}{T_o}\left|\tilde{P}(f)\right|^2 \sum_{m=-\infty}^{\infty} \tilde{S}_a(f - m/T_o), \tag{12}$$

where $\tilde{S}_a(f)$ is shown in Fig. 5a, which contains no spectral lines, and where $\tilde{P}(f)$ is the Fourier transform of $p(t)$. Since there are no spectral lines in $\tilde{S}_a(f)$ (or $\tilde{P}(f)$ since $p(t)$ has finite duration), there are none in $\tilde{S}_x(f)$, as shown in Fig. 7a, regardless of the periodic repetition of pulses in $x(t)$. But, let us look at the square of $x(t)$:

$$y(t) = x^2(t) = \sum_{n=-\infty}^{\infty} b(nT_o)\, q(t - nT_o), \tag{13}$$

where

$$b(nT_o) = a^2(nT_o) \tag{14a}$$

and

$$q(t) = p^2(t). \tag{14b}$$

The PSD for $y(t)$ is given by

$$\tilde{S}_y(f) = \frac{1}{T_o}\left|\tilde{Q}(f)\right|^2 \sum_{m=-\infty}^{\infty} \tilde{S}_b(f - m/T_o), \tag{15}$$

where $\tilde{Q}(f)$ is the Fourier transform of $q(t)$. Because of the spectral line at $f = 0$ in $\tilde{S}_b(f)$, which is shown in Fig. 5c, we have spectral lines in $\tilde{S}_y(f)$ at the harmonics m/T_o (for some integer values of m) of the pulse rate $1/T_o$, as shown in Fig. 7b. Thus, again, we have converted the hidden periodicity in $x(t)$ into first-order periodicity with associated spectral lines by using a quadratic transformation. This is particularly easy to see if $a(nT_o)$ is a random binary sequence with values ±1, because then $b(nT_o) \equiv 1$ and $y(t)$ in (13) is therefore a periodic signal

$$y(t) = \sum_{n=-\infty}^{\infty} q(t - nT_o). \tag{16}$$

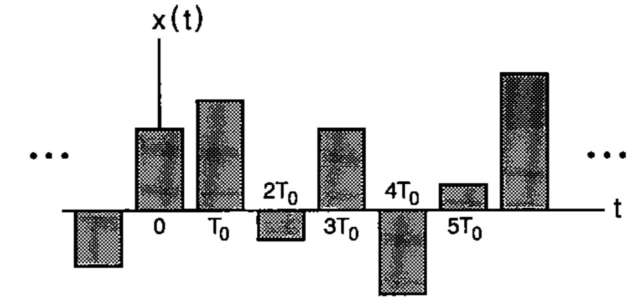

Figure 6: A pulse-amplitude-modulated (PAM) signal with pulse width less than interpulse time.

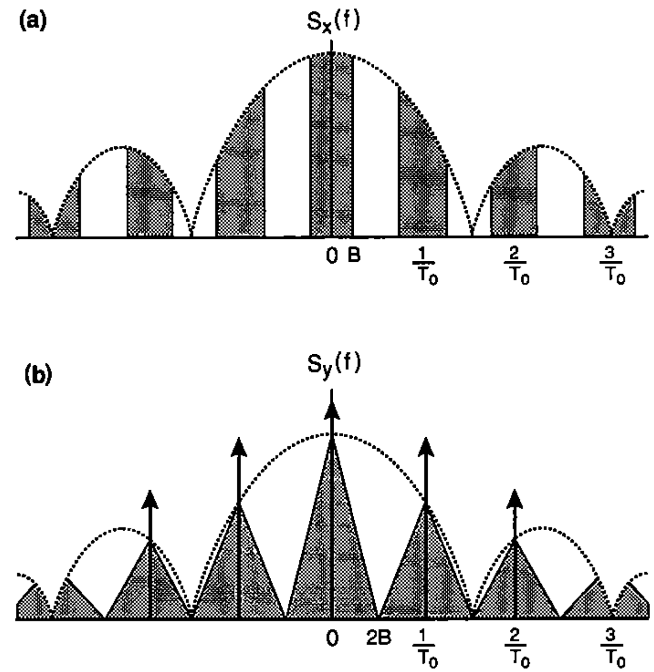

Figure 7: (a) Power spectral density (PSD) of a pulse-amplitude-modulated (PAM) signal with 67% duty-cycle pulses. (b) PSD of the squared PAM signal.

3.2 The Cyclic Autocorrelation Function

Although the squaring transformation works in these examples, a different quadratic transformation involving delays can be required in some cases. For example, if $a(nT_o)$

in Example 2 is again binary, but $p(t)$ is flat with height 1 and width T_o, as shown in Fig. 8, then $y(t) = x^2(t) = 1$, which is a constant for all t. Thus, we have a spectral line at $f = 0$ but none at the harmonics of the pulse rate. Nevertheless, if we use the quadratic transformation

$$y(t) = x(t)x(t-\tau) \tag{17}$$

for any of a number of nonzero delays τ, we will indeed obtain spectral lines at $f = m/T_o$. That is,

$$\begin{aligned} M_y^\alpha &= \langle y(t)\, e^{-i2\pi\alpha t}\rangle \\ &= \langle x(t)x(t-\tau)\, e^{-i2\pi\alpha t}\rangle \neq 0 \end{aligned} \tag{18}$$

for $\alpha = m/T_o$ for some m.

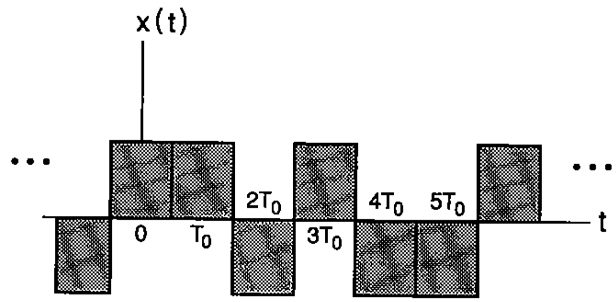

Figure 8: A binary pulse-amplitude-modulated (PAM) signal with full duty-cycle pulses.

The most general time-invariant quadratic transformation of a real time-series $x(t)$ is simply a linear combination of delay products

$$y(t) = \sum_{\tau_1,\tau_2} h(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)$$

for some weighting function $h(\tau_1, \tau_2)$ that is analogous to the impulse-response function for a linear transformation. This motivates us to define the property of *second-order periodicity* as follows: The real signal $x(t)$ contains second-order periodicity if and only if the PSD of the delay-product signal $x(t-\tau_1)x(t-\tau_2)$ for some delays τ_1 and τ_2 contains spectral lines at some nonzero frequencies α. But, this will be so if and only if the PSD of (17) for some delays ($\tau = \tau_2 - \tau_1$) contains spectral lines at some nonzero frequencies $\alpha \neq 0$; that is, if and only if (18) is satisfied.

In developing the continuous-time theory of second-order periodicity it has been found to be more convenient to work with the symmetric delay product

$$y_\tau(t) = x(t+\tau/2)x^*(t-\tau/2). \tag{19}$$

The complex conjugate $*$ is introduced here for generality to accommodate complex-valued signals, but it is mentioned that for some complex-valued signals, the quadratic

transformation without the conjugate can also be useful (Gardner, 1987a, Chapter 10, Sec. C). From (19), the fundamental parameter (18) of second-order periodicity for continuous time becomes

$$\tilde{R}_x^\alpha(\tau) \triangleq \langle x(t+\tau/2)x^*(t-\tau/2)\, e^{-i2\pi\alpha t}\rangle, \tag{20a}$$

which is the Fourier coefficient $M_{y_\tau}^\alpha$ of the additive sine-wave component with frequency α contained in the delay-product signal $y_\tau(t)$. However, for discrete-time signals, delays equal to half the sampling increment are not allowed. Nevertheless, since

$$\langle x(t)x^*(t-\tau)e^{-i2\pi\alpha t}\rangle = \tilde{R}_x^\alpha(\tau)\, e^{-i\pi\alpha\tau}$$

for continuous time, then we can define the fundamental parameter of second-order periodicity for discrete time as follows

$$R_x^\alpha(\tau) \triangleq \langle x(t)x^*(t-\tau)\, e^{-i2\pi\alpha t}\rangle\, e^{i\pi\alpha\tau} \tag{20b}$$

in order to maintain the strongest analogy between the continuous- and discrete-time theories. Observe that since t and τ take on only integer values, then $R_x^\alpha(\tau)$ is periodic in α with period two, and also $R_x^{\alpha+1}(\tau) = R_x^\alpha(\tau)e^{i\pi\tau}$.

The notation $R_x^\alpha(\tau)$ is introduced for this Fourier coefficient because, for $\alpha = 0$, (20) reduces to the conventional autocorrelation function

$$R_x^0(\tau) = \langle x(t)x^*(t-\tau)\rangle,$$

for which the notation $R_x(\tau)$ is commonly used. Furthermore, since $R_x^\alpha(\tau)$ is a generalization of the autocorrelation function, in which a cyclic (sinusoidal) weighting factor $e^{-i2\pi\alpha t}$ is included before the time-averaging is carried out, $R_x^\alpha(\tau)$ is called the *cyclic autocorrelation function*. Also, the *conjugate cyclic autocorrelation* for complex-valued signals obtained from (20) by deleting the conjugate,

$$R_{xx^*}^\alpha(\tau) = \langle x(t)x(t-\tau)\, e^{-i2\pi\alpha t}\rangle\, e^{i\pi\alpha\tau}, \tag{21}$$

is a further modification of the conventional autocorrelation.[15]

Thus, we have two distinct interpretations of $R_x^\alpha(\tau) = M_{y_\tau}^\alpha$. In fact, we have yet a third distinct interpretation, which can be obtained by simply factoring $\exp(-i2\pi\alpha t)$ in order to reexpress (20) as

$$R_x^\alpha(\tau) = \left\langle \left[x(t)\, e^{-i\pi\alpha t}\right]\left[x(t-\tau)\, e^{+i\pi\alpha(t-\tau)}\right]^* \right\rangle. \tag{22}$$

[15] Although some readers will recognize the similarity between the cyclic autocorrelation function and the radar ambiguity function, the relationship between these two functions is only superficial. The concepts and theory underlying the cyclic autocorrelation function, as summarized in this article, have little in common with the concepts and theory of radar ambiguity (cf. (Gardner, 1987a, Chapter 10, Sec. C)). For example, the radar ambiguity function has no meaning relevant to ambiguity (in Doppler) when applied to a real signal, or when applied to a complex signal without the conjugate.

That is, $R_x^\alpha(\tau)$ is actually a conventional crosscorrelation function

$$R_{uv}(\tau) \triangleq \langle u(t)v^*(t-\tau)\rangle = R_x^\alpha(\tau), \tag{23}$$

where

$$u(t) = x(t)\, e^{-i\pi\alpha t} \tag{24a}$$

and

$$v(t) = x(t)\, e^{+i\pi\alpha t} \tag{24b}$$

are frequency translates of $x(t)$. Recall that multiplying a signal by $\exp\{\pm i\pi\alpha t\}$ shifts the spectral content of the signal by $\pm\alpha/2$. For example, the PSDs of $u(t)$ and $v(t)$ are

$$S_u(f) = S_x(f+\alpha/2) \tag{25}$$

and

$$S_v(f) = S_x(f-\alpha/2). \tag{26}$$

It follows from (23) and (24) that $x(t)$ exhibits second-order periodicity ((20) is not identically zero as a function of τ for some $\alpha \neq 0$) if and only if *frequency translates* of $x(t)$ are correlated with each other in the sense that (23) is not identically zero as a function of τ for some $\alpha \neq 0$ in (24). This third interpretation of $R_x^\alpha(\tau)$ suggests an appropriate way to normalize $R_x^\alpha(\tau)$ as explained next.

As long as the mean values of the frequency translates $u(t)$ and $v(t)$ are zero (which means that $x(t)$ does not contain finite-strength[16] additive sine-wave components at frequencies $\pm\alpha/2$ and, therefore, that $S_x(f)$ has no spectral lines at $f = \pm\alpha/2$), the crosscorrelation $R_{uv}(\tau) \equiv R_x^\alpha(\tau)$ is actually a temporal crosscovariance $K_{uv}(\tau)$. That is,

$$\begin{aligned} K_{uv}(\tau) &\triangleq \langle [u(t) - \langle u(t)\rangle]\,[v(t-\tau) - \langle v(t-\tau)\rangle]^*\rangle \\ &= \langle u(t)v^*(t-\tau)\rangle = R_{uv}(\tau) \end{aligned} \tag{27}$$

An appropriate normalization for the temporal crosscovariance is the geometric mean of the two corresponding temporal variances. This yields a temporal correlation coefficient, the magnitude of which is upper bounded by unity. It follows from (24) that the two variances are given by

$$K_u(0) = R_u(0) = \langle |u(t)|^2\rangle = R_x(0) \tag{28a}$$

and

$$K_v(0) = R_v(0) = \langle |v(t)|^2\rangle = R_x(0). \tag{28b}$$

Therefore, the temporal correlation coefficient for frequency translates is given by

$$\frac{K_{uv}(\tau)}{[K_u(0)K_v(0)]^{1/2}} = \frac{R_x^\alpha(\tau)}{R_x(0)} \triangleq \gamma_x^\alpha(\tau). \tag{29}$$

[16] It does contain infinitesimal sine-wave components.

Hence, the appropriate normalization factor for the cyclic autocovariance $R_x^\alpha(\tau)$ is simply $1/R_x(0)$ (and it is the same for the conjugate cyclic autocovariance).

This is a good point at which to introduce some more terminology. A signal $x(t)$ for which the autocorrelation $R_x(\tau)$ exists (e.g., remains finite as the averaging time goes to infinity) and is not identically zero (as it is for transient signals) is commonly said to be *stationary* of second order (in the wide sense). But we need to refine the terminology to distinguish between those stationary signals that exhibit second-order periodicity ($R_x^\alpha(\tau) \not\equiv 0$ for some $\alpha \neq 0$) and those stationary signals that do not ($R_x^\alpha(\tau) \equiv 0$ for all $\alpha \neq 0$). Consequently, we shall call the latter for which $R_x^\alpha(\tau) \equiv 0$ *stationary* of second order (in the wide sense) and the former for which $R_x^\alpha(\tau) \not\equiv 0$ for some values of α that are integer multiples of a single fundamental frequency $1/T$ (corresponding to the period T) *cyclostationary* of second order (in the wide sense). If there is more than one fundamental frequency, then we call the signal *polycyclostationary* of second order (in the wide sense). We shall also call any nonzero value of the frequency parameter α in the principal domain $(-\frac{1}{2}, \frac{1}{2}]$ for which $R_x^\alpha(\tau) \not\equiv 0$ a *cycle frequency*. The discrete set of cycle frequencies is called the *cycle spectrum*. For example, if a signal is cyclostationary, the cycle spectrum contains only harmonics (integer multiples) of the fundamental cycle frequency, which is the reciprocal of the fundamental period. But if the signal is polycyclostationary, then the cycle spectrum contains harmonics of each of the incommensurate fundamental cycle frequencies.

We conclude this section by reconsidering the AM example and determining the cyclic autocorrelation function for the AM signal.

Example 1 continued: AM Let $a(t)$ be a real random stationary signal with zero mean:

$$\langle a(t)\rangle = 0, \tag{30a}$$

$$\langle a(t)a^*(t-\tau)\rangle \not\equiv 0, \tag{30b}$$

$$\langle a(t)a^*(t-\tau)\, e^{-i2\pi\alpha t}\rangle \equiv 0 \quad \text{for all } \alpha \neq 0. \tag{30c}$$

Equation (30c) guarantees that

$$\langle a(t)\, e^{-i2\pi\alpha t}\rangle \equiv 0 \quad \text{for all } \alpha \neq 0. \tag{30d}$$

We consider the amplitude-modulated sine wave

$$\begin{aligned} x(t) &= a(t)\cos(2\pi f_o t + \theta) \\ &= \frac{1}{2}a(t)\left[e^{i(2\pi f_o t+\theta)} + e^{-i(2\pi f_o t+\theta)}\right]. \end{aligned} \tag{31}$$

Because of (30d), $a(t)$ contains no finite-strength additive sine-wave components and, therefore (together with (30a)), $x(t)$ contains no finite-strength additive sine-wave components. This means that its power spectral density contains no spectral lines. However, the quadratic transformation

$$y_\tau(t) = x(t)x^*(t-\tau)$$
$$= a(t)a^*(t-\tau)\frac{1}{4}\left[e^{i2\pi f_o\tau} + e^{-i2\pi f_o\tau} + e^{i(4\pi f_o t+2\theta)}e^{-i2\pi f_o\tau} + e^{-i(4\pi f_o t+2\theta)}e^{i2\pi f_o\tau}\right] \quad (32)$$

does contain finite-strength additive sine-wave components with frequencies $\alpha = \pm 2f_o$, since (30b) renders one or the other of the last two terms in the quantity

$$\begin{aligned}\langle y_\tau(t)\, e^{-i2\pi\alpha t}\rangle &= \frac{1}{4}\, e^{i2\pi f_o\tau}\langle a(t)a^*(t-\tau)e^{-i2\pi\alpha t}\rangle \\ &+\frac{1}{4}\, e^{-i2\pi f_o\tau}\langle a(t)a^*(t-\tau)\, e^{-i2\pi\alpha t}\rangle \\ &+\frac{1}{4}\, e^{i2\theta}\, e^{-i2\pi f_o\tau}\langle a(t)a^*(t-\tau)\, e^{-i2\pi(\alpha-2f_o)t}\rangle \\ &+\frac{1}{4}\, e^{-i2\theta}\, e^{i2\pi f_o\tau}\langle a(t)a^*(t-\tau)\, e^{-i2\pi(\alpha+2f_o)t}\rangle\end{aligned} \quad (33)$$

nonzero for $\alpha = \pm 2f_o$. That these are the only two nonzero cycle frequencies follows from the fact that (30c) renders (33) equal to zero for all α except $\alpha = 0$ and $\alpha = \pm 2f_o$. Thus, the cycle spectrum consists of only the two cycle frequencies $\alpha = \pm 2f_o$ and the degenerate cycle frequency $\alpha = 0$.

