

UNCLASSIFIED

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 31 Jul 90	3. REPORT TYPE AND DATES COVERED final 1 Jul 87 - 30 Apr 89	
4. TITLE AND SUBTITLE Detection and Time Delay Estimation of Non-Gaussian Signals in Noise			5. FUNDING NUMBERS N00014-87-K-0785	
6. AUTHOR(S) Wilson, Gary R.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Applied Research Laboratories The University of Texas at Austin P.O. Box 8029 Austin, Texas 78713-8029			8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-90-25	
8. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of the Chief of Naval Research Department of the Navy Arlington, Virginia 22217-5000			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) Gaussian time series have the property that their higher order spectra are identically zero. In many signal processing applications, the noise fields are predominantly Gaussian. Thus if the signal has a nonzero higher order spectra (i.e., the signal is non-Gaussian), then higher order spectral processing of the combined signal plus noise field may provide processing gains over more traditional processing methods that rely on lower order properties of the signal. (cont'd on reverse side)				
14. SUBJECT TERMS bispectrum cross-bispectrum higher order spectra polyspectra time delay estimation non-Gaussian			15. NUMBER OF PAGES 91	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR	

UNCLASSIFIED

13. (cont'd)

In this report, a detection statistic is defined based on a sample estimate of the bispectrum. It is shown that under the null hypothesis of noise only, the detection statistic is approximately central chi-square distributed and under the alternative hypothesis of signal-plus-noise it is noncentral chi-square distributed. The noncentrality parameter of the noncentral chi-square distribution is derived in terms of the processing parameters of the bispectrum estimate, the degree of non-Gaussianity of the signal (the skewness function) and the signal-to-noise ratio. Since the mean value of the detection statistic is equal to the noncentrality parameter, the performance of the detection statistic as a function of all relevant signal, noise, and processing parameters is clearly demonstrated. A new bispectrum waterfall function is defined for use in displaying the detection results in a familiar lofargram type of display.

The cross-bispectrum is also exploited to define time delay estimators. An estimator based on the phase of the cross-bispectrum is defined and its statistical properties are derived. The dependence of the resolution of the time delay estimator on processing parameters, the signal skewness, the signal-to-noise ratio, and the coherency of the noise is explicitly derived. A new time delay estimator is also defined which casts the two-dimensional bispectrum into a one-dimensional framework much like the bispectrum waterfall used in detection, and allows it to have many of the properties of a crosscorrelation function.

UNCLASSIFIED

TABLE OF CONTENTS

	<u>Page</u>
LIST OF FIGURES.....	v
1. INTRODUCTION	1
2. BISPECTRUM ANALYSIS.....	3
2.1 THEORY OF BISPECTRUM PROCESSING.....	3
2.1.1 Comparison to the Power Spectrum.....	3
2.1.2 Formal Definition and Properties	5
2.1.3 Principal Domain of the Discrete Time Bispectrum.....	8
2.2 ESTIMATION OF THE BISPECTRUM.....	13
2.2.1 The Unnormalized Bispectrum Estimate: Averaging in Frequency and in Time.....	13
2.2.2 The Normalized Bispectrum: Statistical Properties	16
2.3 THE BISPECTRUM WATERFALL	18
3. DETECTION OF NON-GAUSSIAN SIGNALS IN NON-GAUSSIAN NOISE.....	25
4. TIME DELAY ESTIMATION USING THE CROSS-BISPECTRUM	27
4.1 TIME DELAY ESTIMATION BASED ON THE PHASE OF THE CROSS-BISPECTRUM.....	27
4.2 AN ALTERNATIVE TIME DELAY ESTIMATOR.....	29
APPENDIX A DETECTION OF NON-GAUSSIAN SIGNALS IN NON-GAUSSIAN NOISE USING THE BISPECTRA	33
APPENDIX B TIME DELAY ESTIMATION USING THE CROSS- BISPECTRUM.....	59
REFERENCES	87



Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2.1	Symmetries of the Bispectrum	10
2.2	Symmetries of the Bispectrum	11
2.3	Principal Domain of the Bispectrum.....	12
2.4	False Alarm Probability as a Function of Normalized Bispectrum Threshold	19
2.5	Bispectrum Waterfall	21
2.6	Example of Bispectrum Waterfall.....	23
4.1	Principal Domain and Support Set for the Cross-Bispectrum	28

1. INTRODUCTION

Gaussian time series have the property that their higher order spectra are identically zero. In many signal processing applications, the noise fields are predominantly Gaussian. Thus if the signal has a nonzero higher order spectra (i.e., the signal is non-Gaussian), then higher order spectral processing of the combined signal plus noise field may provide processing gains over more traditional processing methods that rely on lower order properties of the signal.

However, any practical signal processing applications of higher order spectra rely on *estimates* of the higher order spectra made from a finite number of data samples, and thus the statistical properties of various estimators of the higher order spectra determine the processing gains that may be achieved in practice for specific signals and noise. Therefore, to determine the utility of higher order spectral processing it is necessary to determine estimators that are appropriate for specific signal processing applications and determine the performance of those estimators for the signal and noise characteristics of the specific application based on the derived statistical properties of the estimators. This report summarizes the work performed under Contract N00014-87-K-0785 in which statistical estimators for the detection and time delay estimation of non-Gaussian signals were defined and their statistical properties in the presence of noise were derived.

This research is being conducted in coordination with other related 6.2 and 6.3 research projects, which serve to identify specific applications of interest to the Navy and serve as a means of transitioning this work to Navy systems. Results from this work are being applied to classified data analysis under these other programs. The results of the work conducted under this project are published in Hinich and Wilson (1989, 1990).

In this report, a detection statistic is defined based on a sample estimate of the bispectrum. It is shown that under the null hypothesis of noise only, the detection statistic is approximately central chi-square distributed and under the alternative hypothesis of signal-plus-noise it is noncentral chi-square distributed. The noncentrality parameter of the noncentral chi-square distribution is derived in terms of the processing parameters of the bispectrum estimate, the degree of non-Gaussianity of the signal (the skewness function) and the signal-to-noise ratio. Since the mean value of the detection statistic is equal to the noncentrality parameter, the performance of the detection statistic as a function of all relevant signal, noise, and processing parameters is clearly demonstrated. A new

bispectrum waterfall function is defined for use in displaying the detection results in a familiar lofargram type of display.

The cross-bispectrum is also exploited to define time delay estimators. An estimator based on the phase of the cross-bispectrum is defined and its statistical properties are derived. The dependence of the resolution of the time delay estimator on processing parameters, the signal skewness, the signal-to-noise ratio (S/N), and the coherency of the noise is explicitly derived. A new time delay estimator is also defined which casts the two-dimensional bispectrum into a one-dimensional framework much like the bispectrum waterfall used in detection, and allows it to have many of the properties of a cross-correlation function.

The results of this study can be used by a system designer to determine the feasibility of higher order spectral processing for specific signal processing applications.

2. BISPECTRUM ANALYSIS

2.1 THEORY OF BISPECTRUM PROCESSING

2.1.1 Comparison to the Power Spectrum

While power spectral analysis of a time series is a concept that is familiar to many, bispectral analysis may not be as familiar. A family of spectra, called polyspectra, has been defined for a stationary time series, of which the power spectrum is only the lowest order member of the family (Brillinger, 1965; Rosenblatt, 1966; Brillinger and Rosenblatt, 1967a,b). The next higher order member of this family is the bispectrum. In the following paragraphs, the power spectrum and bispectrum will be compared and contrasted by explaining their relationships to the Fourier transform of the time series and the correlation (or cumulant) functions of the time series.

For a stationary discrete time series, its power spectrum is given in terms of the Fourier transform of the time series by

$$P(\omega) = \langle X(\omega)X^*(\omega) \rangle = \langle |X(\omega)|^2 \rangle, \quad (2.1)$$

where $X(\omega)$ is the Fourier transform of the time series $X(t)$,

$$X(\omega_i) = \frac{1}{2\pi} \sum_{t_i=-\infty}^{\infty} X(t_i) e^{-i\omega_i t_i}, \quad -\pi < \omega_i < \pi. \quad (2.2)$$

$\langle \cdot \rangle$ denotes the expected value, and $*$ denotes complex conjugate. As can be seen from Eq. (2.1), the power spectrum contains information only about the magnitude of the Fourier transform of a time series.

However, the bispectrum and higher order polyspectra contain information about both the magnitude and specific phase relationships (or coherences) between multiple frequency components of the Fourier transform of a time series. Different polyspectra describe different phase relationships. For this reason, the bispectrum (and higher order polyspectra) contain additional information about a time series that cannot be obtained from

the power spectrum. For a stationary time series, the bispectrum is given in terms of its Fourier transform by

$$B(\omega_1, \omega_2) = \langle X(\omega_1) X(\omega_2) X^*(\omega_1 + \omega_2) \rangle . \quad (2.3)$$

The bispectrum is a two-dimensional function of frequency. The bispectrum is in general a complex function rather than a real function like the power spectrum. The two-dimensional complex bispectrum provides information about phase relationships between the signal components at frequencies ω_1, ω_2 , and $\omega_1 + \omega_2$.

The previous paragraphs have explained mathematically what the bispectrum is but have shed little light on what the bispectrum means and how it can be interpreted. Equation (2.3) represents the bispectrum as the expected value of the product of Fourier transform values at the three frequencies ω_1, ω_2 , and $\omega_1 + \omega_2$. Thus a nonzero bispectrum occurs only when these three frequency components are statistically dependent. For any Gaussian time series, all frequency components are independent, so a Gaussian time series will produce a zero bispectrum. However, if the time series contains a quadratic nonlinearity such that two frequency components interact to produce a sum or difference frequency, then the three frequencies will be statistically dependent. In this case a nonzero value will occur in the bispectrum at the pair of frequencies that define the nonlinear interaction. Thus the bispectrum will indicate the presence of (quadratic) nonlinearities in the time series and identify the frequencies contributing to the nonlinear behavior. This is the basis for the statistical tests for Gaussianity and linearity using the bispectrum described by Hinich (1982) and others. The three frequency components may also be statistically dependent, not from a nonlinear interaction but from linear mechanisms which result in harmonics of some fundamental frequency. In this case the time series can be thought of as a periodically time varying (i.e., nonstationary) process. The bispectrum can indicate the presence of these nonstationarities due to their statistical dependence of distinct frequency components. Thus harmonically related signals can be identified in the bispectrum. Also, if a time series is linearly time varying (e.g., transient), it will have frequency components that are statistically dependent. Therefore, transient signals will produce a nonzero bispectrum. In this brief discussion we have identified that quadratic nonlinear signals, harmonically related signals, and transients can result in nonzero bispectral signatures.

2.1.2 Formal Definition and Properties

Higher order spectra of a random process are often referred to as cumulant spectra. The n th order cumulant function of a random process $X(t)$ is defined in terms of the characteristic function of the random process by

$$[X(t), X(t+\tau_1), \dots, X(t+\tau_{n-1})]_n = (-1)^n \frac{\delta^n \ln \Phi(a_1, \dots, a_n)}{\delta a_1 \dots \delta a_n} \Big|_{a_1 = \dots = a_n = 0} \quad (2.4)$$

where the characteristic function is defined as

$$\Phi(a_1, \dots, a_n) = \langle \exp\{i(a_1 X(t) + a_2 X(t+\tau_1) + \dots + a_n X(t+\tau_{n-1}))\} \rangle \quad (2.5)$$

The square brackets $[]$ denote the cumulant function. One property of the characteristic function is that if any subset of the $X(t_i)$ are independent of the rest of the $X(t_i)$, then the characteristic function factors. For example, if $X(t), \dots, X(t+\tau_{r-1})$ are independent of $X(t+\tau_r), \dots, X(t+\tau_{n-1})$, then the characteristic function is

$$\Phi(a_1, \dots, a_n) = \Phi(a_1, \dots, a_r) \Phi(a_{r+1}, \dots, a_n) \quad (2.6)$$

If the characteristic function factors, then the n th partial derivative of the characteristic function given in Eq. (2.4) is zero and the cumulant function is thus zero. This demonstrates an important property of the cumulant function: the n th order cumulant function is nonzero only when there is statistical dependence between the n elements of the cumulant function.

If the random process $X(t)$ is stationary, then there exists a process $Z(\omega)$ with orthogonal increments such that

$$X(t) = \int_{-\infty}^{\infty} e^{it\omega} dZ(\omega) \quad (2.7)$$

The n th order cumulant of $X(t)$ can then be written as

$$[X(t), X(t+\tau_1), \dots, X(t+\tau_{n-1})]_n = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{i(\omega_1 + \dots + \omega_n)t} e^{i(\omega_1\tau_1 + \dots + \omega_{n-1}\tau_{n-1})} [dZ(\omega_n)]_n \quad (2.8)$$

The n th order cumulant of $X(t)$ is nonzero only when the n th order cumulant of the orthogonal increments $dZ(\omega)$ is nonzero. However, because the increments are orthogonal, their cumulant is nonzero only if all of the n frequencies are not distinct. Furthermore, because $X(t)$ is stationary, its n th order cumulant is independent of t , which is true in general only for the case $\omega_1 + \dots + \omega_n = 0$ due to the presence of the exponential $e^{i(\omega_1 + \dots + \omega_n)t}$ in the integral. This case satisfies the requirement that the n frequencies not be all distinct. Thus the n th order cumulant of $X(t)$ is

$$[X(t), X(t+\tau_1), \dots, X(t+\tau_{n-1})]_n = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{i(\omega_1\tau_1 + \dots + \omega_{n-1}\tau_{n-1})} [dZ(\omega_1), \dots, dZ(\omega_n)]_n \quad (2.9)$$

for $\omega_1 + \dots + \omega_n = 0$, and zero otherwise.

If we now define a function $P_n(\omega_1, \dots, \omega_{n-1})$ such that

$$dP_n(\omega_1, \dots, \omega_{n-1}) = [dZ(\omega_1), \dots, dZ(\omega_{n-1}), dZ(-\omega_1 \dots -\omega_{n-1})]_n \quad (2.10)$$

then if P_n is continuous it can be written as

$$dP_n(\omega_1, \dots, \omega_{n-1}) = p_n(\omega_1, \dots, \omega_{n-1}) d\omega_1 \dots d\omega_{n-1} \quad (2.11)$$

and Eq. (2.9) can be written as

$$\begin{aligned} & [X(t), X(t+\tau_1), \dots, X(t+\tau_{n-1})]_n \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{i(\omega_1\tau_1 + \dots + \omega_{n-1}\tau_{n-1})} p_n(\omega_1, \dots, \omega_{n-1}) d\omega_1 \dots d\omega_{n-1} \\ &= \mathfrak{F}_{n-1}^{-1} \{ p_n(\omega_1, \dots, \omega_{n-1}) \} \end{aligned} \quad (2.12)$$

Thus the n th order cumulant of $X(t)$ is the inverse Fourier transform of the function p_n . This function is referred to as the n th order (cumulant) spectral density function of $X(t)$. If

the cumulant function is integrable, then the nth order spectral density function can be written as the Fourier transform of the cumulant function:

$$p_n(\omega_1, \dots, \omega_{n-1}) = \mathfrak{F}_{n-1} \{ [X(t), \dots, X(t+t_{n-1})]_n \} \quad (2.13)$$

Brillinger (1975) has shown that if $X(\omega)$ is the finite Fourier transform of $X(t)$,

$$X(\omega) = \sum_{t=0}^{N-1} X(t) e^{-i\omega t} \quad (2.14)$$

then the cumulant of the Fourier transforms is equal to the nth order spectral density function plus lower order terms:

$$\begin{aligned} & [X(\omega_1), X(\omega_2), \dots, X(-\omega_1 - \omega_2 - \dots - \omega_{n-1})]_n \\ &= (2\pi)^{n-1} N p_n(\omega_1, \dots, \omega_{n-1}) + O(1) \end{aligned} \quad (2.15)$$

Equation (2.13) is the formal definition of the nth order spectral density function and assumes, among other things, that $X(t)$ is stationary and has an nth order cumulant that is integrable. If in addition $X(t)$ has a finite Fourier transform, then the nth order spectral density function can be expressed as Eq. (2.15). As a result of the independence property of cumulants discussed above, Eq. (2.15) shows that the nth order spectral density function is zero unless the Fourier transforms are statistically dependent.

The lowest order spectral density function ($n=2$) is just the power spectrum. For a zero mean process $X(t)$ the second order cumulant is equal to the second order expected value:

$$[X(t), X(t+\tau)]_2 = \langle X(t) X(t+\tau) \rangle = R(\tau) \quad (2.16)$$

Equation (2.13) then yields the familiar definition of the power spectrum as the Fourier transform of the autocorrelation function $R(\tau)$. Equation (2.15) yields the expression for the power spectrum as the expected value of the magnitude squared of the Fourier transform of the time series, as given in Eq. (2.1). The next higher order spectral density function ($n=3$) is the bispectrum. For a zero mean process the third order cumulant is

equal to the third order expected value, so Eq. (2.15) simplifies to Eq. (2.3) for the case of the bispectrum.

2.1.3 Principal Domain of the Discrete Time Bispectrum

For a real, discrete time series $X(j)$, the bispectrum is defined over the (ω_1, ω_2) plane given by $\{-\pi < \omega_1 < \pi, -\pi < \omega_2 < \pi\}$. However, because of various symmetries in the bispectrum many regions of the bispectrum are equivalent. In this section these symmetries will be described and the principal domain (the region that contains within it no equivalent values) will be defined.

If $X(\omega)$ is the Fourier transform of the discrete time series $X(j)$, then following Eq. (2.15), the bispectrum is

$$B(\omega_1, \omega_2) = \langle X(\omega_1)X(\omega_2)X(\omega_3) \rangle, \quad (2.17)$$

where

$$\omega_3 = 2\pi n - (\omega_1 + \omega_2), \quad n = 0, 1. \quad (2.18)$$

The value of n is restricted to 0 and 1 because for any other values there are no values of ω_1 , ω_2 , and ω_3 in the domain from $-\pi$ to π that satisfy Eq. (2.18). It can be seen from Eq. (2.17) that the value of the bispectrum is not changed if any pair of frequencies is interchanged. This observation leads to the following symmetry relationships:

$$B(\omega_1, \omega_2) = B(\omega_2, \omega_1), \quad (2.19a)$$

$$B(\omega_1, \omega_2) = B(\omega_3, \omega_1), \quad (2.19b)$$

$$B(\omega_1, \omega_2) = B(\omega_3, \omega_2). \quad (2.19c)$$

For $n = 0$, Eqs. (2.19b) and (2.19c) can be written as

$$B(\omega_1, \omega_2) = B(-\omega_1 - \omega_2, \omega_1), \quad (2.19d)$$

$$B(\omega_1, \omega_2) = B(-\omega_1 - \omega_2, \omega_2). \quad (2.19e)$$

For $n = 1$, these symmetry relationships are given by

$$B(\omega_1, \omega_2) = B(2\pi - \omega_1 - \omega_2, \omega_1) , \quad (2.19f)$$

$$B(\omega_1, \omega_2) = B(2\pi - \omega_1 - \omega_2, \omega_2) . \quad (2.19g)$$

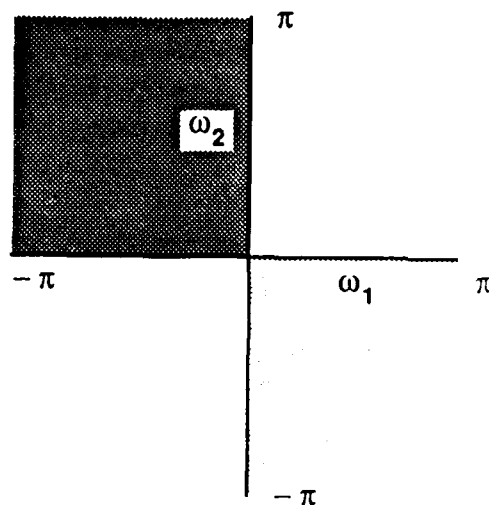
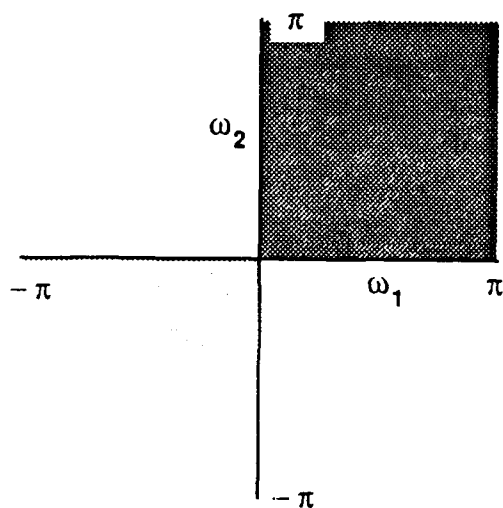
One additional symmetry relationship results from the assumption of a real time series:

$$B(\omega_1, \omega_2) = B^*(-\omega_1, -\omega_2) . \quad (2.19h)$$

Using these symmetries, a principal domain of the bispectrum can be defined. From Eq. (2.19h), it follows that the regions of the (ω_1, ω_2) plane shown in Fig. 2.1(a) are equivalent. Thus it is only necessary to consider the upper half of the plane (positive ω_2). From Eq. (2.19e), the two triangular regions shown in Fig. 2.1(b) are equivalent. Combining Eq. (2.19e) with Eqs. (2.19h) and (2.19a), it can be seen that the two triangular regions in Fig. 2.1(c) are also equivalent. Thus it is necessary to consider only positive values of both ω_1 and ω_2 .

