Technical Report 1666
July 1994

Adaptive Locally Optimum Processing for Interference Suppression from Communication and Undersea Surveillance Signals

James W. Bond
Vincent Broman
David Stein
James Zeidler
NCCOSC RDT&E Division

Stefan Hui
San Diego State University

Approved for public release; distribution is unlimited.

CONTENTS

EXECUTIVE SUMMARY **1**
BACKGROUND 1
INTRODUCTION 1
SUMMARY OF RESULTS 2
SUMMARY OF RESULTS 2
CONCLUSION 3

ADAPTIVE LOCALLY OPTIMUM PROCESSING **3**
DEFLECTION AND LIKELIHOOD RATIOS 7
AMPLITUDE AND PHASE PROCESSING 10
SIGNAL MODELS 15
Signal of Unknown Structure and Uniform Phase 16
Narrowband Nonzero-Mean Frequency Domain Signal 16
Narrowband Zero-Mean Frequency-Domain Signal 17
Summary on Signal Modeling 17
Interference Models 18
Gaussian Noise 18
One-State Amplitude Model 18
One-State Phase Model 20
One-State Model Algorithm Performance 21
Multistate Models for Interference Amplitudes and Phases 23
Implicit Models 28
Processing Gain for Multistate Model Detectors 33
Performance Comparisons Using Deflection 44
Implementation Issues and Summary 47
LINE DETECTION AND CLASSIFICATION 48
Introduction 48
Upper Bounds for Line Detectors 49
Target Tracking 51
Summary 52

REFERENCES **83**

Appendices

A: GLOBAL PROCESSING GAIN BOUNDS FOR MULTISTATE GAUSSIAN MIXTURE MODELS A-1
B: PROCESSING GAIN BOUNDS FOR THE MIDDLETON CLASS A NOISE MODEL B-1
C: PERFORMANCE LOSS DUE TO TARGET MOTION C-1

Figures

1. Signal and noise representations in coordinate system defined by interferer 53
2. Geometry of describing samples of narrowband signal with greater frequency than narrowband interferer 54
3. Complex sample scatter plots for mixture models 55
4. Fuzzy set interpretation of mixture model construction process 56
5. Normalized Bessel function K0 and the Gaussian distribution 56
6. Processing gain upper bound contours for noncentral mixture model first-order detector 57
7. Processing gain upper bound contours for noncentral mixture model second-order detector 58
8. Processing gain upper bound contours for a two-state Gaussian mixture model first-order detector 59
9. Processing gain contours for a two-state Gaussian mixture model first-order detector 59
10. Processing gain losses due to fuzzy set membership for a two-state Gaussian mixture model first-order detector 60
11. Processing gain upper bound contours for a two-state Gaussian mixture model second-order detector 60
12. Processing gain contours for a two-state Gaussian mixture model second-order detector 61
13. Processing gain losses due to fuzzy set membership for a two-state Gaussian mixture model second-order detector 61
14. Processing gain estimates for a class of three-state Gaussian mixture model second-order detectors 62
15. Processing gain upper bounds for a class of three-state Gaussian mixture model second-order detectors 62
16. Deflection comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 4$ 63
17. Processing gain comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 4$ 63
18. Deflection comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$ 64
19. Processing gain comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$ 64
20. Deflection for the two-state Gaussian mixture model second-order detector 65
21. Processing gain for the two-state Gaussian mixtur model second-order detector 65
22. Fuzzy set loss for the two-state Gaussian mixture model second-order detector 66
23. Parameter sensitivity for a two-state Gaussian mixture model for $\sigma_H^2 = 6$ 66

24. Parameter sensitivity for a two-state Gaussian mixture model for $\sigma_H^2 = 10$ 67
25. Parameter sensitivity for a two-state Gaussian mixture model for $\sigma_H^2 = 16$ 67
26. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 16 samples and high-state variance 16 68
27. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 16 samples and high-state variance 16 68
28. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 16 69
29. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 16 69
30. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 64 samples and high-state variance 16 70
31. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 64 samples and high-state variance 16 70
32. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 4 71
33. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 4 71
34. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 64 samples and high-state variance 16 72
35. Implicit model processing gains for a two-state Gaussian mixture model second-order detector 72
36. Implicit modeling loss for a two-state Gaussian mixture model second-order detector 73
37. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 6$ 73
38. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$ 74
39. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 6$ 74
40. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 10$ 75
41. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$ 75
42. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 6,\ 10,$ and 16 76
43. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.5$ and $\sigma_H^2/\sigma_L^2 = 10$ 77
44. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.25$ and $\sigma_H^2/\sigma_L^2 = 10$ 79
45. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.1$ and $\sigma_H^2/\sigma_L^2 = 10$ 81

A-1. Processing gain parametrized by low-state probability A-2
B-1. Processing gain of a first-order detector for Middleton class A noise model B-5
B-2. Processing gain of a second-order detector for Middleton class A noise model B-5
B-3. Comparison of processing gains of the Middleton class A first-order detector and finite state approximations B-7
B-4. Comparison of processing gains of the Middleton class A second-order detector and finite state approximations B-8
C-1. Cumulative distribution of cell-boundary crossings C-2
C-2. Known path detector performance for PD(1) = 0.50 . . . . . C-7
C-3. Known path detector performance for PD(4) = 0.50 . . . . . C-7
C-4. Known path detector performance for PD(M) = 0.50 . . . . . C-8
C-5. Gamma distribution thresholds C-8
C-6. Example of random walk spatial occupation probability C-11
C-7. Dependence of detector on prior velocity variance C-12
C-8. Dependence of detector performance on low estimate of prior velocity variance C-14
C-9. Dependence of detector performance on high estimate of prior velocity variance C-14
C-10. Dependence of detector performance on low estimate of prior velocity variance C-15
C-11. Dependence of detector performance on path length C-17
C-12. Dependence of detector performance on knowledge of track C-18

EXECUTIVE SUMMARY

BACKGROUND

This report and two companion reports have been derived from the chapter on detection theory prepared for the High Gain Initiative (HGI) report. The two companion reports, “Detection Processing for Undersea Surveillance” (Confidential) and “Gaussian Mixture Models for Acoustic Interference” have a undersea surveillance focus, while this report applies to both communications and undersea surveillance. “Detection Processing for Undersea Surveillance” summarizes the status of detection processing for undersea surveillance at the time of the initiation of the HGI program and deals primarily with adaptive filtering. It summarizes the cumulation of many years of effort in the field by J. Zeidler and others. “Gaussian Mixture Models for Acoustic Interference” summarizes the results of simulation and the modeling of actual data collected during a HGI experiment by Gaussian mixture models. To a large extent, it summarizes work accomplished by D. Stein. This report is the cumulation of work by J. Bond, S. Hui, D. Stein, and others on adaptive locally optimum processing for interference suppression and of V. Broman on target tracking.

Initial work related to the work presented in this report by one of the authors (J. Bond) was on the Non–Linear Adaptive Processor (NONAP) program and was funded by SPAWAR 153 beginning in 1986. A parallel program seeking alternatives to NONAP called the Adaptive Locally Optimum Detection (ALOD) program was also funded by SPAWAR beginning in 1989. Theoretical work by Per Kulstam (Paircom Inc.) and Hank Schmidt (Technology Services Incorporated) on the NONAP and ALOD programs also provided valuable insights into the work on adaptive locally optimum processing presented in this report. The first results exponded in this report were obtained by J. Bond and T. Schlosser during 1988 and 1989 under NRaD Independent Engineering Development (IED) funds. All of this early work focused on the use of adaptive locally optimum processing for interference suppression from Very Low Frequency/Low Frequency bandspread communication signals used for submarine communications. The theory was extended to other frequency bands and other waveforms through funding by the Communications and Networking NRaD 6.2 Block managed by Reeve Peterson (NRaD) beginning in 1992 and to undersea surveillance through the funding of the High Gain Initiative (HGI) 6.2 program coordinated by Robert Hearn (NRaD). Adaptive locally optimum processing has a long academic history beginning with the work of David Middleton on Statistical Communication Theory prior to 1960 and this history and other related work outside of that described above is discussed in the body of this report.

INTRODUCTION

Adaptive locally optimum processing applies when the signal is dominated by the interference and is effective when the non-Gaussian component of the interference dominates the Gaussian component. We develop the theory for discrete samples of the baseband representation of a communication signal (or multiple baseband representations of wideband signals) or the analytical representation of an acoustic signal. Practical algorithms have been derived under the assumptions that the interference component of the discrete samples is independent from sample-to-sample and the signal has known phase structure or the signal has unknown phase structure. For the former case, the algorithms involve the calculation of first derivatives of the probability density functions of amplitudes and phase (or symmetric-phase differences) and for the latter case, the algorithms involve the calculation of second derivatives of the probability density functions of amplitudes and phase (or symmetric-phase differences).

We have identified three approaches to the development of practical adaptive locally optimum processing: (a) model the interference statistics by a parametrized family of probability density functions and estimate the parameters in real-time from the data and calculate the corresponding adaptive locally optimum transformation of the data, (b) model the interference statistics as above and identify a family of candidate models for the interference, compare the real-time statistics with each of the stored distributions to identify the best match and then process the data accordingly, and (c) build an implicit model of the interference statistics using the real-time data and process and samples accordingly. The material presented in this report reduces to practice all three of these approaches.

SUMMARY OF RESULTS

The technical results contained in this report include:

1. formulas for optimum processing of complex samples when the signal has known and unknown phase structure,
2. formulas for optimum processing of complex sample amplitude, phases, and symmetric phase-differences when the signal has known and unknown phase structure,
3. formulas for the optimum processing of the frequency domain representation of the signal samples when the signal is narrowband,
4. explicit formulas for processing samples when the statistics are modeled by either non-central or central mixture models for which the parameters are estimated or implicit,
5. processing gain over tradiational processing bounds and estimates in terms of model parameters for the above algorithms when the statistics are modeled by either non-central or central mixture models,
6. comparisons of processing gain bounds and estimates with performance estimates obtained through simulations for two- and three-state mixture models,
7. simulation results for averages of the transformed samples, and
8. results for the loss of performance due to target motion for processing at the output of a beamformer.

SUMMARY OF RESULTS

Central to our approach is the maximization of the deflection of a detector. The form of the detector maximizing deflection is obtained through use of the Cauchy-Schwartz inequality and is the likelihood ratio. A Taylor's expansion of the probability density function of signal plus noise about signal zero is used to express the probability density function of the signal plus noise in terms of the probability density function of noise alone. If the signal has known structure, the optimum detector is dominated by the linear terms in the Taylor's expansion and practical algorithms are obtained by replacing the signal by the signal divided by its norm and using only the linear term of the Taylor expansion; if the signal has unknown structure, it is reasonable to suppose that the mean values of the real and imaginary components of the signal are zero and as a result the quadatic terms of the Taylor's expansion dominate and practical algorithms are obtained by replacing the second order signal terms by signal variances and using only the quadratic term of the Taylor expansion.

Amplitude and phase algorithms are obtained by converting from inphase/quadrature to polar coordinates and assuming that the sample amplitudes and phases are uncorrelated. Amplitude and phase processing then become parallel processes. For wideband signals, the signal can be recovered from a

symmetric phase-difference, so that the preprocessing step to remove interference sample-to-sample phase correlation by replacing the phases by symmetric phase-differences leads to effective phase processing for many cass when processing of phase itself would be ineffective.