Hence, the versions $u(t)$ and $v(t)$ of $x(t)$ obtained by frequency shifting $x(t)$ up and down by $\alpha/2 = f_o$ are correlated. This is not surprising since (31) reveals that $x(t)$ is obtained from $a(t)$ by frequency shifting up and down by f_o and then adding. In conclusion, we have the cyclic autocorrelation function (in the principal domain of α)

$$R_x^\alpha(\tau) = \begin{cases} \frac{1}{4}e^{\pm i2\theta}R_a(\tau) & \text{for } \alpha = \pm 2f_o \\ \frac{1}{2}R_a(\tau)\cos(2\pi f_o\tau) & \text{for } \alpha = 0 \\ 0 & \text{otherwise,} \end{cases} \quad (34)$$

the magnitude of which is graphed in Fig. 9 for a typical autocorrelation $R_a(\tau)$. It follows from (34) that the temporal correlation coefficient is given by

$$\gamma_x^\alpha(\tau) = \begin{cases} \frac{1}{2}e^{\pm i2\theta}\gamma_a^0(\tau) & \text{for } \alpha = \pm 2f_o \\ \gamma_a^0(\tau)\cos(2\pi f_o\tau) & \text{for } \alpha = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (35a)$$

Thus, the strength of correlation between $x(t)\exp(-i\pi\alpha t)$ and $x(t-\tau)\ \exp(i\pi\alpha[t-\tau])$, which is given by

$$\left|\gamma_x^\alpha(\tau)\right| = \frac{1}{2}\left|\gamma_a^0(\tau)\right|, \quad (35b)$$

can be substantial (as large as 1/2 for this amplitude-modulated signal.

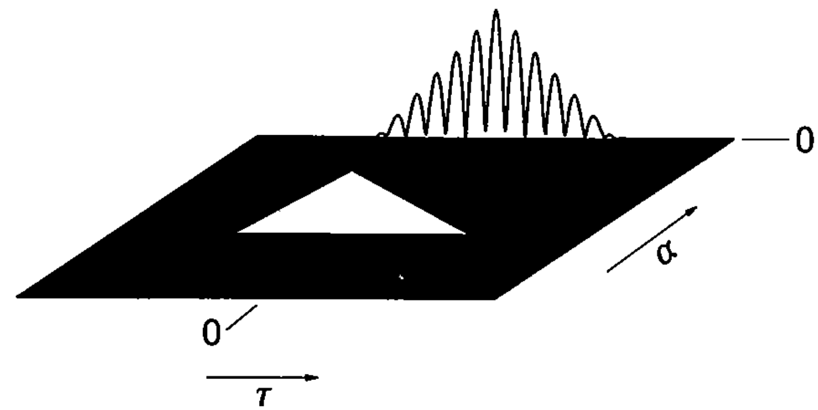

Figure 9: Magnitude of the cyclic autocorrelation function for an AM signal graphed as the height of a surface above the time-frequency plane with coordinates τ and α.

As an especially simple specific example of $a(t)$, we consider as before a random binary sequence, which switches back and forth between $+1$ and -1. If we set $\tau = 0$ in (32), we obtain

$$\begin{aligned} y_0(t) &= |x(t)|^2 = |a(t)|^2\cos^2(2\pi f_o t + \theta) \\ &= \frac{1}{2} + \frac{1}{2}\cos(4\pi f_o t + 2\theta), \end{aligned}$$

which clearly contains finite-strength additive sine-wave components with frequencies $\alpha = \pm 2f_o$. In fact, in this very special case, there is no random component in $y_0(t)$. On the other hand, for $\tau \neq 0$, $y_\tau(t)$ contains both a sine-wave component and a random component.

To illustrate the conjugate cyclic autocorrelation (21), let us consider the analytic signal for AM,

$$z(t) = \frac{1}{2}a(t)\, e^{i(2\pi f_o t+\theta)}.$$

For this signal, we have

$$\begin{aligned} R_{zz^*}^\alpha(\tau) &\triangleq \langle z(t)z(t-\tau)\, e^{-i2\pi\alpha t}\rangle\, e^{i\pi\alpha\tau} \\ &= \frac{1}{4}\langle a(t)a(t-\tau)\, e^{i2\pi(2f_o-\alpha)t}\rangle\, e^{-i[2\pi(f_o-\alpha/2)\tau-2\theta]} \\ &= \begin{cases} \frac{1}{4}R_a(\tau)\, e^{i2\theta} & \text{for } \alpha = 2f_o \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Other examples of cyclostationary and polycyclostationary signals can be similarly viewed as mixtures of stationarity and periodicity. Examples are cited in

Section 1. Typical cycle spectra include harmonics of pulse rates, keying rates, spreading-code chipping rates, frequency hopping rates, code repetition rates, doubled carrier frequencies, and sums and differences of these (Gardner, 1987a, Chapter 12).

3.3 The Spectral Correlation Density Function

In the same way that it is beneficial for some purposes to localize in the frequency domain the average power $\langle |x(t)|^2 \rangle = R_x(0)$ in a stationary random signal, it can be very helpful to localize in frequency the correlation $\langle u(t)v^*(t)\rangle = \langle |x(t)|^2 e^{-i2\pi\alpha t}\rangle \equiv R_x^\alpha(0)$ of frequency-shifted signals $u(t)$ and $v(t)$ for a cyclostationary or polycyclostationary random signal $x(t)$. In the former case of localizing the power, we simply pass the signal of interest $x(t)$ through a narrowband bandpass filter and then measure the average power at the output of the filter. By doing this with many filters whose center frequencies are separated by the bandwidth of the filters, we can partition any spectral band of interest into a set of contiguous narrow disjoint bands. In the limit as the bandwidths approach zero, the corresponding set of measurements of average power, normalized by the bandwidth, constitute the power spectral density (PSD) function. That is, at any particular frequency f (in the principal domain $(-1/2, 1/2]$), the PSD for $x(t)$ is given by

$$S_x(f) \triangleq \lim_{B\to 0} \frac{1}{B}\left\langle \left| h_B^f(t) \otimes x(t) \right|^2 \right\rangle, \tag{36}$$

where $\otimes$ denotes convolution and $h_B^f(t)$ is the discrete-impulse response of a one-sided bandpass filter with center frequency f, bandwidth B, and unity gain at the band center (see Fig. 10).

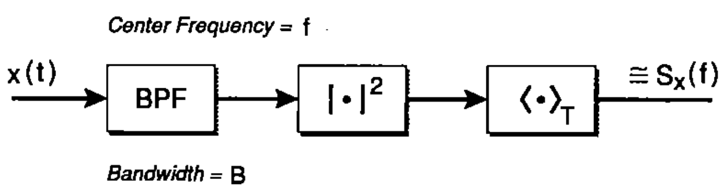

Figure 10: One channel of a spectrum analyzer for measuring the power spectral density (PSD). (The symbol $\cong$ indicates that the output only approximates the ideal function $S_x(f)$ for finite T and B.)

In the latter case of localizing the correlation, we simply pass both of the two frequency translates $u(t)$ and $v(t)$ of $x(t)$ through the same set of bandpass filters that are used for the PSD and then measure the temporal correlation of the filtered signals (see Fig. 11) to obtain

$$S_x^\alpha(f) \triangleq \lim_{B\to 0} \frac{1}{B}\left\langle \left[h_B^f(t) \otimes u(t)\right]\left[h_B^f(t) \otimes v(t)\right]^* \right\rangle, \tag{37}$$

which is called the *spectral correlation density (SCD) function.* This yields the spectral density of correlation in $u(t)$ and $v(t)$ at frequency f, which is identical to

the spectral density of correlation in $x(t)$ at frequencies $f + \alpha/2$ and $f - \alpha/2$ (see Fig. 12). That is, $S_x^\alpha(f)$ is the bandwidth-normalized (i.e., divided by B) correlation of the amplitude and phase fluctuations of the narrowband spectral components in $x(t)$ centered at frequencies $f + \alpha/2$ and $f - \alpha/2$, in the limit as the bandwidth B of these narrowband components approaches zero. For complex-valued signals, the conjugate SCD obtained from (37) by deleting the complex conjugate is also of interest for some signals (Gardner, 1987a, Chapter 10, Sec. C).

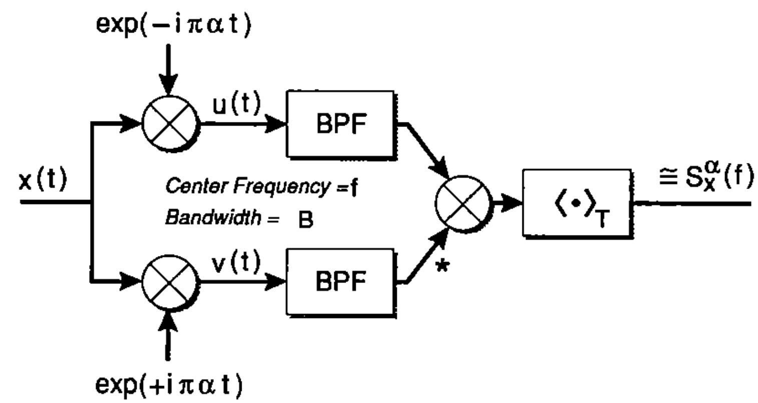

Figure 11: One channel-pair of a spectral correlation analyzer (or a cyclic spectrum analyzer) for measuring the spectral correlation density (or cyclic spectral density).

Strictly speaking, the SCD is not a valid density function in the usual sense, since it is not nonnegative and, in fact, not even real-valued. However, its integral over all frequencies does equal the correlation of $u(t)$ and $v(t)$ and, when $u(t)$ and $v(t)$ are decomposed into narrowband spectral components, the correlation of the components centered at f is indeed the SCD evaluated at f. Because of the lack of the nonnegativity property of the SCD, the correlation of $u(t)$ and $v(t)$ can equal zero without the SCD being identically zero because the integral of the SCD over all f can be zero even though the SCD is not identically zero. Nevertheless, because of the properties that the SCD does share with densities like the PSD, the term *density* is retained.

It is well known (see, for example, (Gardner, 1987a, Chapter 3, Sec. C) for a proof for continuous time) that the PSD obtained from (36) is equal to the Fourier transform of the autocorrelation function,

$$S_x(f) = \sum_{\tau=-\infty}^{\infty} R_x(\tau)\, e^{-i2\pi f\tau}. \tag{38}$$

Similarly, it can be shown (cf. (Gardner, 1987a, Chapter 11, Sec. C) for continuous time) that the SCD (or conjugate SCD) obtained from (37) is the Fourier transform of the cyclic autocorrelation function (or conjugate cyclic autocorrelation),

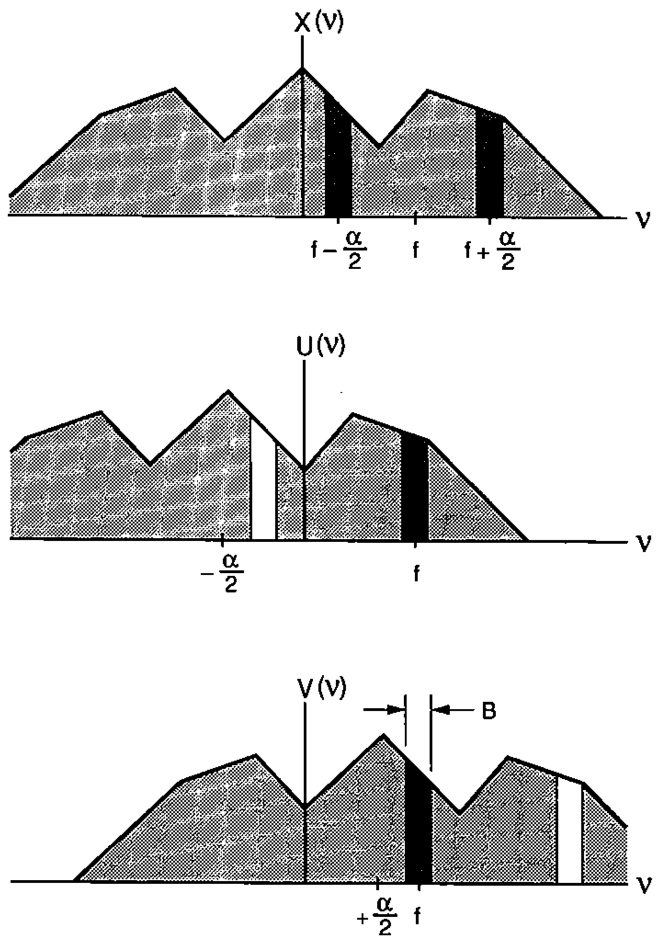

Figure 12: Illustration of spectral bands used in the measurement of the spectral correlation density $S_x^\alpha(f)$. (ν is a dummy frequency variable; the shaded bands are the bands selected by the BPFs.)

$$S_x^\alpha(f) = \sum_{\tau=-\infty}^{\infty} R_x^\alpha(\tau)\, e^{-i2\pi f\tau} \tag{39a}$$

and, therefore, $R_x^\alpha(\tau)$ is given by the inverse transform

$$R_x^\alpha(\tau) = \int_{-1/2}^{1/2} S_x^\alpha(f)\, e^{i2\pi f\tau} df. \tag{39b}$$

Since $R_x^\alpha(\tau)$ is periodic in α with period two, so too is $S_x^\alpha(f)$. Also, since τ takes on only integer values, then $S_x^\alpha(f)$ is periodic in f with period one. Furthermore, since increasing $f \pm \alpha/2$ by ± 1 has no effect on the spectral components at these frequencies, then it follows that $S_x^\alpha(f)$ also exhibits the periodicity $S_x^{\alpha+1}(f+\frac{1}{2}) = S_x^\alpha(f)$. Consequently, the principal domain for $S_x^\alpha(f)$ can be taken to be either the square with vertices $(f,\alpha) = (\pm\frac{1}{2}, \pm\frac{1}{2})$ or the diamond with vertices $(f,\alpha) = (0, \pm 1)$ and $\pm\frac{1}{2}, 0)$. Relation (38) is known as the *Wiener relation* (see, for example, (Gardner, 1987a, Chapter 3, Sec. C)), and (39) is therefore called the *cyclic Wiener relation* (Gardner, 1987a, Chapter 11, Sec. C). The cyclic Wiener relation includes the Wiener relation as the special case of $\alpha = 0$. (In the probabilistic framework of stochastic processes, which is based on expected values [ensemble averages] instead of time averages, the probabilistic counterpart of (38) is known as the *Wiener-Khinchin relation* and, therefore, the probabilistic counterpart of (39) is called the *cyclic Wiener-Khinchin relation* (Gardner, 1990a, Chapter 12, Sec. 12.2).) Because of the relation (39), the SCD is also called the *cyclic spectral density* function (Gardner, 1987a, Chapter 10, Sec. B).

It follows from (39) and the interpretation (23) of $R_x^\alpha(\tau)$ as $R_{uv}(\tau)$ that the SCD is the Fourier transform of the crosscorrelation function $R_{uv}(\tau)$ and is therefore identical to the cross-spectral density function for the frequency translates $u(t)$ and $v(t)$:

$$S_x^\alpha(f) \equiv S_{uv}(f), \tag{40}$$

where $S_{uv}(f)$ is defined by the right-hand side of (37) for arbitrary $u(t)$ and $v(t)$. This is to be expected since the cross-spectral density $S_{uv}(f)$ is known (cf. (Gardner, 1987a, Chapter 7, Sec. A)) to be the spectral correlation density for spectral components in $u(t)$ and $v(t)$ at frequency f, and $u(t)$ and $v(t)$ are frequency-shifted versions of $x(t)$. The identity (40) suggests an appropriate normalization for $S_x^\alpha(f)$: As long as the PSDs of $u(t)$ and $v(t)$ contain no spectral lines at frequency f, which means that the PSD of $x(t)$ contains no spectral lines at either of the frequencies $f \pm \alpha/2$, then the correlation of the spectral components (40) is actually a covariance since the means of the spectral components are zero (Gardner, 1987a, Chapter 11, Sec. C). When normalized by the geometric mean of the corresponding variances, which are given by

$$S_u(f) = S_x(f + \alpha/2) \tag{41a}$$

and

$$S_v(f) = S_x(f - \alpha/2), \tag{41b}$$

the covariance becomes a correlation coefficient:

$$\frac{S_{uv}(f)}{[S_u(f)S_v(f)]^{1/2}} = \frac{S_x^\alpha(f)}{[S_x(f+\alpha/2)S_x(f-\alpha/2)]^{1/2}} \triangleq \rho_x^\alpha(f). \tag{42}$$

Since $|\rho_x^\alpha(f)|$ is bounded to the interval [0, 1], it is a convenient measure of the degree of local spectral redundancy that results from spectral correlation. For example, for $|\rho_x^\alpha(f)| = 1$, we have complete spectral redundancy at $f + \alpha/2$ and $f - \alpha/2$. For conjugate spectral redundancy of complex-valued signals, (42) is modified by replacing the numerator with the conjugate SCD.

Let us now return to the AM example considered previously.

Example 1 continued: AM By Fourier transforming (34) and invoking the cyclic Wiener relation (39), we obtain the following SCD function on the principal domain for the amplitude-modulated signal (31):

$$S_x^\alpha(f) = \begin{cases} \frac{1}{4}e^{\pm i2\theta} S_a(f) & \text{for } \alpha = \pm 2f_o \\ \frac{1}{4}S_a(f+f_o) + \frac{1}{4}S_a(f-f_o) & \text{for } \alpha = 0 \\ 0 & \text{otherwise,} \end{cases} \tag{43}$$

where it has been assumed that $S_a(f \pm f_o) = 0$ for $|f| > 1/2$ to avoid aliasing effects in the principal domain. The magnitude of this SCD is graphed in Fig. 13 as the height of a surface above the bifrequency plane with coordinates f and α. For purposes of illustration, $a(t)$ is assumed to have an arbitrary low pass PSD for this graph. Observe that although the argument f of the SCD is continuous, as it always will be for a random signal, the argument α is discrete, as it always will be since it represents the harmonic frequencies of periodicities underlying the random time-series (the sine-wave carrier in this example).

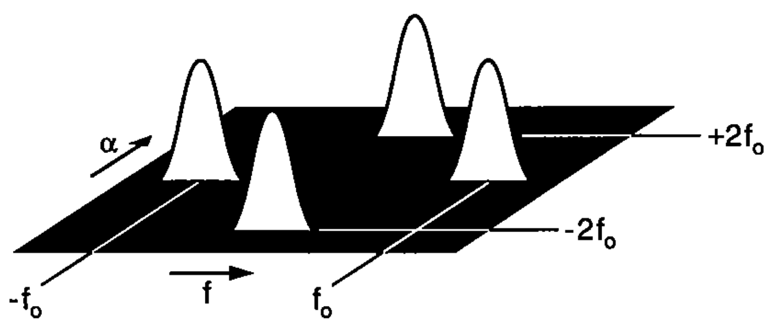

Figure 13: Magnitude of the spectral correlation density function for an AM signal graphed as a height above the bifrequency plane with coordinates f and α.