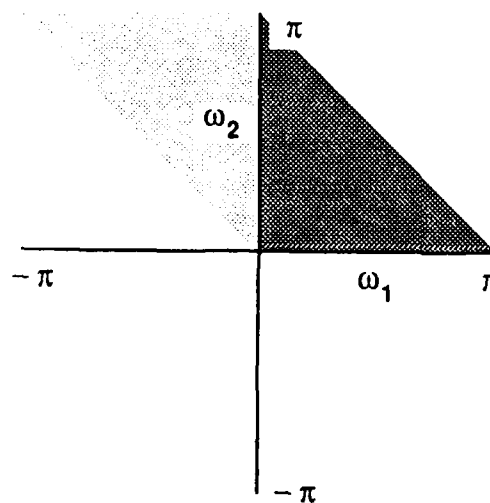
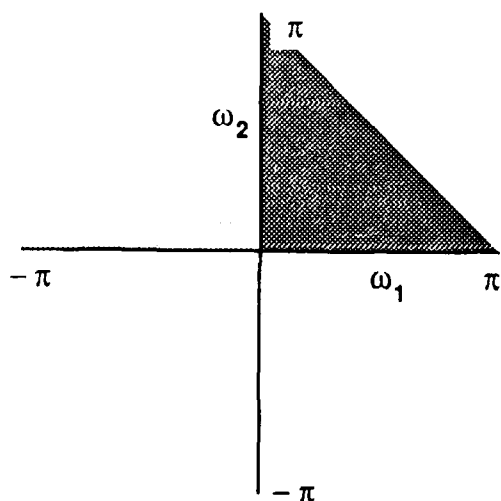
Applying Eq. (2.19a) to the positive (ω_1, ω_2) quadrant results in two halves of the quadrant being equivalent, as shown in Fig. 2.2(a). Equation (2.19f) implies that the two triangles in Fig. 2.2(b) are equivalent, while Eq. (2.19g) implies that the two triangles in Fig. 2.2(c) are equivalent. No other symmetry relationships can be applied to further eliminate equivalent regions in the bispectrum. Thus the principal domain is the triangular region shown in Fig. 2.3(a).

Although this triangular region is the principal domain, if the time series is a stationary, continuous time function that has been low pass filtered prior to digitizing such that there is no energy above the Nyquist frequency (no aliasing), then the bispectrum should only be nonzero in the triangular region shown in Fig. 2.3(b) (Hinich and Wolinsky, 1988). This result is due to the fact that it is only within this triangular region that the sum $\omega_1 + \omega_2$ is less than the Nyquist frequency. Thus with proper filtering it is only necessary to compute the bispectrum over the isosceles triangle shown in Fig. 2.3(b).



(a)

$$B(\omega_1, \omega_2) = B^*(-\omega_1, -\omega_2)$$



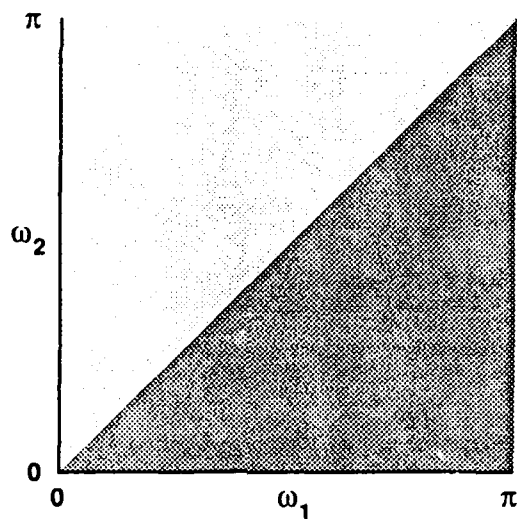
(b)

$$B(\omega_1, \omega_2) = B(-\omega_1 - \omega_2, \omega_2)$$

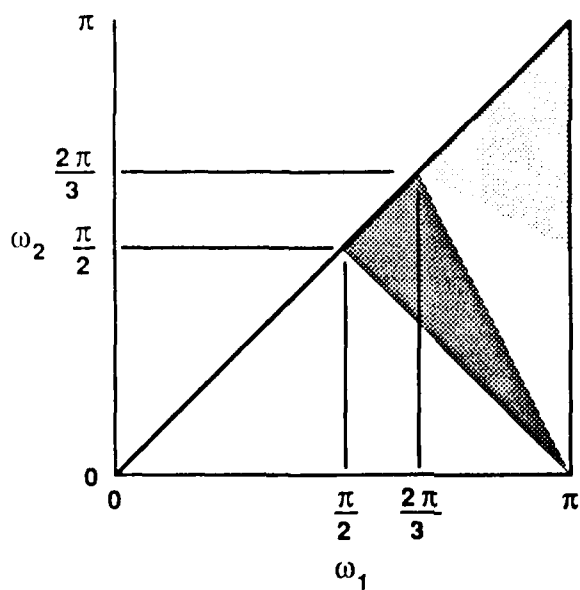
(c)

$$B(\omega_1, \omega_2) = B^*(-\omega_2, \omega_1 + \omega_2)$$

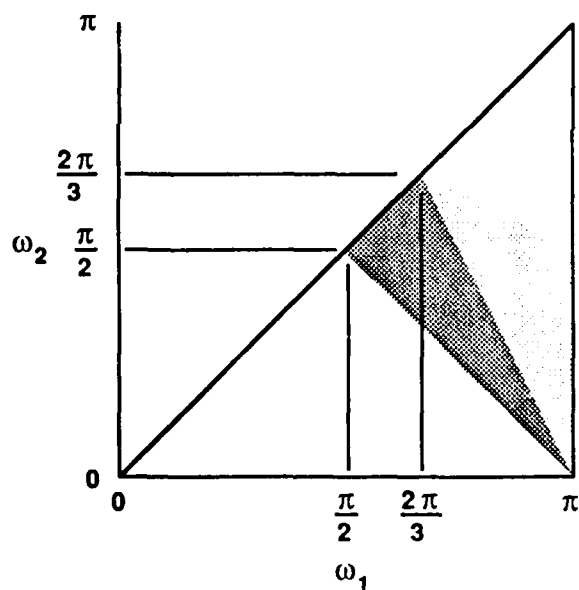
FIGURE 2.1
SYMMETRIES OF THE BISPECTRUM



(a)
 $B(\omega_1, \omega_2) = B(\omega_2, \omega_1)$

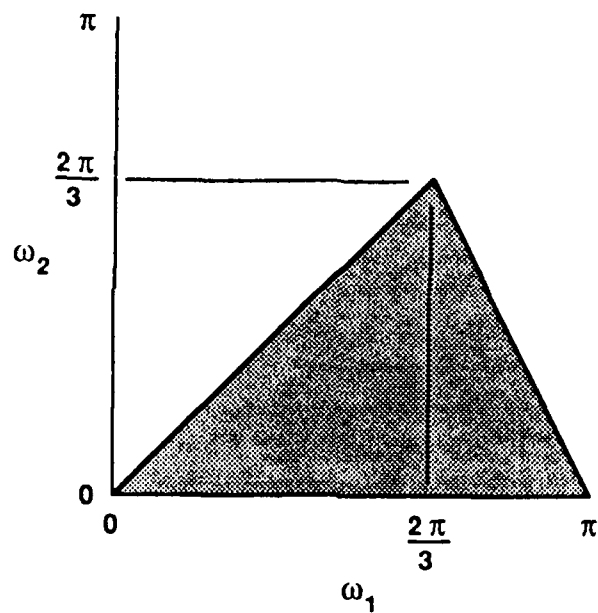


(b)
 $B(\omega_1, \omega_2) = B(2\pi - \omega_1 - \omega_2, \omega_1)$

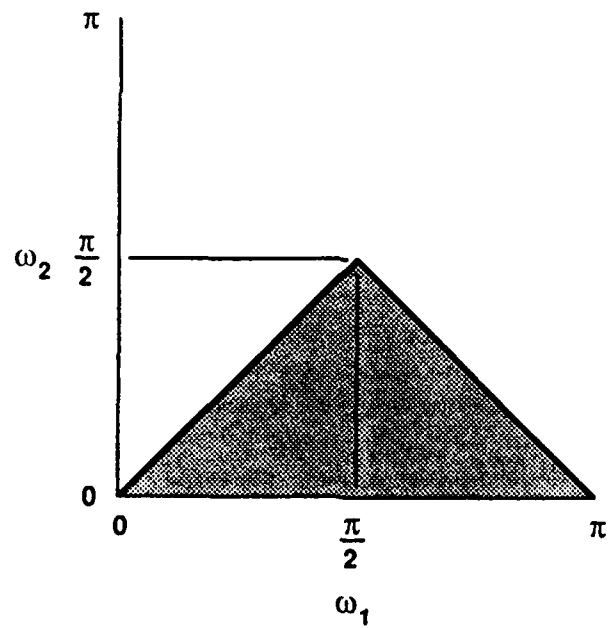


(c)
 $B(\omega_1, \omega_2) = B(2\pi - \omega_1 - \omega_2, \omega_2)$

FIGURE 2.2
 SYMMETRIES OF THE BISPECTRUM



(a)



(b)

FIGURE 2.3
PRINCIPAL DOMAIN OF THE BISPECTRUM

2.2 ESTIMATION OF THE BISPECTRUM

2.2.1 The Unnormalized Bispectrum Estimate: Averaging in Frequency and in Time

The estimator used to estimate the bispectrum for the analysis presented in this report was developed by Hinich (1982) and is based on a discrete Fourier transform of the data. For the discrete time series $x(t)$, $t=0, 1, \dots, N-1$, the discrete Fourier transform is given by

$$X(j) = \frac{1}{N} \sum_{t=0}^{N-1} x(t) e^{-i2\pi jt/N}, \quad j = 0, 1, \dots, N-1. \quad (2.20)$$

The frequency associated with the j th component is

$$\begin{aligned} f_j &= \frac{j}{N} f_s, \quad j = 0, 1, \dots, \frac{N}{2} \\ &= \frac{-(N-j)}{N} f_s, \quad j = \frac{N}{2} + 1, \dots, N-1, \end{aligned} \quad (2.21)$$

where f_s is the rate at which the time series was sampled. The discrete Fourier transform is computed using an FFT algorithm.

A consistent estimate of the bispectrum (one whose expected value approaches the true value and whose variance goes to zero as N approaches infinity) can be constructed using an FFT of the time series. The expected value of the complex function

$$F(j,k) = N^2 X(j) X(k) X^*(j+k) \quad (2.22)$$

is equal to the bispectrum $B(f_j, f_k)$ plus terms on the order of N^{-1} . Thus $F(j,k)$ is an unbiased estimate of the bispectrum. However, it is not a consistent estimate of the bispectrum because its variance increases with N . To obtain a consistent estimate, either the function $F(j,k)$ can be averaged in frequency, or multiple realizations of $F(j,k)$ can be averaged in time, or both.

To average in frequency, the function $F(j,k)$ is averaged over adjacent values in a square of M^2 points centered at the points

$$(g_m, g_n) = \left(\frac{(2m-1)M-1}{2}, \frac{(2n-1)M-1}{2} \right)$$

$$m = 1, \dots, \left[\frac{N}{2M} + .5 \right]$$

$$n = 1, \dots, m \quad n \leq \left[\frac{N}{M} - 3m + \frac{3}{2} + \frac{3}{2M} \right],$$

where $[\cdot]$ denotes the "greatest integer not exceeding" function. Thus the frequency averaged bispectrum estimator is given by

$$\hat{B}(m,n) = M^{-2} \sum_{j=(m-1)M}^{mM-1} \sum_{k=(n-1)M}^{nM-1} F(j,k) \quad \left\{ j,k: 0 < j \leq \frac{N}{2}, 0 < k \leq j, 2j+k \leq N \right\} \quad (2.23)$$

If the bispectrum is slowly varying over the square and if the power spectrum is slowly varying over the band of width M centered at the appropriate frequencies, then the variance of this estimator is given by

$$\text{VAR } \hat{B}(m,n) = N M^{-4} Q(m,n) P(f_{g_m}) P(f_{g_n}) P(f_{g_m+g_n}) + O(M/N) \quad (2.24)$$

where $Q(m,n)=M^2$ if the square is entirely within the principal domain; otherwise it is equal to the number of (j,k) within the square but not on the boundaries $j=k$ or $2j+k=N$ plus twice the number of (j,k) on the boundaries. $P(\cdot)$ is the power spectrum of the time series. From Eq. (2.24) it can be seen that the estimator given by Eq. (2.23) is a consistent estimator for values of M given by

$$\sqrt{N} < M < N \quad (2.25)$$

The bias increases and the variance decreases as M increases.

To average the bispectrum in time also, the time series can be divided into L segments, each of length N , and the bispectrum is estimated (and possibly averaged in frequency) for each segment and all L estimates are averaged together. If each of the L bispectrum estimates is uncorrelated, the variance of this coherent average is just

$$\text{VAR } \hat{B}_L(m,n) = \frac{\text{VAR } \hat{B}(m,n)}{L} . \quad (2.26)$$

In this case, consistency is obtained for values of L and M given by

$$\sqrt{\frac{N}{L}} < M < N . \quad (2.27)$$

Thus consistency can be obtained by averaging in time as well as frequency.

The tradeoffs between frequency and time averaging can best be understood by examining the variance of the bispectrum estimate. From Eqs. (2.24) and (2.26), it can be seen that the variance of the bispectrum estimate is proportional to

$$\frac{N}{LM^2} ,$$

where N is the length of the FFT, L is the number of bispectra that are coherently averaged in time, and M is the size of the square over which the bispectrum is averaged in frequency. The total number of samples used to estimate the bispectrum is NL . The resolution of the bispectrum in frequency is given by

$$\frac{Mf_s}{N} ,$$

where f_s is the sampling rate. If, instead of estimating the bispectrum with FFTs of length N and coherently averaging L bispectra, the bispectrum is estimated with a single FFT of length $N'=NL$ but with the same sampling rate, and then, if this bispectrum is averaged in frequency over a square of size $M'=ML$, the variance and the resolution are the same as the bispectrum estimate using FFTs of size N . Thus, averaging in frequency and averaging in time are equivalent if the appropriate choice of M is made. The advantage of dividing the time series into L segments of length N and averaging in time is that it requires significantly fewer calculations and can thus be performed faster and with less storage requirements than computing one bispectrum using NL samples. The potential liability of this approach is that if the resolution of the FFT is significantly coarser than the bandwidth of the nonlinear component, the presence of the nonlinearity will be reduced in the bispectrum. Thus the best choice of N , M , and L is, as is usually the case, to some extent data dependent.

As can be seen from the above expressions, decreasing the value of M improves the resolution but increases the variance of the bispectrum estimate. This increase in variance can be offset by coherently averaging the bispectrum over a longer time, i.e., increasing L . However, since L is related to the variance by L^{-1} whereas M is related by M^{-2} , decreasing M by a factor of 2, for example, requires increasing L by a factor of 4 to achieve the same variance. In practical situations there are almost always limits to how much L can be increased, the limits being determined by the stationarity of the time series. Thus the frequency resolution of the bispectrum that can be achieved in practice depends on the stationarity of the time series. Furthermore, the improvement in resolution that can be achieved is not linear with the increase in coherent averaging time, but instead goes as the square root of the increase in averaging time.

2.2.2 The Normalized Bispectrum: Statistical Properties

The bispectrum can be normalized to produce a quantity whose asymptotic statistics can easily be calculated. The asymptotic distribution of the estimator given by Eq. (2.23) is complex normal and independent for each frequency pair. Thus the distribution of the normalized bispectrum given by

$$\chi^2(m,n) = \frac{2|\hat{B}(m,n)|^2}{\text{VAR } \hat{B}(m,n)} \quad (2.28)$$

is asymptotically noncentral chi-square with two degrees of freedom and noncentrality parameter

$$\lambda(m,n) = \frac{2}{N M^4 Q(m,n)} \gamma^2(f_{g_m}, f_{g_n}) \quad (2.29)$$

where

$$\gamma(f_{g_m}, f_{g_n}) = \frac{|B(f_{g_m}, f_{g_n})|}{\sqrt{P(f_{g_m}) P(f_{g_n}) P(f_{g_m+g_n})}} \quad (2.30)$$

is called the skewness function. If the time series is Gaussian, its skewness function will be zero for all frequency pairs and the distribution of the normalized bispectrum is just a central chi-square with two degrees of freedom (the noncentrality parameter will be zero).

This implies that an a priori threshold can be estimated for the normalized bispectrum such that values of the normalized bispectrum that exceed the threshold can be said to be representative of a non-Gaussian time series with a probability of false alarm that is related to the threshold. Thus the normalized bispectrum can be used to discriminate between Gaussian and non-Gaussian time series (Hinich, 1982).

In practice, an estimate of the normalized bispectrum given in Eq. (2.28) requires an estimate of the variance of the bispectrum estimate, which in turn requires an estimate of the power spectrum in the expression for the variance. If the power spectrum is estimated by

$$\hat{P}(j) = N|X(j)|^2 \quad (2.31)$$

and is smoothed over a band of at least M adjacent values (or averaged in time), then this estimate of the normalized bispectrum using the estimated power spectrum will also be asymptotically noncentral chi-square distributed with two degrees of freedom and the same noncentrality parameter. If the bispectrum is coherently averaged over L segments, then the distribution of the normalized bispectrum after averaging is still noncentral chi-square with two degrees of freedom but with noncentrality parameter $L\lambda(m,n)$.

A sliding average can be used to average both the power spectrum and bispectrum in time. The temporal average of the power spectrum for L records is given by

$$P_L^{(a)}(j) = \frac{P_L(j) + \alpha \left(\sum_{k=1}^{L-1} \alpha^{k-1} \right) P_{L-1}^{(a)}(j)}{\sum_{k=1}^L \alpha^{k-1}}, \quad (2.32)$$

where $0 \leq \alpha \leq 1$ and $P_0^{(a)} = 0$. A similar expression for the time averaged bispectrum can be obtained. Block averaging results when $(\alpha=1)$.

If a sliding average is applied to the bispectrum, then its variance is given by

$$\text{VAR } B_L^{(a)}(m,n) = \frac{\sum_{k=1}^L (\alpha^{k-1})^2}{\left(\sum_{k=1}^L \alpha^{k-1} \right)^2} \text{VAR } B(m,n) , \quad (2.33)$$

where $\text{VAR } B(m,n)$ is the variance of the bispectrum given by Eq. (2.24) (ignoring terms on the order of M/N) with the averaged power spectrum used as the estimate of the power spectrum. For block averaging ($\alpha=1$), the ratio of the two summations in Eq. (2.33) is just L^{-1} .