The recognition that the use of the Gaussian kernel representations of the probability density function of amplitudes and symmetric phase-differences could be viewed as making use of an implicit non-central Gaussian mixture model, provides the key to unifying algorithms discovered for communications and algorithms developed for undersea surveillance using central Gaussian mixture models. We obtain processing gain bounds defined by the ratio of the deflection of the adaptive locally optimum processing detector deflection and the deflection for the traditional detector in terms of the mixture model parameters. We verified through simulations that for two-state and three-state mixture models, performance measured by deflection and performance defined by probability of detection for a specified probability of false alarm are highly correlated.

Finally, we examined the performance of line detectors obtained by averaging of sample detector values. For traditional undersea surveillance applications, when the processing is used after beamforming, this is closely related to “eye integration” used to detect the presence of narrowband signals in time histories of Fourier coefficient magnitudes represented by a grey scale display as a function of frequency (abscissa) and time (ordinate). For HGI arrays with spatial cells, the spatial resolution of the array will be high and it is desirable to replace “eye integration” by automated processing. For this situation the average detector is an upper bound for obtainable performance and we obtained some preliminary results on the degradation of performance due to unknown target motion for several candidate tracking algorithms.

CONCLUSION

Simulations and processing of real data indicate that adaptive locally optimum processing can provide significant gains over traditional processing. A satisfactory theory now exists that leads to practical algorithms to implement this processing for many communication and undersea surveillance applications.

ADAPTIVE LOCALLY OPTIMUM PROCESSING

Middleton (1960) developed a statistical communication theory that addressed, among other topics, optimum ways to detect weak communication signals in the presence of non-Gaussian interference. His original theory addressed the estimation of a communication signal from the probability density function of received signal plus interference. Later Middleton (1966, 1967, 1977, 1983, 1984, 1991) and Middleton and Spaulding (1983, 1986) extended the theory to include communication and undersea surveillance applications by introducing Gaussian mixture models.

Within a general statistical detection framework, the decision of whether a signal is present or not is usually based on the likelihood ratio $\frac{p_{n+s}(z)}{p_n(z)}$, where the numerator is the probability density function for vectors of complex samples z containing signal plus noise and the denominator is the probability density function for vectors of complex samples z containing noise alone. For many applications, the challenge in implementing optimal processing is to obtain an estimate of $p_n(z)$ given a received signal sample that may contain signal as well as noise. Techniques exist to solve this problem when the signal is stronger than the noise. However, these techniques are of little interest in surveillance because the signals would be detected by using traditional techniques.

Among the approaches to implement detection algorithms are those using the likelihood ratio, approximations to the likelihood ratio, minimax criteria, and nonparametric techniques. A test for the presence of signal based on the likelihood ratio provides the maximum probability of detection at a given false-alarm rate (Poor, 1988; Poor and Thomas, 1978: Whalen, 1971), but it requires knowledge of the signal plus noise and noise only probability densities. The probability density function of noise is difficult to obtain from observed signal-plus-noise samples unless the signal is dominated by the noise.

A minimax approach can be used to account for uncertainty in the class of distributions that describe the noise statistics. Such an approach is based on the designation of a cost function and classes of possible noise densities and signal-plus-noise densities. The Bayes or Neyman-Pearson criteria can be used to define a cost function, and the ε-contaminated class of densities is often used (Kassam and Poor, 1985; Poor, 1988). The minimax algorithm selects the detector that minimizes a maximum cost over the classes of density functions. For example, whereas the matched filter implements a likelihood ratio for known signals in stationary independent Gaussian noise (Berry, 1981; Poor, 1988), the correlator-limiter is the minimax detector for a known signal in stationary independent noise with a distribution belonging to the class of ε-contaminated mixture distributions with a nominal Gaussian distribution (Kassam and Poor, 1985; Poor, 1988). Another minimax approach, perhaps more common, is to minimize the maximum cost (or risk) over the class of prior probabilities (Whalen, p. 135, 1971).

More prosaic techniques are also related to optimal detection. One such technique is the use of various kinds of clippers or automatic gain controls in military radios, designed to reduce the impact of impulsive noise or interference on the reception of communication signals (Blachman, 1964, 1971a, 1971b, 1982, 1992; Arnstein, 1991, 1992). These techiques are especially effective for very low frequency (10 to 30 kHz) and low frequency (30 to 60 kHz) communications, when environmental noise dominated by lightning generated interference can be well modeled as an additive sum of Gaussian noise and high-power short-duration pulses.

One of the most common approaches to obtaining estimates of the probability density function for signals plus noise is to relate this probability density function to the one for noise alone under the assumption that the signal is weaker than the noise. This theory is of interest for detecting communication signals in the presence of jamming and for detection of masked submarine lines in undersea surveillance. The optimum processing techniques developed under the assumption that the signal is weaker than the noise is known as adaptive locally optimum processing techniques.

The basic idea leading to the various locally optimum processing techniques is to expand $p_{n+s}(z)$ in a Taylor series expansion about $s = 0$. By using this expansion, one obtains an approximation for the likelihood ratio $\frac{p_{n+s}(z)}{p_n(z)}$ that is valid when s is small. The application of adaptive locally optimum processing to a particular problem entails (1) estimating the probability density function and (2) calculating the transformation of the data to determine the presence of the signal. We briefly survey the different locally optimum algorithms that have been developed.

The different algorithms arise from the different ways of modeling the signal and the noise. There are three general approaches to modeling the noise: parametric, nonparametric, and model fitting. The parametric and nonparametric techniques are discussed in detail for various mixture models of the noise in this section. The model fitting technique offers an alternative approach. The information required to implement it is developed through an analysis of the parameter estimation techniques.

Parametric estimation proceeds by assuming that the probability density function of the noise is from a family of probability density functions described by a finite set of parameters. Given the data, the parameters are estimated from the data and in this way a probability density function is chosen to represent the samples. The appropriate transformation of the data samples can then be calculated from the chosen probability density function.

Middleton studies probability density functions, called Gaussian mixture models, that are described by infinite sums of Gaussian distributions. The family of probability density functions is described by either two or three parameters. Given these parameters, the optimum transformation of the received signal-plus-interference-plus-noise samples to estimate the signal can be calculated. In particular, Middleton (1966, 1967, 1977, 1983, 1984, 1991) and Middleton and Spaulding (1983, 1986) have formulated adaptive locally optimum detectors based upon Middleton's class A noise model. See appendix B for a discussion of the Middleton class A noise model. The Middleton class A noise model is especially appropriate for the detection of communication signals in impulsive interference, and has been suggested for modeling underwater acoustic noise by Middleton.

Bouvet and Schwartz (1988, 1989) compare the performance of the likelihood ratio detector derived from a two-state Gaussian mixture model, the matched filter, and the correlator-limiter, for detection of known signals in shipping noise measured at sea. They showed that the performance of the matched filter and the correlator-limiter are similar and the likelihood ratio detector based on a Gaussian mixture model provides improved performance at some false-alarm probabilities and signal-to-noise ratios. Baker and Gualtierottio (1986, to appear) have developed likelihood ratio detection algorithms for a general class of signal sin circularly invariant noise, which generalizes the Gaussian mixture model. These generalized Gaussian mixture models play a central role in our theory of adaptive locally optimum for signals of unknown structure.

For the nonparametric approach, an empirical model of the interference probability density function is constructed from samples of the interference. The techniques for estimating the probability density function include fitting polynomials to histograms, using kernel representations, and obtaining the estimates from finite differences of quantities of ordered samples (which represent an estimate of the cumulative probability function of the samples).

Recently, both the Air Force and Navy have funded extensive efforts to investigate the implementation of adaptive locally optimum processing techniques in military radios. Both of these efforts focused on real-time estimation of the probability density function, motivated by the consideration that jamming signals may have structure with very general statistical properties that are under the control of an adversary. The Air Force effort was undertaken by Hazeltine Corporation for Rome Air Development Center (RADC) (Murphy, Tilley, and Torre, 1990) and the Navy effort was initiated by Johns Hopkins Unviersity Applied Physics Laboratory (JHU/APL) (Higbie, 1988).

The Air Force effort focused on processing the real and imaginary components of the baseband samples based on fitting a polynomial to the histogram of received signal samples. This is a very natural approach to describe the statistics and obtaining a differentiable probability density function from which the likelihood function can be calculated. Their work involved a general theory of how best to fit a polynomial to various probability density functions and the performance achievable by using adaptive locally optimum processing.

An alternative approach to estimating statistics for time domain processing has been developed for the Navy by Higbie (1988). He estimates the probability density function of amplitudes and

phase-differences by finite difference of quantiles of the cumulative probability density functions of sample amplitudes and phase-differences. He obtains the quantiles by sorting a block of successive received signal baseband sample amplitudes and phase-differences according to their magnitudes. The algorithm is a sliding block algorithm with the samples used symmetrical around the sample being transformed. Laboratory tests of the algorithms have shown that the techniques provide better detection performance for bandspread communications signals in the presence of wideband interference than any previously implemented techniques.

Another family of locally optimum processing techniques has been developed by Bond (1991) and Bond and Hui (to appear) as a Navy effort. Bond and Hui developed their algorithms from kernel representations of probability density functions and studied the resulting algorithms from an analytical point of view. Bond found that processing amplitudes and phases, as well as phase-differences, could be quite effective. His adaptive locally optimum processing, like that of Higbie, does not require that the probability density function of the noise be described by a few parameters.

The model approach uses preprocessing to identify typical probability density functions describing the statistics of signal plus noise that might be received. A family of these density functions is then selected and stored along with the optimal processing to use for each case. The incoming samples are then processed to obtain the best match between their density function and a stored density function. The samples are then processed by using the processing associated with the stored density function that best fits the incoming data. In this approach, the probability of detection is conditioned on the comparison of the empirical density of received signal plus noise or noise only with the stored probability densities.

Schloz and Giles (1990, 1992) and Schloz (1991, 1992) implemented the model approach by obtaining a family of candidate distributions of the received signal plus interference and noise by preprocessing data samples representative of the channel for which the processing is to be used. Then the distribution of successive blocks of samples of the received signal plus interference and noise are compared to each of the candidate distributions in real time and the processing of these samples is based on the best match. This approach is easy to implement in a digital receiver with substantial processing capability and memory, because the distribution comparisons can be done in parallel and the parameters describing the optimum processing stored in memory.

Adaptive locally optimum processing could be effectively used before beamforming, for moderate or short-range surveillance applications involving few hydrophones, because in these cases the interferer-to-signal ratio at the hydrophone level and at the beamformer output level will often be slightly different. In contrast, for ocean basin surveillance, many of the interfering signals that might mask surveillance signals of interest would not stand out from the general background noise at the hydrophone level, precluding the use of the adaptive locally optimum processing.

Adaptive locally optimum processing techniques could be used after a time domain beamformer and before spectral analysis. For a frequency domain beamformer, the output could be transformed back to the time domain or processed in the frequency domain. The noise statistics depend on the bandwidth of the frequency domain representation when transformed back to the time domain in the first case and the frequency resolution of the spectral analysis in the second case. The noise statistics of the beamformed data for a particular interferer also depend on the propagation modes of the interferer signal to the receiving array. For a matched-field beamformer, the transformation back to the time domain may not always be feasible because the frequency dependence of the beamforming.

Motivated by these considerations, we developed adaptive locally optimum processing techniques suitable for processing either time domain or frequency domain signals.