It follows from (43) that the spectral correlation coefficient is given by

$$\rho_x^\alpha(f) = \frac{S_a(f)\, e^{\pm i2\theta}}{\{[S_a(f+2f_o) + S_a(f)]\,[S_a(f) + S_a(f-2f_o)]\}^{1/2}} \qquad \text{for } \alpha = \pm 2f_o. \tag{44a}$$

Thus, the strength of correlation between spectral component in $x(t)$ at frequencies $f + \alpha/2$ and $f - \alpha/2$ is unity:

$$\left|\rho_x^\alpha(f)\right| = 1 \qquad \text{for } |f| < f_o \text{ and } \alpha = \pm 2f_o, \tag{44b}$$

provided that $a(t)$ is bandlimited to $|f| \le f_o$,

$$S_a(f) = 0 \qquad \text{for } |f| \ge f_o. \tag{45}$$

This is not surprising since the two spectral components in $x(t)$ at frequencies $f \pm \alpha/2 = f \pm f_o$ are obtained from the single spectral component in $a(t)$ at frequency f simply by shifting and scaling. Thus, they are perfectly correlated. That is, the upper (lower) sideband for $f > 0$ carries exactly the same information as the lower (upper) sideband for $f < 0$. Techniques for exploiting this *spectral redundancy* are described in Section 4.

To illustrate the conjugate SCD, we consider the analytic signal $z(t)$ for AM.

$$S_{zz^*}^\alpha(f) = \sum_{\tau=-\infty}^{\infty} R_{zz^*}^\alpha(\tau)\, e^{-i2\pi f\tau} = \begin{cases} \frac{1}{4}\, S_a(f)\, e^{i2\theta} & \text{for } \alpha = 2f_o \\ 0 & \text{otherwise.} \end{cases}$$

Before considering other examples of the SCD, let us first gain an understanding of the effects of some basic signal-processing operations on the SCD. This greatly facilitates the determination of the SCD for commonly encountered man-made signals.

3.4 Filtering

When a signal $x(t)$ undergoes a linear time-invariant (LTI) transformation (i.e., a convolution or a filtering operation),

$$z(t) = h(t) \otimes x(t) \triangleq \sum_{u=-\infty}^{\infty} h(u)x(t-u), \tag{46}$$

the spectral components in $x(t)$ are simply scaled by the complex-valued transfer function $H(f)$, which is the Fourier transform

$$H(f) = \sum_{t=-\infty}^{\infty} h(t)\, e^{-i2\pi ft} \tag{47}$$

of the discrete-impulse-response function $h(t)$ of the transformation. As a result, the PSD gets scaled by the squared magnitude of $H(f)$ (see, for example, (Gardner, 1987a, Chapter 3, Sec. C) or (Gardner, 1990a, Chapter 10, Sec. 10.1) for continuous time)

$$S_z(f) = |H(f)|^2\, S_x(f). \tag{48}$$

Equation (48) can be derived from the definition (36) of the PSD. Similarly, because the spectral components of $x(t)$ at frequencies $f \pm \alpha/2$ are scaled by $H(f \pm \alpha/2)$, the SCD gets scaled by the product $H(f+\alpha/2)H^*(f-\alpha/2)$:

$$S_z^\alpha(f) = H(f+\alpha/2)H^*(f-\alpha/2)S_x^\alpha(f). \tag{49}$$

This result, called the *input-output SCD relation for filtering*, which can be derived from the definition (37) of the SCD, includes (48) as the special case of $\alpha = 0$. Observe that it follows from (49) and the definition (42) that

$$\left|\rho_z^\alpha(f)\right| \equiv \left|\rho_x^\alpha(f)\right|. \qquad (50)$$

That is, the magnitude of the spectral correlation coefficient is unaffected by filtering (if $H(f \pm \alpha/2) \neq 0$).

Example 3: Time Delay As our first example of (49), we consider a filter that simply delays the input by some integer t_o; then $h(t) = \delta(t - t_o)$, where δ is the Kronecker delta, and $H(f) = e^{-i2\pi f t_o}$. Therefore, for $z(t) = x(t - t_o)$, we obtain from the input-output SCD relation (49)

$$S_z^\alpha(f) = S_x^\alpha(f)\, e^{-i2\pi\alpha t_o}, \qquad (51)$$

which indicates that, unlike the PSD, the SCD of a cyclostationary signal is sensitive to the timing or phase of the signal.

Example 4: Multipath Propagation As a second example of (49), if $x(t)$ undergoes multipath propagation during transmission to yield a received signal

$$z(t) = \sum_n a_n x(t - t_n),$$

where a_n and the integer t_n are the attenuation factor and delay of the nth propagation path, we have

$$H(f) = \sum_n a_n e^{-i2\pi f t_n} \qquad (52)$$

and therefore (49) yields

$$S_z^\alpha(f) = S_x^\alpha(f) \sum_{n,m} a_n a_m^* \exp(-i2\pi[f(t_n - t_m) + \alpha(t_n + t_m)/2]). \qquad (53)$$

Example 5: Bandpass Signals As a third example of the utility of the relation (49), let us determine the support region in the (f, α) plane for a bandpass signal with lowest frequency b and highest frequency B. To enforce such a spectrum, we can simply put any signal $x(t)$ through an ideal bandpass filter with transfer function (on the principal domain $(-1/2, 1/2]$)

$$H(f) = \begin{cases} 1 & \text{for } b < |f| < B \\ 0 & \text{otherwise.} \end{cases}$$

It then follows directly from the input-output SCD relation (49) that the SCD for the output of this filter can be nonzero only for $||f| - |\alpha|/2| > b$ and $|f| + |\alpha|/2 < B$:

$$S_z^\alpha(f) = \begin{cases} 0 & \text{for } ||f| - |\alpha|/2| \le b \text{ or } |f| + |\alpha|/2 \ge B, \\ S_x^\alpha(f) & \text{otherwise.} \end{cases} \qquad (54)$$

This shows that the support region in the (f, α) plane for a bandpass signal is the four diamonds located at the vertices of a larger diamond, depicted in Fig. 14a. By letting $b \to 0$, we obtain the support region for a lowpass signal, and by letting $B \to 1/2$, we obtain the support region for a highpass signal. This is shown in Figs. 14b and 14c.

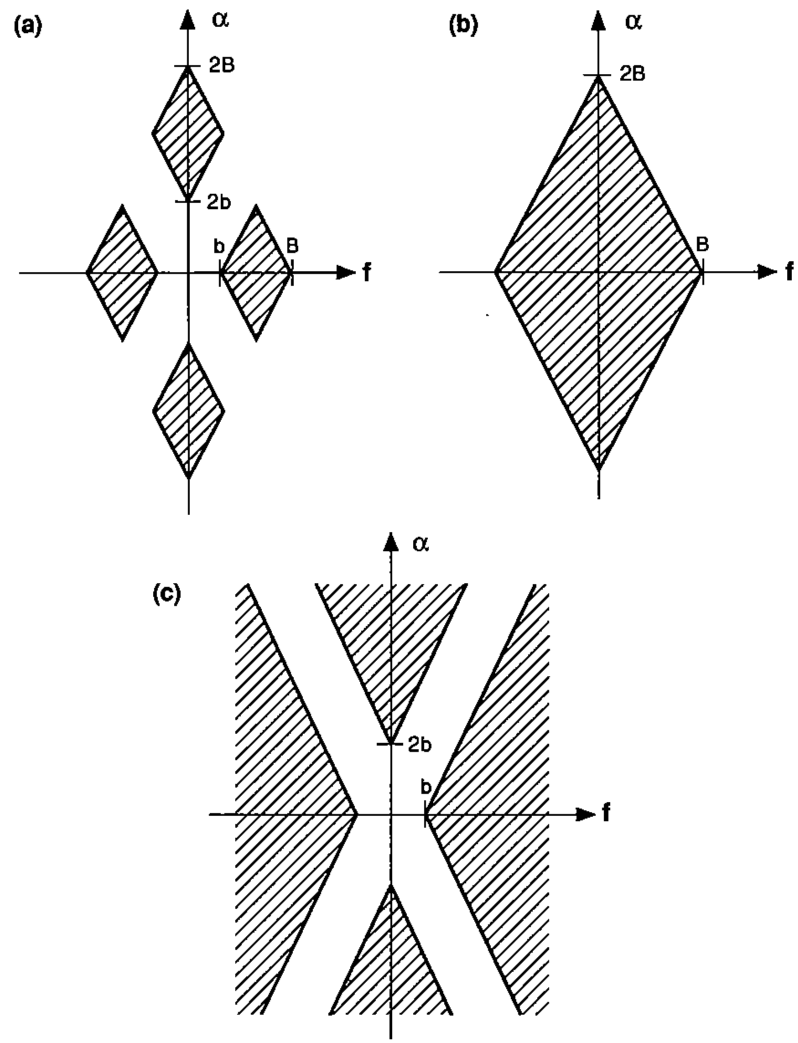

Figure 14: (a) Support region in the bifrequency plane for the spectral correlation density function of a bandpass signal. (b) Support region for a lowpass signal. (c) Support region for a highpass signal (shown over a small fraction of the diamond-shaped principal domain).

3.5 Signal Multiplication and Time Sampling

When two signals are multiplied together, we know from the convolution theorem that their Fourier transforms get convolved. From this, we expect some sort of convolution relation to hold for the SCDs of signals passing through a product modulator. In fact, it can be shown (cf. (Gardner, 1987a, Chapter 11, Sec. C) or (Gardner, 1990a, Chapter 12, exc. 41) for continuous time) that if $x(t)$ is obtained by multiplying together two statistically independent[17] time-series $r(t)$ and $s(t)$,

$$x(t) = r(t)s(t), \tag{55}$$

then the cyclic autocorrelation of $x(t)$ is given by the discrete circular convolution in cycle frequency of the cyclic autocorrelations of $r(t)$ and $s(t)$:

$$R_x^\alpha(\tau) = \sum_{\beta \in (-\frac{1}{2}, \frac{1}{2}]} R_r^\beta(\tau)\, R_s^{\alpha-\beta}(\tau)\,, \tag{56}$$

where, for each α, β ranges over all values in the principal domain $(-\frac{1}{2}, \frac{1}{2}]$ for which $R_r^\beta(\tau) \not\equiv 0$. By Fourier transforming (56), we obtain the input-output SCD relation for signal multiplication:

$$S_x^\alpha(f) = \int_{-1/2}^{1/2} \sum_{\beta \in (-\frac{1}{2}, \frac{1}{2}]} S_r^\beta(\nu) S_s^{\alpha-\beta}(f-\nu)\, d\nu \tag{57}$$

which is a double circular convolution that is continuous in the variable f and discrete in the variable α.

Example 6: Frequency Conversion As an example of (57), if $s(t)$ is simply a sinusoid,

$$s(t) = \cos(2\pi f_o t + \theta),$$

the product modulator becomes a frequency converter when followed by a filter to select either the up-converted version or the down-converted version of $r(t)$. By applying first the input-output SCD relation (57) for the product modulator (which applies since a sinusoid is statistically independent of all time-series (Gardner, 1987a, Chapter 15, Sec. A)), and then (49) for the filter, we can determine the up-converted or down-converted SCD. To illustrate, we first determine the SCD for the sinusoid $s(t)$. By substituting the sinusoid $s(t)$ into the definition of the cyclic autocorrelation, we obtain

$$R_s^\alpha(\tau) = \begin{cases} \frac{1}{2}\cos(2\pi f_o \tau) & \text{for } \alpha = 0 \\ \frac{1}{4} e^{\pm i2\theta} & \text{for } \alpha = \pm 2f_o \\ 0 & \text{otherwise} \end{cases} \tag{58}$$

on the principal domain of α.

[17] Time-series are statistically independent if their joint fraction-of-time probability densities factor into products of individual fraction-of-time probability densities, as explained in (Gardner, 1987a, Chapter 15, Sec. A).

Fourier transforming then yields the SCD

$$S_s^\alpha(f) = \begin{cases} \frac{1}{4}\delta(f - f_o) + \frac{1}{4}\delta(f + f_o) & \text{for } \alpha = 0 \\ \frac{1}{4} e^{\pm i2\theta}\delta(f) & \text{for } \alpha = \pm 2f_o \\ 0 & \text{otherwise} \end{cases} \tag{59}$$

on the principal domain, which is illustrated in Fig. 15a. Using (57), we circularly convolve this SCD with that of a stationary signal $r(t)$, for which

$$S_r^\alpha(f) = \begin{cases} S_r(f) & \text{for } \alpha = 0 \\ 0 & \text{for } \alpha \neq 0 \end{cases} \tag{60}$$

on the principal domain (see Fig. 15b). The result is that the SCD of the stationary signal simply gets replicated and scaled at the four locations of the impulses in the SCD of the sinusoid, as illustrated in Fig. 15c (provided that $S_r(f \pm f_o) = 0$ for $|f| \geq 1/2$ to avoid aliasing effects in the principal domain).

Example 7: Time Sampling Another important signal-processing operation is periodic time sampling. It is known that for a stationary signal $x(t)$, the PSD $S_x(f)$ of the sequence of samples $\{x(nT_s) : n = 0, \pm 1, \pm 2, \ldots\}$ is related to the PSD $\tilde{S}_x(f)$ of the continuous-time waveform by the aliasing formula (cf. (Gardner, 1987a, Chapter 3, Sec. E) or (Gardner, 1990a, Chapter 11, Sec. 11.1))

$$S_x(f) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \tilde{S}_x\left(f - \frac{n}{T_s}\right). \tag{61}$$

It is shown in (Gardner, 1987a, Chapter 11, Sec. C), (Gardner, 1990a, Chapter 12, Sec. 12.4) that this aliasing formula generalizes for the SCD to

$$S_x^\alpha(f) = \frac{1}{T_s} \sum_{m,n=-\infty}^{\infty} \tilde{S}_x^{\alpha+m/T_s}\left(f - \frac{m}{2T_s} - \frac{n}{T_s}\right). \tag{62}$$

Observe that, when $x(t)$ is not stationary (i.e., when $\tilde{S}_x^\alpha(f) \not\equiv 0$ for $\alpha = m/T$ for some nonzero integers m), the conventional PSD aliasing formula (61) must be corrected according to (62) evaluated at $\alpha = 0$:

$$S_x(f) = \frac{1}{T_s} \sum_{m,n=-\infty}^{\infty} \tilde{S}_x^{m/T_s}\left(f - \frac{m}{2T_s} - \frac{n}{T_s}\right). \tag{63}$$

This reflects the fact that, when aliased overlapping spectral components add together, their PSD values add only if they are uncorrelated. When they are correlated, as in a cyclostationary signal, the PSD value of the sum of overlapping aliased components depends on the particular magnitudes and phases of their correlations. The SCD aliasing formula (62) is illustrated graphically in Fig. 16, where the support regions for

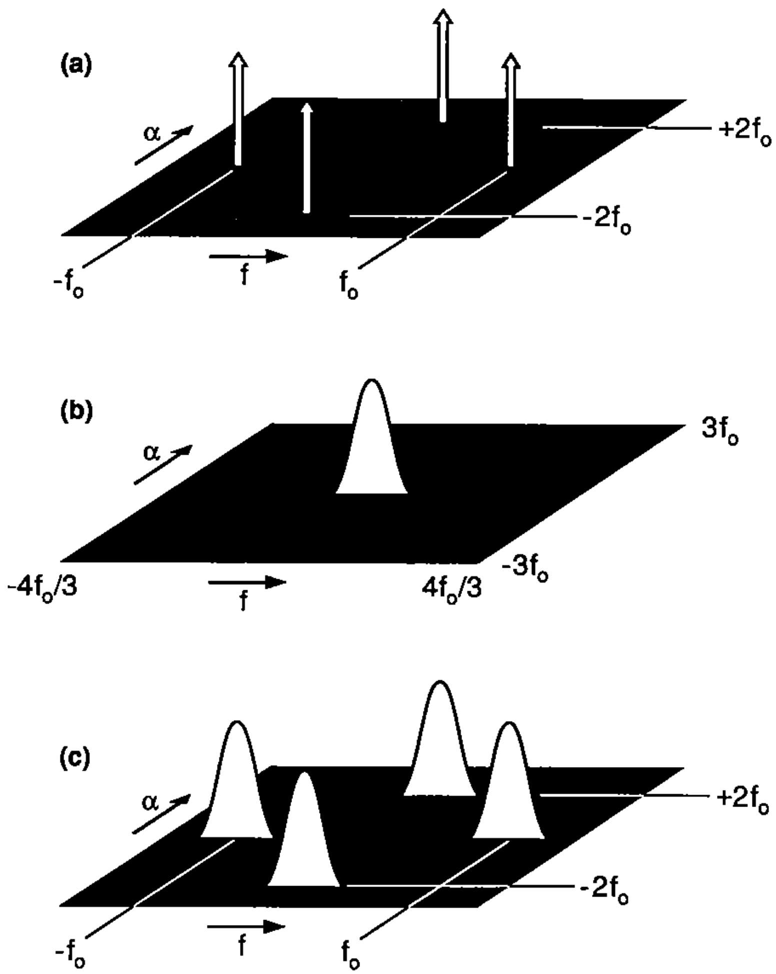

Figure 15: (a) Magnitude of the spectral correlation density (SCD) for a sine wave of frequency f_o. (b) SCD for a lowpass stationary signal. (c) SCD magnitude for the product of signals corresponding to (a) and (b), obtained by convolving the SCDs in (a) and (b).

the SCD $S_x^\alpha(f)$ for the sequence of samples $\{x(nT_s)\}$ is depicted in terms of the single diamond support region for a lowpass waveform $x(t)$, which is shown in Fig. 14b.