From a detection point of view, the issue is how large the normalized bispectrum statistic given by Eq. (2.28) has to be in order to confidently reject the Gaussian noise only hypothesis and assert that a non-Gaussian signal is present. Shown in Fig. 2.4 is a plot of the probability of false alarm as a function of the normalized bispectrum value based on the central chi-squared distribution function with two degrees of freedom. The plot is approximately linear on a log false alarm scale with a slope of 4.6 per decade. To operate at a false alarm rate of 10^{-3} , one would reject the hypothesis that noise only is present (the normalized bispectrum statistic is central chi-square distributed rather than noncentral chi-square distributed) for values of the normalized bispectrum statistic that are 13.8 or larger. Since the mean value of the normalized bispectrum statistic is equal to the noncentrality parameter, then the noncentrality parameter must be 13.8 or larger to achieve a probability of detection of 0.5 at this false alarm rate. This result is based on detection at a single point in the bispectrum. Summing the bispectrum over its principal domain can also be used to detect the presence of a non-Gaussian signal (Hinich, 1982).

2.3 THE BISPECTRUM WATERFALL

It is typical to view the time history of the power spectrum in a waterfall display format, where the vertical axis is time and the horizontal axis is frequency. Because the bispectrum is a function of two frequencies, it is not possible to directly display it in a waterfall format. However, it is possible to reduce the two-dimensional bispectrum to a form that is amenable to a waterfall display format. This is done by summing over the principal domain all values of the normalized, averaged bispectrum $\chi^2(m,n)$ such that the sum of the two frequencies (f_m, f_n) is constant

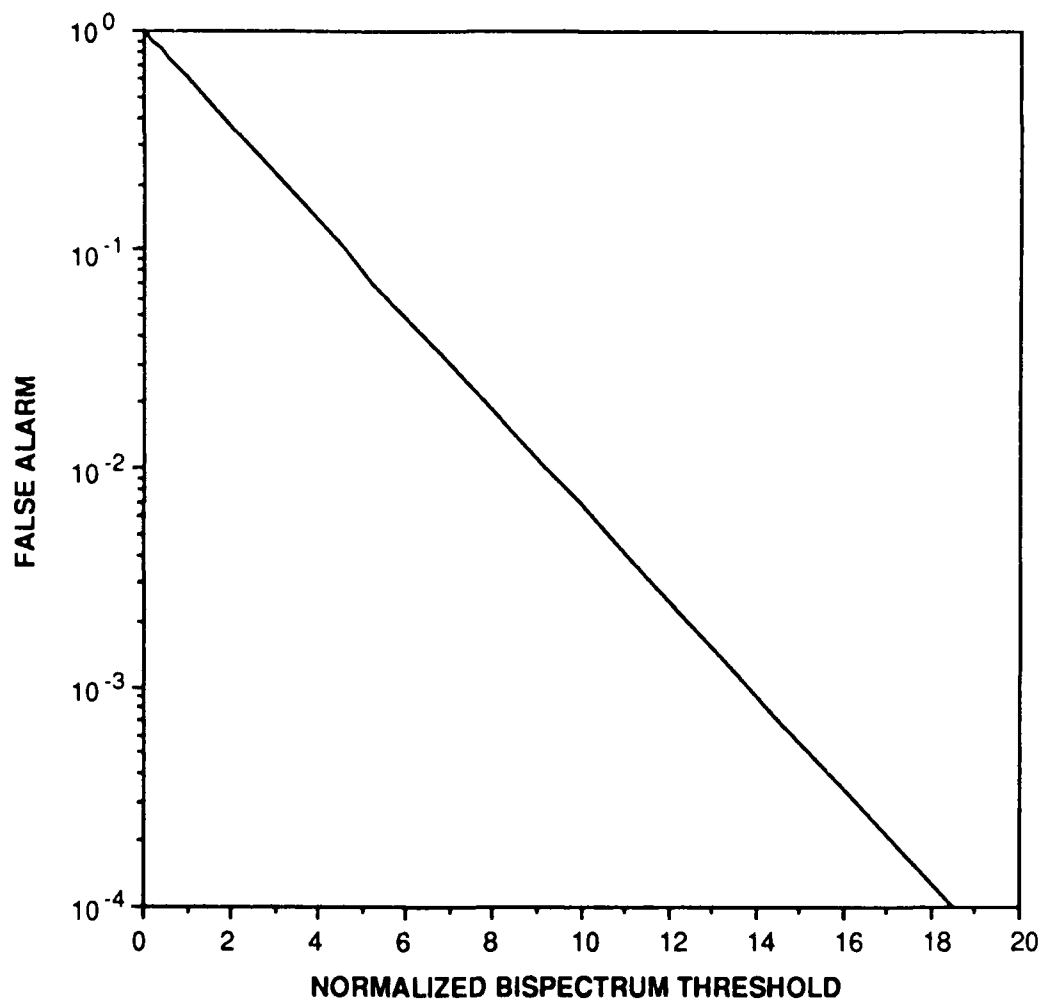


FIGURE 2.4
FALSE ALARM PROBABILITY AS A FUNCTION OF
NORMALIZED BISPECTRUM THRESHOLD

$$\chi^2(i) = \sum_{m,n: m+n=i} \chi^2(m,n) \quad (2.34)$$

The frequency pairs that satisfy this constraint lie along a diagonal as shown in Fig. 2.5. This summation is repeated for all frequencies (f_m, f_n) that are represented in the Fig. 2.5 bispectrum. In this way the two-dimensional bispectrum has been collapsed to one dimension and can then be displayed (with loss of information) as a function of time in a waterfall format, much like a power spectrum waterfall.

Since the bispectral estimate at each frequency pair is asymptotically independent, then under the assumption of Gaussianity each value in this waterfall will asymptotically have a central chi-square distribution with $2I$ degrees of freedom, where I is the number of bispectral values included in the summation in Eq. (2.34). Since the mean and variance of a chi-square distributed random variable are proportional to the degrees of freedom, values of the waterfall at higher frequency values will have a larger mean and variance than values at lower frequency values. This tends to produce an uneven display that hinders detection. The summation in Eq. (2.34) can be scaled by the number of terms in the summation, thus causing the mean values of the waterfall to be independent of frequency and resulting in a variance that decreases at higher frequencies. The summation can also be restricted to include only those values that exceed a specified threshold, so that the number of degrees of freedom is determined by the number of large bispectral values rather than by the frequency. This latter approach is illustrated in Fig. 2.6, where a signal with a bispectrum at a single point has been added to Gaussian noise. The presence of the bispectral signal is much more obvious in the waterfall than in the full bispectrum, demonstrating the usefulness of the waterfall presentation for detecting the presence of bispectral signals in noise.

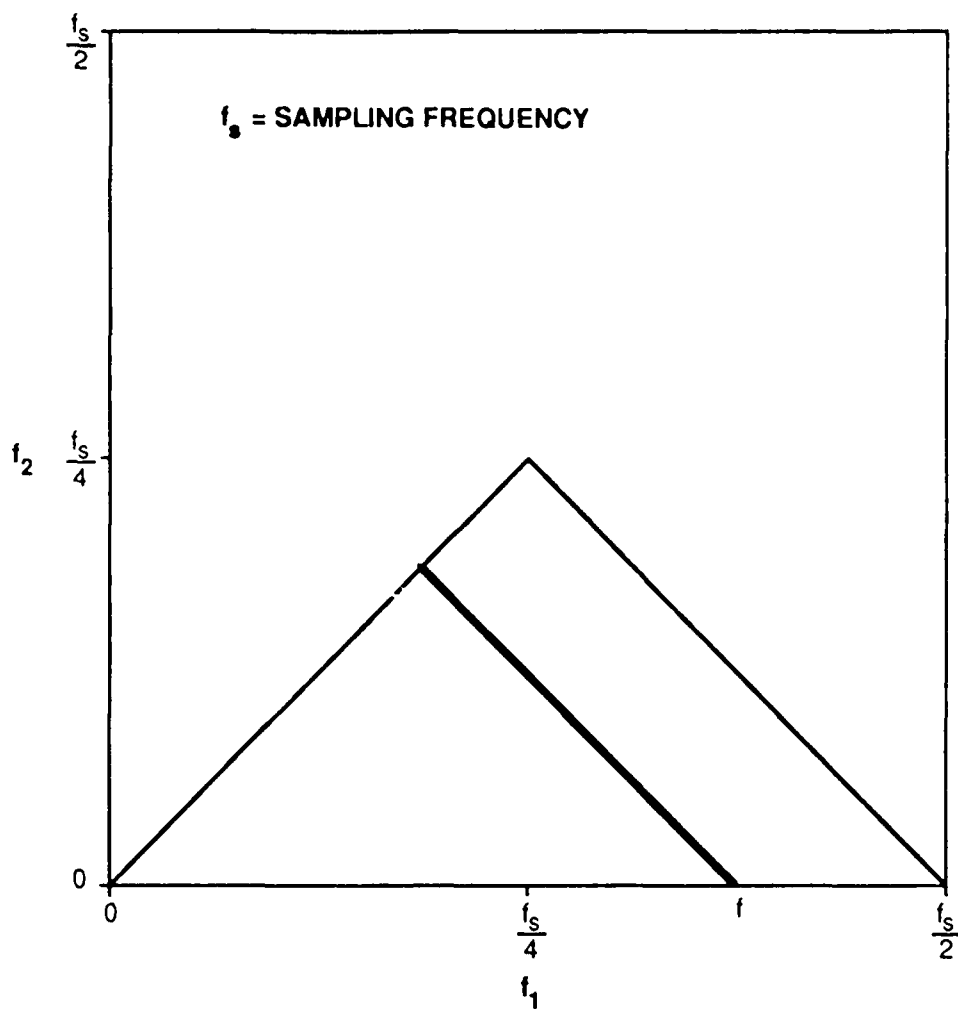
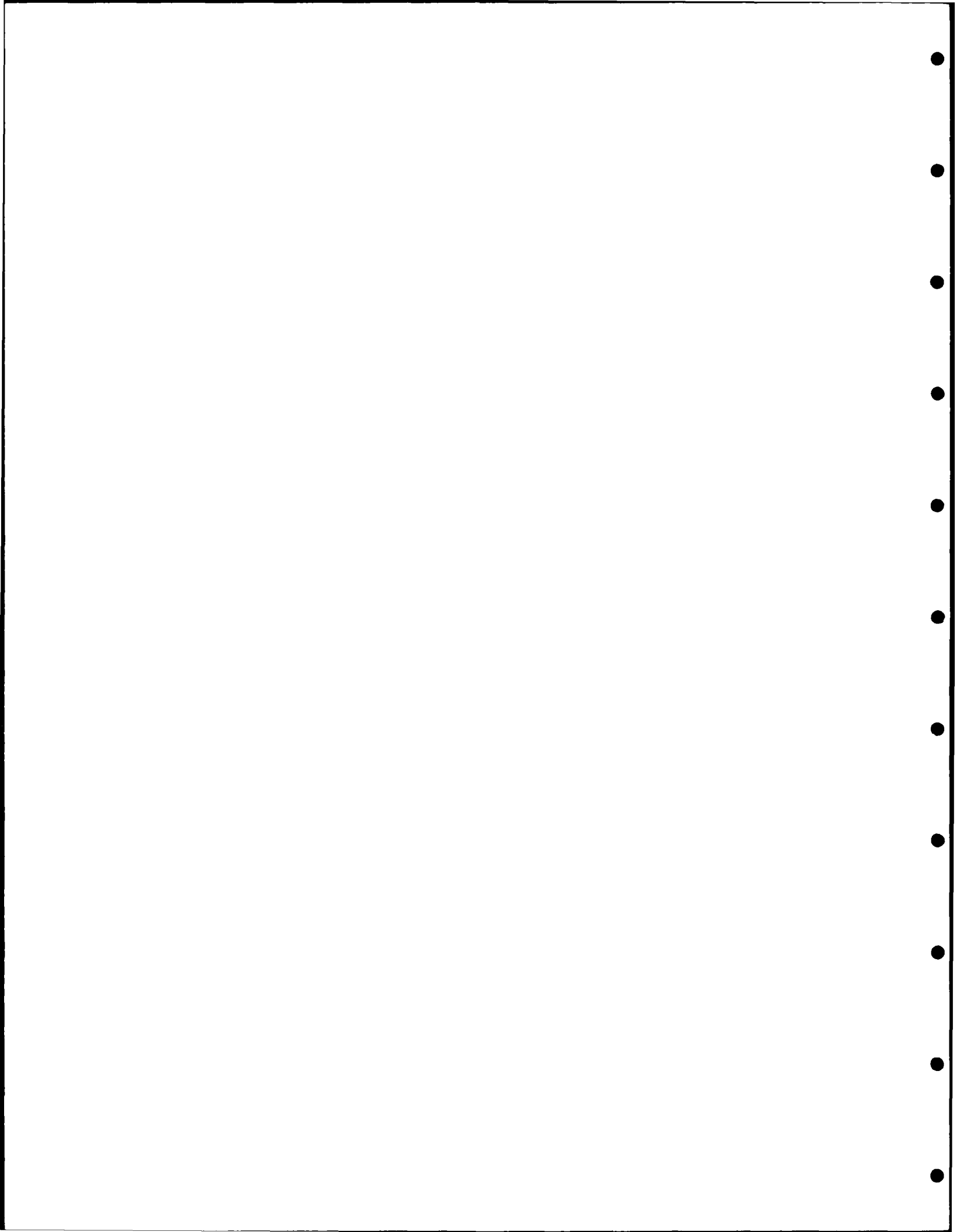


FIGURE 2.5
BISPECTRUM WATERFALL



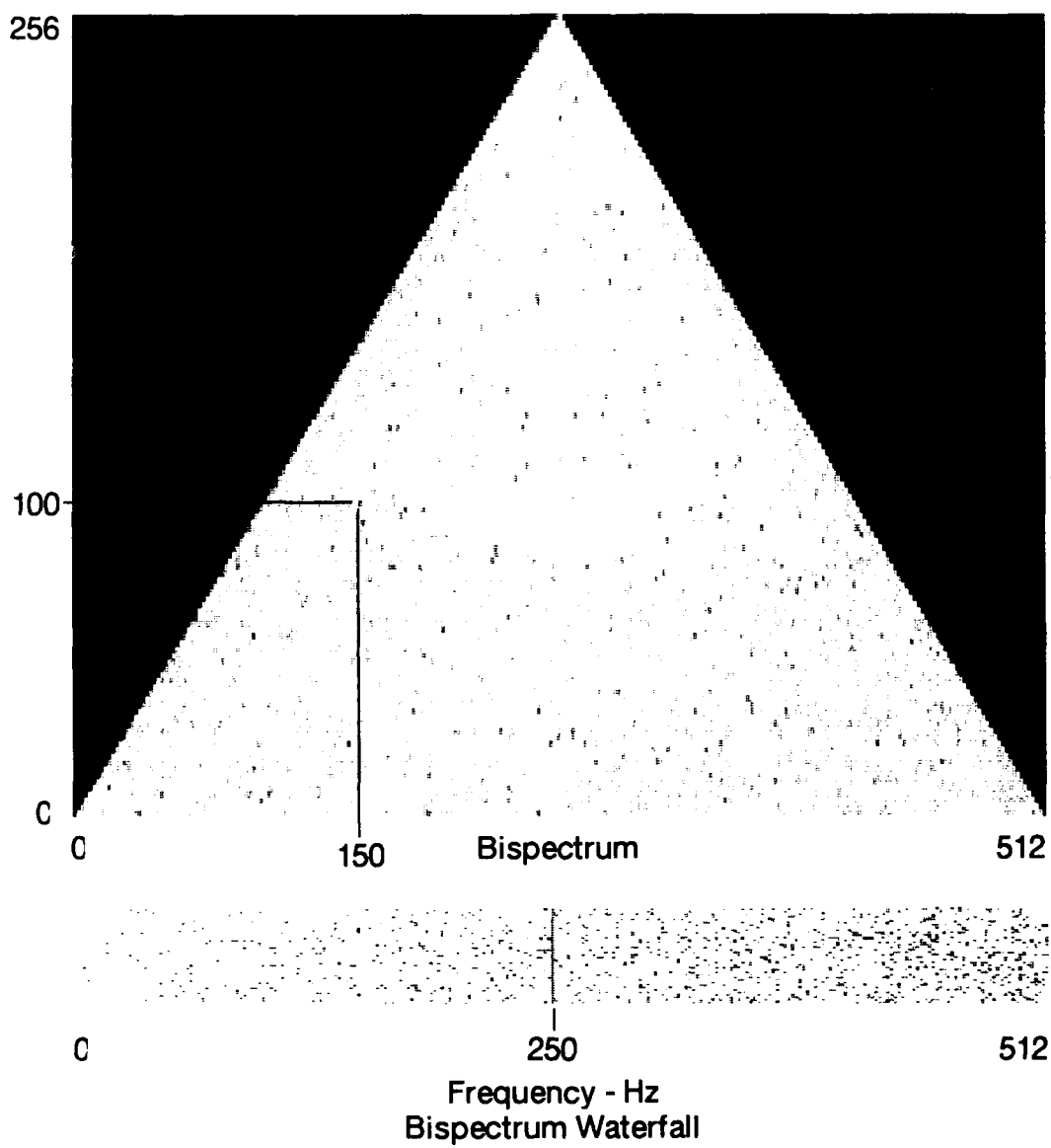


FIGURE 2.6
EXAMPLE OF BISPECTRUM WATERFALL

3. DETECTION OF NON-GAUSSIAN SIGNALS IN NON-GAUSSIAN NOISE

It is shown in Hinich and Wilson (1990) that the statistic

$$CH(m,n) = \frac{2LM^4Q^{-1}(m,n)}{N} \frac{|B^{(N)}(m,n) - B_n(m,n)|^2}{P_{s+n}(f_{g_m})P_{s+n}(f_{g_n})P_{s+n}(f_{g_m+g_n})} , \quad (3.1)$$

where B_n is the bispectrum of the noise, P_{s+n} is the signal-plus-noise power spectrum, and $B^{(N)}$ is a consistent estimate of the measured bispectrum such as given by Eq. (2.23), is approximately a central chi-square random variable with two degrees of freedom under the null hypothesis of noise only. Under the alternative hypothesis of signal plus noise it is a noncentral chi-square random variable with noncentrality parameter given by

$$\lambda(m,n) = \frac{2 L M^4 Q^{-1}(m,n)}{N (1+\rho^{-1}(f_{g_m}))(1+\rho^{-1}(f_{g_n}))(1+\rho^{-1}(f_{g_m+g_n}))} \gamma_s^2(f_{g_m}, f_{g_n}) , \quad (3.2)$$

where γ_s is the skewness function of the signal and ρ is the signal-to-noise power ratio:

$$\rho(f) = \frac{P_s(f)}{P_n(f)} . \quad (3.3)$$

In Eqs. (3.1) and (3.2) it is assumed that a block average in time of L records is applied. If a sliding average is applied, the factor L is replaced by the appropriate ratio of summations that appears in Eq. (2.33). Detection occurs when the value of the statistic given by Eq. (3.1) exceeds a threshold determined by the central chi-square distribution (see Fig. (2.5)). Since the mean value of Eq. (3.1) is equal to Eq. (3.2), the value of the noncentrality parameter given by Eq. (3.2) determines the detectability of the signal.

From Eq. (3.2) it can be seen that several factors contribute to the value of the noncentrality parameter. One is the skewness function γ_s which is a characteristic of the signal. one is the S/N ρ , which is a characteristic of the signal and noise power levels, and the rest are characteristics of the processing. Given that a signal has a nonzero skewness function, then there is a tradeoff between S/N and processing parameters that determines if the nonzero skewness will result in a sufficiently large noncentrality parameter to allow its detection at a given false alarm rate. For small values of the skewness function, the

processing parameters and S/N have to be such that their product, given by Eq. (3.2), remains large enough to produce a sufficiently large noncentrality parameter for detection. The noncentrality parameter has a linear dependence on the number of temporal averages L , a quadratic dependence on the number of frequency averages M (Q is equal to M^2 except on the boundaries of the principal domain), and an approximately cubic dependence on S/N (for low S/N). This implies that if the S/N decreases by a factor of 2 (3 dB), then it is necessary to either increase the number of temporal averages by a factor of 8, increase the number of averages in frequency by a factor of $\sqrt{8}$, or increase both the frequency and time averaging such that the product LM^2 increases by a factor of 8, in order to retain the same level of detectability.