In the subsequent sections, we present an integrated theory of adaptive locally optimum processing suitable for processing beamformed time domain and frequency domain data. We proceed from first principles by showing how maximizing the deflection, a natural measure of detector performance (defined below), leads to likelihood ratios. Imposition of the small signal hypothesis allows the likelihood ratios to be evaluated in terms of derivatives of the probability density function of interference and noise. These formulas involving derivatives applied to various interference models lead to algorithms for processing the data. Our treatment includes both parametric and nonparametric modeling of the interference. Formulas for first-order and second-order detectors are described depending on whether the signal has known or unknown structure. The performance of the various algorithms is then established by analysis and simulation.

DEFLECTION AND LIKELIHOOD RATIOS

A unified treatment of nonlinear processing for time domain and frequency domain beamformed data is obtained by treating received signal samples as complex numbers. For time domain outputs, the complex samples are obtained by replacing the real signal with its analytical signal. The analytical signal is mathematically determined from the real signal by using the Hilbert transform (Papoulis, 1977). If $s(t)$ is the signal and $\hat{s}(t)$ is its Hilbert transform, then $s_a(t) = s(t) + i\hat{s}(t)$ is the analytic signal. In communication systems, the signal is often modulated by a fixed frequency after reception and then highpass filtered. In either case, the analytic signal is often called the *baseband representation* with the real part called the *in-phase component* and the imaginary part called the *quadrature component* of the baseband sample. Some authors call samples of the analytic signals *complex samples*.

Throughout this section, the input to the nonlinear processing algorithms are complex samples of either an analytic signal or complex Fourier coefficients. We adopt the following notation. Let z_j denote the j-th complex sample with x_j and y_j denoting its in-phase and quadrature components, respectively, and $z_j^* = x_j - iy_j$ denote the complex conjugate of z_j. The complex number x_j, y_j also can be represented as a vector (x_j, y_j). We define the norm of $x_j + iy_j$, or equivalently the length of (x_j, y_j), by

$$|x_j + iy_j| = |(x_j, y_j)| = \sqrt{x_j^2 + y_j^2} \ .$$

A common starting point for the development of adaptive locally optimum processing algorithms for processing time domain and frequency domain beamformer outputs is provided by considering the maximization of the deflection for sequences of samples. A powerful idea in functional analysis is to optimize a functional on a space of functions (Rudin, 1973). It often turns out that the optimal function has many other desirable properties. We have found this to be the case for deflection and make use of it throughout this subsection to obtain many useful adaptive locally optimum processing results. In the next subsection we relate deflection to probability of line detection for a given probability of false alarm.

Deflection measures (in noise standard deviation units) how different the expected values of the detector is for signal plus noise and noise alone. The deflection of a detector is a natural extension of detector output signal-to-noise for cases when the expected value of the detector under noise alone may be nonzero. Even though deflection for a particular detector only involves second-order statistics, maximization of deflection over a class of functions involves all the moments of the probability density functions of signal plus noise and noise alone.

Suppose u is any real-valued detection variable. Then the deflection $\delta(u)$ of u is

$$\delta(u) = \frac{E_{n+s}(u) - E_n(u)}{\sigma_n(u)}$$

where $E_{n+s}(u)$, $E_n(u)$, and $\sigma_n(u)$ denote the expected value of the detection quantity for signal plus noise, the expected value of the detection quantity for noise, and the standard deviation of the detection quantity for noise, respectively. It can happen that the detection problem is trivial, for example if the known distributions of signal plus noise and noise only do not intersect or if the noise is zero. We preclude these cases throughout this section by assuming that $p_n(x, y)$ does not vanish except on a set with probability 0.

For the time being, we restrict our treatment to real-valued detectors $D : R^2 \rightarrow R$, that is to real-valued functions of pairs of real numbers which are the real and imaginary components of complex samples of the analytical signal. Futhermore, we assume that the detectors D are functions with finite deflection. Our strategy is to see how much we can learn about the functional form of a detector $D(x, y)$ with finite deflection by imposing the condition that it maximizes the absolute value of the deflection. The information available about the noise is the probability density function $p_n(x, y) = p(z|n)$. Calculation of the expectation of D given that it contains signal and noise requires the probability density function $p_{n+s}(x, y)$.

We first obtain the form of the detector that maximizes deflection. Let D be a detector with finite deflection and let $\hat{D}(x, y) = D(x, y) - E_n(D(x, y))$. It is clear from the definition that $\delta(D) = \delta(\hat{D})$ and $E_n(\hat{D}(x, y) = 0$. The square of the denominator of the deflection is simply

$$\sigma_n^2(\hat{D}(x, y)) = \int\int_{R^2} \hat{D}^2(x, y) p_n(x, y) dxdy \, .$$

The Cauchy-Schwartz inequality provides the mechanism to determine the functional form of the optimal $\hat{D}(x, y)$. In its general form, the Cauchy-Schwartz inequality applies to any measure μ and for finite integrals defined by the measure

$$(\int fgd\mu)^2 \leq \int f^2 d\mu)(\int g^2 d\mu)$$

with equality if and only if $f = cg$ for some constant c. We have

$$\begin{aligned} & E_{n+s}(\hat{D}(x, y)) - E_n(\hat{D}(x, y)) \\ & = \int\int_{R^2} \hat{D}(x, y)[p_{n+s}(x, y) - p_n(x, y)] dxdy \end{aligned}$$

and letting $\Delta p(x, y) = \dfrac{p_{n+s}(x, y) - p_n(x, y)}{\sqrt{p_n(x, y)}}$,

$$\begin{aligned} & (\int\int_{R^2} \hat{D}(x, y)[p_{n+s}(x, y) - p_n(x, y)] dxdy)^2 \\ & = \left\{ \int\int_{R^2} [\hat{D}(x, y)\sqrt{p_n(x, y)}] \Delta p(x, y) dxdy \right\}^2 \\ & , = [\sigma_n^2(\hat{D}(x, y))][\int\int_{R^2} [\Delta p(x, y)] sup2 dxdy] \, . \end{aligned}$$

Therefore,

$$\delta^s(\hat{D}) \leq \int\int_{R^2} \frac{[p_{n+s}(x,y) - p_n(x,y)]^2}{p_n(x,y)} dxdy$$

with equality if, and only if, for some $c \neq 0$,

$$\hat{D}(x,y)\sqrt{p_n(x,y)} = c\,\frac{p_{n+s}(x,y) - p_n(x,y)}{\sqrt{p_n(x,y)}}$$

or equivalently,

$$\hat{D}(x,y) = c\left(\frac{p_{n+s}(x,y)}{p_n(x,y)} - 1\right).$$

Note that the deflection is independent of $c \neq 0$ and the constant 1 and we can conclude that

$$D(x,y) = \frac{p_{n+s}(x,y)}{p_n(x,y}$$

is also a detector which maximizes deflection, hereafter referred to as an optimum detector.

$$D(x,y) = \frac{p_{n+s}(x,y)}{p_n(x,y)}$$

If $s = a + ib$ and the samples consist of signal plus additive roles, then $p_{n+s}(x,y) = p_n(x-a, y-b)$. If we assume that the signal is small, a Taylor's expansion can be used to express this latter probability in terms of $p_n(x,y)$ and its derivatives:

$$\begin{aligned} & p_n(x-a, y-b) \\ &= p_n(x,y) + [-\frac{\partial p(x,y)}{\partial x}a - \frac{\partial p(x,y)}{\partial y}b] \\ &+ \frac{1}{2}[\frac{\partial^2 p(x,y)}{\partial x}a^2 + 2\frac{\partial^2 p(x,y}{\partial x \partial y}ab + \frac{\partial^2 p(x,y)}{\partial y^2}b^2 \\ &+ \textit{higher order terms in a and b.} \end{aligned}$$

The optimum detector given the signal $a + ib$ is, up to second-order terms,

$$\begin{aligned} D(x,y) &= -\frac{1}{p_n(x,y)}\frac{\partial p_n(x,y)}{\partial x}a \\ &- \frac{1}{p_n(x,y)}\frac{\partial p_n(x,y)}{\partial y}b + \frac{1}{2}\frac{1}{p_n(x,y)}\frac{\partial^2 p_n(x,y)}{\partial x^2}a^2 \\ &+ \frac{1}{p_n(x,y)}\frac{\partial^2 p_n(x,y)}{\partial x \partial y}ab + \frac{1}{2}\frac{1}{p_n(x,y)}\frac{\partial^2 p_n(x,y)}{\partial y^2}b^2 . \end{aligned}$$

We call the sum of the first-order terms in the above detector the first-order detector and the sum of the second-order terms the second-order detector.

We next consider the case when the signal is not fixed but is given by $x = X + iY$, where X and Y are random variables. It can be shown, by using techniques similar to the known signal case, that the optimum detector still has the form and the terms of the form $D(x,y) = \frac{p_n(x,y)}{p_n(x,y)} - 1$ in the Taylor series expansion should be replaced by the expected values $E(X^m Y^n)$. If X and Y are independent, then a^m and b^m can be replaced by the $m-th$ moments of X and Y , respectively. In particular, if X and Y are independent and identically distributed with zero mean and variance σ^2, the optimum detector has, up to second-order terms, the form

$$D(x,y) = \frac{1}{p_n(x,y)}[\frac{\partial^2 p_n(x,y)}{\partial x^2} + \frac{\partial^2 p_n(x,y)}{\partial y^2}] .$$

The detector output for a number of samples, say , which need not be independent, can be combined to give a new detection quantity

$$\frac{1}{N} \sum_{1}^{N} D(x_k, y_k) .$$

We call this multisample detector the *line detector* in anticipation of its use to detect narrowband signals of interest in ocean basin surveillance. For a sequence of independent samples, the probability density function is the product of the probability density functions of each sample:

$$\begin{aligned} & p(x_1, y_1, x_2, ..., x_N, y_N) \\ = \; & p(x_1, y_1)p(x_2, y_2)...p(x_N, y_N) . \end{aligned}$$

We have

$$\begin{aligned} & \frac{1}{p(x_1, y_1)p(x_2, y_2)...p(x_N, y_N)} \\ & x \frac{\partial(p(x_1, y_1)p(x_2, y_2)...p(x_N, y_N))}{\partial x_k} \\ & = \sum_{h=1}^{N} \frac{1}{p(x_h, y_h)} \frac{\partial p(x_h, y_h)}{\partial x_k} \\ & = \frac{1}{p(x_k, y_k)} \frac{\partial p(x_k, y_k)}{\partial x_k} \end{aligned}$$

Thus, the optimum detector $D(z_1, z_2, ..., z_N)$ given the same vector $(z_1, z_2, ..., z_N)$ with independent $z_1, z_2, ..., z_N$ is up to second-order terms,

$$\begin{aligned} D(z_1, z_2, ..., z_N) = & - \sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial p(x_k, y_k)}{\partial x} a_k \\ & - \sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial p(x_k, y_k)}{\partial y} b_k \\ & + \frac{1}{2} \sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial^2 p(x_k, y_k)}{\partial x^2} a_k^2 \\ & + \sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial^2 p(x_k, y_k)}{\partial x \partial y} a_k b_k \\ & + \frac{1}{2} \sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial^2 p(x_k, y_k)}{\partial y^2} b_k^2 , \end{aligned}$$

which is a constant times the line detector. Therefore, the line detector is optimum for independent samples. The performance of the line detector is discussed later in this section. The remainder of this subsection treats first-order and second-order detectors.