When we subsample a discrete-time signal $x(t)$ with sampling rate $1/T_s$ for some integer T_s to obtain the signal $z(t) = x(tT_s)$, we obtain the discrete-time analog of (62),

$$S_z^\alpha(f) = \frac{1}{T_s} \sum_{q \in P_\alpha} S_x^{(\alpha+q)/T_s}\left(\frac{f+q/2}{T_s}\right), \tag{64}$$

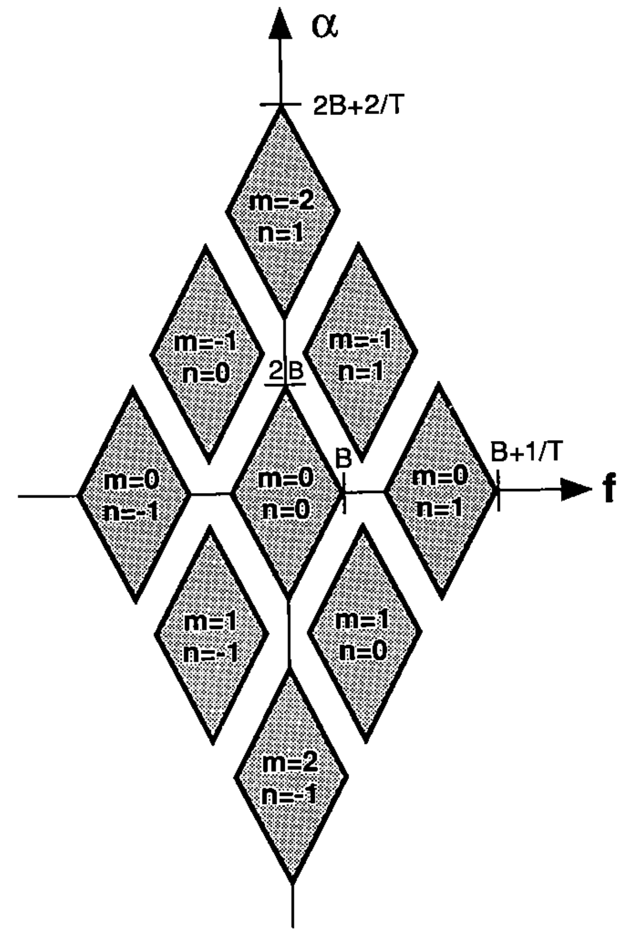

Figure 16: Illustration of support regions in the bifrequency plane for the spectral correlation densities that are aliased by periodic time sampling.

where P_α is the set of all integers $q = \beta T_s - \alpha$ for which $\alpha \in (-\frac{1}{2}, \frac{1}{2}]$. Similarly, when we resample a discrete-time signal $x(t)$, by (effectively) interpolating back to a continuous-time waveform and then time sampling at the new rate $1/T_s$ to obtain $z(t)$, the SCD is given by

$$S_z^\alpha(f) = \frac{1}{T_s} \sum_{m,n=-\infty}^{\infty} \tilde{S}_x^{\alpha+m/T_s}\left(f - \frac{m}{2T_s} - \frac{n}{T_s}\right), \tag{65}$$

where $\tilde{S}_x^\alpha(f)$ is the SCD $S_x^\alpha(f)$ restricted to its principal domain.

3.6 Periodically Time-Variant Filtering

Many signal-processing devices such as pulse and carrier modulators, multiplexors, samplers, and scanners, can be modeled as periodically time-variant filters, especially if multiple incommensurate periodicities (i.e., periodicities that are not harmonically related) are included in the model. By expanding the periodically time-variant discrete-impulse-response function in a Fourier series as explained shortly, any such system can be represented by a parallel bank of sinusoidal product modulators followed by time-invariant filters. Consequently, the effect of any such system on the SCD of its input can be determined by using the SCD relations for filters and product modulators. In particular, it can be shown (cf. (Gardner, 1987a, Chapter 11, Sec. D) for continuous time) that the SCD of the output $z(t)$ of a multiply-periodic system with input $x(t)$ is given by

$$S_z^\alpha(f) = \sum_{\beta,\gamma\in A} G_\beta(f+\alpha/2)G_\gamma^*(f-\alpha/2)S_x^{\alpha-\beta+\gamma}\left(f-\frac{\beta+\gamma}{2}\right), \tag{66}$$

provided that $S_x(f+\beta)=0$ for $|f|\geq 1/2$ for all $\beta\in A$ to avoid aliasing effects in the principal domain, where $G_\beta(f)$ are the transfer functions of the filters and A is the set of sinusoid frequencies associated with the product modulators in the system representation. More specifically, for the input-output equation

$$z(t) = \sum_{u=-\infty}^{\infty} h(t,u)x(u), \tag{67}$$

the multiply-periodic discrete-impulse-response function $h(t,u)$ can be expanded in the Fourier series

$$h(t+\tau,t) = \sum_{\beta\in A} g_\beta(\tau)\, e^{i2\pi\beta t}, \tag{68}$$

where the Fourier coefficients (for each τ) are given by

$$g_\beta(\tau) = \langle h(t+\tau,t)\, e^{-i2\pi\beta t}\rangle. \tag{69}$$

It follows from (67) and (68) that the filter output can be expressed as

$$z(t) = \sum_{\beta\in A}\left[x(t)\, e^{i2\pi\beta t}\right]\otimes g_\beta(t), \tag{70}$$

where $g_\beta(t)$ are the discrete-impulse-response functions of the filters with corresponding transfer functions $G_\beta(f)$. Thus, periodically time-variant filters perform time-invariant filtering on frequency-shifted versions $x(t)e^{i2\pi\beta t}$ of the input. This results in summing scaled, frequency-shifted, cycle-frequency-shifted versions of the SCD for the input $x(t)$ to obtain the SCD for the output $z(t)$, as indicated in (66).

Let us now consider some additional examples of modulation types, making use of the results obtained in the preceding paragraphs to determine SCDs. However, in

the interest of realism and for the sake of analytical simplicity, continuous-time signal models are used.

Example 2 continued: PAM Let $\{a_n\}$ be a stationary random sequence, and let us interpret these random variables as the time samples of a continuous-time random waveform, $a_n = a(nT_o)$, with PSD $\tilde S_a(f)$. We consider the continuous-time PAM signal

$$x(t) = \sum_{n=-\infty}^{\infty} a_n\, p(t-nT_o+\varepsilon), \tag{71}$$

where $p(t)$ is a deterministic finite-energy pulse and ε is a fixed pulse-timing phase parameter. To determine the SCD of $x(t)$, we can recognize that $x(t)$ is the output of a periodically time-variant linear system with input $a(t)$, and impulse response

$$h(t,u) = \sum_{n=-\infty}^{\infty} p(t-nT_o+\varepsilon)\,\delta(u-nT_o),$$

where δ is the Dirac delta. We can then use the continuous-time counterpart of the input-output SCD relation (66), which is identical in form except that continuous-time Fourier transforms are used (cf. (Gardner, 1987a, Chapter 11, Sec. D)). Or we can recognize that this particular periodically time-variant system is composed of a product modulator that implements an impulse sampler, followed by a linear time-invariant pulse-shaping filter with impulse-response function $h(t)=p(t)$, as shown in Fig. 17. We can then use the continuous-time counterpart of the input-output SCD relation (57), which is identical in form except the convolutions are linear (cf. (Gardner, 1987a, Chapter 11, Sec. C)), as it applies to impulse sampling, together with the relation (49) for filtering. The result is

$$\tilde S_x^\alpha(f) = \frac{1}{T_o^2}\tilde P(f+\alpha/2)\tilde P^*(f-\alpha/2)\sum_{m,n=-\infty}^{\infty}\tilde S_x^{\alpha+m/T_o}\left(f-\frac{m}{2T_o}-\frac{n}{T_o}\right)e^{i2\pi\alpha\varepsilon}. \tag{72}$$

Using the SCD aliasing formula (62) for $a(t)$ we can reexpress (72) as

$$\tilde S_x^\alpha(f) = \frac{1}{T_o}\tilde P(f+\alpha/2)\tilde P^*(f-\alpha/2)S_a^\alpha(f)\, e^{i2\pi\alpha\varepsilon}, \tag{73}$$

where $S_a^\alpha(f)$ is the SCD for the pulse-amplitude sequence $\{a_n\}$. Having assumed that $\{a_n\}$ is stationary, and using the periodicity property

$$S_a^\alpha(f) = \begin{cases} S_a(f+\alpha/2) & \text{for } \alpha = k/T_o \\ 0 & \text{otherwise} \end{cases} \tag{74}$$

for $k=0,\pm1,\pm2,\ldots,$ we can express (73) as

$$\tilde S_x^\alpha(f) = \begin{cases} \dfrac{1}{T_o}\tilde P(f+\alpha/2)\tilde P^*(f-\alpha/2)S_a(f+\alpha/2)\, e^{i2\pi\alpha\varepsilon} & \text{for } \alpha = k/T_o \\ 0 & \text{otherwise.} \end{cases} \tag{75}$$

A graph of the magnitude of this SCD for the full-duty-cycle rectangular pulse

$$p(t) = \begin{cases} 1 & \text{for } |t| \le T_o/2 \\ 0 & \text{otherwise} \end{cases} \tag{76}$$

and a white-noise amplitude sequence with PSD

$$S_a(f) = 1 \tag{77}$$

is shown in Fig. 18.

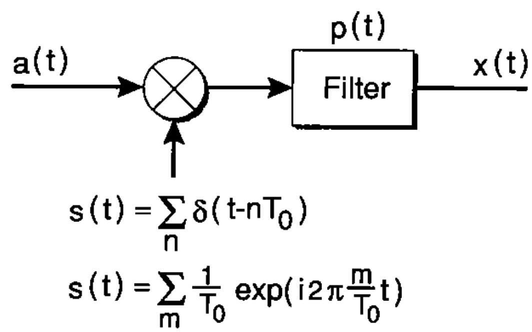

Figure 17: Interpretation of PAM signal generator as the cascade of an impulse sampler and a pulse-shaping filter.

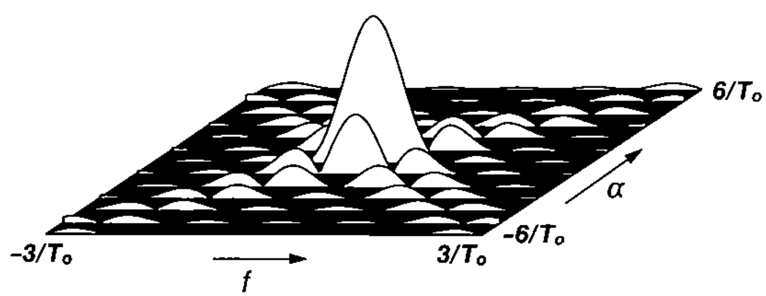

Figure 18: Magnitude of the spectral correlation density for a PAM signal with full-duty-cycle rectangular pulses.

It follows from (77) that for all $\alpha = k/T_o$ for which $S_a(f \pm \alpha/2) \ne 0$ and $\tilde{P}(f+\alpha/2)\tilde{P}^*(f-\alpha/2) \ne 0$, the spectral correlation coefficient $\rho_x^\alpha(f)$ is unity in magnitude:

$$\left|\rho_x^\alpha(f)\right| = 1. \tag{78}$$

Thus, all spectral components outside the band $|f| < 1/2T_o$ are completely redundant with respect to those inside this band. Techniques for exploiting this spectral redundancy are described in Section 4.

The conjugate SCD for the PAM signal (71) is given by (73) with $\tilde{P}^*(f-\alpha/2)$ replaced by $\tilde{P}^*(\alpha/2 - f)$ and $S_a^\alpha(f)$ replaced by $S_{aa^*}^\alpha(f)$. For a real PAM signal, the conjugate SCD is identical to the SCD; however, for complex PAM the conjugate SCD is, in general, different and is, in fact, zero for the complex PAM that models the complex envelopes of most digital QAM signals, including QPSK. This follows from the fact that $\langle a_n a_{n+m}\rangle = 0$ for all m for such signals; consequently, $S_{aa^*}^\alpha(f) = 0$ for all α.

By inverse Fourier transforming the SCD (75), we obtain the cyclic autocorrelation function

$$\tilde{R}_x^\alpha(\tau) = \begin{cases} \dfrac{1}{T_o} \displaystyle\sum_{n=-\infty}^{\infty} \tilde{R}_a(nT_o) r_p^\alpha(\tau - nT_o)\, e^{i2\pi\alpha\varepsilon} & \text{for } \alpha = k/T_o \\ 0 & \text{otherwise,} \end{cases} \tag{79}$$

where

$$r_p^\alpha(\tau) \triangleq \int_{-\infty}^{\infty} p(t+\tau/2)p^*(t-\tau/2)e^{-i2\pi\alpha t}dt. \tag{80}$$

For a white-noise amplitude-sequence as in (77), (79) reduces to

$$\tilde{R}_x^\alpha(\tau) = \frac{1}{T_o} r_p^\alpha(\tau)\, e^{i2\pi\alpha\varepsilon} \quad \text{for } \alpha = k/T_o, \tag{81}$$

and, for a rectangular pulse as in (76), this yields the temporal correlation coefficient

$$\tilde{\gamma}_x^\alpha(\tau) = \frac{\sin[\pi\alpha(T_o - |\tau|)]}{\pi\alpha T_o} e^{i2\pi\alpha\varepsilon} \quad \text{for } |\tau| \le T_o, \tag{82}$$

the magnitude of which is shown in Fig. 19. This correlation coefficient peaks for $\alpha = 1/T_o$ at $\tau = T_o/2$, where it takes on the value

$$\left|\tilde{\gamma}_x^\alpha(T_o/2)\right| = 1/\pi \quad \text{for } \alpha = 1/T_o. \tag{83}$$

That is, the strongest possible spectral line that can be regenerated in a delay-product signal for this particular PAM signal occurs when the delay equals half the pulse period. In contrast to this, when the more bandwidth-efficient pulse whose transform is a raised cosine is used, the optimal delay for sine-wave regeneration is zero.

An especially simple example of a sequence of pulse amplitudes $\{a_n\}$ is a binary sequence with values ± 1. If we consider $\tau = 0$ in the delay-product signal, then we obtain

$$y_0(t) = |x(t)|^2 = \sum_{m,n=-\infty}^{\infty} a_n a_m\, p(t - nT_o + \varepsilon)p(t - mT_o + \varepsilon).$$

If the pulses do not overlap (i.e., if $p(t) = 0$ for $|t| \ge T_o/2$), this reduces to

$$\begin{aligned} y_0(t) &= \sum_{n=-\infty}^{\infty} a_n^2\, p^2(t - nT_o + \varepsilon) \\ &= \sum_{n=-\infty}^{\infty} p^2(t - nT_o + \varepsilon), \end{aligned}$$

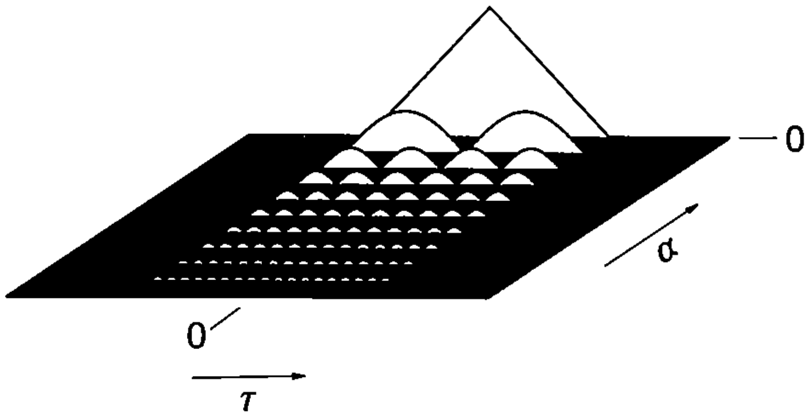

Figure 19: Magnitude of the cyclic autocorrelation function (normalized to form a correlation coefficient) for a PAM signal with full duty-cycle rectangular pulses.

which is periodic with period T_o and therefore contains finite-strength additive sine-wave components with frequencies k/T_o (except when $p(t)$ is flat as in (76)). In this very special case where $\{a_n\}$ is binary and the pulses do not overlap, there is no random component in $y_0(t)$; but, for $\tau \neq 0$, $y_\tau(t)$ contains both sine-wave components and random components (even when $p(t)$ is flat).

Example 8: ASK and PSK By combining the amplitude-modulated sine wave and the digital amplitude-modulated pulse train, we obtain the amplitude-shift-keyed (ASK) signal

$$x(t) = a(t)\cos(2\pi f_o t + \theta), \tag{84}$$

where

$$a(t) = \sum_{n=-\infty}^{\infty} a_n\, p(t - nT_o + \varepsilon), \tag{85}$$

and $\{a_n\}$ are digital amplitudes. By using the continuous-time counterpart of the SCD relation (57) for signal multiplication and the result (75) for the SCD of $a(t)$, we can obtain the SCD for the signal (84) by simply convolving the SCD functions shown in Figs. 15a and 18. The result is shown in Fig. 20a, where the cycle frequencies shown are $\alpha = \pm 2f_o + m/T_o$ and $\alpha = m/T_o$ for integers m, and where $f_o = 3.3/T_o$. When f_oT_o is irrational, the ASK signal is polycyclostationary with fundamental periods T_o and $1/2f_o$.

For a binary sequence with each $a_n = \pm 1$, this amplitude-shift-keyed signal, with the pulse (76), is identical to the binary phase-shift-keyed (BPSK) signal

$$x(t) = \cos\left[2\pi f_o t + \theta + \sum_{n=-\infty}^{\infty} \phi_n\, p(t - nT_o)\right], \tag{86}$$

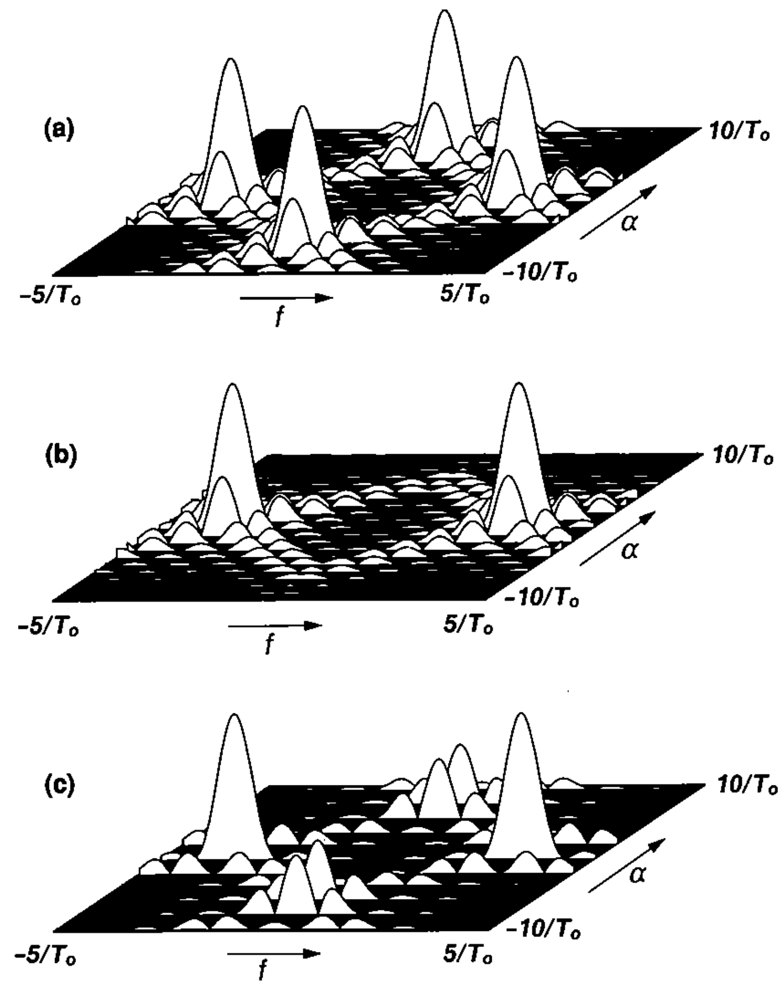

Figure 20: Magnitude of spectral correlation densities. (a) BPSK, (b) QPSK, and (c) SQPSK. (Each signal has a rectangular keying envelope.)

where $\phi_n \triangleq (a_n - 1)/2$, since shifting the phase of a sine wave by 0 or π is the same as multiplying its amplitude by 1 or -1. Other commonly used types of phase-shift-keyed signals include quaternary phase-shift keying (QPSK) and staggered QPSK (SQPSK). The details of these signal types are available in the literature (see, for example (Gardner, 1987a, Chapter 12, Sec. E) or (Gardner, 1990a, Chapter 12, Sec. 12.5)). Only their SCD-magnitude surfaces are shown here in Fig. 20b, c, where again $f_o = 3.3/T_o$.