Equation (3.2) demonstrates the essential relationship between signal characteristics, noise characteristics, and processing parameters that determines the detection performance of the bispectrum. To determine the viability of bispectrum processing for detection of non-Gaussian signals, it is essential to know the skewness function of signals of interest. Given a skewness function, the processing parameters necessary to achieve detection as a function of S/N can then be determined. It should also be noted that Eq. (3.2) is relevant for "narrowband" detection, i.e., detection at a single point in the bispectrum. One can also consider "broadband" detection in which the detection statistic is based on bispectrum values over the entire principal domain. Examples of the detection performance as a function of these parameters for the broadband case can be found in Hinich and Wilson (1990), which is included in Appendix A of this report.

4. TIME DELAY ESTIMATION USING THE CROSS-BISPECTRUM

4.1 TIME DELAY ESTIMATION BASED ON THE PHASE OF THE CROSS-BISPECTRUM

Following the formal definition of higher order spectra for a single process in Section 2.1.2, the cross-bispectrum can be written as

$$B_{112}(\omega_1, \omega_2) = \mathfrak{I}_2\{[X_1(t+\tau_1) X_1(t+\tau_2) X_2(t)]\} \quad , \quad (4.1)$$

where $X_1(t)$ and $X_2(t)$ are two stationary zero mean random processes. By arguments similar to those given in Section 2.1.3, it can be shown that the principal domain of the cross-bispectrum is given by the region in Fig. 4.1(a). The support set is given by Fig. 4.1(b). The cross-bispectrum can be estimated in a manner completely analogous to the bispectrum estimate given by Eqs. (2.22) and (2.23).

If $X_1(t)$ is just a time delayed version of $X_2(t)$, then

$$B_{112}(m,n) = B_{111}(m,n) e^{i2\pi f_{m+n}\tau} \quad . \quad (4.2)$$

Thus B_{112} is just a phase shifted version of B_{111} :

$$\theta_{112}(m,n) = \theta_{111}(m,n) + 2\pi f_{m+n}\tau \quad , \quad (4.3)$$

where θ_{112} and θ_{111} are the phases of the cross- and auto-bispectrum, respectively. Several non-parametric methods for estimating this time delay have been put forth that essentially rely upon estimates of the cross- and auto-bispectrum phases. (Nikias and Pan, 1988).

The obvious estimator of the phase of the cross-bispectrum is just

$$\Phi_{112}(m,n) = \tan^{-1} \left[\frac{B_{112}^{(I)}(m,n)}{B_{112}^{(R)}(m,n)} \right] \quad , \quad (4.4)$$

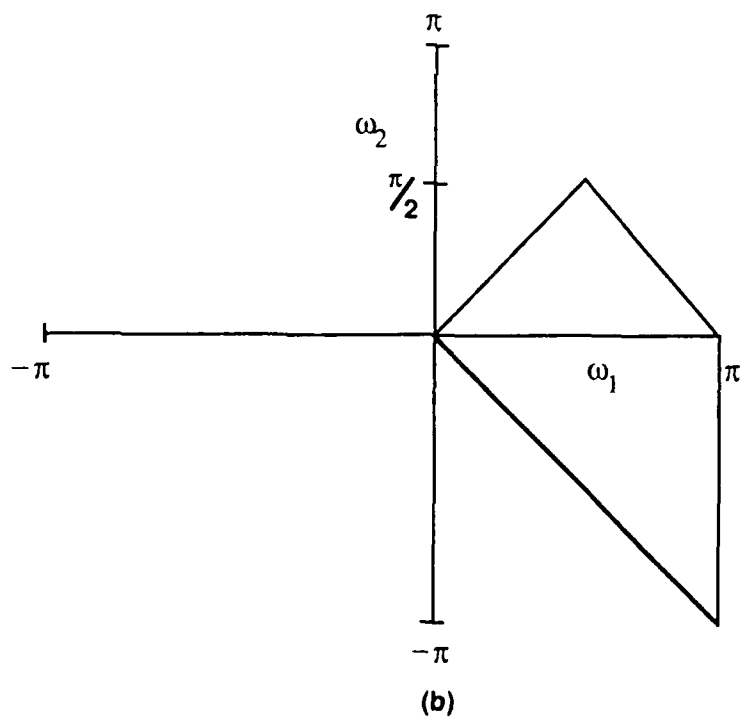
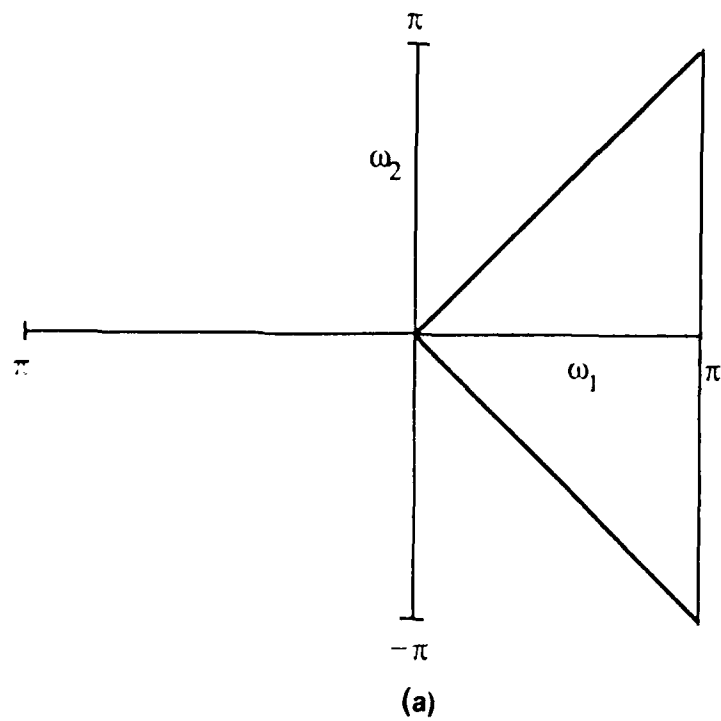


FIGURE 4.1
PRINCIPAL DOMAIN AND SUPPORT SET FOR THE CROSS-BISPECTRUM

where $B_{112}^{(I)}$ and $B_{112}^{(R)}$ are just the imaginary and real parts of a consistent estimate of the cross-bispectrum. The estimate of the auto-bispectrum phase is similar. In Hinich and Wilson (1989) we derived the variance and distribution of this phase estimator and showed that the exact distribution could be approximated by a Gaussian distribution in most cases.

If the time delay is estimated from Eq. (4.3) as

$$\hat{\tau}(m,n) = \frac{\Phi_{112}(m,n) - \Phi_{111}(m,n)}{2\pi f_{m+n}} \quad , \quad (4.5)$$

then it can also be approximated by a Gaussian distribution since it is just the scaled difference between two approximately Gaussian random variables. The variance of the phase difference in additive Gaussian noise was shown in Hinich and Wilson (1989) to be

$$\begin{aligned} \text{VAR}\{\Phi_{112}(m,n) - \Phi_{111}(m,n)\} &= \frac{2NM^2Q^{-1}(m,n)}{L\gamma_{111}^2(f_m, f_n)} [1 + \rho^{-1}(f_m)][1 + \rho^{-1}(f_n)]\rho^{-1}(f_{m+n}) \\ &\times \left\{ 1 - \text{Re}[W_{12}^{(N)}(f_{n+m}) e^{-i2\pi f_{m+n}\tau}] \right\} \quad , \end{aligned} \quad (4.6)$$

where $W_{12}^{(N)}(f)$ is the coherency spectrum of the noise:

$$W_{12}(f) = \frac{P_{12}^{(N)}(f)}{\sqrt{P_{11}^{(N)}(f) P_{22}^{(N)}(f)}} \quad . \quad (4.7)$$

The effects of signal skewness, processing parameters, S/N, and noise coherence on the variance of the time delay estimate are clearly demonstrated in Eq. (4.6). Specific examples of the effects of these parameters are described in Hinich and Wilson (1989), which is included as Appendix B of this report.

4.2 AN ALTERNATIVE TIME DELAY ESTIMATOR

If we define an estimator $\hat{S}_B(i)$ such that

$$\hat{S}_B(i) = \sum_{(m,n): m+n=i} \hat{B}_{112}(m,n) \hat{B}_{111}^*(m,n) \quad , \quad (4.8)$$

where $\hat{B}_{112}$ and $\hat{B}_{111}$ are consistent estimators of the cross- and auto-bispectrum respectively, then

$$E\{\hat{S}_B(i)\} = \sum_{(m,n):m+n=i} B_{112}(f_m, f_n) \dot{B}_{111}(f_m, f_n) + NP_1(f_m)P_1(f_n)\dot{P}_{12}(f_{m+n}) . \quad (4.9)$$

This equation derives from the expression of $E\{X_1(m)X_1(n)X_2(m+n)X_2(m)X_1(n)X_2(m+n)\}$ in terms of its cumulants and the relationship between the cumulants of DFTs and cumulant spectra given by Brillinger (1975). If we define

$$\hat{S}_\Gamma(i) = \frac{\hat{S}_B(i)}{P_1(f_m)P_1(f_n)P_{12}(f_{m+n})} \quad (4.10)$$

and

$$\hat{\Gamma}_{112}(m,n) = \frac{\hat{B}_{112}(m,n)}{\sqrt{P_1(f_m)P_1(f_n)P_2(f_{m+n})}} \quad (4.11)$$

and similarly for Γ_{111} , then we have

$$E\{\hat{S}_\Gamma(i)\} = S_\Gamma(f_i) + NI(i)W_{12}(f_i) , \quad (4.12)$$

where

$$\begin{aligned} S_\Gamma(f_i) &\equiv \sum_{(m,n):m+n=i} \Gamma_{112}(f_m, f_n) \dot{\Gamma}_{111}(f_m, f_n) \\ &\equiv \text{cross-skewness spectrum} , \end{aligned} \quad (4.13)$$

and $I(i)$ is the number of terms in the sum in Eq. (4.8). We can see from Eq. (4.2) that if $X_1(t)$ is a time delayed version of $X_2(t)$, then the phase of the cross-skewness spectrum is just $2\pi f_{m+n} \tau'$ and its magnitude is the sum of the square of the skewness function over all $(m,n):m+n=i$. Thus the time delay can be estimated directly from the phase of the cross-skewness spectrum and the magnitude of the cross-skewness spectrum at each frequency is related to the square of the skewness function of the signal.

The cross-skewness spectrum thus has characteristics similar to the cross-power spectrum. If in analogy to the cross-skewness spectrum we define a *cross-skewness* correlation $r_f(\tau)$ as the Fourier transform of the cross-skewness spectrum, then $r_f(\tau)$ will have many of the properties of a crosscorrelation function. For a broadband non-Gaussian signal, the cross-skewness correlation will have a maximum at the time delay τ' and be near zero elsewhere. Thus the cross-skewness spectrum and correlation can serve as useful tools for measuring the time delay of broadband non-Gaussian sources.

APPENDIX A

DETECTION OF NON-GAUSSIAN SIGNALS IN NON-GAUSSIAN NOISE USING THE BISPECTRA (Hinich and Wilson (1990))

(This paper has been published in IEEE Trans. Acoust, Speech,
and Signal Proc. 38(7), pp. 1126-1131 (1990)).

Detection of non-Gaussian signals in non-Gaussian noise using the bispectra

Melvin Hinich and Gary R. Wilson, Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

Bispectrum analysis provides a means of testing a stochastic time series for non-Gaussianity (including nonlinearity). For cases of practical interest, the non-Gaussian time series may be partially masked by either Gaussian or non-Gaussian noise. In this paper we cast the problem of detecting a non-Gaussian time series in the presence of additive Gaussian or non-Gaussian noise into a classical hypothesis testing framework, using the sample bispectrum as the test statistic. The power of the test is demonstrated as a function of signal-to-noise ratio, the degree of skewness of the signal, and processing parameters. The results are compared to the power of a classical energy detection test.

INTRODUCTION

The bispectrum is becoming a useful and practical tool for non-Gaussian time series analysis and diagnosis in fields such as biomedicine (Barnett, et al, 1971), fluid mechanics (Lii, et al, 1976), oceanography (Hasselmann, et al, 1963), plasma physics (Kim and Powers, 1979), geophysics (Hinich and Clay, 1968), and economics (Hinich and Patterson, 1985). A recent tutorial on bispectrum estimation has been provided by Nikias and Raghuveer (1987). Since most realistic

measurements are corrupted by interfering noise, and in some cases dominated by noise, application of the bispectrum to experimental measurements of a time series requires an understanding of the effects of interfering noise on the bispectral measurements.

In this paper we are only concerned with the effects of noise on the ability of the bispectrum to detect the presence of a non-Gaussian time series. The effects of noise on the accuracy of a bispectrum estimate is not directly addressed. Using the bispectrum as a test for Gaussianity of a time series has been described by Hinich (1982) for the signal only case. In this paper the ability to test for Gaussianity of a time series in the presence of independent noise is demonstrated as a function of signal-to-noise ratio (SNR). In Section I the problem is posed in a classical hypothesis testing framework. In Section II the bispectrum is defined, followed in Section III by development of the bispectrum test statistic to address the hypothesis test specified in Section I. The power of this test for a specific signal and noise case is demonstrated in Section IV, and compared to the power of a standard energy detection test in Section V.

I. SIGNAL MODEL

Suppose that we observe a signal $\{s(t)\}$ plus noise $\{n(t)\}$, where the $\{s(t)\}$ is a non-Gaussian (possibly nonlinear) stationary random process and $\{n(t)\}$ is in general also a non-Gaussian stationary process. Let $s(t)$ have a power spectral density $S_s(f)$ and bispectrum $B_s(f_j, f_k)$. Similarly, let the noise $n(t)$ have a power spectral density $S_n(f)$ and bispectrum $B_n(f_j, f_k)$. We will use a classical hypothesis testing framework to

determine the detectability of this non-Gaussian signal in additive noise from a finite sample of the input signal.

Let $\{y(t)\}$ denote the input signal that we observe for a finite time using an equally spaced discrete-time sampling method that avoids any aliasing. We use the standard convention that t is an integer, which means that the sampling interval is set equal to one. Under the null hypothesis that there is no signal present in the observed record, $y(t)=n(t)$ for $t=1,\dots,N$ where N is the sample size. If the signal is present, $y(t)=s(t)+n(t)$. We present a test of these compound hypotheses for a given false alarm probability α under the following assumptions about the signal and noise.

(1) All the moment functions exist for the signal and noise.

(2) The signal and noise are M -dependent. That is, for any $t_1,\dots,t_n$, $\{s(t_1),\dots,s(t_n)\}$ and $\{s(t_1+m),\dots,s(t_n+m)\}$ are independent for $m>M$, and similarly for $n(t)$.

This is a strong form of a "short memory" assumption that is needed to obtain the asymptotic Gaussian distribution that we need. It is, however, the most intuitive of the standard mixing conditions used to prove asymptotic results. These two assumptions guarantee that the signal and noise spectrum and the polyspectra of all orders exist. We will use the estimated bispectrum of a sample of the observed $y(t)$'s to detect the non-Gaussian signal. Let us now review the definition of a bispectrum and its use as a tool to study non-Gaussian random signals.

II. THE BISPECTRUM

The bispectrum of a discrete-time signal is a periodic function in two frequency indices. Let $\{y(t)\}$ denote a real zero mean stationary

continuous-time random signal. Assume all expected values, sums, and integrals used below hold. The bicovariance function of the process is $c(u,v) = E y(t) y(t+u) y(t+v)$, which does not depend on t since the process is stationary. Its Fourier transform

$$B(f,g) = \int_{-\infty}^{\infty} c(u,v) \exp[-i2\pi(fu+gv)] du dv \quad (1)$$

is called the bispectrum. The bispectrum's symmetry lines are $f=g$, $f=h(2f=-g)$, and $g=h(2g=-f)$. Another symmetry holds since $c(u,v)$ is real, namely $B(-f,-g,-h) = B^*(f,g,h)$, where $*$ denotes complex conjugation. This symmetry yields the symmetry line $f=-g$ (Fig. 1). Thus the pointed cone $C = \{f,g: 0 \leq f, g \leq f\}$ is a principal domain of this continuous time bispectrum in the (f,g) plane.

Now consider the discrete-time sequence $\{y(n)\}$ where the sampling rate f_0 is greater than $1/2$. The bicovariance function of this sampled version of $\{y(t)\}$ is really an array $\{c(j,k): j,k=0, (+/-)1, (+/-)2, \dots\}$. Then the bispectrum of the sampled data is defined, analogous with (1), to be given by the Fourier transform in two indices:

$$B_0(f,g) = \sum_j \sum_k c(j,k) \exp[-i2\pi(f+jg)] \quad (2)$$

The sampling introduces an infinite set of parallel symmetry lines $2f+g=n$, $2f-g=n$, $f+2g=n$, and $f-2g=n$. The cone C is only cut by the symmetries $2f+g=n$, and thus the principal domain of B_0 is the triangle $\{f,g: 0 \leq f \leq 1/2, g \leq f, 2f+g \leq 1\}$ in the cone C . If the discrete-time sequence is stationary and unaliased, then the bispectrum of the sampled data can be non-zero only in the triangle defined by $\{f,g: 0 \leq f \leq 1/2, g \leq f, f+g \leq 1/2\}$ (Hinich and Wolinsky, 1988). Let us now drop the zero subscript on B to simplify the notation.

III. BISPECTRUM TEST STATISTICS

Given a sample of the signal $\{y(1), \dots, y(N)\}$, let $B^{(N)}(f_j, f_k)$ denote a consistent estimator of $B(f_j, f_k)$ for a sample of size N of $\{y(n)\}$. This estimator can be computed directly by smoothing the sample bicovariance (Rao, 1983), smoothing the sample bispectrum in the bifrequency domain (Hinich, 1982), or by dividing the sample into pieces and averaging the piecewise sample bispectra and then doing bifrequency smoothing (Lii and Rosenblatt, 1982). Parametric methods can also be used to estimate the bispectrum (see Nikias and Raghuveer, 1987).

Let Δ_N denote the bandwidth of the bispectrum for the smoothing method used. For example, if the sample of length N is divided into L pieces and the L piecewise sample bispectra are averaged and the result is then smoothed in bifrequency over a square whose sides are of length M , then $\Delta_N = \frac{LM}{N}$. Assume (to satisfy consistency) that $\Delta_N = O(1/\sqrt{N})$. Under some short memory restriction which holds for the M -dependence stated in assumption (2), Brillinger and Rosenblatt (1967) or Rosenblatt (1985) show that for large N , the distribution of

$$\Delta_N^{1/2} [B^{(N)}(f_j, f_k) - B(f_j, f_k)] / [S_Y(f_j) S_Y(f_k) S_Y(f_{j+k})]^{1/2} \quad (3)$$

is approximately complex Gaussian $N_c(0, 1)$ where S_Y is the spectrum of the observed signal. This approximation is of order $O(1/N)$. Thus if H_0 is true, $\Delta_N^{1/2} B^{(N)}(f_j, f_k) / [S_Y(f_j) S_Y(f_k) S_Y(f_{j+k})]^{1/2}$ is approximately a complex Gaussian variate whose mean is $B_N(f_j, f_k)$ and variance is 1.

Another equally important large sample result that follows from the asymptotic results developed by Brillinger and Rosenblatt is that these statistics can be treated as independent random variables over the grid in the principal domain if the grid width is larger than or equal to the

bandwidth, i.e., $B^{(N)}(f_j, f_k)$ and $B^{(N)}(f_r, f_s)$ are independent for $j \neq k$ or $k \neq s$ if $|f_{j+1} - f_j| \geq \Delta_N$ or $|f_{s+1} - f_s| \geq \Delta_N$.

The asymptotic independence and Gaussianity imply that the statistic

$$CH(f_j, f_k) = 2\Delta_N^2 |B^{(N)}(f_j, f_k) - B_n(f_j, f_k)|^2 / S_Y(f_j) S_Y(f_k) S_Y(f_{j+k}) \quad (4)$$

is approximately a central chi-square variate with 2 degrees of freedom if H_0 is true. If the signal is present and its bispectrum B_S is nonzero, the $CH(f_j, f_k)$ is approximately a noncentral chi-square statistic with 2 degrees of freedom and noncentrality parameter

$$\Lambda(j, k) = 2\Delta_N^2 |B_S(f_j, f_k)|^2 / S_Y(f_j) S_Y(f_k) S_Y(f_{j+k}) \quad (5)$$

Artificial data analyses indicate that this approximation holds for samples as small as $N=256$ if $\Delta_N \approx 1/\sqrt{N}$ (Ashley, Patterson, and Hinich, 1986).