AMPLITUDE AND PHASE PROCESSING

For many applications, it is more natural to model the amplitudes and phases of complex samples than their real and imaginary components. Let $A = \sqrt{x^2 + y^2}$ and $\theta = \arg(x + iy)$ be the amplitude and phase of the sample $z = x + iy$). The phase θ should be assigned to avoid discontinuities of 2π, which is always possible if $A > 0$. The process of assigning phase in a continuous manner is known as

phase unwrapping (Oppenheim and Shafer, 1989; Tribolet, 1977). Hereafter, we assume that all phases are unwrapped phases because for some of the processing algorithms discussed later it is important to avoid unnecessary discontinuities in the phase.

Suppose $p(x,y) = \frac{p(A)}{A} q(\theta)$, where $p(A)$ is the probability density function of the amplitudes A of the complex samples z, and $q(\theta)$ is the probability density function of the phases θ of the complex samples z. Note that the factor A is necessary so that $p(x,y)$ is a probability density when $p(A)$ is a probability density and when $q(\theta)$ is a probability density because

$$\int\int_{R^2} p(x,y)dxdy = \int_0^{2\pi}\int_0^{\infty} p(A)p(\theta)dAd\theta .$$

It is quite informative to express the optimum detector in terms of the partial derivatives of $p(A)$ and $q(\theta)$ that we obtained previously in terms of $p(x,y)$. The algebra is simpler if we introduce $\hat{p}(A) = \frac{p(A)}{A}$. Observe that $\frac{\partial A}{\partial x} = \frac{x}{A}, \frac{\partial A}{\partial y} = \frac{y}{A}, \frac{\partial \theta}{\partial x} = \frac{y}{A^2}$, and $\frac{\partial \theta}{\partial y} = \frac{x}{A^2}$. We next calculate the partial derivatives occurring in the first- and second-order detectors.

$$\begin{aligned}
\frac{\partial p(x,y)}{\partial x} &= \left[\hat{p}'(A)\frac{x}{A}\right]q(\theta) + \hat{p}(A)\left[q'(\theta)\frac{-y}{A^2}\right] \\
\frac{\partial p(x,y)}{\partial y} &= \left[\hat{p}'(A)\frac{y}{A}\right]q(\theta) + \hat{p}(A)\left[q'(\theta)\frac{x}{A^2}\right], \\
\frac{\partial^2 p(x,y)}{\partial x^2} &= \left[\hat{p}''(A)\frac{x^2}{A^2} + \hat{p}'(A)\frac{y^2}{A^3}\right]q(\theta) \\
&\quad + 2\left[\hat{p}'(A)\frac{x}{A}\right]\left[q'(\theta)\frac{-y}{A^2}\right] \\
&\quad + \hat{p}(A)\left[q''(\theta)\frac{y^2}{A^4} + \frac{2xy}{A^4}q'(\theta)\right] \\
\frac{\partial^2 p(x,y)}{\partial y \partial x} &= \left[\hat{p}''(A)\frac{xy}{A^2} - \frac{xy}{A^3}\hat{p}'(A)\right]q(\theta) \\
&\quad + 2\left[\hat{p}'(A)q'(\theta)\left(\frac{x^2-y^2}{A^3}\right)\right] \\
&\quad + \hat{p}(A)\left[q''(\theta)\frac{x^2}{A^4} + \left(\frac{y^2-x^2}{A^4}\right)'(\theta)\right],
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 p(x,y)}{\partial y^2} &= \left[\hat{p}''(A)\frac{y^2}{A^2} + \hat{p}'(A)\frac{x^2}{A^3}\right]q(\theta) \\
&\quad + 2\left[\hat{p}'(A)\frac{y}{A}\right]\left[q'(\theta)\frac{x}{A^2}\right] \\
&\quad + \hat{p}(A)\left[q''(\theta)\frac{x^2}{A^4} - \frac{2xy}{A^4}q'(\theta)\right].
\end{aligned}$$

The first-order detector is

$$
\begin{aligned}
&-\frac{1}{\hat{p}(A)q(\theta)}\left\{a\frac{\partial p(x,y)}{\partial x}+b\frac{\partial p(x,y)}{\partial y}\right\}= \\
&-\left\{\left(\frac{xa+by}{A}\right)\frac{\hat{p}'(A)}{\hat{p}(A)}+\left(\frac{-ya+xb}{A^2}\right)\frac{q'(\theta)}{q(\theta)}\right\}.
\end{aligned}
$$

and the second-order detector is

$$
\begin{aligned}
&\frac{1}{\hat{p}(A)q(\theta)} \\
&x\left[a^2\frac{\partial^2 p(x,y)}{\partial x^2}+2ab\frac{\partial^2 p(x,y)}{\partial x\partial y}+b^2\frac{\partial^2 p(x,y)}{\partial y^2}\right]= \\
&\qquad\left(\frac{xa+by}{A}\right)^2\frac{\hat{p}''(A)}{\hat{p}(A)}+\left(\frac{-ya+xb}{A}\right)^2\frac{\hat{p}'(A)}{A\hat{p}(A)} \\
&\qquad+\left(\frac{-ya+xb}{A}\right)^2\frac{q''(\theta)}{A^2q(\theta)} \\
&+2\left(\frac{xy(a^2-b^2)+ab(y^2-x^2)}{A^3}\right)\left(\frac{1}{A}-\frac{\hat{p}'(A)}{\hat{p}(A)}\right)\left(\frac{q'(\theta)}{q(\theta)}\right).
\end{aligned}
$$

The quotients $\frac{\hat{p}'(A)}{\hat{p}(A)}$ and $\frac{\hat{p}''(A)}{\hat{p}(A)}$ are related to $\frac{p'(A)}{p(A)}$ and $\frac{p''(A)}{p(A)}$ by the following relations:

$$
\frac{\hat{p}'(A)}{\hat{p}(A)}=\frac{p'(A)}{p(A)}-\frac{1}{A}
$$

and

$$
\frac{\hat{p}''(A)}{\hat{p}(A)}=\frac{p''(A)}{p(A)}-\frac{2}{A}\left[\frac{p'(A)}{p(A)}-\frac{1}{A}\right].
$$

We make the following observation about the first-order and second-order detectors for a small signal. When a small signal is present, the linear terms are expected to dominate the second-order and higher-order terms so that the optimum detector is closely approximated by the terms involving only first partial derivatives. Indeed, for the case of uniform phase when the probability density function factors into probability density functions of amplitude and phase, the second-order term has expected value 0 with signal present up to fourth-order signal terms, in contrast to the linear term which has nonzero expected value with signal terms of second order. An outline of the argument follows.

Observe that

$$
\begin{aligned}
p_{n+s}(x,y) &= p_n(x-a,y-b) \\
&= \hat{p}\left(\sqrt{(x-a)^2+(y-b)^2}\right)q\left(\arctan\frac{y-b}{x-a}\right) \\
&\cong \hat{p}\left(A-\frac{ax+by}{A}\right)q\left(\theta-\frac{bx-ay}{A^2}\right) \\
&\cong \left[\hat{p}(A)-\frac{ax+by}{A}\hat{p}'(A)\right]\left[q\left(\theta)-\frac{bx-ay}{A^2}\right)q'(\theta)\right] \\
&= \left[\hat{p}(A)-(a\cos\theta+b\sin\theta)\hat{p}'(A)\right] \\
&\quad\times\left[q(\theta)-\frac{b\cos\theta-a\sin\theta}{A}q'(\theta)\right].
\end{aligned}
$$

Now we are ready to calculate

$$E_{n+s}[D_2(x,y)] = E_{n+s}\left[\left(\frac{ax+by}{A}\right)^2 \frac{\hat{p}''A}{\hat{p}(A)} + \left(\frac{bx-ay}{A}\right)^2 \frac{\hat{p}'(A)}{A\hat{p}(A)}\right],$$

the expected value of the second-order terms under the assumption that $q(\theta) = \frac{1}{2\pi}$. By making use of the fact that $E_n[D_2(x,y)] = 0$,

$$\begin{aligned} E_{n+s}\{D_2(x,y)] = & \\ & -\int_0^\infty \int_0^{2\pi} l[(a\cos\theta + b\sin\theta)^2 \frac{\hat{p}''(A)}{\hat{p}(A)} \\ & + (b\cos\theta - a\sin\theta)^2 \frac{\hat{p}'(A)}{A\hat{p}(A)}] \\ & \times \left(a\cos\theta + b\sin\theta\right)\hat{p}'(A)Ad\theta dA = 0 \end{aligned}$$

because

$$\int_0^{2\pi} [(a\cos\theta + b\sin theta)^3 d\theta = 0$$

and

$$\int_0^{2\pi} [(a\cos\theta + b\sin\theta)(b\cos\theta - a\sin\theta)^2 d\theta = 0\,.$$

We now give a geometric interpretation of the first-order detector. Recall that the first-order detector has the form:

$$D_1(x,y) = \left(\frac{p'(A)}{p(A)} - \frac{1}{A}\right)\left(\frac{ax+by}{A}\right) + \frac{q'(\theta)}{q(\theta)}\frac{-ay+bx}{A}$$

Call the operation in the first term of the above sum *amplitude processing* and the second term *phase processing*. Let $s = a + ib$ and $z = x + iy$. Then $(ax + by)/A$ and $(-ay + bx)/A$ are the projections of s onto z and $z^\perp = iz$, respectively. Note that $\frac{ax+by}{A} = \mathrm{Re}\left[\frac{z}{|z|}s^*\right]$ and $\frac{-ay+bx}{A} = \mathrm{Im}\left[\frac{z}{|z|}s^*\right]$, where $s = a + ib$ and $z = x + iy$. Let $g(A) = \frac{\hat{p}'(A)}{\hat{p}(A)} = \frac{p'(A)}{p(A)} + \frac{1}{A}$ and $h(\theta) = -\frac{q'(\theta)}{q(\theta)}$. The amplitude processing can be viewed as consisting of a first step $z_k \to g(|z_k|)\frac{z_k}{|z_k|}$, where the "nonlinearity" $G(|z_k|)$ is used to 1"weight" the samples, followed by a second signal reconstruction step:

$$g(|z_k|)\frac{z_k}{|z_k|} \to g(|z_k|)\,\mathrm{Re}\left[\frac{z_k}{|z_k|}\,\frac{s_k^*}{|s_k^*|}\right].$$

Similarly, the phase processing can be viewed as a two-step process: weighting the samples and reconstructing the signal. The real part contains all the signal information, so taking the real part is a natural thing to do in either the amplitude or phase processing. Note further that the amplitude and phase processing naturally complement each other; amplitude processing depends on the projection of the signal onto the received signal-plus-interference baseband sample and phase processing depends on the projection of the signal onto the received signal-plus-interference baseband sample rotated counterclockwise by 90°.

As for in-phase and quadrature quantities, the above result can be extended to unknown signals by replacing the various powers of a and b with the movements of the real and imaginary parts of the desired signal.

We have seen that the optimal detectors for amplitude and phase involve projections of the signal onto the received signal and the received signal rotated counterclockwise by 90°. Considerable insight into adaptive locally optimum processing is gained by characterizing the signal information contained in these projections. Toward this end, consider a baseband sample $z = n + n + s$ with u and n structured and random components of the noise, respectively. The structured component of the noise is often referred to as interference, while the random component is referred to as noise.