It is emphasized that the three signals BPSK, QPSK, and SQPSK differ only in their carrier phase shifts and pulse timing and, as a result, they have identical PSDs, as shown in Fig. 20 (consider $\alpha = 0$). However, as also shown in Fig. 20, these differences in phase and timing result in substantially different SCDs (consider $\alpha \neq 0$). That is, the phase-quadrature component present in QPSK but absent in BPSK results in cancellation of the SCD at cycle frequencies associated with the carrier frequency (viz., $\alpha = \pm 2f_o + m/T_o$ for all integers m) in QPSK. Similarly, the pulse staggering by $T_o/2$ (between the in-phase and quadrature components) present in SQPSK but absent in QPSK results in the SCDs being cancelled at $\alpha = \pm 2f_o + m/T_o$ only for even integers m, and at $\alpha = m/T_o$ only for odd integers m in SQPSK. This again illustrates the fact that the SCD contains phase and timing information not available in the PSD. In fact, as formulas (43) and (75) reveal, the carrier phase θ in (31) and the pulse timing ε in (71) are contained explicitly in the SCDs for these carrier- and pulse-modulated signals.

3.7 Measurement of Spectral Correlation

The ideal SCD function (37) is derived by idealizing the practical spectral correlation measurement depicted in Fig. 11, by letting the averaging time T in the correlation measurement approach infinity and then letting the spectral resolving bandwidth B approach zero. Consequently, the practical measurement with finite parameters T and B can be interpreted as an estimate of the ideal SCD. This estimate will be statistically reliable only if $TB \gg 1$, and it will approach the ideal SCD only for sufficiently large T and sufficiently small B. Numerous alternative methods for making this practical measurement are described in (Gardner, 1986c; Gardner, 1987a, Chapter 13), and computationally efficient digital algorithms and architectures for some of these, which are developed in (Roberts et al., 1991; Brown and Loomis, 1992), are presented in Article 6 in this volume. The statistical behavior (bias and variance) of such estimates is analyzed in detail in (Gardner, 1986c; Gardner, 1987a, Chapter 15, Sec. B; Brown and Loomis, 1992), and in Chapter 6 in this volume. For the purpose of making the applications described in Section 4 more concrete, it suffices here to simply point out that because the SCD $S_x^\alpha(f)$ is equivalent to a particular case of the conventional cross spectral density $S_{uv}(f)$ (cf. (40)), one can envision any of the conventional methods of cross spectral analysis as being used in the applications.

Example 9: QPSK As an example, the result of using the Wiener-Daniell method (Gardner, 1987a, Chapter 7, Sec. D), based on frequency smoothing of the cross-periodogram of $u(t)$ and $v(t)$ (the conjugate product of their FFTs), is illustrated in Fig. 21 for a QPSK signal with carrier frequency $f_o = 1/4T_s$ and keying rate $1/T_o = 1/8T_s$, where $1/T_s$ is the sampling rate. An FFT of length 128 ($T = 128T_s$) was used in Fig. 21a, and only four frequency bins were averaged together ($B = 4/T$) to produce each output point, whereas, in Fig. 21b, the FFT length used was 32,768 ($T = 32{,}768T_s$) and 1,024 bins were averaged together ($B = 1{,}024/T$). It is easily seen by comparing with the ideal SCD in Fig. 20b that without adequate spectral smoothing the variability of the SCD estimate can be very large.

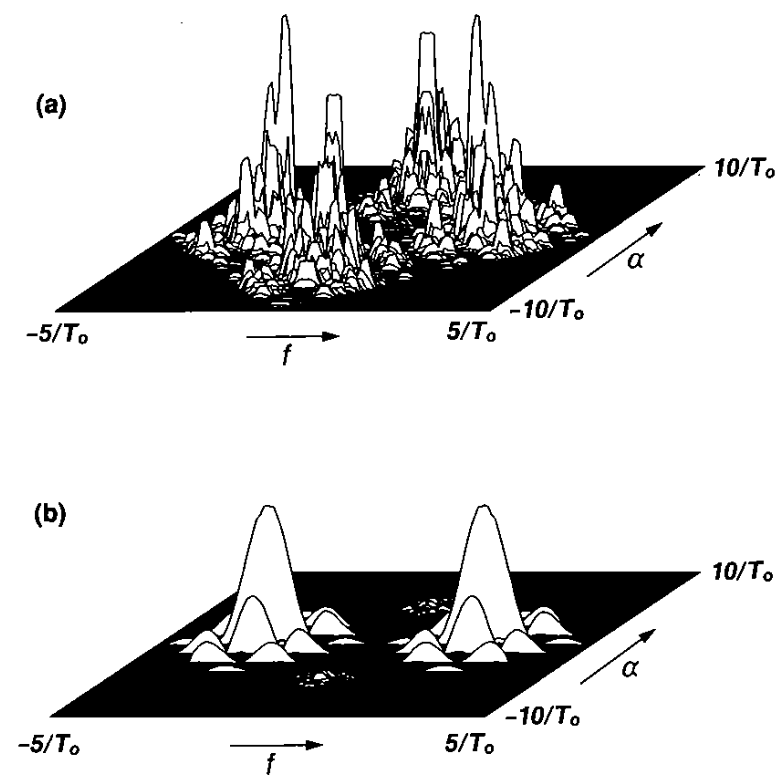

Figure 21: Magnitude of a spectral correlation density (SCD) estimate obtained from a finite-length data record for the QPSK signal whose ideal SCD is shown in Fig. 20b. (a) Record length is 128 time samples, and four adjacent frequency (f) bins are averaged together. (b) Record length is 32,768 and 1,024 adjacent frequency (f) bins are averaged together. (The sampling rate in both (a) and (b) is $10/T_o$, where $1/T_o$ is the keying rate of the QPSK signal.)

4 EXPLOITATION OF CYCLOSTATIONARITY

This section describes some ways of exploiting sine-wave generation and the inherent spectral redundancy associated with the spectral correlation in cyclostationary signals to perform various signal-processing tasks. These include detecting the presence of signals buried in noise and/or severely masked by interference; recognizing such corrupted signals according to modulation type; estimating parameters such as time-difference-of-arrival at two reception platforms and direction of arrival at a reception array on a single platform; blind-adaptive spatial filtering of signals impinging on a reception array; reduction of signal corruption due to cochannel interference and/or channel fading for single-receiver systems; linear periodically time-variant prediction; and identification of linear and nonlinear systems from input and output measurements. The descriptions include brief explanations of how and why the signal processors that exploit sine-wave generation or spectral redundancy can outperform their more conventional counterparts that ignore cyclostationarity. References to more detailed treatments are given throughout. It should be clarified at this point that although the classical theory of statistical inference and decision is certainly applicable in principle to cyclostationary signals (modeled either as stochastic processes or as nonstochastic time-series with fraction-of-time probability models, cf. Section 2), the types of problems where exploitation of cyclostationarity can really pay off are often not amenable to the classical theories (e.g., Bayes minimum risk and maximum likelihood) because of analytical intractability and implementational complexity. Thus, the techniques surveyed here are for the most part ad hoc, but nevertheless very powerful.

4.1 Spectral Redundancy

The existence of correlation between widely separated spectral components (separation equal to a cycle frequency α) can be interpreted as *spectral redundancy*. The meaning of the term *redundancy* that is intended here is essentially the same as that used in the field of information theory and coding. Specifically, multiple randomly fluctuating quantities (random variables) exhibit some redundancy if they are statistically dependent, for example, correlated. In coding, undesired redundancy is removed from data to increase the efficiency with which it represents information, and redundancy is introduced in a controlled manner to increase the reliability of storage and transmission of information in the presence of noise by enabling error detection and correction.

Here, redundancy that is inadvertently introduced into signals by the modulation process is to be exploited to enhance the accuracy and reliability of information gleaned from the measurements of corrupted signals, but the term *information* is interpreted in a broad sense. For instance, it includes the eight examples outlined in Section 1.2. In all of these examples, the performance of the signal processors that make the decisions and/or produce the estimates can be substantially improved by suitably exploiting spectral redundancy. The degree of improvement relative to the performance of more commonly used signal processors that ignore spectral redundancy depends on both the severity of the signal corruption (noise, interference, distortion) and the degree of redundancy in the signal $x(t)$, as measured by the magnitude of the spectral correlation coefficient $|\rho_x^\alpha(f)|$ (or its conjugate counterpart) defined in Section 3. The degree of improvement also depends on the amount of data available for processing (the collection time). The utility of exploiting spectral redundancy can also be enhanced by intentionally designing the signal to exhibit a sufficient amount of spectral redundancy.

The primary feature of spectral redundancy that enables it to be readily exploited is its distinctive character. That is, most man-made signals exhibit spectral redundancy, but most noise (all noise that is not cyclostationary) does not and, more importantly, when multiple signals of interest and signals not of interest (interference) overlap in both time and frequency, their spectral redundancy functions are nonoverlapping because their cycle frequencies α are distinct. This is a result of signals having distinct carrier frequencies and/or pulse rates or keying rates, even when occupying the same spectral band.

The distinctive character of spectral redundancy makes signal-selective measurements possible. Specifically, for the received composite signal

$$x(t) = \sum_{\ell=1}^{L} s_\ell(t) + n(t), \tag{87}$$

where the set $\{s_\ell(t)\}_1^L$ includes both signals of interest and interference—all of which are statistically independent of each other—and where $n(t)$ is background noise, we have the SCD (for measurement time $T \to \infty$)

$$S_x^\alpha(f) = \sum_{\ell=1}^{L} S_{s_\ell}^\alpha(f) + S_n^\alpha(f). \tag{88}$$

But if the only signal with the particular cycle frequency $\alpha = \alpha_k$ is $s_k(t)$, then (for $T \to \infty$) we have

$$S_x^\alpha(f) = S_{s_k}^\alpha(f), \tag{89}$$

regardless of the temporal or spectral overlap among $\{s_\ell(t)\}_1^L$ and also $n(t)$. This perfect signal selectivity of ideal SCDs implies that practical measurements of SCDs or their parameters can be made signal selective for measurement times τ that are long enough.

Example 1: BPSK Signal in Multiple AM Interference and Noise To illustrate the concept of signal selectivity, let us consider the situation in which a broadband BPSK signal of interest is received in the presence of white noise and five interfering AM signals with narrower bandwidths that together cover the entire band of the BPSK signal. The noise and each of the five interfering signals have equal average power. Therefore, the total signal-to-interference-and-noise ratio (SINR) is approximately -8 dB. The BPSK signal has carrier frequency $f_o = 0.25/T_s$ and keying rate $\alpha_o = 0.0625/T_s$. It has full-duty-cycle half-cosine envelope, which results in an approximate bandwidth of $B_o = 0.1875/T_s$. The five AM signals have carrier frequencies $f_1 = 0.156/T_s$, $f_2 = 0.203/T_s$, $f_3 = 0.266/T_s$, $f_4 = 0.313/T_s$, $f_5 = 0.375/T_s$, and bandwidths $B_1 = 0.04/T_s$, $B_2 = 0.05/T_s$, $B_3 = 0.045/T_s$, $B_4 = 0.04/T_s$, $B_5 = 0.08/T_s$. With the use of the same measurement parameters (FFT length = 32,768) as in Example 9 in Section 3 for the measurement of the SCD of QPSK, the SCD for these six signals in noise was measured. The resultant SCD magnitude is shown in Fig. 22a. Also shown in Fig. 22b, c are the SCD magnitudes for the BPSK signal alone and for the five AM interferences plus noise alone. Although all six signals exhibit strong spectral redundancy $\left(|\rho_{s_\ell}^\alpha(f)| = 1\right)$, the cycle frequencies α at which this redundancy exists are distinct because the carrier frequen-

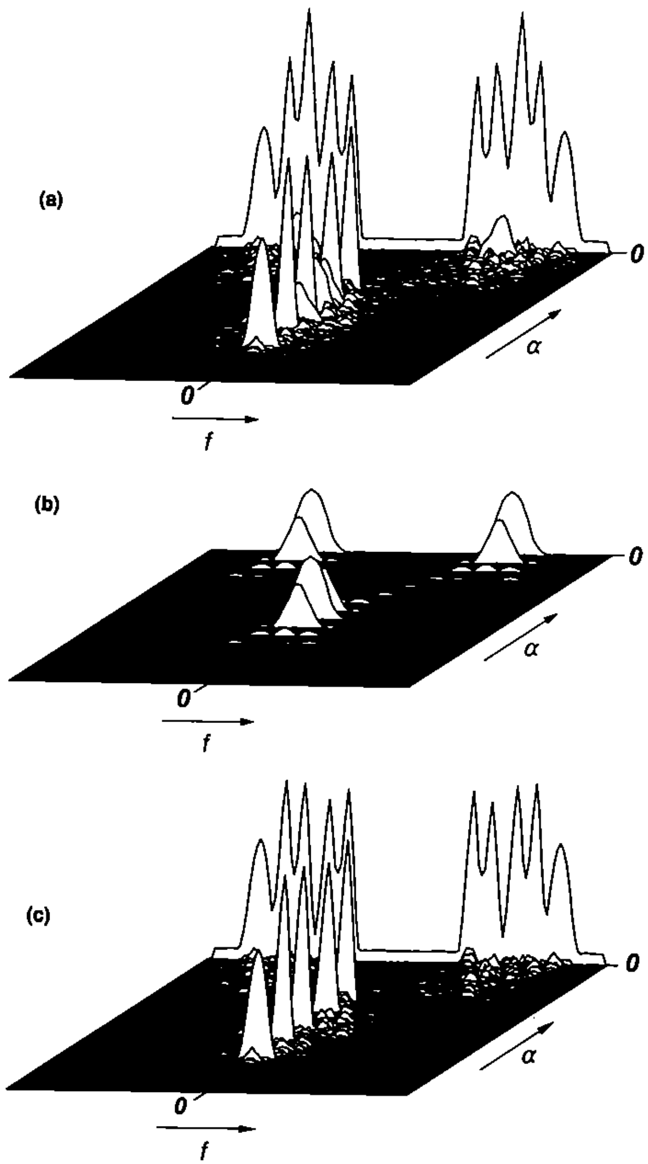

Figure 22: Magnitudes of estimated spectral correlation densities (SCDs). (a) SCD magnitude for a BPSK signal corrupted by white noise and five AM interferences. (b) SCD magnitude for the BPSK signal alone. (c) SCD magnitude for the white noise and five AM interferences. (The power levels, center frequencies, and bandwidths for the signals and noise are specified in the text; the record length used is 32,768 time samples, and 1,024 adjacent frequency (f) bins are averaged together.)

cies are all distinct. Thus, an accurate estimate of the SCD for the BPSK signal is easily extracted from the SCD for the corrupted measurements. Similarly, accurate estimates of the SCDs for each of the five AM signals can be extracted. Consequently, any information contained in these SCDs can be reliably extracted.

In connection with this example, let us briefly consider some of the signal-processing tasks outlined in Section 1.2.

4.2 Detection and Classification

We can see from Fig. 22 that knowing the particular pattern of the SCDs for BPSK and AM signals (see Figs. 13 and 20) enables us to detect the presence of six signals and to classify them according to modulation type. This would be impossible if only PSD (SCD at $\alpha = 0$) measurements were used. One approach to exploiting the spectral redundancy of a signal to detect its presence is to generate a spectral line at one of its cycle frequencies and then detect the presence of the spectral line (cf. Section 3). It has been shown that the maximum-SNR spectral-line generator for a signal $s(t)$ in additive Gaussian noise and interference with PSD $S_n(f)$ produces the detection statistic (cf. (Gardner, 1987a, Chapter 14, Sec. E) for continuous time)

$$z = \int_{-1/2}^{1/2} \hat{S}_x^\alpha(f) \frac{S_s^\alpha(f)^*}{S_n(f+\alpha/2)S_n(f-\alpha/2)}\, df \tag{90}$$

for comparison to a threshold. In (90), $\hat{S}_x^\alpha(f)$ is a crude estimate of $S_x^\alpha(f)$ obtained by deleting the time-averaging operation $\langle\cdot\rangle$ and the limiting operation from (37) and choosing B equal to the reciprocal of the record length of $x(t)$. It can be shown that (90) is equivalent to whitening the noise and interference using a filter with transfer function $1/[S_n(f)]^{1/2}$, and then correlating the measured SCD for the noise-and-interference-whitened data with the ideal SCD of the signal, as transformed by the whitener, to be detected (Gardner, 1987a, Chapter 14, Sec. E). Equivalently, for noise consisting of a white component plus strong narrowband components, (90) corresponds to attenuating the narrowband components well below the white-noise component—i.e., excising the narrowband components—using a filter with transfer function $1/S_n(f)$, and then correlating the measured SCD for the narrowband-excised data with the ideal SCD of the signal (untransformed by the excision filter).

A detailed study of both optimum (e.g., maximum-likelihood and maximum-SNR) and more practical suboptimum detection on the basis of SCD measurement is reported in (Gardner, 1988b), and receiver operating characteristics for these detectors obtained by simulation are presented in (Gardner and Spooner, 1992a, 1993). See also (Zivanovich and Gardner, 1991).

4.3 Parameter Estimation

Once the six signals have been detected and classified, their carrier frequencies and phases and the keying rate and phase of the BPSK signal can—with sufficiently long signal duration—be accurately estimated from the magnitude and phase of the SCD (cf., f_o, θ in (43) and T_o, ε in (75)) (Gardner and Spooner, 1993). It is clear from the theory discussed in Section 3 that SCD measurement is intimately related to the measurement of the amplitudes and phases of sine waves generated by quadratic transformations of the data. Thus, the fact that an SCD feature occurs at $\alpha = 2f_o$ for each carrier frequency f_o is a direct result of the fact that a sine wave (spectral line) with frequency $\alpha = 2f_o$ and phase 2θ can be generated by putting the data through a quadratic transformation. Similarly, for the SCD feature at $\alpha = 1/T_0$, where $1/T_0$ is the keying rate, a spectral line with frequency $\alpha = 1/T_0$ and phase ε can be quadratically generated. Consequently, SCD measurement is useful either directly or indirectly for estimation of synchronization parameters (frequencies and phases) required for the operation of synchronized receivers. The link between synchronization problems and spectral redundancy is pursued in (Gardner, 1986a) and also in Article 2 in this volume.