If in expression (4) we replace $B_n(f_j, f_k)$ with $B_n^{(N)}(f_j, f_k)$, (assuming that we have a noise-only sample of the data) and the spectrum S_Y at the frequencies $\{f_k\}$ with a consistent spectrum estimator denoted $S_Y^{(N)}$ whose rms error is $O(1/N)$, we then have a statistic that is approximately chi-square 2 under the null hypothesis. Summing these approximately chi-square statistics for all bifrequency pairs in the principal domain yields a statistic whose distribution is approximately chi-square with $2K$ degrees of freedom, where K is the number of bifrequency pairs for the nonzero bispectral values in the principal domain (approximately $N^2/16$). For $\Delta_N \approx 1/\sqrt{N}$, K is given approximately by

$$K \approx \left(\frac{1}{4\Delta_N} \right)^2 \quad (6)$$

The detection test is to reject H_0 if the test statistic

$$TCH = 2\Delta_N^2 \sum_{(j,k)} |B^{(N)}(f_j, f_k) - B_n^{(N)}(f_j, f_k)|^2 / S_Y^{(N)}(f_j) S_Y^{(N)}(f_k) S_Y^{(N)}(f_{j+k}) \quad (7)$$

is greater than a threshold T_α where $\Pr(\chi_{2K}^2 > T_\alpha) = \alpha$ and T_α can be determined from tables of the χ^2 distribution. The summation over the (j,k) is over the support set of the bispectrum given by $\{j,k: 0 \leq j \leq 1/2, k \leq j, j+k \leq 1/2\}$. The false alarm probability is approximately α since the distribution of TCH is approximately that of a χ_{2K}^2 variate for large N . Note that the detection test requires an estimate of the noise-only bispectrum and an estimate of the spectrum of the data (or a priori knowledge of the two). The statistical power of this test to detect a non-Gaussian signal is a function of the noncentrality parameter Λ which is the sum over (j,k) in the principal domain of the $\Lambda(j,k)$ given by (5). We discuss the power of the test in the next section.

IV. POWER OF THE TEST

From (5), the noncentrality parameter for the test statistic is

$$\Lambda = 2N\Delta_N^2 \sum_{(j,k)} \Gamma_s(f_j, f_k) / (1+\rho^{-1}(f_j)) (1+\rho^{-1}(f_k)) (1+\rho^{-1}(f_{j+k})) \quad , \quad (8)$$

where

$$\Gamma_s(f_j, f_k) = |B_s(f_j, f_k)|^2 / S_s(f_j) S_s(f_k) S_s(f_{j+k}) \quad (9)$$

and $\rho(f) = S_s(f) / S_n(f)$ is the signal-to-noise power ratio at the frequency f . $\Gamma_s(f_j, f_k)$ is called the skewness function of the signal. Let us define the weighted average skewness of the signal as Γ_s , where

$$\Gamma_s = \Delta_N^2 \sum_{(j,k)} \Gamma_s(f_j, f_k) w(f_j) w(f_k) w(f_{j+k}) \quad , \quad (10)$$

where $w(f) = (\rho(f) + \rho(f)\rho^{-1}) / (\rho(f) + 1)$ and ρ is an average signal-to-noise ratio (SNR) across all the frequencies. The function $w(f)$ weights the skewness function of the signal by the SNR at each of the three

frequencies f_j , f_k , and f_{j+k} relative to the average SNR. If $\rho(f)=\rho$, then $w(f)=1$. Thus the noncentrality parameter is

$$\Lambda = \frac{2N}{(1+\rho^{-1})^3} \Gamma_S \quad (11)$$

It is clear from (11) that Λ is linear in Γ_S and approximately cubic in ρ (for low SNR). It is also clear that the power of the test is dependent on the skewness of the signal but independent of the noise skewness, assuming that it is known or can be estimated.

The probability of detection is the probability that the detection statistic given by (7) will exceed the threshold T_α , where the detection statistic under the alternative hypothesis is noncentral chi-square with $2K$ degrees of freedom and noncentrality parameter given by (11). We will now examine the probability of detection as a function of average SNR ρ , weighted average skewness Γ_S , and processing bandwidth Δ_N . In all cases, the probability of false alarm α was set to 10^{-3} .

Shown in Fig. 2 is a plot of the probability of detection as a function of average SNR for several values of the weighted average skewness Γ_S . For these cases, $N=10^4$, $\Delta_N=0.01$, and $K=625$. Since the noncentrality parameter given by Eq. 11 has a linear dependence on Γ_S and approximately a cubic dependence on ρ , at low average SNR it is necessary for Γ_S to increase by a factor of 8 to achieve a 3 dB improvement in detection performance. This behavior can be observed in Fig. 2. Also because of the cubic dependence on ρ , the detection curves exhibit a rapid increase in probability of detection from near 0 probability to near a probability of 1 over a small SNR range of only 4-5 dB.

A slightly different display of the relationship between the weighted

average skewness and the SNR is shown in Fig. 3. In this figure, the value of the weighted average skewness needed to achieve a probability of detection of 0.5 is plotted as a function of average SNR for the same values of N , Δ_N , and K as given above. The average SNR needed to achieve a probability of detection of 0.5 is often referred to as the minimum detectable level (MDL). At low average SNR the cubic dependence can be observed in the slope of the curve.

The noncentrality parameter also has a linear dependence on N , the sample size. Thus it is necessary to increase the number of samples by a factor of 8 for a fixed value of Γ_S to achieve a 3 dB improvement in detection performance at low average SNR. Figure 4 shows a plot of N as a function of average SNR for a probability of detection of 0.5. For this case, $\Delta_N=0.01$, $K=625$, and $\Gamma_S=8$.

It can be seen from the previous figures that the bispectrum detector can detect non-Gaussian signals at low average SNR for reasonable sample sizes and reasonable (in our experience) values of the weighted average skewness. Thus it appears that the bispectrum can be used to detect a non-Gaussian signal even in the presence of quite low signal-to-noise ratios. We will now compare this performance to an energy detector.

V. COMPARISON TO ENERGY DETECTION

If we wish to detect the presence of the signal without using its non-Gaussian property, a reasonable choice for the detector would be an energy detector. Energy detection relies on the power of the signal only and does not attempt to make use of any possible non-Gaussian properties of the signal. Thus comparing the performance of the bispectrum detector

to the energy detector will serve to illustrate the effect on detection performance of utilizing the non-Gaussian property of the signal.

The noise will be assumed to be Gaussian with zero mean and variance σ_N^2 . The signal is also assumed to have zero mean and variance $\rho\sigma_N^2$, where ρ is the average SNR, as before. Since the energy detector does not make use of the non-Gaussianity of the signal, the signal will also be assumed to be Gaussian. Thus the signal plus noise is zero mean Gaussian with variance $(1+\rho)\sigma_N^2$. For the observed time series $y(t)$, $t=1,2,\dots,N$, the test statistic for the energy detector is

$$TE = \sum_{t=1}^N y^2(t) / \sigma_N^2 \quad . \quad (12)$$

The energy detector requires a knowledge of the variance of the noise process. If the variance of the noise is not known a priori, it can be replaced by its sample estimate, assuming a sample of the noise only process is available. Under the null hypothesis of noise only, TE is a central chi-square variate with N degrees of freedom. Under the alternative hypothesis of signal plus noise, $TE/(1+\rho)$ is central chi-square N. The detection test is to reject the null hypothesis if the test statistic TE is greater than a threshold T_α , where $\Pr(\chi_N^2 > T_\alpha) = \alpha$, and α is the probability of false alarm. The probability of detection is the probability that $TE/(1+\rho)$ will exceed T_α .

Figure 5 shows the probability of detection as a function of SNR for the energy detector for $\alpha=10^{-3}$ and several values of the sample size N. Because of the linear dependence on SNR, in contrast to the cubic dependence of the bispectrum detector, the transition region from low to high probability of detection is more gradual. For a probability of detection of 0.5, Fig. 6 contains a plot showing the sample size required

to achieve detection at a specified SNR. For low SNR, the sample size is required to increase by a factor of 4 to achieve detection at a 3 dB lower SNR.

A comparison of the detection performance of the bispectrum detector and the energy detector can be made from Fig. 7. The SNR required to achieve a probability of detection of 0.5 at a false alarm rate of 10^{-3} for a specified sample size for the energy detector and the bispectrum detector is shown in this figure. The detection performance of the bispectrum detector depends not only on the sample size and the SNR, as the energy detector does, but also on the value of the weighted average skewness. Detection performance curves are given for the bispectrum detector for two values of the weighted average skewness in Fig. 7. Obviously, the larger the weighted average skewness, the better the bispectrum detector will perform. This example just demonstrates that if the signal is sufficiently non-Gaussian, the bispectrum detector can detect it at a lower SNR than the energy detector could detect a Gaussian signal with the same sample size. If the energy detector were presented with a non-Gaussian signal or non-Gaussian noise, its detection performance is likely to be even worse than the optimum results presented here (Machell and Penrod, 1989). On the other hand, the bispectrum will completely fail to detect a Gaussian signal.

VI. SUMMARY AND CONCLUSIONS

Because bispectrum processing is becoming a more widely used tool in time series analysis, it is useful to understand the behavior of the bispectrum in the presence of practical measurement problems such as interfering noise. In this paper we have focused on the ability of the

bispectrum to detect a non-Gaussian time series when that time series is corrupted by non-Gaussian (or Gaussian) noise. The dependence on weighted average skewness, sample size, and average SNR have been explicitly identified and demonstrated. It was shown that for reasonable values of weighted average skewness and sample size, the bispectrum can detect non-Gaussian signals at quite low average SNR. It was also shown that for reasonable values of skewness and sample size, the bispectrum will detect a non-Gaussian signal at lower SNR than energy detection will detect a Gaussian signal. Thus it is concluded that the bispectrum can be used effectively to detect non-Gaussian signals in the presence of interfering noise, and may perform better, depending on the degree of non-Gaussianity, than energy detection. The results presented in this paper can be used to determine under what conditions of weighted average skewness, sample size, and average SNR the bispectrum can be used to detect a non-Gaussian signal.

ACKNOWLEDGMENTS

This work was funded by Naval Air Systems Command and the Office of Naval Research. The authors would also like to acknowledge the contributions of Yen-Len Tang through the Applied Research Laboratories DoD Science and Engineering Apprenticeship Program.

REFERENCES

- Ashley, R., D. Patterson, and M. Hinich (1986). "A diagnostic test for nonlinear serial dependence in time series fitting errors," *J. Time Series Analysis*, Vol. 7, 165-178.
- Barnett, T., L. C. Johnson, P. Naitoh, N. Hicks, and C. Nute (1971). "Bispectrum analysis of electroencephalogram signals during waking and sleeping," *Science*, Vol. 172, 401-402.
- Brillinger, D.R. and M. Rosenblatt (1967). "Asymptotic Theory of Estimates of k-th Order Spectra," in Spectral Analysis of Time Series, B. Harris, ed., New York: Wiley, 153-188.
- Hasselmann, K., W. Munk, and G. MacDonald (1963). "Bispectra of ocean waves," in Time Series Analysis, M. Rosenblatt, ed., New York: Wiley, 125-139.
- Hinich, Melvin J. and Murray Wolinsky (1988). "A test for aliasing using bispectral analysis," *J. Am. Stat. Assoc.* 83(402), 499-502.
- Hinich Melvin J. and D. M. Patterson (1985). "Evidence of nonlinearity in daily stock returns," *J. Business and Economic Statistics*, Vol. 3, 69-77.

- Hinich, Melvin J. (1983). "Testing for Gaussianity and linearity of a stationary time series," J. Time Series Analysis, Vol. 3, 169-176.
- Hinich, Melvin J. and C. S. Clay (1968). "The application of the discrete Fourier transform in the estimation of power spectra, coherence, and bispectra of geophysical data," Reviews of Geophysics, Vol. 6, 347-363.
- Kim, Y. C. and E. J. Powers (1979). "Digital bispectral analysis and its applications to nonlinear wave interactions," IEEE Trans. Plasma Science, Vol. PS-7, 120-131.
- Lii, K. S. and M. Rosenblatt (1982). "Deconvolution and estimation of transfer function phase and coefficients for nonGaussian linear processes," Ann. Statistics, Vol. 10(4), 1195-1208.
- Lii, K. S., M. Rosenblatt, and C. W. Van Atta (1976). "Bispectral measurements in turbulence," J. Fluid Mechanics, Vol. 77, 45-62.
- Machell, F. W. and C. S. Penrod (1989). "Energy Detection in the Ocean Acoustic Environment," in Topics in Non-Gaussian Signal Processing, E. J. Wegman, S. C. Schwartz, and J. B. Thomas, eds., New York: Springer-Verlag, 228-235.

Nikias, Chrysostomos L. and Mysore R. Raghuveer (1987). "Bispectrum estimation: a digital signal processing framework," Proc. IEEE, Vol. 75, 869-891.

Rosenblatt, M., (1985). Stationary Sequences and Random Fields, Boston: Birkhauser.

Subba Rao, T. (1983). "The Bispectral Analysis of Nonlinear Stationary Time Series with Reference to Bilinear Time Series Models," in Handbook of Statistics, (Vol. 3), D. Brillinger and P. Krishnaiba, eds., Amsterdam: North-Holland, Chpt. 14.

LIST OF FIGURES

Figure 1 - Symmetry lines and principal domain of the continuous time bispectrum.

Figure 2 - Probability of detection as a function of SNR for the bispectrum detector.

Figure 3 - Weighted average skewness required to achieve a specified minimum detectable level for the bispectrum detector.

Figure 4 - Sample size required to achieve a specified minimum detectable level for the bispectrum detector.

Figure 5 - Probability of detection as a function of SNR for the energy detector.

Figure 6 - Sample size required to achieve a specified minimum detectable level for the energy detector.

Figure 7 - Comparison of the minimum detectable level achieved by the bispectrum detector and the energy detector.