The amplitude and phase of a complex baseband sample z can be decomposed into vector components parallel and perpendicular to the interferer vector u as shown in figure 1,* provided that the signal and noise are much less than the interferer. Let $s_{||}$ and $s_{\perp}$ denote the projection of the vector s onto the vector u and the vector u rotated counterclockwise by 90° and let $n_{||}$ and $n_{\perp}$ denote the projections of the vector n onto the vector u and the vector u rotated counterclockwise by 90°. Let $S_{||}$, $S_{\perp}$, $N_{||}$, and $N_{\perp}$ denote the real numbers defined by $s_{||=S_{\perp}\frac{u_{\perp}}{|u|}}$, $s_{\perp} = S_{\perp}\frac{u_{\perp}}{|u|}$, $n_{||} = N_{||}\frac{u}{|u|}$, and $n_{\perp} = N_{\perp}\frac{u_{\perp}}{|u|}$, where $u_{\perp}$ denotes the vector u rotated counterclockwise by 90°. Then $|r| \approx |u| + S_{||} + N_{||}$ and $\theta \approx \phi + \frac{S_{\perp}}{|U|} + \frac{N_{\perp}}{|u|}$, where ϕ denotes the phase of the interferer.

Each of the projections of the signal contains half of the available information on the signal. This follows from the fact that

$$S_{||} = \frac{s + e^{i2\phi}(s*)}{2} \text{ and } s_{\perp} = \frac{s - e^{92\phi}(s*)}{2} ,$$

where $s*$ is the complex conjugate of s. For a known signal, if good estimates of both $s_{||}$ and $s_{\perp}$ are available, their sum provides a good estimate of the signal up to its magnitude. In some cases, adaptive locally optimum processing of amplitudes may be successful, while that of phases is unsuccessful, and in other cases processing of phases may be successful, while that of amplitudes is unsuccessful. For a known signal, if a good estimate of either $s_{||}$ or $s_{\perp}$ is available, the signal can also be recovered if the output is despread for a bandspread signal, spectrally analyzed for a narrowband signal, or beamformed. In each of these cases, the signal processing following adaptive locally optimum processing separates the desired signal term s from the undesired signal distorting term $s * e^{i2\phi}$ because of the presence of ϕ.

As an example, let us consider direct sequence bandspread communications in some detail. The known signal detection by correlation in this case is called despreading. This constitutes the situation for which most of the theoretical work on adaptive locally optimum processing for communications has been apllied. Let s be the spreading sequence (the signal of known structure). Let B denote a set of successive baseband samples of the received signal associated with a given bit. Then despreading consists of the complex sample correlation:

$$\sum\nolimits_{z_j \text{ in B}} z_j \frac{s_j^*}{|s_j|} .$$

Observe then that the outputs of the amplitude and phase algorithms after despreading are $\sum_{z_j \, in \, B} \frac{s_j + e^{i2\phi_j}(s_j^*)}{2} \frac{s_j^*}{|s_j|}$ and $\sum_j \frac{s_j - e^{i2\phi_j}(s_j^*}{2} \frac{s_j^*}{|s_j|}$, respectively. The undesirable term $\sum_{z_j \, in \, B} \frac{e^{i2\phi_j}(s_j^*)}{2} \frac{s_j^*}{|s_j|}$ is small

* All figures appear at the end of the report.

compared with the desirable term $\sum_{z_j \text{ in } B} \frac{s_j}{2} \frac{s_j^*}{|s_j|}$, provided the phase of u^2 is uncorrelated with the phase of s_j^*, a technical condition that is usually met, and the number of chips spreading a bit is at least 10, another condition that is traditionally far exceeded by existing bandspread communication signals. This same argument applies to a narrowband signal whose structure is known when the interference has phase that can be modeled as random. This is one reason why detection techniques developed for communications have application to undersea surveillance.

For a signal of unknown structure, there is no question of recovering the signal. However, it may still be of considerable interest to detect its presence, such would be the case in the frequency domain when the phase of successive complex Fourier coefficients for a given frequency cannot be estimated and it is desirable to decide whether there is a narrowband signal present at that frequency. However, even for the case of unknown signal, it is still of interest to perform both amplitude and phase processing, if possible, because amplitude processing provides energy proportional to the projection of the signal onto the received vector. For a signal vector rotating around the interferer, this leads to a detector output over time with periods indicating the presence of the signal separated by periods in which it does not appear. Successful processing of phase tends to fill in the segments when the signal does not manifest itself after amplitude processing.

From our discussions above, we see that the processing of phase uses information about the projection of the signal onto the interferer rotated counterclockwise by $90°$ and, therefore, depends on the relative phases of the signal to the interferer component (or the dominant noise component when interferer is not present). Furthermore, for the phase decomposition, signal phase and noise phase divided by amplitude are approximated by the projection of the signal vector onto the dominant interferer vector rotated counterclockwise by $90°$. For this reason, the performance of phase processing is always somewhat dependent on received signal baseband amplitude statistics. Another consideration related to amplitude and phase processing is that the processing is nearly optimum when the interferer is much stronger than the desired signal. However, in many cases, an interferer-to-signal ratio of two is sufficient for the processing to be more effective than traditional processing. When applicable, it is always better to perform both amplitude and phase processing.

SIGNAL MODELS

Different models for the signal lead to different processing algorithms. We discuss the following:

1. signal of known structure and unknown received power, and
2. signal of unknown received power and independent amplitude and uniform phase.

Case 1 includes processing bandspread communication signals and the use of spectral analysis to detect a narrowband signal of unknown frequency. For bandspread signals, when each bit is spread by a known chip sequence, the timing of the received signal and the internally stored chip sequence is aligned through the process of synchronization. Synchronization is usually done through processing of known bits. Information bits are detected by correlation over the bit duration of the received signal with the stored chip sequence for that time interval. Thus case 1 applies to those adaptive locally optimum processing techniques used to detect bandspread communication signals. For detecting a narrow band signal by using spectral analysis, we can view the Fourier transform as the simultaneous correlation of the received signal with a family of candidate signals of known structure. In this sense, case 1 applies to locally optimum processing of time domain signals that are then spectrally analyzed to detect the presence of narrowband signals.

Case 2 applies, in particular, to the detection of broadband signals through energy detection. For such signals, it is reasonable to suppose that the phase is uniformly distributed on $(-\pi, \pi)$ and as a result the mean values of the real and imaginary signal components are zero. Case 2 can be reduced to case 1 for the detection of the presence of a narrowband signal for which the center frequency of a Fourier frequency bin is nearly the same as the frequency of a received narrowband signal received in an single mode. In this case the signal component of the complex Fourier coefficient can be modeled as a fixed unit vector of unknown phase times an unknown constant.

For a signal of known structure, we have seen in the previous sections how the optimum detector can be implemented given a probability density function of the interference. We have also discussed the implementation given the probability density functions of interferer amplitude and phase under the assumption that they are independent. We now briefly discuss the unknown signal structure case.

Signal of Unknown Structure and Uniform Phase

In this case, $\frac{s_k}{|s_k|}$ is assumed unknown. To implement the optimum detector, we replace a_k by $E(a_k) = 0$, b_k by $E(b_k) = 0$, a_k^2 by $E(a_k^2)$, $a_k b_k$ by $E(a_k b_k)$, and b_k^2 by $E(b_k^2)$. In addition, if the amplitude and phase of the signal are assumed independent, a natural assumption for signals with uniform phase, the following hold

$$\begin{aligned} E(a_k^2) &= E(|s_k|^2 \cos^2 \psi_k) = \frac{1}{2} E(|s_k|^2), \\ E(a_k b_k) &= E(|s_k|^2 \cos \psi_k \sin \psi_k) = 0, \text{ and} \\ E(b_k^2) &= E(|s_k)^2 \sin^2 \psi_k) = \frac{1}{2} E(|s_k|^2) . \end{aligned}$$

In any case, the only natural way to implement the second-order detector is to treat the unknown nonzero parameters as equal. With the above assumption, the second-order detector, with $c = \frac{1}{2E(|s|^2)}$, reduces to $\frac{1}{p(x,y)}\left[\frac{\partial^2 p(x,y)}{\partial x^2} + \frac{\partial^2 p(x,y)}{\partial y^2}\right]$, which can be implemented given that an estimate of the probability density function $p(x,y)$ is available.

If in addition, the interferer is assumed to have independent amplitude and phase and the phase uniform, the second-order detector becomes $\frac{\hat{p}''(A)}{\hat{p}(A)} + \frac{\hat{p}'(A)}{A\hat{p}(A)} = \frac{p''(A)}{p(A)} - \frac{1}{A}\left(\frac{p'(A)}{p(A)} - \frac{1}{A}\right)$.

Narrowband Nonzero-Mean Frequency Domain Signal

If adaptive locally optimum detection is to be applied in the frequency domain, and the desired signal is narrowband and stable, this information can be used to reduce the general detector for frequency domain signals to two cases that correspond closely to the two cases already discussed for time domain signals. One case occurs when the center frequency of the Fourier transform frequency bin containing the signal and the frequency of the signal are very nearly the same, so that the Fourier coefficient of the signal is nearly stationary. In this case, the signal can be modeled by a unit vector u times the expected value of the signal magnitude. The second case occurs when the complex Fourier coefficient of the signal rotates around the origin so that during the detection interval the expected values of the real and imaginary components of the signal can be taken as zero. We treat the nonzero-mean case in this subsection and the zero-mean case in the next section.

Let the signal vector be approximated by $(u, v)|s_k|$ with (u, v) an unknown, but fixed, unit vector. Let $\{(u_m, v_m)|1 \leq m, + M\}$ be a set of unit vectors corresponding to equally spaced angles between 0 and π. After the processing, the magnitude of the result is the quantity examined to determine the presence of the narrowband signal so that $-(u, v)$ and (u, v) would provide the same performance. Then, this case is reduced to the signal of known structure case as follows. Let

$$D(x_1, y_1, ..., x_N y_N) = \max_m \frac{1}{N} \left| \sum_{k=1}^{N} D_m(x_k, y_k) \right|,$$

where

$$\frac{1}{N} \sum_{k=1}^{N} D_m(x, y) = \frac{1}{N} [\sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial p(x_k, y_k)}{\partial x} u_m + \sum_{k=1}^{N} \frac{1}{p(x_k, y_k)} \frac{\partial p(x_k, y_k)}{\partial y} v_m]$$

is a line detector. A minimal set of unit vectors to implement the above detector would be

$$\{(u_m, v_m)|1 \leq m \leq 4\} = \left\{ (1, 0), (0, 1), \frac{1}{\sqrt{2}}(1, 1), \frac{1}{\sqrt{2}}(1, -1) \right\}.$$

The argument presented for a signal of known structure can then be applied with $s_k = (u_m, v_m)$ and with m chosen as the value leading to the maximum value of the line detector $\frac{1}{N} \sum_{k=1}^{N} D_m(x, y)$. In particular, for independent amplitude and uniform phase, the second partial derivatives term of the detector has expected valued zero-order up to fourth-order in the signal strength, while the detector given here has nonzero expected value with second-order terms in signal strength.