4.4 Time-difference-of-arrival Estimation

The cross SCD $S^\alpha_{wx}(f)$ for two signals $x(t)$ and $w(t)$ is defined in a way that is analogous to the definition (37) and (24) of the auto SCD $S^\alpha_x(f)$. That is, $x(t)$ in (24a) is simply replaced with $w(t)$. If we were to compute the cross SCD for two sets of corrupted measurements obtained from two reception platforms, then the cross SCD magnitude would look very similar to that in Fig. 22 (except that the low flat feature at $\alpha = 0$, which represents the PSD of the receiver noise, would be absent), but the phase of the cross SCD would contain a term linear in f at each value of α where the auto SCD of one of the six signals is nonzero. The slope of this linear phase is proportional to the time-difference-of-arrival (TDOA) of the wavefront at the two platforms for the particular signal with that feature. That is, for $x(t)$ from one platform given by (87) and $w(t)$ from the other platform given by

$$w(t) = \sum_{\ell=1}^{L} a_\ell s_\ell(t - t_\ell) + m(t) \tag{91}$$

where $\{t_\ell\}$ are the TDOAs, we have

$$S^\alpha_{wx}(f) = S^\alpha_{s_\ell}(f) a_\ell e^{-i2\pi(f+\alpha/2)t_\ell} \tag{92}$$

provided that $s_\ell(t)$ is the only signal with cycle frequency α. Consequently, accurate estimates of the TDOAs of each of these signals can be obtained from the cross SCD measurement, regardless of temporal and spectral overlap or of the closeness of the individual TDOAs. In other words, the signal selectivity in the α domain eliminates the problem of resolving TDOAs of overlapping signals.

For example, it follows from (89) and (92) that

$$\frac{S^\alpha_{wx}(f)}{S^\alpha_x(f)} = a_\ell\, e^{-i2\pi(f+\alpha/2)t_\ell} \tag{93}$$

over the support band of $S^\alpha_{s_\ell}(f)$. This suggests doing a weighted least-squares fit, with respect to a_ℓ and t_ℓ, of a measurement of the left side of (93) to the right side:

$$\min_{\hat{a}_\ell, \hat{t}_\ell} \left\{ \int_{-1/2}^{1/2} \left| W_\alpha(f) \left[\frac{\hat{S}^\alpha_{wx}(f)}{\hat{S}^\alpha_x(f)} - \hat{a}_\ell\, e^{-i2\pi(f+\alpha/2)\hat{t}_\ell} \right] \right|^2 df \right\} \tag{94}$$

where $W_\alpha(f)$ is some weighting function. After minimization with respect to $\hat{a}_\ell$, this reduces to

$$\max_{\hat{t}_\ell} \left\{ \int_{-1/2}^{1/2} |W_\alpha(f)|^2 \frac{S^\alpha_{wx}(f)}{S^\alpha_x(f)} e^{-i2\pi(f+\alpha/2)\hat{t}_\ell}\, df \right\}. \tag{95}$$

The two algorithms corresponding to the two choices $W_\alpha(f) = S^\alpha_x(f)$ (which yields the SPECCOA method) and $W_\alpha(f) = 1$ over some band (which yields the SPECCORR method), along with several other related algorithms are studied in detail in (Gardner and Chen, 1992; Chen and Gardner, 1992; Gardner and Spooner, 1993), where excellent robustness to unknown and/or varying noise and interference is demonstrated. It is also shown in Article 3 in this volume that this approach is easily generalized to the problem of multipath channel identification where multiple t_ℓ and a_ℓ for a single signal are to be estimated using the least-squares criterion (94) with a sum over ℓ included (provided that the multiple t_ℓ are resolvable, i.e., spaced farther apart than the width of the inverse discrete Fourier transform of $|W_\alpha(f)|^2\, S^\alpha_{wx}(f)/S^\alpha_x(f)$).

4.5 Spatial Filtering

Continuing in the same vein, we consider receiving these same six signals in noise with an antenna array. Then we can use the signal selectivity in α to blindly adapt (without any training information other than knowledge of the cycle frequency α of each signal of interest) adapt a linear combiner of the complex-valued outputs from the elements in the array to perform spatial filtering. Specifically, by directing the linear combiner to enhance or restore spectral redundancy (or conjugate spectral redundancy) in its output at a particular cycle frequency α, the combiner will adapt to null out all other signals (if there are enough elements in the array to make this nulling possible). This behavior of the combiner can be seen from the fact that the spectral correlation coefficient for $x(t)$ in (87) is (from (89))

$$\rho^\alpha_x(f) = \frac{S^\alpha_{s_\ell}(f)}{[S_x(f+\alpha/2)S_x(f-\alpha/2)]^{1/2}}, \tag{96}$$

where

$$S_x(f) = \sum_{k=1}^{L} S_{s_k}(f) + S_n(f), \tag{97}$$

and, similarly, the temporal correlation coefficient for the frequency-shifted versions of $x(t)$ is

$$\gamma^\alpha_x(\tau) = \frac{R^\alpha_{s_\ell}(\tau)}{R_x(0)}, \tag{98}$$

where

$$R_x(0) = \sum_{k=1}^{L} R_{s_k}(0) + R_n(0). \qquad (99)$$

Thus, nulling signals other than $s_\ell(t)$ in the output $x(t)$ of the linear combiner and attenuating the noise $n(t)$ in $x(t)$ reduces the denominators in (96) and (98) but not the numerators. Hence, $|\rho_x^\alpha(f)|$ and $|\gamma_x^\alpha(\tau)|$ can be increased by attenuating the noise and nulling any of the signals other than $s_\ell(t)$. Moreover, the linear combiner needs no knowledge of the reception characteristics of the array (no calibration) to accomplish this attenuation and nulling. A thorough study of spectral-coherence-restoral algorithms that perform this blind adaptive spatial filtering is reported in (Schell and Agee, 1988; Agee et al., 1990; Schell and Gardner, 1993b) and a tutorial discussion is given in Chapter 3 in this volume.

4.6 Direction Finding

We can take this approach one step further if we do indeed have calibration data for the reception characteristics of an antenna array because we can then also exploit signal selectivity in α to perform high-resolution direction finding (DF) without some of the drawbacks (described below) of conventional methods for high-resolution DF, such as subspace fitting methods (Schell and Gardner, 1993a), that do not exploit spectral redundancy. In particular, let us consider the narrowband model

$$\boldsymbol{x}(t) = \sum_{\ell=1}^{L} \boldsymbol{a}(\theta_\ell) s_\ell(t) + \boldsymbol{n}(t) \qquad (100)$$

for the analytic signal (or complex envelope) $\boldsymbol{x}$ of the received data vector of dimension r, where $\boldsymbol{a}(\theta_\ell)$ is the direction vector associated with the ℓ-th received signal $s_\ell(t)$, and the function $\boldsymbol{a}(\cdot)$ is specified by the calibration data for the array. Then, by working with the magnitude and phase information contained in the $r \times r$ cyclic correlation matrix

$$\boldsymbol{R}_x^\alpha(\tau) = \boldsymbol{R}_{s_k}^\alpha(\tau) = \boldsymbol{a}(\theta_k) R_{s_k}^\alpha(\tau) \boldsymbol{a}^\dagger(\theta_k) \qquad (101)$$

for some fixed τ (where $\dagger$ denotes conjugate transpose), instead of working with the information contained in the conventional correlation matrix

$$\boldsymbol{R}_x(0) = \sum_{\ell=1}^{L} \boldsymbol{R}_{s_\ell}(0) + \boldsymbol{R}_n(0) = \sum_{\ell=1}^{L} \boldsymbol{a}(\theta_\ell) \tilde{R}_{s_\ell}(0) \boldsymbol{a}^\dagger(\theta_\ell) + \boldsymbol{R}_n(0) \qquad (102)$$

we can avoid the need for advance knowledge of the correlation properties of the noise $\boldsymbol{R}_n(0)$ and interference $R_{s_\ell}(0)$ for $\ell \neq k$, and we can avoid the constraint imposed by conventional methods that the number of elements in the array exceed the total number L of signals impinging on the array. Also, by resolving signals in α, we need not resolve them in direction of arrival. Consequently, superior effective spatial resolution is another advantage available through the exploitation of spectral redundancy. As an example of a cyclostationarity-exploiting DF method, we can use

the fact that the $r \times r$ matrix in (101) has a rank of unity and the $(r-1)$-dimensional null space of this matrix is orthogonal to $\boldsymbol{a}(\theta_k)$. Therefore, we can choose as our estimate of θ_k that value $\hat{\theta}_k$ which renders $\boldsymbol{a}(\hat{\theta}_k)$ most nearly orthogonal to the null space of an estimate of the matrix $\boldsymbol{R}_x^\alpha(\tau)$ obtained from finite-time averaging. Similar remarks apply to the conjugate cyclic autocorrelation matrix. A thorough study of this approach to signal-selective DF is reported in (Schell et al., 1989; Schell, 1990; Schell and Gardner, 1991), where various algorithms are introduced and their performances are evaluated, and a tutorial discussion is given in Chapter 3 in this volume.

In the preceding paragraphs of this Section 4, the signal-processing tasks (with the exception of spatial filtering) involve decision or parameter estimation, but do not involve estimating (or extracting) an entire signal or an information-bearing message carried by the signal. Nevertheless, for the signal-extraction problem, the utility of spectral redundancy is just as apparent, as explained in the following paragraphs.

4.7 Signal Extraction

Spectrally redundant signals that are corrupted by other interfering signals can be more effectively extracted in some applications by exploiting spectral correlation through the use of periodic or multiply-periodic linear time-variant filters, instead of the more common time-invariant filters. These time-variant filters enable spectral redundancy to be exploited for signal extraction, because such filters perform frequency-shifting operations (cf. (70)) as well as the frequency-dependent magnitude-weighting and phase-shifting operations performed by time-invariant filters. The utility of this is easily seen for the simple example in which interference in some portions of the spectral band of the signal is so strong that it overpowers the signal in those partial bands. In this case, a time-invariant filter can only reject both the signal and the interference in those highly corrupted bands, whereas a time-variant filter can replace the rejected spectral components of the signal of interest with spectral components from other uncorrupted (or less corrupted) bands that are highly correlated with the rejected components from the signal.

AM is an obvious example of this because of the complete redundancy that exists between its upper sideband (above the carrier frequency) and its lower sideband (below the carrier frequency). Although this redundancy is exploited in the conventional double sideband demodulator to obtain a 3-dB gain in SNR performance, it is seldom exploited properly when partial-band interference is present. The proper exploitation in this case is illustrated in Fig. 23. Figure 23a shows the spectral content (Fourier transform magnitude of a finite segment of data) for an AM signal with partial-band interference in the upper sideband. Figure 23b shows the spectral content after the interference has been rejected by time-invariant filtering. The signal distortion caused by rejection of the signal components along with the interference can be completely removed by simply shifting replicas of perfectly correlated components from the lower sideband into the upper sideband, and then properly adjusting their magnitudes and phases, as suggested in Fig. 23c.

A less easily explained example involves two spectrally overlapping linearly modulated signals such as AM, PAM, ASK, PSK, or digital QAM (quadrature AM).

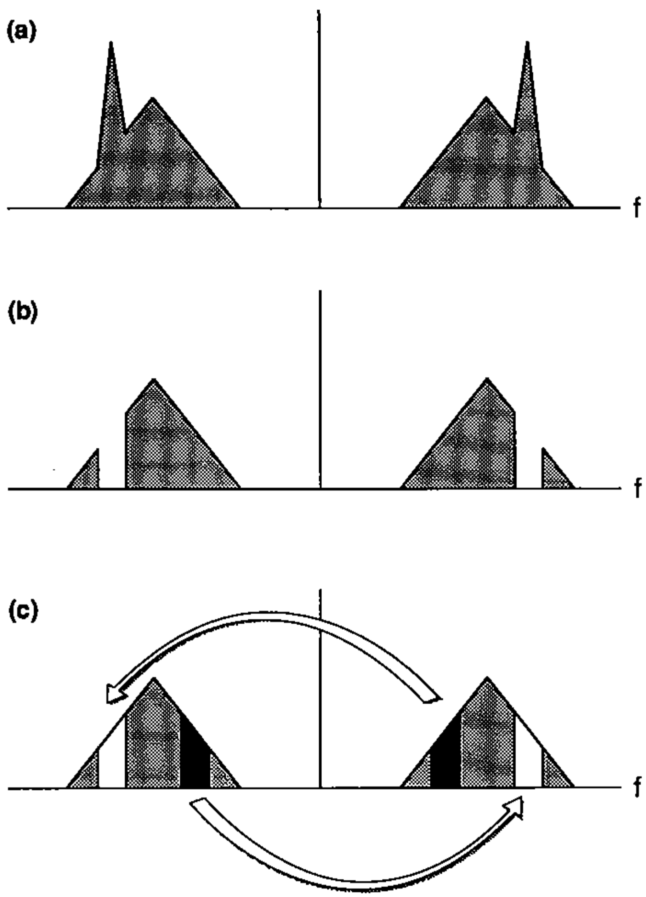

Figure 23: Illustration of power spectral densities (PSDs) for cochannel-interference removal with minimal signal distortion. (a) PSD for AM signal plus interference. (b) PSD after interference removal by time-invariant filtering. (c) PSD after distortion removal by frequency-shifting.

It can be shown that, regardless of the degree of spectral and temporal overlap, each of the two interfering signals can be perfectly extracted by using frequency shifting and complex weighting, provided only that they have either different carrier frequencies or phases (AM, ASK, BPSK) or different keying rates or phases (PAM, ASK, PSK, digital QAM) and at least 100% excess bandwidth (bandwidth in excess of the minimum Nyquist bandwidth for zero intersymbol interference). In addition, when the excess bandwidth is $(L-1)100\%$, L spectrally overlapping signals can be separated if they

have the same keying rate but different keying phases or carrier frequencies. Also, when broadband noise is present, extraction of each of the signals can in many cases be accomplished without substantial noise amplification. To illustrate the potential for signal separation in this case, consider L digital QAM signals with $(L-1)100\%$ excess bandwidth, all sharing the same carrier frequency and keying rate, but with distinct keying phases. Then for any particular frequency f in the Nyquist band the received spectral component at that frequency is a weighted sum of the L spectral components of the L individual signals at that same frequency f and the same is true at the $L-1$ additional frequencies separated by the keying rate, except that the sets of L weights in each of these L weighted sums are distinct, although the L sets of L spectral components are all identical (because the spectral correlation coefficients are unity in magnitude). Thus, for each frequency within the Nyquist band, we have L equations with L unknowns. In practice, the $L \times L$ array of weights also will be unknown and will have to be adaptively learned.

This particular problem of separating multiple digital QAM signals sharing the same carrier frequency (or baseband PAM signals) and sharing the same keying rate is explored in Article 1 in this volume.

To gain additional insight into how spectrally overlapping signals can be separated by frequency-shift filtering, we consider the case of two QPSK (quadrature-phase-shift-keyed) signals with unequal carrier frequencies and unequal keying rates and 100% excess bandwidth. The graphs in Fig. 24 show the overlapping spectra for these two signals. Starting from the top of this figure, each pair of graphs illustrates the result of one filtering and frequency-shifting stage. The subband shaded with a single set of parallel lines represents spectral components from one signal that are not corrupted by the other signal. These components are selected and complex-weighted by a filter and then frequency-shifted to cancel the components in another subband, which is identified by crosshatched shading. The result of this cancellation is shown in the second graph (which contains no shading) of each pair. After five such stages, a full sideband of each of the two QPSK signals has been completely separated. In each stage the complete spectral redundancy between components separated by the keying rate is being exploited, and this same spectral redundancy can be used to reconstruct the entire QPSK signal from either one of its sidebands.

The five cascaded stages of filtering, frequency-shifting, and adding operations can be converted into one parallel connection of frequency-shifters, each followed by a filter, simply by using standard system-transformations to move all frequency-shifters to the input.

Further insight into how spectrally overlapping signals can be separated by frequency-shift filtering can be gained by considering the case of two double sideband AM signals with suppressed carrier (or, equivalently, two ASK signals, or one AM and one ASK) with different carrier frequencies and any amount of spectral overlap. For each of these signals the upper sideband is completely redundant with the lower sideband. Consequently, if we were to reflect the complex spectrum about its center—its downconverted carrier frequency—say f_1, by replacing frequency f with

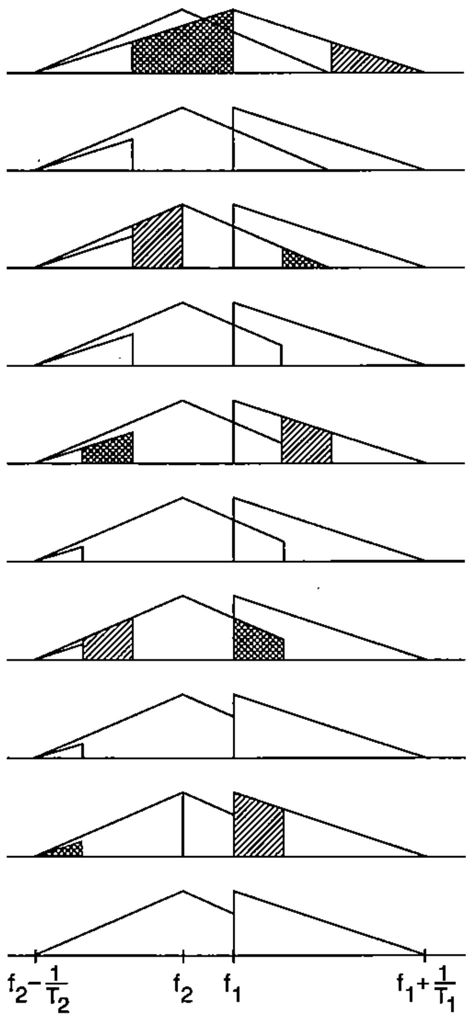

Figure 24: Illustration of power spectral densities for cochannel-QPSK-signal separation. The keying rates of the two signals are different and the carrier frequencies also are different. Each QPSK signal has a positive-frequency bandwidth equal to twice its keying rate.