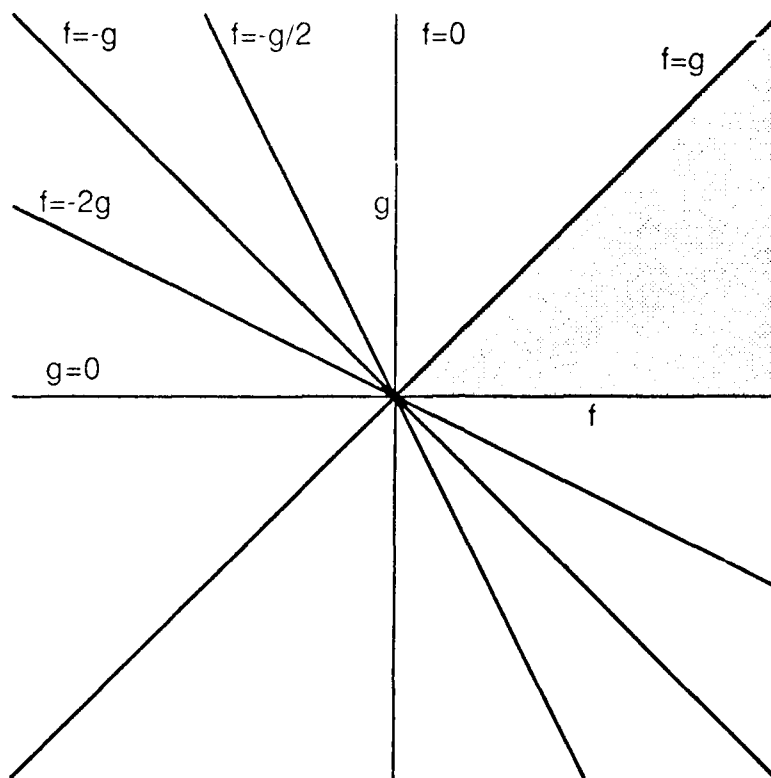


Fig. 1
Hinich/Wilson

ARL:UT
AS-87-952
MJH - GA
10-12-87

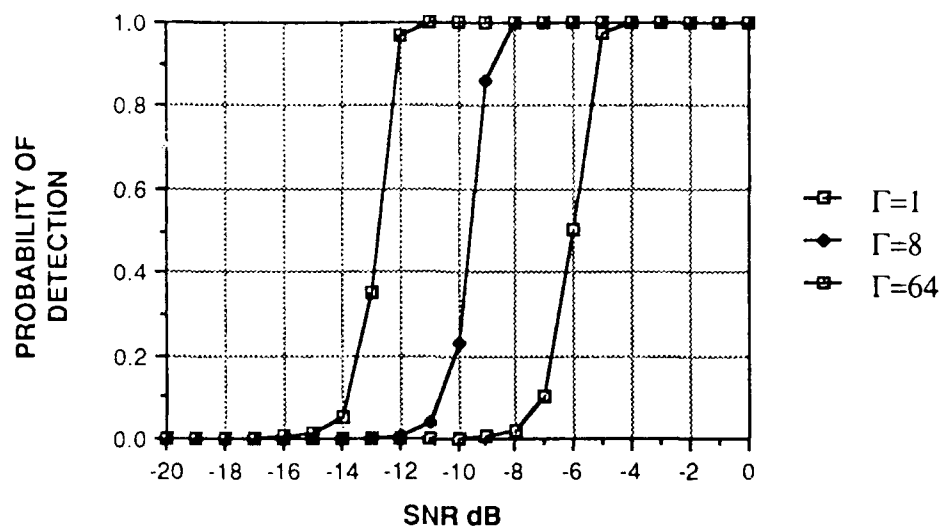


Fig. 2
Hinich/Wilson

ARL:UT
AS-87-953
MJH - GA
10-12-87

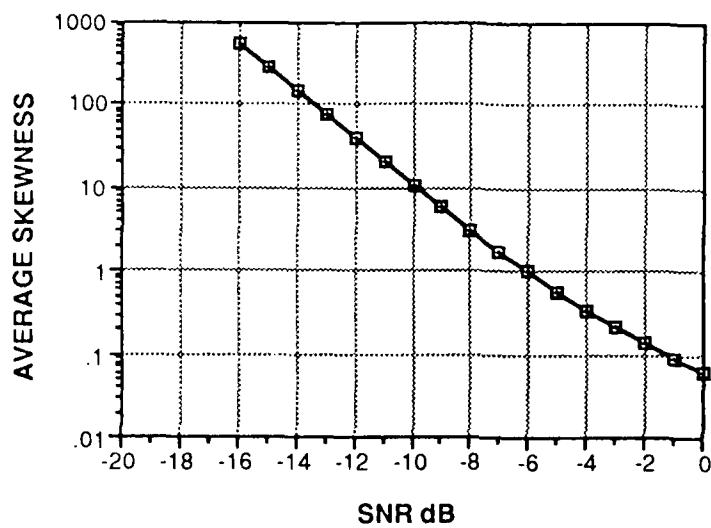


Fig. 3
Hinich/Wilson

ARL:UT
AS-87-954
MJH-GA
10-12-87

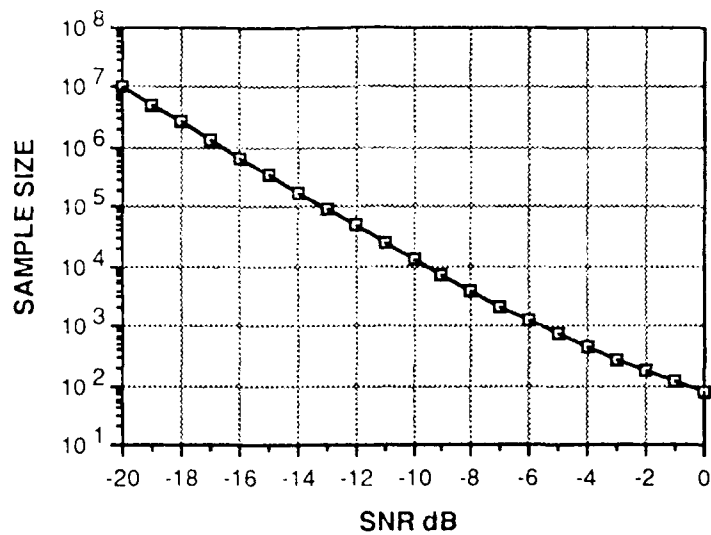


Fig. 4
Hinich/Wilson

ARL:UT
AS-87-955
MJH - GA
10-12-87

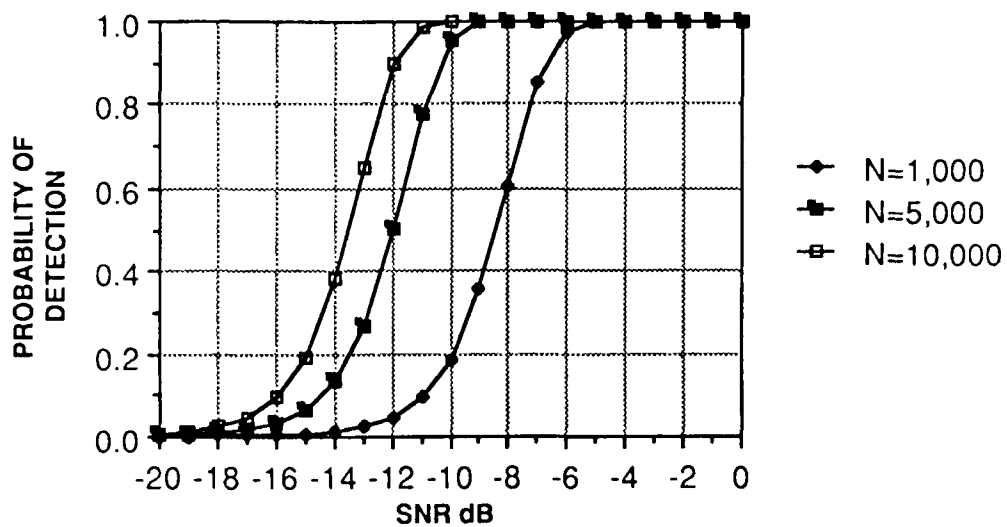


Fig. 5
Hinich/Wilson

ARL:UT
AS-87-957
MJH - GA
10-12-87

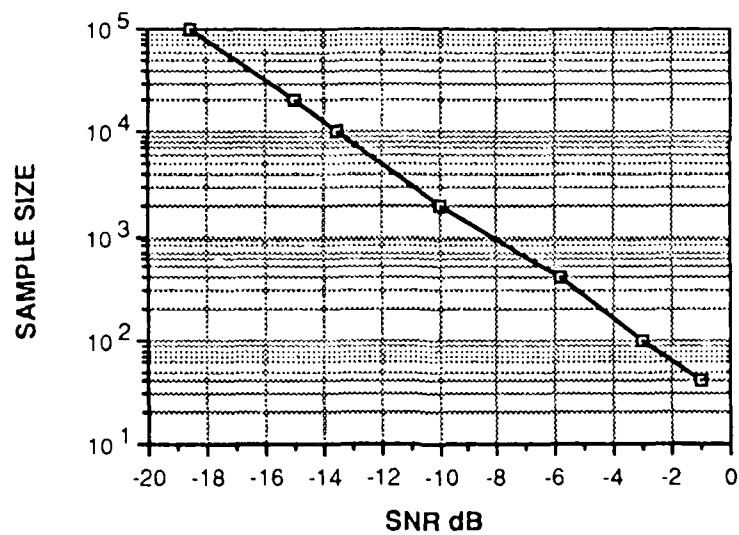


Fig. 6
Hinich/Wilson

ARL:UT
AS-87-958
MJH-GA
10-12-87

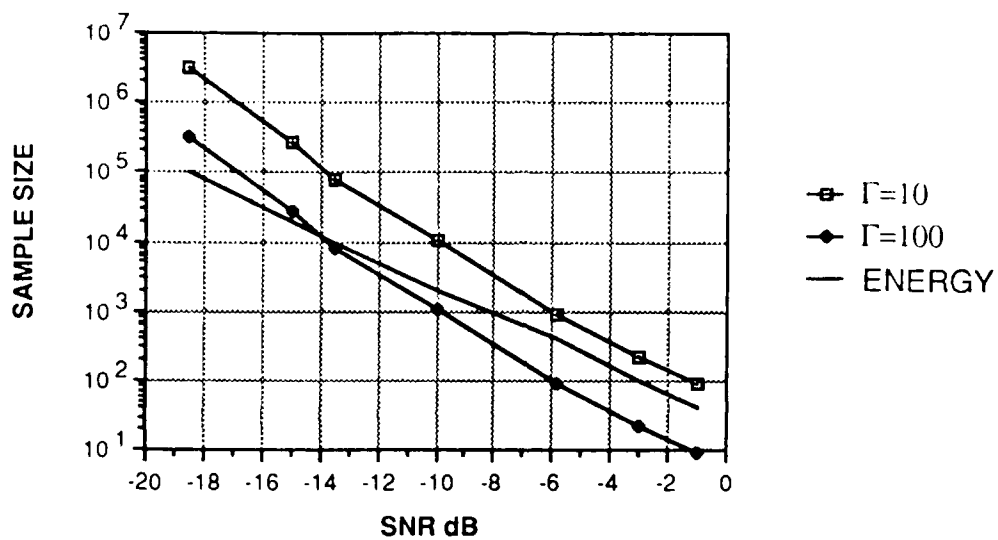


Fig. 7
Hinich/Wilson

ARL:UT
AS-87-959
MJH-GA
10-12-87

APPENDIX B

TIME DELAY ESTIMATION USING THE CROSS-BISPECTRUM (Hinich and Wilson (1989))

(This paper has been accepted for publication in
IEEE Trans. Acoust., Speech, and Signal Proc.)

Time delay estimation using the cross-bispectrum

Melvin J. Hinich and Gary R. Wilson, Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

The cross bispectrum phase can be effectively used to estimate the time required for a non-Gaussian signal to propagate between a pair of spatially separated sensors in the presence of highly correlated Gaussian noise. In this paper we present a consistent estimator of the phase of the cross bispectrum, derive the exact distribution of the phase of a complex Gaussian sample bispectrum, and show that in most cases the exact distribution can be approximated by a Gaussian distribution. Using this Gaussian approximation, we derive the variance of the time delay estimate computed from the sample cross bispectrum of a signal in additive correlated noise. This results allows the performance of time delay estimators based on the cross bispectrum phase to be quantified as a function of the sample size, the skewness of the signal, the signal-to-noise ratio (SNR), and the noise correlation.

INTRODUCTION

The estimation of the time required for a signal to propagate between two spatially separated receivers is a fundamental approach for measuring the direction of arrival of the signal relative to the axis of the sensors. Estimating direction of arrival of a propagating signal is a standard signal processing task in geophysics, acoustics, and astronomy, as well as in radar and sonar systems.

The most widely used approach to estimating time delay is to find the peak of the sample crosscorrelation function of the outputs of the two sensors (Hamon and Hannan, 1974, and Knapp and Carter, 1976). This approach works well if the signal is highly correlated and the noise is uncorrelated. In cases where the noise is correlated, ambiguous results are often produced from this crosscorrelation approach.

If the signal is a stationary NON-GAUSSIAN time series, then higher order spectra can be used to unambiguously estimate the time delay even in the presence of correlated noise. Several methods for estimating the time delay using the cross bispectrum have been presented by Nikias and Pan (1988). The non-parametric methods essentially rely upon an estimate of the phase of the sample cross bispectrum computed from the sensor outputs. The statistical properties of such estimators have to be derived.

In this paper we present a consistent estimator of the phase of the cross bispectrum, derive the exact distribution of the phase of a complex Gaussian sample bispectrum, and show that in most cases the exact distribution can be approximated by a Gaussian distribution. Using this Gaussian approximation, we derive the variance of the time delay estimate computed from the sample cross bispectrum of a signal in additive correlated noise. This results allows the performance of time delay estimators based on the cross bispectrum phase to be quantified as a function of the sample size, the skewness of the signal, the signal-to-noise ratio (SNR), and the noise correlation.

This paper is organized as follows. Section I provides an introduction to the cross-bispectrum, followed by a presentation of a statistically consistent estimator of the cross-bispectrum phase in Section II. Section III gives the derivation of the distribution of the phase estimator, and Section IV compares this distribution with an approximate distribution. Sections V and VI apply these results to the specific application of time delay estimation of a non-Gaussian signal, and Section VII includes the effects of additive Gaussian noise. Section VIII compares the cross-bispectrum time delay estimation to time delay estimation based on the cross-power spectrum, and Section IX summarizes the results of the paper.

I. THE CROSS-BISPECTRUM

Let $\{x\}$ denote a two dimensional zero mean vector random process in continuous time whose components are denoted $\{x_1\}$ and $\{x_2\}$. Assume that (1) $\{x\}$ is strictly stationary, and (2) the joint density of $[x(t_1), \dots, x(t_n)]$ for any finite sequence $(t_1, \dots, t_n)$ has bounded moments, and (3) the process has finite

memory. By finite memory we mean that there exists a time shift D such that for any $(t_1, \dots, t_n)$, $[x(t_1), \dots, x(t_n)]$ and $[x(t_1+d), \dots, x(t_n+d)]$ are independent for all $d > D$.

These three assumptions are more stringent than those used by Brillinger (1975) and Rosenblatt (1985) to prove the asymptotic properties of sample bispectra of vector processes that we use in this paper. The more restrictive assumptions that we employ are easier to understand than the more general conditions and other restrictions on the joint distributions of the process which Brillinger and Rosenblatt use.

Suppose that the two processes are simultaneously observed and sampled using a standard method for obtaining unaliased discrete-time sampled data. The processes are filtered by a bandpass filter whose effective cutoff frequency is denoted f_0 , and the series are sampled at the sampling rate $2f_0$. Let us simplify notation by setting the time interval to the sampling interval $1/2f_0$ and by using n rather than $t_n = n/2f_0$ as the n th time point.

The 1,1,2 cross-bispectrum $B_{112}(f,g)$ is defined as follows.

$$B_{112}(f,g) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} c_{112}(r,s) \exp[-i2\pi(fr+gs)] , \quad (1)$$

where

$$c_{112}(r,s) = E[x_1(n+r)x_1(n+s)x_2(n)] \quad (2)$$

is the cross-bicovariance. The following inverse Fourier integral transform of the cross-bispectrum yields the relationship

$$c_{112}(r,s) = \int_{\Omega} B_{112}(f,g) \exp(i2\pi(fr+gs)) , \quad (3)$$

where Ω is the square $\{f,g: -1/2 \leq f \leq 1/2, -1/2 \leq g \leq 1/2\}$.

The symmetry lines of this cross-bispectrum are easily derived from the Cramer representation

$$x_k(n) = \int \exp(i2\pi fn) dA_k(f) \quad \text{for } k=1,2 \quad (4)$$

where $E[dA_1(f)dA_1(g)dA_2(h)] = B_{112}(f,g)dfdg$ if $h = -f-g$, and zero otherwise. Clearly there is only one symmetry line, namely $f=g$. Since the processes are real, there is one conjugate symmetry line in the (f,g) plane, the line defined by $f+g=0$ ($h=0$). It thus follows that the principal domain PD of the 1,1,2 cross-bispectrum is the union of two triangles $T1 = \{f,g: 0 \leq f \leq 1/2, 0 \leq g \leq f\}$ and $T2 = \{f,g: 0 \leq f \leq 1/2, -f \leq g \leq 0\}$ (see Fig. 1).

The support set for the amplitude of the cross-bispectrum is a proper subset of the PD. B_{112} is identically zero in the triangle $OT = \{f,g: 0 \leq f \leq 1/2, 1/2-f \leq g \leq f\}$ by the same argument as in Hinich and Wolinsky (1988) for bandlimited signals.

II. ESTIMATING THE PHASE OF THE CROSS-BISPECTRUM

In this section we discuss a simple approach to obtaining an asymptotically unbiased and Gaussian estimator of the real and imaginary parts of B_{112} . These estimators are then transformed to yield an asymptotically unbiased and Gaussian estimator of the phase function $\phi(f,g)$, where in polar form

$$B_{112}(f,g) = |B_{112}(f,g)| \exp[i\phi(f,g)] \quad (5)$$

Let N denote the sample size for the observation of $x(n)$. Select a block length L which is approximately $\sqrt{N}$. The number of whole blocks in the record of length N is then $\lfloor N/L \rfloor$, which is also approximately $\sqrt{N}$. Let $X_{kp}(f_m)$ denote the $f_m = m/L$ term of the discrete Fourier transform of the p th piece $\{x_k(1+L^*(p-1)), \dots, x_k(L+L^*(p-1))\}$ of data ($k=1,2$). The "raw" 1,1,2 cross-bispectral estimator is the set products

$$F_p(m,n) = L^{-1} X_1(-f_m) X_1(-f_n) X_2(f_{m+n}) \quad (6)$$

for $\{f_m, f_n\}$ in the PD.

The expected value of $F_p(m, n)$ is approximately $B_{112}(f_m, f_n)$ with an error of order $o(1/L)$. Denoting the spectrum of $\{x_k(n)\}$ by $S_k(f)$, the large sample variance of the real and imaginary parts of $(2/L)^{1/2}F_p(m, n)$ is

$$[1 + \delta(m-n)]S_1(f_m)S_1(f_n)S_2(f_{m+n}) + O(L^{-1}) , \quad (7)$$

where $\delta(0)=1$ and $\delta(k)=0$ for nonzero k . These results follow from the asymptotic theory presented in Chapter 4 of Brillinger (1975). Moreover, the real and imaginary parts are Gaussian and asymptotically independent as $L \rightarrow \infty$.

The estimator S_{112} of B_{112} which we use is the average of these F_p values over the $[N/L]$ non-overlapping pieces. The resolution bandwidth of this estimator is $\Delta_N = 1/L$. If L is greater than the finite memory length D , then the variances of $\text{Re}S_{112}(f_m, f_n)$ and $\text{Im}S_{112}(m, n)$, the real and imaginary parts of $S_{112}(f_m, f_n)$, respectively, are equal to $1/[N/L]$ times the large sample variance of the real and imaginary parts of $F_p(m, n)$, which is proportional to the expression in Eq. 7. In other words, the large sample variances of the real and imaginary parts are equal to the product of the three spectra multiplied by $\sigma^2 = (2N\Delta_N^2)^{-1}$. If we assume that $L = o(\sqrt{N})$, then the bandwidth Δ_N and the large sample variance go to zero as $N \rightarrow \infty$. Moreover, the real and imaginary parts of S_{112} are asymptotically Gaussian and independent. We will assume that N and L are sufficiently large and the processes are sufficiently well behaved, that we may apply these asymptotic results as large sample approximations.

There is one more large sample approximation that plays an essential part in the application of statistical polyspectra theory. Returning to the raw cross spectral values, $F_p(f_m, f_n)$ and $F_p(f_r, f_s)$ are approximately INDEPENDENT for large L . This implies that the same holds for the S_{112} at different bifrequencies.

Given these estimators of S_{112} , the obvious estimator of the phase function at the Fourier frequencies (f_m, f_n) is

$$\Phi(f_m, f_n) = \tan^{-1} \left[S_I(m, n) / S_R(m, n) \right] , \quad -\frac{\pi}{2} \leq \Phi(f_m, f_n) \leq \frac{\pi}{2} , \quad (8)$$

where we now use the convenient abbreviations $S_R(m, n)$ and $S_I(m, n)$ for the real and imaginary parts, respectively, of the ESTIMATOR of the 1,1,2 cross-bispectrum. We will use B_R and B_I as abbreviations of the real and imaginary parts of the TRUE cross-bispectrum.

III. THE DISTRIBUTION OF THE PHASE ESTIMATOR

For each (m, n) pair, $S_R(m, n)$ and $S_I(m, n)$ are asymptotically Gaussian and independent with expected values $B_R(m, n)$ and $B_I(m, n)$ respectively. We will now derive the exact distribution of the phase estimator when they are exactly Gaussian. It can be shown (Thomas, 1969) that the probability density function of $\Phi(f_m, f_n)$ given by Eq. 8 is

$$f(\Phi) = \frac{1}{\pi} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} + \frac{\Gamma_{112}}{\sqrt{2\pi}\sigma} \cos(\Phi - \theta) e^{-\frac{\Gamma_{112}}{2\sigma^2} \sin^2(\Phi - \theta)} \operatorname{erf} \left(\frac{\Gamma_{112} \cos(\Phi - \theta)}{\sqrt{2}\sigma} \right) , \quad (9)$$

where the dependence on f_m and f_n has been dropped for compactness of notation. The function $\Gamma_{112}(f_m, f_n)$ is called the cross-skewness function and is given by

$$\Gamma_{112}(f_m, f_n) = \frac{|B_{112}(f_m, f_n)|}{\sqrt{S_1(f_m)S_1(f_n)S_2(f_{m+n})}} . \quad (10)$$

The parameter θ is just the true phase of the cross-bispectrum and is given by

$$\theta(f_m, f_n) = \tan^{-1} \left(\frac{B_I(f_m, f_n)}{B_R(f_m, f_n)} \right) . \quad (11)$$

It can be seen from Eq. 9 that if Γ_{112} is zero, the phase estimator is uniformly distributed, as expected. On the other hand, if Γ_{112} is large, then the second term in the sum in Eq. 9 dominates and the phase estimator has an approximately Gaussian distribution for Φ near θ .

The variance of the phase estimator is given by the following. Center Φ by replacing $\Phi - \theta$ in Eq. 9 with Φ . Then,

$$E\{\Phi^2\} = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \Phi^2 f(\Phi) d\Phi \quad (12)$$

$$= \frac{\pi^2}{12} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} + \frac{2}{\pi} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} \sum_{j=1}^{\infty} \frac{2^j \left(\frac{\Gamma_{112}^2}{2\sigma^2}\right)^j}{(2j-1)!!} \int_0^{\frac{\pi}{2}} \Phi^2 \cos^{2j}\Phi d\Phi,$$

where the series expansion for the error function has been used to rewrite the density function given by Eq. 9 in the equivalent form:

$$f(\Phi) = \frac{1}{\pi} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} \left[1 + \sum_{j=1}^{\infty} \frac{2^j \left(\frac{\Gamma_{112}^2}{2\sigma^2}\right)^j}{(2j-1)!!} \cos^{2j}\Phi \right]. \quad (13)$$

By repeated substitution, the remaining integral can be evaluated as

$$\int_0^{\frac{\pi}{2}} \Phi^2 \cos^{2j}\Phi d\Phi = \frac{(2j-1)!!}{(2j)!!} \frac{\pi}{2} \left[\frac{\pi^2}{12} - \frac{1}{2} \sum_{k=1}^j \frac{1}{k^2} \right]. \quad (14)$$

The expression $(2j-1)!! = 1 \cdot 3 \cdot 5 \cdots (2j-1)$ and $(2j)!! = 2 \cdot 4 \cdot 6 \cdots (2j)$. Note that $(2j)!! = 2^j(j!)$. Substituting this equation into Eq. 12 and making use of the series expansion for the exponential gives for the variance

$$E\{\Phi^2\} = \frac{\pi^2}{12} - \frac{1}{2} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} \sum_{j=1}^{\infty} \frac{\left(\frac{\Gamma_{112}^2}{2\sigma^2}\right)^j}{j!} \sum_{k=1}^j \frac{1}{k^2}. \quad (15)$$

This is the exact expression for the variance of the phase estimator using the asymptotic distribution of the cross-bispectrum. For $\Gamma_{112} = 0$ the variance is $\frac{\pi^2}{12}$, as it should be for a uniform distribution. Since

$$\sum_{k=1}^j \frac{1}{k^2} \leq \frac{\pi^2}{6}$$

for all j , then

$$\sum_{j=1}^{\infty} \frac{\left(\frac{I_{112}^2}{2\sigma^2} \right)^j}{j!} \sum_{k=1}^j \frac{1}{k^2} \leq e^{\frac{\Gamma_{112}^2}{2}} (\pi^2/6) ,$$

which implies that $E\{\Phi^2\} \geq 0$, as it should.

The variance of the phase estimator under the Gaussian approximation can be determined by the following.