Narrowband Zero-Mean Frequency-Domain Signal

There are two ways that this case can occur; one way is that the signal vector is rotating around the origin at a fixed rate. This case is not of much interest to us, for presumably the center frequency of the Fourier bin containing the signal will have frequency quite close to that of the signal, and as long as the integration time is not too long, the above case will occur. The other way is that the signal vector can be modeled as having random phase. In either case, the signal can be treated as unknown as in the second case considered. In addition, the linear partial derivative terms have zero expected value when a signal is present.

Summary on Signal Modeling

The signal model can be taken as either of known structure or of unknown structure. Furthermore, for processing independent amplitudes and uniform phases, the two cases are complementary. For the known structure signals, the detector given by the first partial derivative terms dominates the detector, while for the unknown signal case, the detector given by the second partial derivatives dominate.

For processing frequency domain beamformed data, the natural approach is to perform processing under both assumptions and add the results. In this manner, regardless of which assumption on the received signal, nonzero-mean real or nonzero-mean imaginary part, or zero mean real and zero-mean imaginary parts, best describe the signal, and depending to some extent on the interference first and second partial relative magnitudes, the combined processing should improve detection of the signal.

Interference Models

In the following section, we discuss the adaptive locally optimum algorithms that arise for different interference models for the cases of signals of either known structure or unknown structure. We first discuss the case of Gaussian noise to show that the adaptive locally optimum processing reduces to traditional processing. We then show that appropriate preprocessing can sometimes transform a baseband sample to a quantity, still containing a recoverable signal term, which is Gaussian even though the original sample contains structured interference. This approach leads to important algorithms for undersea surveillance, in particular to algorithms that can be used to detect weak narrowband signals masked by stronger narrowband signals of nearly the same frequency. The algorithms based on preprocessing can be viewed as special cases of a more general class of adaptive locally optimum processing algorithms that are derived from modeling the interference plus noise by multistate models. The preprocessing is used to assign the samples to an interference model with a single state, while for the more general multistate models, the samples are assigned probabilistically to the states.

Gaussian Noise

The simplest interference model for a real variable x is that the probability density function of x is $p(x) = \frac{1}{\sqrt{2}\pi\sigma} e^{-\frac{x^2}{2\sigma^2}}$ with σ^2 the known variance of the Gaussian noise. For this case,

$$D_1(x) = -\frac{d \ln (p(x))}{dx} = \frac{x}{\sigma^2} \text{ and } D_2(x) = \frac{1}{p(x)} \frac{d^2 p(x)}{dx^2}$$

and as a result, the first-order detector reduces to a linear transformation and the second-order detector to its square up to the constant $\frac{1}{\sigma^2}$.

When the successive samples contain an interferer term that is correlated, it is sometimes possible to remove this correlation by processing appropriate linear combinations of the original samples with the result that the only interference remaining is Gaussian, that is the remaining interference can be modeled by a one-state Gaussian mixture model. We present two examples, an amplitude case and a phase case, for which the results are surprising, have important applications to undersea surveillance, and are new.

One-State Amplitude Model

Suppose that a small signal is received in the presence of a large constant amplitude interferer and modest levels of Gaussian noise. Suppose further that successive sample signal components are uncorrelated and independent of the sample interferer components. Consider the amplitude preprocessing step

$$A_j \to A_j - \frac{1}{2N} \sum_{k=-N, k\neq 0}^{N} proj_{z_{j+k}} A_{j+k} .$$

The signal term in the preprocessor output is

$$proj_{z_j}(s_j) - \frac{1}{2N} \sum_{k=-N, k\neq 0}^{N} proj_{z_{j+k}}(s_{j+k}) .$$

The desired signal term in this expression depends in general on N, the sample rate, and the relative phase of the signal and interference. The term $\frac{1}{2N} \sum_{k=-N, k\neq 0}^{N} proj_{z_{j+k}}(s_{j+k})$. can be viewed as a signal

distortion term. This term can be shown to be small under quite general conditions. We outline the argument for two cases: (1) random phase angles between the signal and interference and (2) linear phase difference between the signal and interference. The first case usually applies when either the signal or interference is a broadband signal, and the second case applies when both the signal and interference are narrowband signals.

The signal distortion term can be viewed as the distance from the origin of a one-dimensional random walk when the relative phase of the signal and interfernence are random. The individual steps are the projections $proj_{z_{j+k}}(s_{j+k})$. The expected distance from the origin is roughly proportional to the square root of the number of steps 2N times an average step size. For the amplitude case, this distance is

$$\left(\sqrt{2N}\right) avg\left(\frac{|proj_{z_{j+k}}(s_{j+k})|}{2N}\right) \cong \frac{1}{2\sqrt{2N}} avg|s| ,$$

because

$$avg(|proj_{z_{j+k}}(s_{j+k})|) = \frac{1}{2} avg|s| .$$

This means that the distortion energy is expected to be $\frac{1}{2N}$ times the signal energy. Thus, a desire to neglect the signal distortion term imposes a mild restraint on N. Consider the case when all the weights are about the same size, then $M = 2N$. Here, $N = 4$ would lead to about 12-1/2% distortion, so that $N = 4$ is a practical lower limit on the value of N. More generally, for adaptive weight cases, $N = 8$ is a practical lower limit on the value of N.

For the case of constant frequency signal and interference, the magnitude of the signal distortion term depends on the frequency difference between signal and interferer and the sample rate. We exploit the fact that the signal vector rotates around the interferer vector at a fixed frequency. The geometry governing this situation is shown in figure 1. The analysis makes use of the coordinate system defined relative to the interferer vector. Observe that $proj_{z_{j+k}}(s_{j+k}) = proj_{\hat{u}_{j+k}}(\hat{s}_{j+k})$, where $\hat{u}_{j+k} = R_{-kf_u}(u_{j+k})$ and $\hat{s}_{j+k} = R_{-kf_u}(s_{j+k})$, where f_u is the ratio of the frequency and $R_\psi(u)$ is a rotation of u through ψ radians. Next, observe that $\hat{u}_{j+k} = u_j$ so that the projections resulting from the vectors $\hat{s}_{j+k}$ onto u_j are identical to those of s_{j+k} onto u_{j+k}. Let ψ denote the angle between s_j and u_j. Then

$$proj_{z_j}(s_j) - \frac{1}{2N} \sum_{k=-N,k\neq 0} proj_{z_{j+k}}(s_{j+k}) \cong |s| \cos\psi - \frac{1}{2N} \sum_{k=-N,k\neq 0}^{N} proj_{\hat{u}_{j+k}}(\hat{s}_{j+k\|} .$$

Note that

$$proj_{\hat{u}_{j+k}}(\hat{s}_{j+k\|}) = |s| \cos(k(f_s - f_u) + \psi) ,$$

as seen from figure 2, where f_u is the ratio of the signal frequency to the sampling frequency. It follows that

$$\sum_{k=-N,k\neq 0}^{N} proj_{\hat{u}_{j+k}}(\hat{s}_{j+k\|} = \sum_{k=-N,k\neq 0}^{N} |s| \cos(k(f_s - f_u) + \psi) .$$

This last sum can be viewed as an approximation to an integral after adding back in the k = 0 term. Let $\Delta f = f_s - f_u$ and suppose that $(2N + 1)\Delta f \leq 1$. Then

$$
\begin{aligned}
&\cos\psi + \sum_{k=-N,k\neq 0}^{N} \cos(k(f_s - f_u) + \psi) \\
&\cong \int_{-N}^{N} \cos(t\Delta f + \psi)dt \\
&= \frac{1}{\Delta f}[\sin(N\Delta f + \psi) - \sin(-N\Delta f + \psi)] \\
&= \frac{1}{\Delta f}(\sin N\Delta f)\cos\psi\,.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&proj_{z_{j+k}}(s_j - \frac{1}{2N} \sum_{k=-N,k\neq 0}^{N} proj_{z_{j+k}}(s_{j+k} \\
&\cong |s|\cos\psi + \frac{|s|\cos\psi}{2N} - \frac{1}{N\Delta f}(\sin N\Delta f)|s|\cos\psi \\
&= proj_{z_j}(s_j)(1 + \frac{1}{2N} - \frac{1}{N\Delta f}(\sin N\Delta f))\,.
\end{aligned}
$$

It follows that N should be taken large enough so that the term involving $\sin N\Delta f$ can be neglected. In general, N should be chosen at least as large as the sampling frequency divided by the difference of the signal and interferer frequencies, i.e., as the frequency resolution of the spectral analysis.

For circularly distributed Gaussian noise, the variance of the Gaussian noise in the output of the processing algorithm under discussion is

$$
(1 + 2N\left(\frac{1}{2N}\right)^2 var(n) = \left(1 + \frac{1}{2N}\right)var(n)
$$

with $var(n)$ equal to the variance of the real and imaginary components of the Gaussian noise component of the received baseband samples.

One-State Phase Model

Suppose that the interference has linear phase (that is, it has constant frequency) and the signal and Gaussian noise are uncorrelated. Consider the preprocessing step to remove the phase of the interferer phase contribution

$$
\theta_j \rightarrow \Delta_k\theta_j = \theta_j - \frac{1}{2}(\theta_{j-k} + \theta_{j+k})
$$

followed by

$$
\theta_j \rightarrow \frac{1}{N}\sum_{k=1}^{N}\Delta_k\theta_j = \theta_j - \frac{1}{2N}\sum_{k=1,k\neq 0}^{N}\theta_{j+k}\,.
$$

The signal term in this equation involves projections onto iz divided by A. For phase processing, the signal and Gaussian noise after this processing depends on the values of the ratios $\frac{A_j}{A_{j+k}}$ as well as N.

The signal term in the output of the algorithm is approximately

$$
\left(\frac{proj_{iz_j}(s_j}{A_j} - \frac{1}{2N}\sum_{k=-N,k\neq 0}\frac{proj_{iz_{j+k}}(s_{j+K}}{A_{j+k}}\right)A_j\,.
$$

The signal distortion term is expected to be small when the signal-to-interferer relative phase is random. The random walk argument used for the amplitude case still applies provided that the value of $\frac{A_j}{A_{j+k}}$ is uncorrelated with the relative phase ψ_j. The distortion estimate using the random walk is as before except that the average step is multiplied by the average value of the absolute values of the ratios $\frac{A_j}{A_{j+k}}$.

The linear relative phase case, that is the case for narrowband signal and interference, poses some difficulties when the amplitude of the interferer varies from sample to sample. However, the case when the interferer amplitude is a constant follows as before by replacing u by iu in the argument. In particular, we obtain the approximation

$$\begin{aligned} & proj_{iz_j}(s_j - \frac{1}{2N} \sum_{k=-N, l \neq 0} proj_{z_{j+k}}(s_{j+k}) \\ & \cong proj_{iz_j}(s_j)(1 + \frac{1}{2N} - \frac{1}{N\Delta f}(\sin N\Delta f)) \,. \end{aligned}$$

The Gaussian noise variance of the one-state phase algorithm for sample j is

$$\left[1 + \left(\frac{1}{2N}\right)^2 \sum_{k=-N, k \neq 0}^{N} \left(\frac{A_j}{A_{j+k}}\right)^2 \right] var(n) \,.$$

This reduces to the estimate of the variance of the Gaussian noise for the one-state amplitude algorithm in a constant amplitude interferer.

This example shows that even though the original samples had uniform phase and the processing of them under the assumption of independent samples led to no processing gain, a preprocessing step to remove the correlation of the interference phases and then adaptive locally optimum processing (here trivial, but in multistate model state to be treated next, not so) can lead to considerable processing gains.