$2f_1 - f$ for all f, and we were to shift its phase so that the downconverted carrier phase becomes zero, and we were to conjugate this reflected phase-shifted spectrum, then we would obtain precisely the original spectrum. Thus, if we subtracted the conjugated, phase-shifted, reflected spectrum from the original spectrum, we would cancel the signal. This cancellation in the frequency domain is equivalent to simply

downconverting the signal to a carrier frequency and phase of zero and then conjugating the time-domain signal and subtracting it from the unconjugated downconverted time-domain signal. However, in the process of cancelling the signal downconverted to zero frequency, we introduce severe distortion to the other signal present, which is downconverted to frequency $f_2 - f_1 \neq 0$: there will be two replicas of the signal present, with carrier frequencies of $f_2 - f_1$ and $f_1 - f_2$. Nevertheless, by downconverting the processed data so that one of the two replicas present (either the original signal or its conjugate that was added in the first stage of processing) has carrier frequency and phase of zero, we can again subtract the conjugate of this processed data to cancel one of the two replicas present. This will introduce new distortion; that is, there will again be two replicas of the signal present, but this time the carrier frequencies will be $\pm 2\,|f_2 - f_1|$. By proceeding through $N = B/2\,|f_2 - f_1| - 1$ stages of downconverting, conjugating, and subtracting (where B is the bandwidth of the signal with original carrier frequency f_2), we end up with two nonoverlapping replicas of the signal, which can be separated with a filter.

When signal distortion due to convolution (e.g., from passage through a channel) is present, this procedure will still work, in principle, provided that a filtered version of the conjugated data is subtracted at each stage. The challenge in practice is to find a way to adapt the filter needed to obtain effective cancellation.

A final example involves the reduction of the signal distortion due to frequency-selective fading caused by multipath propagation. Straightforward amplification in faded portions of the spectrum using a time-invariant filter suffers from the resultant amplification of noise. In contrast to this, a periodically time-variant filter can replace the faded spectral components with stronger highly correlated components from other bands. If these correlated spectral components are weaker than the original components before fading there will be some noise enhancement when they are amplified. But the amount of noise enhancement can be much less than that which would result from the time-invariant filter, which can only amplify the very weak faded components.

Detailed studies of the principles of operation and the mean-squared-error performance of both optimum and adaptive frequency-shift filters are reported in (Gardner, 1987a, Chapter 14, Secs. A, B; Gardner, 1990a, Chapter 12, Sec. 12.8; Gardner and Brown, 1989; Reed and Hsia, 1990; Gardner, 1993). See also (Zivanovich and Gardner, 1991).

4.8 Prediction and Causality

If a signal is correlated with time-shifted versions of itself (i.e., if it is not a white-noise signal), then its past can be used to predict its future. The higher the degree of temporal coherence $|\gamma_x^0(\tau)|$, the better the prediction. A signal that exhibits cyclostationarity is also correlated with frequency-shifted versions of itself. Consequently, its future can be better predicted if frequency-shifted versions of its past also are used, so that its spectral coherence as well as its temporal coherence can be exploited. For example, if $x(t)$ has cycle frequencies $\{\alpha_1, \ldots, \alpha_{N-1}\}$ then we can estimate the future value

$x(t+\tau)$ for some $\tau > 0$ using a linear combination of the present and past values of the N signals

$$x_q(t) = x(t)\, e^{i2\pi\alpha_q t} \qquad \text{for } q = 0, \ldots, N-1. \tag{103}$$

That is, the predicted value is given by

$$\hat{x}(t+\tau) = \sum_{u=0}^{M-1} \sum_{q=0}^{N-1} h_q(u)\, x_q[t-u], \tag{104}$$

where M is the memory-length of the predictor. The set of MN prediction coefficients that minimize the time-averaged (over t) squared magnitude of the prediction error $\hat{x}(t+\tau) - x(t+\tau)$ can be shown to be fully specified by the cyclic correlation functions for the N cycle frequencies. Specifically, the set of MN coefficients $\{h_q(u)\}$ is the solution to the set of MN simultaneous linear equations

$$\sum_{u=0}^{M-1} \sum_{q=0}^{N-1} h_q(u) R_x^{\alpha_p - \alpha_q}(t-u)\, e^{i\pi(\alpha_p - \alpha_q)(t-u)} = R_x^{\alpha_p}(t+\tau)\, e^{i\pi\alpha_p(t+\tau)} \tag{105}$$

for $t = 0, \ldots, M-1$ and $p = 0, \ldots, N-1$. Also, the percent accuracy of prediction is determined solely by the temporal coherence functions (29) for the frequency translates. It can be shown that for each cycle frequency α_q exploited, there is a corresponding increase in the percent accuracy of the prediction.

In the same way that time-invariant autoregressive model-fitting of stationary time-series data is mathematically equivalent to time-invariant linear prediction (Gardner, 1987a, Chapter 9, Sec. B), it can be shown that frequency-shift (or polyperiodic time-variant) autoregressive model-fitting is mathematically equivalent to frequency-shift linear prediction. Studies of this problem are reported in (Brelsford, 1967; Pagano, 1978; Miamee and Salehi, 1980; Tiao and Grupe, 1980; Sakai, 1982, 1983, 1990, 1991; Vecchia, 1985; Obeysekera and Salas, 1986; Li and Hui, 1988; Anderson and Vecchia, 1992). Also, the univariate prediction problem for cyclostationary (not polycyclostationary) time-series is equivalent to the multivariate prediction problem for stationary time-series (Pagano, 1978). This follows from the representation of univariate cyclostationary time-series in terms of multivariate stationary time-series (Gladyshev, 1961; Gardner and Franks, 1975). A survey of recent results in prediction theory for cyclostationary processes is given in Article 7 in this volume.

A measure of the degree to which one time-series causes another time-series is the degree to which the present and past of the former can linearly predict the future of the latter. If the two time-series are jointly cyclostationary, then cyclic as well as constant causality is possible. In fact, by considering only time-invariant predictors, it is possible to conclude for some pairs of time-series that no causality exists when, in fact, one time-series is perfectly cyclically caused by the other. An example of this is $x(t) = z(t)$ and $y(t) = z(t-\tau)\cos(t)$, where $\tau > 0$ and $z(t)$ is an independent identically distributed sequence. The best linear time-invariant predictor of $y(t)$ using the past of $x(t)$ is $\hat{y}(t) = 0$, whereas the best linear periodically time-variant predictor is $\hat{y}(t) = x(t-\tau)\cos(t) = y(t)$, which yields perfect prediction. Moreover, if $z(t)$ takes on values of only ± 1, it is particularly easy to show for this example that

$$\langle x^n(t-s) y^m(t) \rangle = \langle x^n(t-s) \rangle \langle y^m(t) \rangle$$

for all positive integers m, n, and s. Consequently, even the best *nonlinear* time-invariant predictor is $\hat{y}(t) = 0$.

4.9 Linear and Nonlinear System Identification

The cyclostationarity of signals passing through linear time-invariant systems can be exploited in several ways for the purpose of using input/output or output-only measurements to identify the system. In the case where the input/output measurements are corrupted by additive noise and/or interfering signals, the signal selectivity property associated with cyclostationarity can be used to obtain (asymptotically) complete immunity to this corruption. As explained in more detail in (Gardner, 1990b), the transfer function $H(f)$ of a system with corrupted input $w(t)$ and corrupted output $x(t)$ is given by

$$\frac{S_{xw}^{\alpha}(f-\alpha/2)}{S_w(f-\alpha/2)}$$

regardless of the additive corruption in $w(t)$ and $x(t)$, provided only that α is a cycle frequency of the uncorrupted system-input and is not a cycle frequency of the corruption, and that the support of the SCD $S_{xw}^{\alpha}(f-\alpha/2)$ in f covers the whole passband of the system.

Also, in the case where corrupted output-only measurements are available, the fact that the spectral correlation function $S_x^{\alpha}(f)$ of the system output contains information about the phase as well as the magnitude of the transfer function $H(f)$ (cf. (49)) means that blind identification of the system using only second-order statistics (SCD and PSD) is possible (Gardner, 1991c). One particularly simple scheme for blind channel equalization for digital QAM signals (or PAM signals) uses the fact that over each and every symbol interval, the channel output is the sum of noise and a linear combination of the same functions (viz., $\{h(t-nT) : n = 0, \pm 1, \pm 2, \ldots\}$ for $0 \le t < T$, where T is the length of the symbol interval, and $h(t)$ is the combined impulse response function of the transmitter's pulse-shaping filter and the channel. Consequently, the first term of an empirical Karhunen-Loeve expansion of the channel output over one symbol interval obtained from an eigendecomposition of the empirical output autocovariance matrix over one symbol interval (measured by performing synchronized averaging over multiple symbol intervals) can be used to equalize the channel. (This is particularly so when the symbol sequence and noise are both white.) That is, the eigenvector φ_1 corresponding to the largest eigenvalue will tend to be colinear with the channel output pulse over one symbol interval and orthogonal to the tails within this interval from the pulses centered in other symbol intervals. Thus, the inner product

$$\hat{a}_n = \sum_{t=(n-1)T}^{nT-1} w(t)\,\varphi_1^*(t)$$

of this eigenvector with the channel output (for the nth symbol interval with length T) can suppress intersymbol interference and provide an accurate estimate of the symbol in each interval.

The aforementioned synchronized average that provides the empirical autocovariance matrix is given by

$$\frac{1}{2N+1} \sum_{n=-N}^{N} x(t+nT)\,x^*(s+nT)$$

for the ts element of the $T \times T$ matrix (assuming the sampling increment is $T_s = 1$).

This straightforward approach is the special case of a more general approach (corresponding to the choice of parameters $K = L = 0$) proposed by Schell in Chapter 3 in this volume, where the results of a simulation of this special case are presented. Schell's more general method is derived from another approach based on least-squares filtering of a cyclostationary signal model to the channel output. Other approaches are described in Articles 4 and 5 in this volume.

A popular approach to the identification of nonlinear dynamical systems from input-output measurements is to model the system in terms of the Volterra series, which is a generalization of the power-series (or polynomial) representation of a memoryless system to systems with memory, and then to identify one-by-one the Volterra kernels, each one of which characterizes one term in the series representation. The first kernel is the impulse response of the linear part of the system. The second kernel is a two-dimensional generalization of the impulse response of the quadratic part of the system, and so on. Common approaches to identifying the kernels are based on crosscorrelation measurements between the unknown-system output and specially designed nonlinear functions of the system input.

Although the fundamental theory of this crosscorrelation approach to nonlinear system identification is built on the foundation of stationary random processes or time-series (Schetzen, 1989), it has recently been shown (Gardner and Archer, 1993) that substantial advantages can be gained by using cyclostationary inputs to the unknown system and cyclic crosscorrelations. In particular, desirable orthogonality (zero-correlation) properties between the system output and nonlinear functions of the input that are not possible for stationary inputs are possible for cyclostationary inputs, and this leads to particularly convenient designs for the inputs and the nonlinear functions. Moreover, this approach of exploiting cyclostationarity to identify time-invariant systems has recently been generalized to identify polyperiodic nonlinear systems (Gardner and Paura, 1992). In (Gardner and Archer, 1993; Gardner and Paura, 1992), the basic theory of this new approach is presented for both a time-domain method, which directly identifies the Volterra kernels or their polyperiodic counterparts, and a frequency-domain method, which directly identifies the multidimensional Fourier transforms of the kernels—the Volterra transfer functions, and several examples of cyclostationary inputs and corresponding nonlinear functions are given. This work exploits higher-than-second-order cyclostationarity, the principles of which are given in Chapter 2 in this volume.

REFERENCES

Agee, B. G., S. V. Schell, and W. A. Gardner (1987) "Self-coherence restoral: A new approach to blind adaptation of antenna arrays," *Proceedings of the Twenty-First Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 1987, pp. 589–593.

Agee, B. G., S. V. Schell, and W. A. Gardner (1988) "The SCORE approach to blind adaptive signal extraction: An application of the theory of spectral correlation," *Proceedings of the IEEE/ASSP Fourth Workshop on Spectral Estimation and Modeling*, 3-5 August 1988, Minneapolis, MN, pp. 277–282.

Agee, B. G., S. V. Schell, and W. A. Gardner (1990) "Spectral self-coherence restoral: A new approach to blind adaptive signal extraction," *Proceedings of the IEEE*, Vol. 78, pp. 756–767.

Albuquerque, J. P. A., O. Shimbo, and L. N. Ngugen (1984) "Modulation transfer noise effects from a continuous digital carrier to FDM/FM carriers in memoryless nonlinear devices," *IEEE Transactions on Communications*, Vol. COM-32, pp. 337–353.

Alekseev, V. G. (1988) "Estimating the spectral densities of a Gaussian periodically correlated stochastic process," *Problems of Information Transmission*, Vol. 24, pp. 109–115.

Alekseev, V. G. (1991) "Spectral density estimators of a periodically correlated stochastic process," *Problems of Information Transmission*, Vol. 26, pp. 286–288.

Anderson, P. L., and A. V. Vecchia (1992) "Asymptotic results for periodic autoregessive moving average processes," *Journal of Time Series Analysis*, Vol. 13.

Barnett, T. P. (1983) "Interaction of the monsoon and pacific trade wind system at interannual time scales. I. The equatorial zone," *Monthly Weather Review*, Vol. III, pp. 756–773.

Barnett, T. P., H. D. Heinz, and K. Hasselmann (1984) "Statistical prediction of seasonal air temperature over Eurasia," *Tellus Ser. A* (Sweden), Vol. 36A, pp. 132–146.

Bartolini, P., J. D. Salas, and J. T. B. Obeysekera (1988) "Multivariate periodic ARMA (1, 1) processes," *Water Resources Research*, Vol. 24, pp. 1237–1246.

Bennett, W. R. (1958) "Statistics of regenerative digital transmission," *Bell System Tech. Jour.*, Vol. 37, pp. 1501–1542.

Bittanti, S. (1987) "The periodic prediction problem for cyclostationary processes–An introduction," *Proc. of the NATO Advanced Research Workshop*, Groningen 1986, pp. 239–249.

Bittanti, S., and G. De Nicolao (1993) "Spectral factorization of linear periodic systems with application to the optimal prediction of periodic ARMA models," *Automatica*, Vol. 29, March.

Blum, J. R., and D. L. Hanson (1966) "On quasi-periodic probability measures," *Journal of Math. Mech.*, Vol. 15, pp. 291–303.

Boyles, R. A., and W. A. Gardner (1983) "Cycloergodic properties of discrete-parameter non-stationary stochastic processes," *IEEE Transactions on Information Theory*, Vol. IT-29, pp. 105–114.

Brelsford, W. M. (1967) "Probability predictions and time series with periodic structure," Ph.D. Dissertation, Johns Hopkins University, Baltimore, Maryland.

Brown, W. A. (1987) "On the theory of cyclostationary signals," Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, University of California, Davis, California.

Brown, W. A., and H. H. Loomis, Jr. (1992) "Digital implementation of spectral correlation analyzers," *IEEE Transactions on Signal Processing*, Vol. 40, pp. 703–720.

Campbell, J. C., A. J. Gibbs, and B. M. Smith (1983) "The cyclostationary nature of crosstalk interference from digital signals in multipair cable—Part I: Fundamentals," *IEEE Transactions on Communications*, Vol. COM-31, pp. 629–637.

Chen, C. K. (1989) "Spectral correlation characterization of modulated signals with application to signal detection and source location," Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, University of California, Davis, California.

Chen, C. K., and W. A. Gardner (1992) "Signal-selective time-difference-of-arrival estimation for passive location of manmade signal sources in highly corruptive environments. Part II: Algorithms and performance," *IEEE Transactions on Signal Processing*, Vol. 40, pp. 1185–1197.

Dehay, D. (1991) "Contributions to the spectral analysis of nonstationary processes," Ph.D. Dissertation, Universite des Sciences et Techniques de Lille Flandres Artois, June 1991.

Dragan, Ya. P., and I. N. Yavorskii (1982) *Rhythmics of Sea Waves and Underwater Acoustic Signals*, Kiev: Naukova Dumka (Russian).

Dragan, Ya. P., V. A. Rozhkov, and I. N. Yavorskii (1984) "Applications of the theory of periodically correlated random processes to the probabilistic analysis of oceanological time series," *Probabilistic Analysis and Modeling of Oceanological Processes* (Russian) (V. A. Rozhkov, ed.), pp. 4–23, Leningrad: Gidrometeoizdat.

Dragan, Ya. P., V. A. Rozhkov, and I. N. Yavorskii (1987) *Methods of Probabilistic Analysis of Oceanological Rhythmics* (Russian), Leningrad: Gidrometeoizdat.

Franks, L. E. (1969) *Signal Theory*, Englewood Cliffs, NJ: Prentice-Hall.

Franks, L. E., and J. Bubrouski (1974) "Statistical properties of timing jitter in a PAM timing recovery scheme," *IEEE Transactions on Communications*, Vol. COM-22, pp. 913–930.

Franks, L. E. (1980) "Carrier and bit synchronization in data communication—A tutorial review," *IEEE Transactions on Communications*, Vol. COM-28, pp. 1107–1121.

Gardner, W. A. (1972) "Representation and estimation of cyclostationary processes," Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of Massachusetts, reprinted as *Signal and Image Processing Lab Technical Report No. SIPL-82-1*, Department of Electrical and Computer Engineering, University of California, Davis, California, 95616 (1982).

Gardner, W. A., and L. E. Franks (1975) "Characterization of cyclostationary random signal processes," *IEEE Transactions on Information Theory*, Vol. IT-21, pp. 4–14.

Gardner, W. A. (1978) "Stationarizable random processes," *IEEE Transactions on Information Theory*, Vol. IT-24, pp. 8–22.

Gardner, W. A. (1985) *Introduction to Random Processes with Applications to Signals and Systems*, New York: Macmillan.

Gardner, W. A. (1986a) "The role of spectral correlation in design and performance analysis of synchronizers," *IEEE Transactions on Information Theory*, Vol. COM-34, pp. 1089–1095.

Gardner, W. A. (1986b) "The spectral correlation theory of cyclostationary time-series," *Signal Processing* (EURASIP), Vol. 11, pp. 13–36.

Gardner, W. A. (1986c) "Measurement of spectral correlation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, pp. 1111–1123.

Gardner, W. A. (1987a) *Statistical Spectral Analysis: A Nonprobabilistic Theory*, Englewood Cliffs, NJ: Prentice-Hall.

Gardner, W. A. (1987b) "Spectral correlation of modulated signals, Part I—Analog modulation," *IEEE Transactions on Communications*, Vol. COM-35, pp. 584–594.

Gardner, W. A., W. A. Brown, and C.-K. Chen (1987) "Spectral correlation of modulated signals, Part II—Digital modulation," *IEEE Transactions on Communications*, Vol. COM-35, pp. 595–601.

Gardner, W. A. (1988a) "Simplification of MUSIC and ESPRIT by exploitation of cyclostationarity," *Proceedings of the IEEE*, Vol. 76, No. 7, pp. 845–847.