Let e_R and e_I denote the random error in S_R and S_I , i.e.,

$$S_R(m,n) = B_R(m,n) + e_R(m,n) , \quad S_I(m,n) = B_I(m,n) + e_I(m,n) . \quad (16)$$

Expanding Eq. 8, we have

$$\begin{aligned} \tan^{-1}(S_I/S_R) - \tan^{-1}(B_I/B_R) &= (1+B_I^2/B_R^2)^{-1} (B_R^{-1}e_I - B_R^{-2}B_Ie_R) + o(\sigma) \\ &= (B_R^2+B_I^2)^{-1} (B_Re_I - B_Ie_R) + o(\sigma) , \end{aligned} \quad (17)$$

where $\sigma^2 = (2N\Delta^2N)^{-1}$ is the variance of e_R and e_I . Thus the large sample variance of $\Phi(f_m, f_n)$ for $m \neq n$ is

$$\text{Var } \Phi(f_m, f_n) = \sigma^2 \Gamma_{112}^{-2}(f_m, f_n) . \quad (18)$$

When $m=n$, the large sample variance is $2\sigma^2 \Gamma_{112}^{-2}(f_m, f_n)$.

The important feature to note is that the statistics of the phase estimator are determined by the cross-skewness function and the sample size, and, in particular, the variance of the phase estimator is inversely proportional to the square of the cross-skewness function.

The distribution function of the phase estimator is given by

$$\begin{aligned}
 F_{\Phi}(\phi) &= \int_{-\frac{\pi}{2}}^{\phi} f(\Phi) d\Phi \\
 &= \int_{-\frac{\pi}{2}}^{\phi} \frac{1}{\pi} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} \left[1 + \sum_{j=1}^{\infty} \frac{2^j \left(\frac{\Gamma_{112}^2}{2\sigma^2} \right)^j}{(2j-1)!!} \cos^{2j} \Phi \right] d\Phi \\
 &= \frac{1}{\pi} e^{-\frac{\Gamma_{112}^2}{2\sigma^2}} \left[\left(\phi + \frac{\pi}{2} \right) + \sum_{j=1}^{\infty} \frac{2^j \left(\frac{\Gamma_{112}^2}{2\sigma^2} \right)^j}{(2j-1)!!} \int_{-\frac{\pi}{2}}^{\phi} \cos^{2j} \Phi d\Phi \right] \\
 &= \frac{1}{2} + \frac{\phi}{\pi} + \sin \phi \sum_{j=1}^{\infty} \frac{(\Gamma_{112}^2)}{2^j (2j-1)!!} \left\{ \cos \phi^{2j-1} + \sum_{k=1}^{j-1} \frac{(2j-1)(2j-3) \dots (2j-2k+1)}{2^k (j-1)(j-2) \dots (j-k)} \cos \phi^{2j-2k-1} \right\}
 \end{aligned} \tag{19}$$

IV. COMPARISON OF THE EXACT AND APPROXIMATE DISTRIBUTION OF THE PHASE ESTIMATOR

A Q-Q plot of the exact distribution of the phase estimator given by Eq. 19 and its Gaussian approximation with variance $(\Gamma_{112}/\sigma)^{-2}$ is shown in Fig. 2 for each of three different values of the ratio Γ_{112}/σ . For $\Gamma_{112}/\sigma=1$, the exact distribution has smaller tails than the Gaussian distribution, but as Γ_{112}/σ increases the tails become larger than the Gaussian distribution before converging to the normal distribution. A relatively small value of Γ_{112}/σ results in good agreement between the exact and approximate distributions.

Figure 3 shows a plot of the percent difference between the exact and approximate distributions as a function of Γ_{112}/σ for three values of the distribution near the tail. Again, good agreement is achieved at relatively small values of Γ_{112}/σ .

The percent difference between the exact variance given by Eq. 15 and approximate variance given by $(\Gamma_{112}/\sigma)^{-2}$ is shown in Fig. 4 as a function of the ratio Γ_{112}/σ . For small values of the ratio, the percent difference is positive, indicating that the exact distribution has smaller tails, as was shown in Fig. 2. As the ratio gets larger, the tails become larger than the normal and the percent difference becomes negative.

V. LINEAR RELATIONSHIP BETWEEN THE SERIES

Suppose that $\{x_2(n)\}$ is the output of a linear filter applied to $\{x_1(n)\}$ with impulse response $\{h(n)\}$, i.e., $x_2(n) = \sum_m h(m)x_1(n-m)$. We only assume that the filter is stable. Then $dA_2(f) = H(f)dA_1(f)$ where $H(f)$ is the filter's transfer function. Thus $S_2(f) = |H(f)|^2 S_1(f)$ and

$$\begin{aligned} B_{112}(f,g)dfdg &= E[dA_1(f)dA_1(g)dA_1(-f-g)] \\ &= H(-f-g)B_{111}(f,g)dfdg \end{aligned} \quad (20)$$

where B_{111} is the bispectrum of $\{x_1(n)\}$. Because of the linear relationship between X_1 and X_2 , the regions T1 and T2 in the cross-bispectrum (see Fig. 1) contain the same information. In the T2 region, Eq. 20 becomes

$$B_{112}(f,-g) = H(g-f)B_{111}(f-g,g) .$$

Thus it is only necessary to use the T1 region to estimate time delay. Applying Eq. 20 and the above result for the spectrum of x_2 to Eq. 10, $\Gamma_{112} = \Gamma_{111}$ where Γ_{111} is the skewness function of x_1 . In other words, the cross-skewness is the skewness of each process if the two are linearly related. Thus from Eq. 18, the large sample variance of the phase is

$$\text{Var } \Phi(f_m, f_n) = \sigma^2 \Gamma_{111}^{-2}(f_m, f_n) . \quad (21)$$

VI. TIME DELAY ESTIMATION

Suppose further that the linear relationship is given by

$$x_1(t) = x_2(t - \tau) = \begin{cases} x_2(t - \tau) & \text{if } t \geq \tau \\ 0 & \text{if } t < \tau \end{cases} \quad (22)$$

i.e., x_1 is just a time delayed version of x_2 . Then,

$$B_{112}(f_m, f_n) = B_{111}(f_m, f_n) e^{i2\pi(f_m + f_n)\tau} \quad (23)$$

Thus B_{112} is just a phase shifted version of B_{111} :

$$\theta_{112}(f_m, f_n) = \theta_{111}(f_m, f_n) + 2\pi(f_m + f_n)\tau \quad (24)$$

The time delay can thus be estimated as the scaled difference between the phase estimates of the cross and auto bispectra:

$$\hat{\tau} = \frac{\Phi_{112}(f_m, f_n) - \Phi_{111}(f_m, f_n)}{2\pi(f_m + f_n)} \quad (25)$$

In practice, this time delay estimate can be improved by averaging over frequency pairs that have large bispectral values or other equivalent averaging methods (Nikias and Pan, 1983).

Since the bispectra phases are asymptotically Gaussian for large skewness values, the time delay estimate is also asymptotically Gaussian. Using the Taylor series expansion for the bispectrum phase, (Eq. 17) gives for the phase differences

$$\Phi_d(f_m, f_n) - \theta_d(f_m, f_n) = \frac{-1}{2i} \left[\frac{B_{112}(f_m, f_n) e_{112}^*(f_m, f_n) - B_{112}^*(f_m, f_n) e_{112}(f_m, f_n)}{|B_{112}(f_m, f_n)|^2} \right] \quad (26)$$

$$- \frac{B_{111}(f_m, f_n) e_{111}^*(f_m, f_n) - B_{111}^*(f_m, f_n) e_{111}(f_m, f_n)}{|B_{111}(f_m, f_n)|^2} \right]$$

where

$$\Phi_d(f_m, f_n) = \Phi_{112}(f_m, f_n) - \Phi_{111}(f_m, f_n) , \quad (27)$$

and

$$\theta_d(f_m, f_n) = \theta_{112}(f_m, f_n) - \theta_{111}(f_m, f_n) . \quad (28)$$

The asymptotic variance of Φ_d can now be calculated from Eq. 26 by making use of the following relationships.

$$E \left\{ |e_{111}(f_m, f_n)|^2 \right\} = \text{VAR} \{ S_{111}(f_m, f_n) \} = 2\sigma^2 S_1(f_m) S_1(f_n) S_1(f_{m+n}) + O\left(\frac{L}{N}\right) , \quad (29)$$

$$\begin{aligned} E \left\{ |e_{112}(f_m, f_n)|^2 \right\} &= \text{VAR} \{ S_{112}(f_m, f_n) \} \\ &= 2\sigma^2 S_1(f_m) S_1(f_n) S_2(f_{m+n}) + O\left(\frac{L}{N}\right) , \end{aligned} \quad (30)$$

$$\begin{aligned} E \{ e_{111}^*(f_m, f_n) e_{112}(f_m, f_n) \} &= \text{cov} \{ S_{111}^*(f_m, f_n), S_{112}(f_m, f_n) \} \\ &= 2\sigma^2 S_1(f_m) S_1(f_n) S_{12}(f_{m+n}) + O\left(\frac{L}{N}\right) , \end{aligned} \quad (31)$$

$$\begin{aligned} E \{ e_{112}^*(f_m, f_n) e_{111}(f_m, f_n) \} &= \text{cov} \{ S_{112}^*(f_m, f_n), S_{111}(f_m, f_n) \} \\ &= 2\sigma^2 S_1(f_m) S_1(f_n) S_{12}^*(f_{m+n}) + O\left(\frac{L}{N}\right) , \end{aligned} \quad (32)$$

where $S_{12}(f)$ is the cross-power spectrum.

The variance of Φ_d is then

$$\text{VAR} \{ \Phi_d(f_m, f_n) \} = \frac{\sigma^2}{\Gamma_{112}^2(f_m, f_n)} + \frac{\sigma^2}{\Gamma_{111}^2(f_m, f_n)} \quad (33)$$

$$- 2 \operatorname{Re} \left\{ \frac{B_{112}(f_m, f_n)}{|B_{112}(f_m, f_n)|} \frac{B_{111}(f_m, f_n)}{|B_{111}(f_m, f_n)|} \frac{\sigma^2}{G_{112}(f_m, f_n)} \right\},$$

where

$$G_{112}(f_m, f_n) = \frac{|B_{112}(f_m, f_n)| |B_{111}(f_m, f_n)|}{S_1(m)S_1(n)S_{12}(m+n)} \quad (34)$$

This expression for the variance of the phase difference holds without making the time-delay restriction implied by Eq. 22.

VII. EFFECTS OF NOISE

Now suppose that we observe the two processes with additive Gaussian noise processes that are independent of the signals. Then the cross-bispectrum of the signals plus noise is the SAME as that of the signals only. This result holds if certain cross-bispectra of the signals and noise are zero, but we will stick with the stronger independence assumption to simplify exposition.

If we denote the ratio of the signal and noise power spectra of the i th channel by $\rho_i(f)$, then the variance of the phase difference for the signal plus noise case is

$$\begin{aligned} \operatorname{VAR}\{\Phi_d^{S+N}(f_m, f_n)\} &= \frac{\sigma^2}{I_{112}^2(f_m, f_n)} - [1 + \rho_1^{-1}(f_m)][1 + \rho_1^{-1}(f_n)][1 + \rho_2^{-1}(f_{m+n})] \\ &+ \frac{\sigma^2}{I_{111}^2(f_m, f_n)} [1 + \rho_1^{-1}(f_m)][1 + \rho_1^{-1}(f_n)][1 + \rho_1^{-1}(f_{m+n})] \end{aligned} \quad (35)$$

$$- 2 \operatorname{Re} \left\{ \frac{B_{112}(f_m, f_n)}{|B_{112}(f_m, f_n)|} \frac{B_{111}(f_m, f_n)}{|B_{111}(f_m, f_n)|} \frac{\sigma^2}{G_{112}(f_m, f_n)} [1 + \rho_1^{-1}(f_m)][1 + \rho_1^{-1}(f_n)] \left[\frac{W_{12}^{(N)}(f_{m+n})}{W_{12}^{(S)}(f_{m+n})} \sqrt{\rho_1^{-1}(f_{m+n})\rho_2^{-1}(f_{m+n})} \right] \right\}.$$

$W_{12}(S)$ and $W_{12}(N)$ are the signal and noise (complex) coherency spectrum defined as

$$W_{12}(f) = \frac{S_{12}(f)}{\sqrt{S_{11}(f)S_{22}(f)}} .$$

Equation 35 is the general expression for the variance of the phase difference for signal and noise of arbitrary coherence.

If we now invoke the assumption of Eq. 22 that the signals are time delayed versions of each other and the additional assumption that the noises have approximately the same power spectral densities, then the variance of the phase difference can be simplified to

$$\text{VAR}\{\Phi_d^{S+N}(f_m, f_n)\} = \frac{2\sigma^2}{I_{111}^2(f_m, f_n)} \left\{ [1+\rho_1^{-1}(f_m)][1+\rho_1^{-1}(f_n)] \rho_1^{-1}(f_{m+n}) \left\{ 1 - \text{Re}[W_{12}^{(N)}(f_{n+m})e^{-i2\pi f_{m+n}\tau}] \right\} \right\} . \quad (36)$$

This is the expression for the variance of the phase difference for correlated noise. If the noise processes are time delayed versions of one another with delay time τ_n , then the term in the braces $\{ \}$ is just $1 - \cos(2\pi f_{m+n}(\tau_n - \tau))$, i.e., the variance is dependent upon the difference in time delays between the signal and noise. On the other hand, if the noise processes are independent of one another, then the term in the braces $\{ \}$ is 1 and the variance of the phase differences reduces in this case to

$$\text{VAR}\{\Phi_d^{S+N}(f_m, f_n)\} = \frac{2\sigma^2}{I_{111}^2(f_m, f_n)} [1+\rho_1^{-1}(f_m)][1+\rho_1^{-1}(f_n)] \rho_1^{-1}(f_{m+n}) . \quad (37)$$

Note that the asymptotic variance of the phase difference for the bifrequency pair (f_m, f_n) decreases with the square of the skewness and with (approximately) the cube of the signal-to-noise ratio. The term $\sigma^2 = (N\Delta_N^2)^{-1}$ determines the rate of convergence of the estimator to the true phase difference as $N \rightarrow \infty$.

VIII. COMPARISON TO THE CROSS-POWER SPECTRUM PHASE

The more common approach to estimating a time delay is by computing the cross-power spectrum and its inverse Fourier transform, the cross-correlation. The parallels to Eqs. 23 and 25 for the cross-power spectrum are

$$S_{12}(f) = S_1(f)e^{i2\pi ft} \quad (38)$$

and

$$T = \frac{\Phi_{12}(f)}{2\pi f}, \quad (39)$$

where $\Phi_{12}(f)$ is the estimate of the phase of the cross-power spectrum. If the cross-power spectrum is estimated in a manner similar to the bispectrum estimate presented in Section II, and if the cross-power spectrum phase is estimated as the arctangent of the ratio of the imaginary to the real parts of the cross-power spectrum, then from the Taylor series expansion of the arctangent (similar to Eq. 17) it can be shown that the variance of the cross-power spectrum phase is

$$\text{VAR } \Phi_{12}(f) = \frac{\sigma_{p_{12}}^2}{|W_{12}(f)|^2} \quad (40)$$

where $\sigma_{p_{12}}^2 = (2N\Delta_N)^{-1}$.

For the case of signal-plus-noise, the variance is given by

$$\text{VAR } \Phi_{12}^{S+N}(f) = \quad (41)$$

$$\sigma_{p_{12}}^2 \left[\frac{|W_{12}^S|^2}{(1+\rho_1^{-1}(f))(1+\rho_2^{-1}(f))} + \frac{|W_{12}^N|^2}{(1+\rho_1(f))(1+\rho_2(f))} + \frac{W_{12}^S(W_{12}^N)^* + (W_{12}^S)^* W_{12}^N}{\sqrt{(1+\rho_1^{-1}(f))(1+\rho_2^{-1}(f))(1+\rho_1(f))(1+\rho_2(f))}} \right]^{-1}$$

This is analogous to Eq. 35 for the cross-bispectrum.

If the signal is assumed to be perfectly coherent and the noises have approximately the same power spectral densities, then in analogy to Eq. 36, the variance of the cross-power spectrum phase simplifies to

$$\text{VAR } \Phi_{12}^{S+N}(f) = \sigma_{p,2}^2 \left[\frac{1}{(1+\rho_1^{-1}(f))^2} + \frac{|W_{12}^N|^2}{(1+\rho_1(f))^2} + \frac{(W_{12}^N)^* + W_{12}^N}{(1+\rho_1^{-1}(f))(1+\rho_1(f))} \right]^{-1} \quad (42)$$

Finally, if the noises are incoherent, then the variance further reduces to

$$\text{VAR } \Phi_{12}^{S+N}(f) = \sigma_{p,2}^2 (1+\rho_1^{-1}(f))^2 \quad (43)$$

This is to be compared with Eq. 37 for the cross-bispectrum phase difference.

The question that now arises is in the case of uncorrelated noise, are there conditions under which the cross-bispectrum time delay estimation method will provide a more accurate estimation of the time delay than the cross-power spectrum method. This essentially involves a comparison of Eqs. 37 and 43. Note that the variance of the cross-bispectrum phase difference is approximately (for low signal-to-noise ratios) inversely proportional to the cube of the signal-to-noise ratio, whereas for the cross-power spectrum phase the variance is inversely proportional to the square of the signal-to-noise ratio. This implies that for low signal-to-noise ratios the variance will increase more rapidly for the cross-bispectrum phase difference than for the cross-power spectrum phase. However, the variance of the cross-bispectrum phase difference is also inversely proportional to the square of the skewness function of the signal, implying that larger values of the skewness function can partially offset the more pronounced effect that lower signal-to-noise ratios have on the cross-bispectrum phase difference.

Sample size also contributes differently to the two variances. For the cross-bispectrum phase difference, $\sigma^2 = (2N\Delta^2N)^{-1}$, whereas for the cross-power spectrum phase, $\sigma_{p,2}^2 = (2N\Delta N)^{-1}$. If ΔN is chosen to be on the order of $N^{-1/2}$ in both cases, then $\sigma^2 \sim O(1)$ while $\sigma_{p,2}^2 \sim O(N^{-1/2})$. Thus the variance will

reduce more rapidly as sample size increases for the cross-power spectrum phase than for the cross-bispectrum phased difference.

Shown in Fig. 5 are plots of the variances of the cross-bispectrum phase difference and cross-power spectrum phase as a function of signal-to-noise ratio. For Eq. 37, the signal-to-noise ratios are taken to be the same at the three frequencies, and Γ^2 was set to 1. For both Eqs. 37 and 43, σ^2 was set to 1. For this case, the comparison shown in Fig. 5a demonstrates that for low signal-to-noise ratios the cross-bispectrum phase difference can have a much larger variance than the cross-power spectrum phase due to its cubic dependency on signal-to-noise ratio. However, at high signal-to-noise ratios its variance is slightly smaller than the variance of the cross-power spectrum phase.

In interpreting these results, it should be kept in mind that the variance of the cross-bispectrum phase difference is inversely proportional to Γ^2 , so that larger values of Γ^2 than presented in Fig. 5 will reduce its variance. On the other hand, σ^2 and $\sigma_{p_{12}}^2$ were both set to 1 for the two cases, even though $\sigma_{p_{12}}^2$ can be made to reduce more rapidly than σ^2 and still maintain consistency as sampling size increases. Thus an appropriate choice of averaging can result in a smaller cross-power spectrum phase variance than presented in Fig. 5. Clearly, for low signal-to-noise ratios and uncorrelated noise, the cross-bispectrum will not usually provide any performance advantage. The obvious advantage of the cross-bispectrum is when correlated Gaussian noise is present, which results in a bias for the cross-power spectrum.

VII. SUMMARY

In this paper we have derived the statistical properties of a consistent time delay estimator for non-Gaussian signals in correlated noise. We have shown the dependence of the performance of the time delay estimator on the skewness of the signal, the correlation of the noise, the signal-to-noise ratio, and the sample size. The asymptotic variance of the time delay estimator is inversely proportional to the square of the skewness and approximately inversely proportional to the cube of the signal-to-noise ratio.

ACKNOWLEDGEMENT

The work was sponsored by the Office of Naval Research.