One-State Model Algorithm Performance

For a constant amplitude, constant frequency interferer with the signal and noise satisfying the assumptions above, the signal can be reconstructed by forming:

$$\begin{aligned} (A_j, \theta_j) \rightarrow & \left[A_j - \frac{1}{2N} \sum_{k=-N, k \neq 0} A_{j+k} \right] \frac{z_j}{|z_j|} \\ & + \left[\theta_j - \frac{1}{2N} \sum_{k=1, k \neq 0}^{N} \theta_{j+k} \right] iz_j \,. \end{aligned}$$

If the signal is of known structure, this preprocessing step is followed by taking the real part of this quantity after multiplication by $\frac{s_j^*}{|s_j|}$. If the signal is of unknown structure, this preprocessing step is followed by taking the norm of the quantity.

For a known signal, the amplitude and phase terms provide the projections of the signal onto z and iz. The signal is reconstructed with distortion depending on N, small if N is large. This is important

for undersea surveillance applications because it means that clues used to classify signals are preserved by the processing.

The variance of the Gaussian noise increases from $2\sigma^2$ to $2\sigma^2\left(1 + \frac{1}{2N}\right)$ as a result of the processing. This is the minimal variance for any filter of the complex samples using 2N weights whose sum is 1. Therefore, the one-state processing represents a nonadaptive implementation of a Wiener filter. See Bond (1992) for a more general treatment of the relationship between adaptive locally optimum processing and adaptive Wiener filtering.

If only the one-state amplitude algorithm or the one-state phase algorithm is effective, the insertion loss, the difference in dB between the output signal-to-Gaussian noise and input signal-to-Gaussian noise, for the algorithm used is about 3 dB, because the expected signal term energy is reduced by a factor of four while the Gaussian noise variance is reduced by a factor of two.

The expected output signal-to-noise ratio can be calculated for the second-order one-state detector $A \rightarrow D_2(A) = (A - \mu)^2$ by using the expansion of the baseband samples into the amplitude of the interferer plus the projections of the signal and Gaussian noise onto the interferer:

$$\begin{aligned} E_{n+s}(D_2(A)) &\cong E_{n+s}\left[(proj_u(s) + proj_u(n))^2\right] \\ &= E_{n+s}\left(proj_u^2(s)\right] + E_{n+s}\left[proj_u^2(n)\right] \end{aligned}$$

under the assumption that $proj_u(s) = |s|\cos\phi$ and $proj_u(n) = |n|\cos\psi$ with ϕ and ψ independent random variables whose values are uncorrelated with $|s|$ and $|n|$. Also, $E_n(D_2(A)) \cong E_n[proj_u^2(n))]$ so that $E_{n+s}(D_2(A)) - E_n(D_2(A)) = E_n[proj_u^2))] = \frac{1}{2}|s|^2$. Note that $\sigma_n^2(D_2(A)) \cong E[proj_u^4(n)] = E[|n|^4]E[\cos^4\psi] = 3\sigma^4$ using $E[|n|^4] = E[(n_x^2 + n_y^2)^2] = E[n_x^4 + 2n_x^2 n_y^2 + n_y^4] = 8\sigma^4$, where n_x and n_y are the real and imaginary components of the Gaussian noise n and assuming the relative phase ψ is uniformly distributed. It follows that the signal-to-noise ratio after the processing is approximately $\frac{1}{2\sqrt{3}}\frac{|s|^2}{\sigma^2}$, where σ^2 is the variance of the real and imaginary components of the Gaussian noise component of the received signal.

Consider estimating the expected preprocessor output signal-to-noise ratio for the one-state symmetric phase-difference model in a similar way to that for amplitude one-state model. For the second-order detector,

$$\Delta\theta \rightarrow D_2(\Delta\theta)A^2 = (\Delta\theta - \lambda)^2 A^2$$

under the assumption that ϕ and ψ are independent random variables whose values arde uncorrelated with $|s|$ and $|n|$.

$$\begin{aligned} &E_{n+s}(D_2(\Delta\theta_j)A_j^2) - E_n(D_2(\Delta\theta_j)A^2) \\ &\cong E_n[(\Delta(proj_{u\perp}(s))^2], \end{aligned}$$

where

$$\begin{aligned} \Delta(proj_{u\perp}(s)) &= |s_j|\cos\phi_j \\ &- \frac{1}{2}\left(\frac{A_j}{A_{j+1}}|s_{j-1}|\cos\phi_{j-1} + \frac{A_j}{A_{j+1}}|s_{j+1}|\cos\phi_{j+1}\right). \end{aligned}$$

In general, the expectation on the right-hand side of the equation depends on signal levels and baseband sample amplitude correlation properties as well as the relative phase correlation properties for

the three successive samples j-1, j, and j+1. In practice, the fact that the signal component can be small in this expression limits the use of the symmetric phase difference algorithm.

Assume that $\frac{A_j}{A_{j-1}} \cong 1$ and $\frac{A_j}{A_{j+1}} \cong 1$, $|s_{j-1}| \cong |s_j|$, $|s_{j+1}| \cong |s_j|$, and the relative phases are independent random and uniformly distributed. Then

$$E_n[(\Delta(proj_{u\perp}(s))^2] \cong |s|^2\left(\frac{1}{2\pi}\right)^3$$
$$\times \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\cos\phi - \frac{\cos\psi + \cos\varphi}{2}\right)^2 d\phi d\psi d\varphi$$
$$= \frac{|s|^2}{4}\,.$$

Assume that $\frac{A_j}{A_{j-1}} \cong 1$ and $\frac{A_j}{A_{j+1}} \cong 1$. Treating the three Gaussian noise terms in the detector as independent, identically distributed, and independent of coordinate system, it follows that

$$\sigma_n^2(D_2(\Delta\theta)) \cong E\left[\left(n_x - \frac{1}{2}(\hat{n}_x + \tilde{n}_x)\right)^4\right]$$

with n_x, $\hat{n}_x$ and $\tilde{n}_x$ zero-mean random variables with variance σ^2. Note that

$$E\left[\left(n_x - \frac{1}{2}(\hat{n}_x + \tilde{n}_x)\right)^4\right] =$$
$$E(n_x^4) + \frac{6}{4}E\left[n_x^2\left(\hat{n}_x^2 + \tilde{n}_x^2\right)\right] + \frac{1}{16}E\left[\hat{n}_x^4 + 6\hat{n}_x^2\tilde{n}_x^2 + \tilde{n}_x^4\right]$$
$$= \left[3 + \frac{6}{4} + \frac{6}{4} + \frac{1}{16}(3 + 6 + 3)\right]\sigma^4 = \frac{27}{4}\sigma^4\,.$$

It follows that for this special case the expected signal-to-noise ratio for the symmetric phase-difference one-state second-order detector is approximately $\frac{2}{3\sqrt{3}}\,\frac{|s|^2}{\sigma^2}$. For most cases, this is probably the best that can be expected from the phase processing.

The insertion loss for the second-order amplitude algorithm alone is 5.4 dB and for the second-order symmetric phase-difference algorithm is 4.1 dB. Combining the results should lead to improved line detection over either alone by producing a better defined line, but it is not clear how to calculate the signal-to-noise ratio associated with this noncoherent combining process.

Multistate Models for Interference Amplitudes and Phases

Gaussian mixture models have been proposed for modeling ship-generated acoustic noise. Gaussian kernel representations of probability density function have been used to model jamming plus background noise for communications applications. Recall that the Gaussian mixture model arises from modeling the interference plus background noise of the real and imaginary components of the baseband samples as independent identical zero-mean Gaussian distributions. Interference states are distinguished by the variance of the Gaussian interference plus noise. The probability density function for a complex sample is obtained by treating the in-phase and quadrature samples as independent. We introduce another mixture model related to kernel representations of probability density

functions. This model involves distributions with nonzero means and we will call it a noncentral mixture model.

Different noncentral mixture models are used to model amplitude and phase or symmetric phase differences. In this context, the sample amplitudes and phases are treated as independent and each is then modeled by a noncentral mixture model. For these models, the interference-plus-noise samples are assumed to contain deterministic and random components. The states of the model are determined by the deterministic interference with the random component treated as stationary for modeling. For a noncentral mixture model of sample amplitudes, the deterministic interferer is assumed to take on discrete values of amplitude that define the states of the model. The probability density function of the amplitudes of the samples in a state is a noncentral truncated Gaussian distribution with variance after truncation equal to the variance of the background noise.

Figure 3 presents scatter plots for Gaussian noise and a Gaussian noise-plus-CW interferer. The Gaussian noise case represents the scatter of complex samples for a typical Gaussian mixture model state, while the Gaussian noise-plus-CW case represents the scatter plot of complex samples for a typical noncentral mixture model state. Observe that the scatter plots presented in the companion report to this report, “Gaussian Mixture Models for Undersea Surveillance,” for the MDA hydrophone data are closer in appearance to that of Gaussian noise than Gaussian noise plus CW interferer. Nevertheless, it seems reasonable to suppose that under some conditions acoustic interference might be better modeled by a noncentral model than by a Gaussian mixture model, and this is one reason the theory of noncentral mixture models is discussed. For communication applications, interference is often better modeled by noncentral mixture models than by central mixture models. In any case, the discussion of noncentral mixture models together with Gaussian mixture models provides additional insight into the theory of adaptive locally optimum processing for ocean basin surveillance and for communications.

The states of a Gaussian mixture model have probability density

$$\begin{aligned} p_k(x,y) &= p(x)p(y) \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{x^2}{2\sigma_k^2}}\right)\left(\frac{1}{\sqrt{2\pi}\,\sigma_k}\frac{e^{-y^2}}{2\sigma_k^2}\right) = \frac{1}{2\pi\sigma_k^2}\left(e^{\frac{-A^2}{2\sigma_k^2}}\right). \end{aligned}$$

The corresponding density functions for A and θ are $p_k(A) = \frac{A}{\sigma_k^2}\left(e^{-\frac{A^2}{2\sigma_k^2}}\right)$ and $q(\theta) = \frac{1}{2\pi}$ for $-\pi < \theta \le \pi$, which satisfy $p_k(x,y) = \frac{p(A)}{A}q(\theta)$.

The probability density function of the sample amplitudes in the k-th state for a noncentral mixture model has probability density function

$$p_k(A) \cong \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(A-\mu_k)^2}{2\sigma^2}},$$

where μ_k is the mean value of the amplitudes of the samples in the k-th state. The fact that $A > 0$ imposes a constraint on the Gaussian component of the interference in the k-th state, namely that $\sigma_n < \mu_k$. This means that the noncentral mixture model is only applicable when the small signal hypothesis applies for all the discrete values of the interference.

The probability density function for the amplitudes of the samples in an S state mixture model, Gaussian or noncentral, has the form

$$p(A) = \sum_{k=1}^{S} p_k(A) p_k ,$$

where $p_k(A)$ is the probability density function of the amplitudes of the samples in the k-th state and p_k is the probability that a sample is in the k-th state.