Gardner, W. A. (1988b) "Signal interception: A unifying theoretical framework for feature detection," *IEEE Transactions on Communications*, Vol. COM-36, pp. 897–906.

Gardner, W. A. (1988c) "Exploitation of spectral correlation in cyclostationary signals," *Proc. Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling*, 3-5 August 1988, Minneapolis, MN, pp. 1–6.

Gardner, W. A., and C-K. Chen (1988) "Interference-tolerant time-difference-of-arrival estimation for modulated signals," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-36, pp. 1385–1395.

Gardner, W. A., and W. A. Brown (1989) "Frequency-shift filtering theory for adaptive co-channel interference removal," *Conference Record, Twenty-Third Asilomar Conference on Signals, Systems and Computers*, 30 Oct.-1 Nov. 1989, Pacific Grove, CA, pp. 562–567.

Gardner, W. A. (1990a) *Introduction to Random Processes with Applications to Signals and Systems*, 2nd ed., New York: McGraw-Hill.

Gardner, W. A. (1990b) "Identification of systems with cyclostationary input and correlated input/output measurement noise," *IEEE Transactions on Automatic Control*, Vol. 35, pp. 449–452.

Gardner, W. A. (1990c) "Spectral characterization of n-th order cyclostationarity," *Proceedings of the Fifth IEEE/ASSP Workshop on Spectrum Estimation and Modeling*, 10-12 Oct. 1990, Rochester, NY, pp. 251–255. (Invited paper)

Gardner, W. A., and S. Venkataraman (1990) "Performance of optimum and adaptive frequency-shift filters for co-channel interference and fading," *Proceedings of the Twenty-Fourth Annual Asilomar Conference on Signals, Systems, and Computers*, 5-7 Nov. 1990, Pacific Grove, CA, pp. 242–247.

Gardner, W. A., and C. M. Spooner (1990) "Higher order cyclostationarity, cyclic cumulants, and cyclic polyspectra," *Proceedings of the 1990 International Symposium on Information Theory and Its Applications*, 27-30 Nov. 1990, Honolulu, HI, pp. 355–358.

Gardner, W. A. (1991a) "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, Vol. 8, pp. 14–36.

Gardner, W. A. (1991b) "Two alternative philosophies for estimation of the parameters of time-series," *IEEE Transactions on Information Theory*, Vol. 37, pp. 216–218.

Gardner, W. A. (1991c) "A new method of channel identification," *IEEE Transactions on Communications*, Vol. 39, pp. 813–817.

Gardner, W. A., and W. A. Brown (1991) "Fraction-of-time probability for time-series that exhibit cyclostationarity," *Signal Processing*, Vol. 23, No. 3, pp. 273–292.

Gardner, W. A. (1992) "A unifying view of coherence in signal processing," *Signal Processing* (EURASIP), Vol. 29, pp. 113–140.

Gardner, W. A., and Chen, C. K. (1992) "Signal-selective time-difference-of-arrival estimation for passive location of manmade signal sources in highly corruptive environments. Part I: Theory and method," *IEEE Transactions on Signal Processing*, Vol. 40, pp. 1168–1184.

Gardner, W. A., and L. Paura (1992) "Identification of polyperiodic nonlinear systems," Technical Report, December 1992, Department of Electrical and Computer Engineering, University of California, Davis.

Gardner, W. A., S. V. Schell, and P. Murphy (1992) "Multiplication of cellular radio capacity by blind adaptive spatial filtering," accepted for publication in *1992 International Conference on Selected Topics in Wireless Communications*, June 25-26 and *Proceedings of the 22nd Annual IEEE Communication Theory Workshop*, June 28-July 21, 1992.

Gardner, W. A., and C. M. Spooner (1992a) "Signal interception: Performance advantages of cyclic feature detectors," *IEEE Transactions on Communications*, Vol. 40, No. 1, pp. 149–159.

Gardner, W. A., and C. M. Spooner (1992b) "The cumulant theory of cyclostationary time-series, Part I: Foundation," Technical Report, March 1992, Department of Electrical and Computer Engineering, University of California, Davis, and *IEEE Transactions on Signal Processing* (in press).

Gardner, W. A. (1993) "Cyclic Wiener filtering: Theory and method," *IEEE Transactions on Communications*, Vol. 41, pp. 151–163.

Gardner, W. A., and T. L. Archer (1993) "Exploitation of cyclostationarity for identifying the Volterra kernels of nonlinear systems," *IEEE Transactions on Information Theory*, Vol. 39, pp. 535–542.

Gardner, W. A., and C. M. Spooner (1993) "Detection and source location of weak cyclostationary signals: Simplifications of the maximum-likelihood receiver," *IEEE Transactions on Communications*, Vol. 41 (in press).

Genossar, M. J., H. Lev-Ari, and T. Kailath (1993) "Consistent estimation of the cyclic autocorrelation," *IEEE Transactions on Signal Processing*, Vol. 41 (in press).

Gladyshev, E. G. (1961) "Periodically correlated random sequences," *Soviet Math. Dokl.*, Vol. 2, pp. 385–388.

Gladyshev, E. G. (1963) "Periodically and almost periodically correlated random processes with continuous time parameter," *Theory Prob. Appl.*, Vol. 8, pp. 173–177.

Gudzenko, L. I. (1959) "On periodic nonstationary processes," *Radio Eng. Electron. Phys.* (Russian), Vol. 4, pp. 220–224.

Hasselmann, K., and T. P. Barnett (1981) "Techniques of linear prediction for systems with periodic statistics," *Journal of Atmospheric Science*, Vol. 38, pp. 2275–2283.

Hofstetter, E. M. (1964) "Random processes," Chapter 3 in *The Mathematics of Physics and Chemistry*, Vol. II, H. Margenau and G. M. Murphy eds., Princeton, NJ: D. Van Nostrand Co.

Honda, I. (1982) "On the spectral representation and related properties of periodically correlated stochastic process," *The Transactions of the IECE of Japan*, Vol. E65, pp. 723–729.

Honda, I. (1990) "On the ergodicity of Gaussian periodically correlated stochastic processes," *Transactions of the Institute of Electronics, Information and Communication Engineers E*, Vol. E73, pp. 1729–1737.

Hurd, H. L. (1969) "An investigation of periodically correlated stochastic processes," Ph.D. Dissertation, Duke University, Durham, North Carolina.

Hurd, H. L. (1974a) "Periodically correlated processes with discontinuous correlation functions," *Theory Prob. Appls.*, Vol. 19, pp. 804–807.

Hurd, H. L. (1974b) "Stationarizing properties of random shifts," *SIAM Journal of Applied Math.*, Vol. 26, pp. 203–212.

Hurd, H. L. (1989a) "Nonparametric time series analysis for periodically correlated processes," *IEEE Transactions on Information Theory*, March 1989, Vol. 35, pp. 350–359.

Hurd, H. L. (1989b) "Representation of strongly harmonizable periodically correlated processes and their covariances," *Jour. Multivariate Anal.*, Vol. 29, pp. 53–67.

Hurd, H. L. (1991) "Correlation theory of almost periodically correlated processes," *Jour. Multivariate Anal.*, Vol. 37, No. 1, pp. 24–45.

Hurd, H. L., and J. Leskow (1992a) "Estimation of the Fourier coefficient functions and their spectral densities for Φ-mixing almost periodically correlated processes," *Statistics and Probability Letters*, Vol. 14, pp. 299–306.

Hurd, H. L., and J. Leskow (1992b) "Strongly consistent and asymptotically normal estimation of the covariance for almost periodically correlated processes," *Statistics and Decision*, Vol. 10, pp. 201–225.

Izzo, L., L. Paura, and G. Poggi (1989) "Multiple-source localization: A new method exploiting the cyclostationarity property," *Proceedings of Douzième Colloque sur le Traitement du Signal et des Images*, Juan-les-Pins, France, pp. 481–484.

Izzo, L., A. Napolitano, L. Paura (1990) "Interference-tolerant estimation of amplitude and time-delay parameters of a composite signal," *Proceedings of Fifth European Signal Processing Conference* (EUSIPCO '90), Barcelona, 18-21 September 1990, pp. 103–106.

Izzo, L., L. Paura, and M. Tanda (1992) "Signal interception in non-Gaussian noise," *IEEE Transactions on Communications*, Vol. 40, pp. 1030–1037, June 1992.

Izzo, L., L. Paura, and G. Poggi (1992) "An interference-tolerant algorithm for localization of cyclostationary-signal sources," *IEEE Transactions on Signal Processing*, Vol. 40, pp. 1682–1686, July 1992.

Jacobs, K. (1958) "Almost periodic Markov processes," *Mathematics Annual*, Vol. 134, pp. 408–427 (German).

Johnson, C. M., P. Lemke, and T. P. Barnett (1985) "Linear prediction of sea ice anomalies," *Journal of Geophys. Res.*, Vol. 90, pp. 5665–5675.

Li, W. K., and Y. V. Hui (1988) "An algorithm for the exact likelihood of periodic autoregressive moving average models," *Comm. Stat., Simulation Comput.*, Vol. 17, pp. 1483–1494.

McLeod, A. I., D. J. Noakes, K. W. Hipel, and R. M. Thompstone (1987) "Combining hydrologic forecasts," *Journal of Water Resources Planning and Management*, Vol. 113, pp. 29–41, January 1987.

Miamee, A. G., and H. Salehi (1980) "On the prediction of periodically correlated stochastic processes," *Multivariate Analysis-V*, P. R. Krishnaiah (ed.), Amsterdam: North Holland Publishing Company, pp. 167–179.

Moeneclaey, M. (1982) "Linear phase-locked loop theory for cyclostationary input disturbances," *IEEE Transactions on Communications*, Vol. COM-30, pp. 2253–2259.

Moeneclaey, M. (1983) "The optimum closed-loop transfer function of a phase-locked loop used for synchronization purposes," *IEEE Transactions on Communications*, Vol. COM-31, pp. 549–553.

Moeneclaey, M. (1984) "A fundamental lower bound on the performance of practical joint carrier and bit synchronizers," *IEEE Transactions on Communications*, Vol. COM-32, pp. 1007–1012.

Nedoma, J. (1963) "Über die ergodizität and r-ergodizität stationärer wahrscheinlichkeitsmasse," *Z. Wahrscheinlichkeitstheorie*, Vol. 2, pp. 90–97.

Newton, H. J. (1982) "Using periodic autoregressions for multiple spectral estimation," *Technometrics*, Vol. 24, pp. 109–116.

Obeysekera, J. T. B., and J. D. Salas (1986) "Modeling of aggregated hydrologic time series," *Journal of Hydrol.* Vol. 86, pp. 197–219.

Pagano, M. (1978) "On periodic and multiple autoregressions," *Ann. Statist.*, Vol. 6, pp. 1310–1317.

Parzen, E., and M. Pagano (1979) "An approach to modeling seasonally stationary time-series," *Journal of Econometrics*, Vol. 9, North Holland Publishing Company, pp. 137–153.

Petersen, B. R. (1992) "Equalization in Cyclostationary Interference," Ph.D. Dissertation, Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada.

Pourahmadi, M., and H. Salehi (1983) "On subordination and linear transformation of harmonizable and periodically correlated processes," *Probability Theory on Vector Spaces, III* (Lublin), Berlin-New York: Springer, pp. 195–213.

Rao, M. M., and D. K. Chang (1988) "Special representations of weakly harmonizable processes," *Stochastic Analysis and Applications*, Vol. 6, pp. 169–190.

Reed, J. H., and T. C. Hsia (1990) "The performance of time-dependent adaptive filters for interference rejection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 38, pp. 1373–1385.

Roberts, R. S. (1989) "Architectures for digital cyclic spectral analysis," Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, University of California, Davis, California.

Roberts, R. S., W. A. Brown, and H. H. Loomis, Jr. (1991) "Computationally efficient algorithms for cyclic spectral analysis," *IEEE Signal Processing Magazine*, Vol. 8, pp. 38–49.

Rootenberg, J., and S. A. Ghozati (1977) "Stability properties of periodic filters," *International Journal of Systems and Science*, Vol. 8, pp. 953–959.

Rootenberg, J., and S. A. Ghozati (1978) "Generation of a class of nonstationary random processes," *International Journal of Systems and Science*, Vol. 9, pp. 935–947.

Sakai, H. (1982) "Circular lattice filtering using Pagano's method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-30, pp. 279–287.

Sakai, H. (1983) "Covariance matrices characterization by a set of scalar partial autocorrelation coefficients," *The Annals of Statistics*, Vol. 11, pp. 337–340.

Sakai, H. (1990) "Spectral analysis and lattice filter for periodic autoregressive processes," *Electronics and Communications in Japan, Part 3*, January 1990, Vol. 73, pp. 9–15.

Sakai, H. (1991) "On the spectral density matrix of a periodic ARMA process," *Journal of Time Series Analysis*, Vol. 12, pp. 73–82.

Salas, J. D. (1972) "Range analysis for storage problems of periodic stochastic processes," Ph.D. Dissertation, Department of Civil Engineering, Colorado State University, Fort Collins, CO.

Salas, J. D., and R. A. Smith (1980) "Correlation properties of periodic AR(ρ) models," *Proceedings of the 3rd International Symposium on Stochastic Hydraulics*, 5-7 August 1980, Tokyo, pp. 107–115.

Schell, S. V., and B. G. Agee (1988) "Application of the SCORE algorithm and SCORE extensions to sorting in the rank-L spectral self-coherence environment," *Conference Record, Twenty-Second Asilomar Conference on Signals, Systems and Computers*, 31 Oct.-2 Nov. 1988, Pacific Grove, CA, Vol. 1, pp. 274–278.

Schell, S. V., R. A. Calabretta, W. A. Gardner, and B. G. Agee (1989) "Cyclic MUSIC algorithms for signal-selective direction estimation," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Glasgow, 23-26 May 1989, pp. 2278–2281.

Schell, S.V., and W. A. Gardner (1989) "Signal-selective direction finding for fully correlated signals," *Abstracts of Sixth Multidimensional Signal Processing Workshop of the IEEE/ASSP Society*, 6-8 Sept. 1989, Pacific Grove, CA, pp. 139–140.

Schell, S. V. (1990) "Exploitation of spectral correlation for signal-selective direction finding," Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, University of California, Davis, California.

Schell, S. V., and W. A. Gardner (1990a) "Signal-selective high-resolution direction finding in multipath," *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, NM, 3-6 April 1990, pp. 2667–2670.

Schell, S. V., and W. A. Gardner (1990b) "Progress on signal-selective direction finding," *Proceedings of the Fifth IEEE/ASSP Workshop on Spectrum Estimation and Modeling*, 10-12 Oct. 1990, Rochester, NY, pp. 144–148.

Schell, S. V., and W. A. Gardner (1990c) "Detection of the number of cyclostationary signals in unknown interference and noise," *Proceedings of the Twenty-Fourth Annual Asilomar Conference on Signals, Systems, and Computers*, 5-6 Nov. 1990, Pacific Grove, CA, pp. 473–477.

Schell, S. V., and W. A. Gardner (1991) "Estimating the directions of arrival of cyclostationary signals—Part I: Theory and methods," Technical Report, November 1991, Department of Electrical and Computer Engineering, University of California, Davis.

Schell, S. V., and W. A. Gardner (1992) "Cramér-Rao lower bound for parameters of Gaussian cyclostationary signals," *IEEE Transactions on Information Theory*, Vol. 38, pp. 1418–1422, July 1992.

Schell, S. V., and W. A. Gardner (1993a) "High resolution direction finding," Chapter in *Handbook of Statistics*, N. K. Bose and C. R. Rao (Eds.), Amsterdam: Elsevier.

Schell, S. V., and W. A. Gardner (1993b) "Blind adaptive spatio-temporal filtering for wideband cyclostationary signals," *IEEE Transactions on Signal Processing*, Vol. 41, pp. 1961–1964.

Schell, S. V., W. A. Gardner, and P. A. Murphy (1993) "Blind adaptive antenna arrays in cellular communications for increased capacity" in *Proc. of 3rd Virginia Tech. Symp. on Wireless Personal Comm.*, June 1993.

Schetzen, M. (1989) *The Volterra and Wiener Theories of Nonlinear Systems*, Second Edition, Malabar, FL: Krieger Publ. Co.

Sherman, P. J. (1992) "Random processes from rotating machinery," *Proceedings of the Workshop on Nonstationary Stochastic Processes and Their Applications*, Hampton, VA, 1-2 Aug. 1991, Singapore: World Scientific.

Spooner, C. M., and W. A. Gardner (1991) "Estimation of cyclic polyspectra," *Proceedings of Twenty-Fifth Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 4-6 Nov. 1991, pp. 370–376.

Spooner, C. M. (1992) "Theory and application of higher-order cyclostationarity," Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of California, Davis.

Spooner, C. M., and W. A. Gardner (1992a) "An overview of higher-order cyclostationarity," *Proceedings of the Workshop on Nonstationary Stochastic Processes and Their Applications*, Hampton, VA, 1-2 Aug. 1991, Singapore: World Scientific.

Spooner, C. M., and W. A. Gardner (1992b) "The cumulant theory of cyclostationary time-series, Part II: Development and applications," Technical Report, March 1992, Department of Electrical and Computer Engineering, University of California, Davis, and *IEEE Transactions on Signal Processing* (in press).

Spooner, C. M., and W. A. Gardner (1992c) "Performance evaluation of cyclic polyspectrum estimators," *Proceedings of Twenty-Sixth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 26-29 Oct. 1992.

Thompstone, R. M., W. Hipel, and A. I. McLeod (1985) "Grouping of periodic autoregressive models," *Time Series Analysis: Theory and Practice 6*, edited by O. D. Anderson, J. K. Ord, and E. A. Robinson, pp. 35–49, Amsterdam: Elsevier, North-Holland.

Tiao, G. C., and M. R. Grupe (1980) "Hidden periodic autoregressive-moving average models in time series data," *Biometrika*, Vol. 67, pp. 365–373.

Vecchia, A. V. (1983) "Aggregation and estimation for periodic autoregressive-moving average models," Ph.D. Dissertation, Colorado State University, Fort Collins.

Vecchia, A. V. (1985) "Maximum likelihood estimation for periodic autoregressive moving average models," *Technometrics*, Vol. 27, pp. 375–384.

Wold, H. O. A. (1948) "On prediction in stationary time-series," *Ann. Math. Stat.*, Vol. 19, pp. 558–567.

Xu, G., and T. Kailath (1992) "Direction-of-arrival estimation via exploitation of cyclostationarity—A combination of temporal and spatial processing," *IEEE Transactions on Signal Processing*, Vol. 40, pp. 1775–1786.

Zivanovich, G. D., and W. A. Gardner (1991) "Degrees of cyclostationarity and their application to signal detection and estimation," *Signal Processing* (EURASIP), Vol. 22, pp. 287–297.