REFERENCES

- Brillinger, David R. (1975). Time Series Data Analysis and Theory (New York: Holt, Rinehart and Winston, Inc.
- Hamon, B. Y. and Hannan, E. J. (1974). "Spectral Estimation of Time Delay for Dispersive and Non-Dispersive Systems," App. Statist. 23 (2), 134-142.
- Hinich, M. J. and Wolinsky, M. A. (1988). "A Test for Aliasing using Bispectral Analysis," J. Am. Statist. Assoc. 83 (402), 499-502.
- Knapp, C. H. and Carter, G. C. (1976). "The Generalized Correlation Method for Estimation of Time Delay," IEEE Trans. Acoust., Speech, Signal Processing, ASSP-24, 320-327.
- Nikias, C. L. and Pan, R. (1988). "Time Delay Estimation in Unknown Gaussian Spatially Correlated Noise," IEEE Trans. ASSP 36, 1706-1714.
- Rosenblatt, M. (1985). Stationary Sequences and Random Fields (Boston: Birkhauser).
- Thomas, J. B. (1969). An Introduction to Statistical Communications Theory (New York: Wiley).

LIST OF FIGURES

- Figure 1 Principal Domain of the Continuous Time Cross-Bispectrum
- Figure 2 Q-Q Plot Comparison of the Exact and Normal Distributions of the Phase Estimator
- Figure 3 Percent Difference Between the Exact Distribution and the Gaussian Approximation of the Phase Estimator for Three Specific Quantiles
- Figure 4 Percent Difference Between the Variance of the Exact and Normal Distributions of the Phase Estimator
- Figure 5 Comparison of the Variance of Cross-Bispectrum Phase Difference and Cross Power Spectrum Phase with Uncorrelated Noise

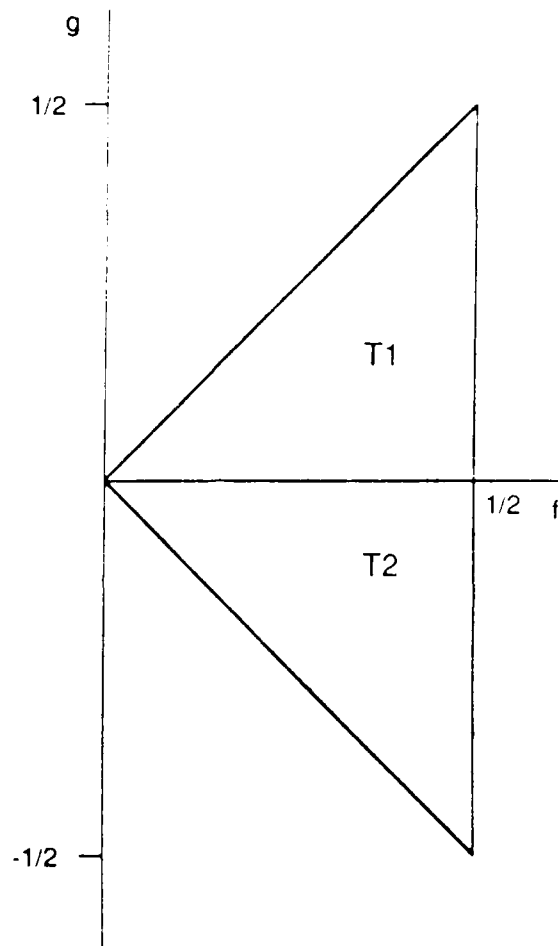


Figure 1
Wilson

ARL:UT
AS-90-594
GRW - DS
8 - 17 - 90

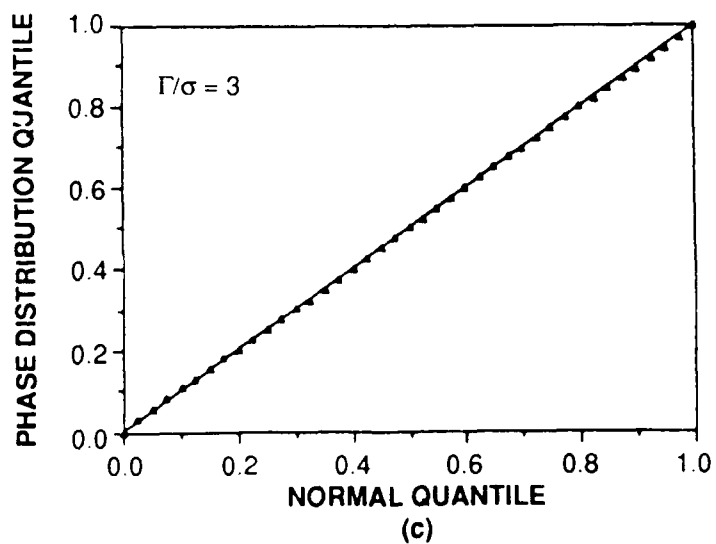
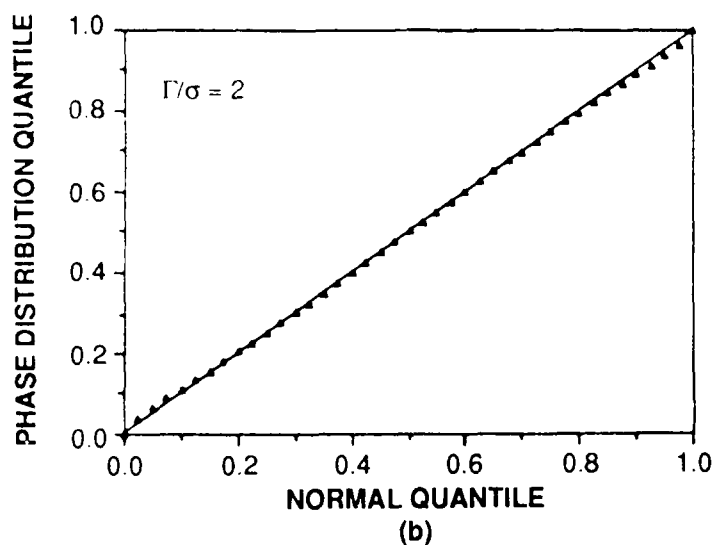
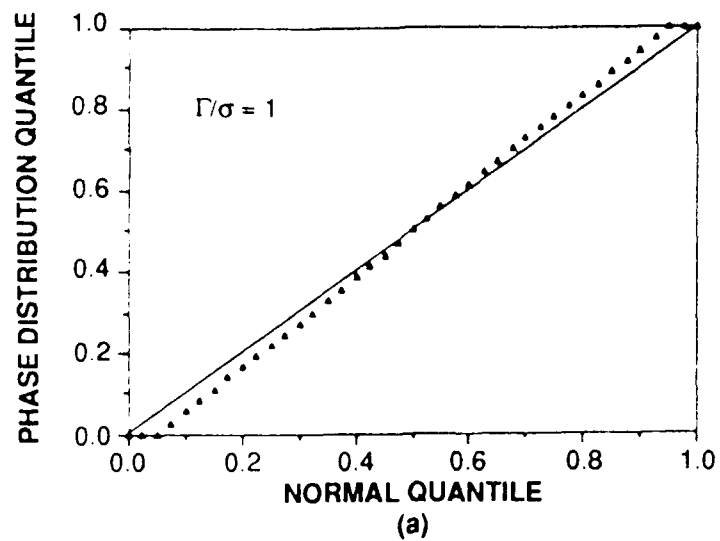


Figure 2
Wilson

ARL:UT
AS-90-595
GRW - DS
8 - 17 - 90

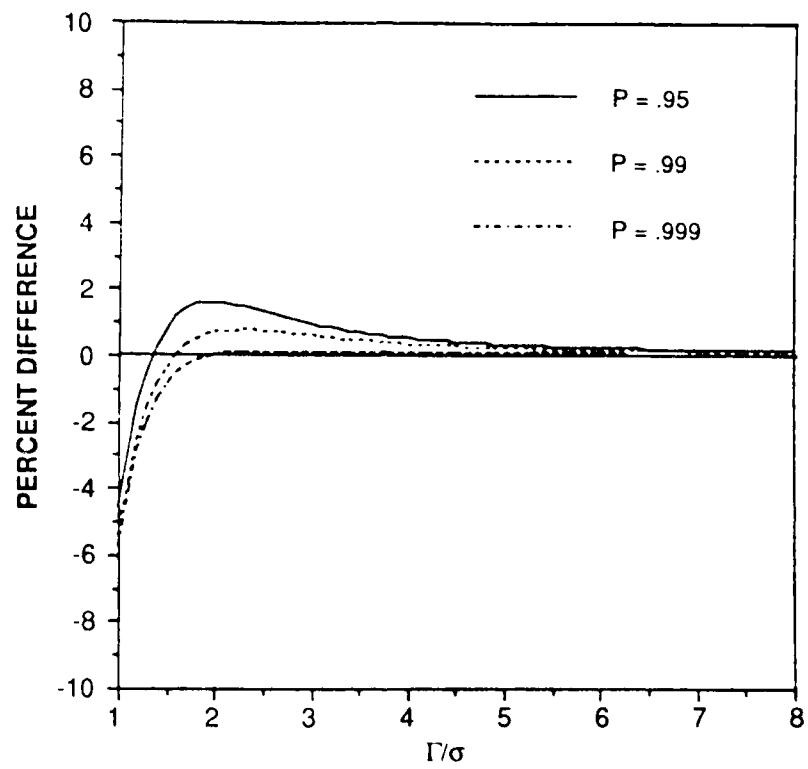


Figure 3
Wilson

ARL:UT
AS-90-596
GRW - DS
8 - 17 - 90

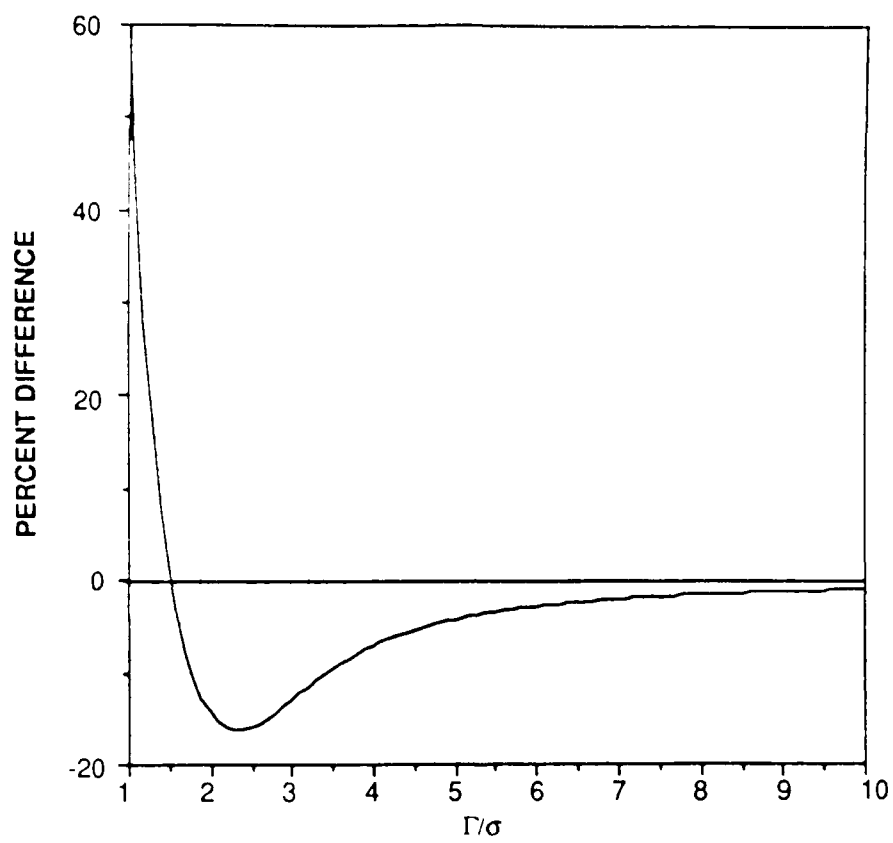


Figure 4
Wilson

ARL:UT
AS-90-597
GRW - DS
8 - 17 - 90

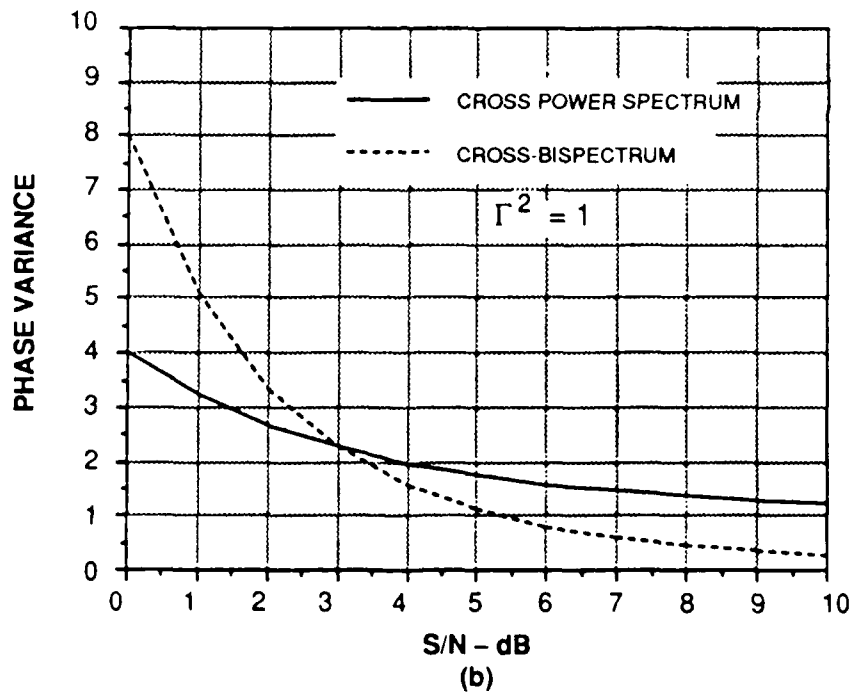
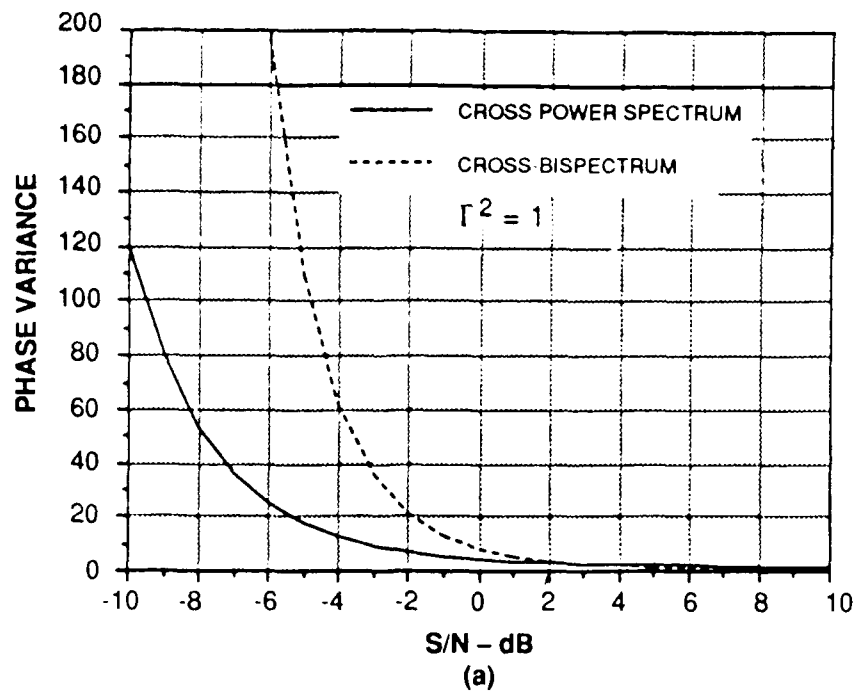


Figure 5
Wilson

ARL:UT
AS-90-598
GRW - DS
8 - 17 - 90

REFERENCES

- Brillinger, D. R. (1965). "An Introduction to Polyspectra," *Annals of Mathematical Statistics* 36, 1351 - 1374.
- Brillinger, D. R., and M. Rosenblatt (1967a). "Asymptotic Theory of Estimates of k th Order Spectra," in *Spectral Analysis of Time Series*, B. Harris, ed. (John Wiley and Sons, Inc., New York), pp. 153-188.
- Brillinger, D. R., and M. Rosenblatt (1967b). "Computation and Interpretation of k th Order Spectra," *Spectral Analysis of Time Series*, B. Harris, ed. (John Wiley and Sons, Inc., New York), pp. 189-232.
- Brillinger, David R. (1975). *Time Series Data Analysis and Theory* (Holt, Rinehart and Winston, Inc., New York).
- Hinich, Melvin J. (1982). "Testing for Gaussianity and Linearity of a Stationary Time Series," *J. Time Series Analysis* 3, 169 - 176.
- Hinich, Melvin J., and Murray Wolinsky (1988). "A Test for Aliasing Using Bispectral Analysis," *J. Am. Stat Assoc.* 83 (402), 499-502.
- Hinich, Melvin J., and Gary R. Wilson (1989). "Time Delay Estimation Using the Cross-Bispectrum," submitted for publication in *IEEE Trans. Acoust., Speech, and Signal Proc.*
- Hinich, Melvin J., and Wilson, Gary R. (1990). "Detection of Non-Gaussian Signals in Non-Gaussian Noise Using the Bispectra," *IEEE Trans. Acoust., Speech, and Signal Proc.* 38(7), 1126-1131.
- Nikias, C. L., and Pan, R. (1988). "Time Delay Estimation in Unknown Gaussian Spatially Correlated Noise," *IEEE Trans. ASSP* 36, 1706-1714.
- Rosenblatt, M. (1966). "Remarks on Higher Order Spectra," in *Multivariate Analysis*, P. R. Krishnaiah, ed. (Academic Press, New York), pp. 383-389.

31 July 1990

**DISTRIBUTION LIST FOR
ARL-TR-90-25
FINAL REPORT UNDER CONTRACT N00014-87-K-0785**

Copy No.

	Office of the Chief of Naval Research Department of the Navy Arlington, VA 22217-5000
1	Attn: J. Smith (Code 1211)
2	G. Remmers (Code 233)
	Office of Naval Intelligence Department of the Navy Washington, DC 20350-2000
3	Attn: L. Long (OP924C4)
	Commanding Officer Naval Technical Intelligence Center 4301 Suitland Road Washington, DC 20390
4	Attn: R. Kalny (Code 211)
	Commanding Officer David Taylor Research Center Bethesda, MD 20084-5000
5	Attn: J. Valentine (Code 1932)
6	B. Douglas (Code 0113)
7	J. O'Donnell (Code 193)
	Commanding Officer David Taylor Research Center Attachment Puget Sound Naval Shipyard Bremerton, WA 98314-5212
8	Attn: D. Groutage (Code 1910.2)
	Commander Space and Naval Warfare Systems Command Department of the Navy Washington, D.C. 20363-5100
9	Attn: L. Parish (PMW181T)
	Commanding Officer Naval Underwater Systems Center New London Laboratory New London CT 06320
10	Attn: R. Tozier (Code 2121)

Distribution List for ARL-TR-90-25 under Contract N00014-87-K-0785
(cont'd)

Copy No.

11	Commanding Officer Naval Ocean Systems Center San Diego, CA 92152 Attn: T. Albert (Code 632)
12	Planning Systems, Inc. 7925 Westpark Drive McLean, VA 22102 Attn: C. Halpeny
13	Commander Submarine Development Squadron Twelve Submarine Base New London Groton, CT 06349-5200 Attn: H. Netzer, NSAP Consultant (Code N2A)
14	Commander Submarine Force U.S. Atlantic Fleet Norfolk, VA 23511 Attn: J. Martins, Science Advisor (Code 013)
15 - 16	Commanding Officer and Director Defense Technical Information Center Cameron Station, Building 5 5010 Duke Street Alexandria, VA 22314
17	Bendix Oceanics 15825 Roxford Street Sylmar, CA 91342 Attn: W. Eugene Brown
18	Martin L. Barlett, ARL:UT
19	Kevin W. Baugh, ARL:UT
20	Joseph F. England, ARL:UT
21	Douglas J. Fox, ARL:UT
22	Keith R. Hardwicke, ARL:UT
23	Melvin H. Hinich, ARL:UT
24	James L. Lamb, ARL:UT
25	Fredrick W. Machell, ARL:UT
26	Jack H. Sheehan, ARL:UT

Distribution List for ARL-TR-90-25 under Contract N00014-87-K-0785
(cont'd)

Copy No.

27	Ronald O. Stearman, ARL:UT
28	Robert T. Trochta, ARL:UT
29	Gary R. Wilson, ARL:UT
30	Signal Physics Group, ARL:UT
31	Library, ARL:UT
32 - 39	Reserve, ARL:UT