For a noncentral mixture model, the first-order detector for sample j is

$$\left(-\frac{p'(A_j)}{p(A_j)} + \frac{1}{A_j} \right) \frac{\mathrm{Re}[z_j s_j^*}{|z_j \| s_j|}$$
$$= \left[\frac{1}{\sigma^2} \sum_{k=1}^{S} (A_j - \mu_k) w(j,k) + \frac{1}{A_j} \right] \frac{\mathrm{Re}[z_j s_j^*}{|z_j \| s_j|}$$

and the second-order detector for sample j is

$$\frac{p''(A_j)}{p(A_j)} + \frac{1}{A_j} \left(-\frac{p'(A_j)}{p(A_j)} + \frac{1}{A_j} \right)$$
$$= \frac{1}{\sigma^2} \sum_{k=1}^{S} \left[\frac{(A_j - \mu_k)^2}{\sigma^2} - 1 \right] w(j,k)$$
$$+ \frac{1}{A_j \sigma^2} \sum_{k=1}^{S} (A_j - \mu_k) w(j,k) + \frac{1}{A_j^2}$$

with

$$w(j,k) = p_k \frac{e^{-\frac{A-\mu_k)^2}{2\sigma^2}}}{\sum_{h=1}^{S} p_h e^{-\frac{A-\mu_k)^2}{2\sigma^2}}} .$$

For a Gaussian mixture model, the first-order detector is

$$\left(-\frac{\hat{p}'(A_j}{\hat{p}(A_j)} \right) \frac{Re[z_j s_j^*]}{|z_j \| s_j|} = \left[\sum_{k=1}^{S} \frac{1}{\sigma_k^2} w(j,k) \right] \frac{Re[z_j s_j^*]}{|s_j|}$$

and the second-order detector is

$$\frac{\hat{p}''(A_j)}{p(A_j)} = \sum_{k=1}^{S} \left[\frac{A_j^2}{\sigma_k^4} - \frac{1}{\sigma_k^2} \right] w(j,k) ,$$

when the first-order detector vanishes. In either case, the weights are

$$w(j,k) = p_k \frac{\frac{1}{\sigma_k^2} e^{-\frac{A_j^2}{2\sigma_k^2}}}{\sum_{h=1}^{s} \frac{p_h}{\sigma_h^2} e^{-\frac{A_j^2}{2\sigma_h^2}}} .$$

The first-order and second-order Gaussian mixture model detectors can also be viewed as functions of the complex sample norms. This was the viewpoint for the original derivation of the detectors by Stein, Bond, and Zeidler (1993).

Mixture models for phase are not discussed, since rarely would such models provide significant performance other than for a narrowband interferer received with linear phase, which can better be modeled by a one-state model. A Gaussian mixture model for symmetric phase differences could model the symmetric phase differences of the interferer component of the interference plus noise as zero or uniformly distributed. This is equivalent to modeling the interference as having linear phase at some times and random phase at other times. Another option is to model symmetric phase differences by a noncentral mixture model. In this case, the deterministic interferer is assumed to take on discrete values of phase difference. These values determine the states of the model and as a result the model would have similar structure to that of a noncentral mixture model for amplitudes.

A noncentral mixture model for the probability density function of the sample symmetric phase differences in the k-th state has probability density function

$$p_k(\Delta\theta) \cong \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(\Delta\theta-\mu_k)^2}{2\sigma^2}} ,$$

where μ_k is the mean value of the symmetric phase differences of the samples in the k-th state. Note that, treating the projections of the Gaussian noise onto the interferer vector rotated counterclockwise by 90° as Gaussian is an approximation, which only makes sense if $\frac{proj_{iu}n}{A}$ is small relative to ϕ, i.e., that the small signal hypothesis applies to the Gaussian noise term relative to the interference as well as to the signal term relative to the interference.

A Gaussian mixture model for the probability density function of the sample symmetric phase differences in the k-th state has probabilty density function

$$p_k(\Delta\theta) \cong \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{\Delta\theta^2}{2\sigma^k}} .$$

The probability density function of symmetric phase differences in an S state mixture model, either Gaussian or noncentral, has the form

$$p(\Delta\theta) = \sum_{k=1}^{S} p_k(\Delta\theta) p_k ,$$

where $p_k(\Delta\theta)$ is the probability density function of the symmetric phase differences of the samples in the k-th state and p_k is the probability that a sample is in the k-th state.

For a noncentral mixture model the first-order detector for symmetric phase differences for sample j is

$$\left(-\frac{p'(\Delta\theta_j)}{p(\Delta\theta_j)}\right)\frac{\mathrm{Re}[iz_j s_j^*]}{|z_j|^2|s_j|}$$
$$= \left(\frac{1}{\sigma^2}\sum_{k=1}^{S}(\Delta\theta_j 0 - \mu_k) w(j,k)\right)\frac{Re[iz_j s_j^*]}{|z_j|^2|s_j|}$$

and the second-order detector for sample j is

$$\left(-\frac{p''(\Delta\theta_j)}{p(\Delta\theta_j)}\right)\frac{1}{|z_j|^2|}$$
$$= \left(\frac{1}{\sigma^2}\sum_{k=1}^{S}\left[\frac{(\Delta\theta_j-\mu_k)^2}{\sigma^2} - 1\right] w(j,k)\right)\frac{1}{|z_j|^2} ,$$

with

$$w(j,k) = p_k \frac{e^{\frac{-(\Delta\theta_j - \mu_k)^2}{2\sigma^2}}}{\sum_{h=1}^{S} p_h e^{-\frac{\Delta\theta_j^2}{2\sigma_h^2}}}$$

For a Gaussian mixture model,

$$\left(-\frac{\hat{p}(\Delta\theta_j)}{\hat{p}(\Delta\theta_j)} \right) \frac{Re[iz_j s_j^*]}{|z_j|^2 s_j \backslash}$$
$$= \left(\sum_{k=1}^{S} \frac{1}{\sigma_k^2} w(j,k) \right) \frac{Re[iz_j s_j^*]}{|z_j|^2 |s_j|}$$

and

$$\left(\frac{\hat{p}''(\Delta\theta_j)}{p(\Delta\theta_j)} \right) \frac{1}{|z_j|^2} = \left(\sum_{k=1}^{S} \left[\frac{\Delta\theta_j^2}{\sigma_k^4} - \frac{1}{\sigma_k^2} \right] w(j,k) \right) \frac{1}{|z_j|^2} ,$$

with

$$w(j,k) = p_k \frac{\frac{1}{\sigma_k} e^{-\frac{\Delta\theta_j^2}{2\sigma_k^2}}}{\sum_{h=1}^{s} \frac{p_h}{\sigma_h} e^{-\frac{\Delta\theta_j^2}{2\sigma_h^2}}} .$$

For all the mixture models discussed for amplitude and symmetric phase differences, the weights satisfy the conditions

1. $w(j,k) > 0, k \neq 0$ and
2. $\sum_{k=1}^{S} w(j,k) = 1.$

Furthermore, the weights can be interpreted as the probability that the j-th sample belongs to the k-th state modeling the interference.

The detectors use a fuzzy set model (Klir and Folger, 1988) of sample amplitudes and symmetric phase differences. Each sample z_j is assumed to be a member of one and only one state of the model as shown in figure 4. The model is constructed using one set of baseband samples to represent interference. The model is then used to process another (possibly, the same) set of baseband samples to detect a signal in the presence of the interference. Due to the presence of Gaussian noise, it is desirable to assign set membership using probabilities. In this sense, the locally optimum processing algorithms involve fuzzy set modeling.

All the detectors described in this subsection can be implemented in two ways. One way is to assume a given model and estimate the parameters of the model. The other way is to correlate an empirical distribution of the interference samples with a predetermined family of distributions with known parameters and choose the distribution that has highest correlation with the empirical distribution as a model of the interference. The samples used to construct the interference model in either of

these ways could be drawn from frequency bins adjacent to the bin or two adjacent bins to the bin containing the sample processed for signal. In this manner, the samples used to construct the model are signal free. A third way to obtain detectors involves implicit multistate modeling of the interference. The motivation for approach is the NSE algorithm used by several undersea surveillance systems.

Implicit Models

Noise Equalization Model. The NSE algorithm is closely related to adaptive locally optimum processing. The following discussion provides motivation for adaptive locally optimum processing algorithms obtained through use of implicit models.

The quantity appearing in the Gaussian mixture algorithms has the form

$$F_k(z_j = \frac{|z_j|^2 - 2\sigma_k^2}{\sigma_k^4}\sigma_n^2$$

which can be written as

$$\left[\frac{|z_j|^2 - 2\sigma_k^2}{\sigma_k^2}\right]\frac{\sigma_n^2}{\sigma_k^2}\ .$$

The quantity in brackets, used in the NSE algorithm, is the predicted signal energy normalized by the predicted noise variance. The NSE algorithm can be interpreted using a mixture model in the following way. The states are defined by a prescribed set of possible variances, $\sigma_1^2, \sigma_2^2, \ldots$, of the broadband background noise. That is, the k-th state has variance σ_k^2. Then the implicit model is that all the samples in a given state has independent real and imaginary components with zero-mean Gaussian distributions with variances equal to the variances for the state. Thus the probability density function for state k is

$$p_k(|z|^2) = \frac{1}{2\pi\sigma_k^2}e^{-\frac{|z|^2}{2\sigma_k^2}}.$$

Observe that $p(k|z_j) = 0$ or 1 is implicitly assumed in the NSE algorithm. In other words, a sample is assigned to one and only one state for the NSE algorithm in contrast to the probabilistic assignments for the Gaussian mixture model. Thus the NSE algorithm involves sets rather than fuzzy sets.

The results of the SOSUS noise equalization algorithm are presented to a human on a LOFARGRAM using a grey scale whose intensity is proportional to the logarithm base of $\frac{|z_j|^2 - 2\sigma_k^2}{2\sigma_k^2}$. The logarithmic scale is used to match the displayed grey scale to the response characteristics of the human eye. For broadband noise changing slowly relative to the FFT samples duration, the omission of the factor $\frac{\sigma_{bn}^2}{\sigma_k^2}$ probably has little impact on eye integration.

Implicit Noncentral Mixture Models. The adaptive locally optimum processing techniques developed for communication applications arise out of an implicit representation of a noncentral mixture model. Recall that the assumption of independent amplitude and phase leads to parallel processing of baseband sample amplitudes and phases.

Different adaptive locally optimum processing algorithms arise from the different ways of estimating the true probability density function from the received signal statistics. Many of these algorithms

result from an implicit representation of a noncentral multistate mixture model. The resulting algorithms are practical to implement and have proven particularly effective for communication applications when the interference is non-stationary.

Probability density estimation has been widely studied by mathematical statisticians during the last 20 years (Silverman, 1986). One powerful and efficient approach is to represent the probability density function as a sum of Gaussian kernels defined by the discrete samples. Gaussian kernels can be used to recursively implement adaptive locallly optimum processing algorithms in the following way.

Suppose that X_j is to be processed using the samples $\left\{X_{j+k} | -N \le k \le N, k \ne 0\right\}$ and their statistics. The constant N is chosen depending on the stationarity of the interference and the sample rate. Typical values of N are between 8 and 32 when sampling at the Nyquist rate. The Gaussian kernel estimate of the probability density function is

$$p(X) = \frac{1}{\sqrt{2\pi}\,\sigma_j}\left[\sum_{k=-N,k\ne 0}^{N} e^{-\frac{(X-X_{j+k})^2}{2\sigma_j^2}}\right],$$

where σ_j^2 is the variance of $\left\{X_{j+k} | -N \le k \le N, k \ne 0\right\}$. Given this representation, the probability density quotient involved in the first-order detector for a signal of known structure takes the particularly simple form

$$\left[-\frac{p'(X)}{p(X)}\right]_{X=X_j} = \frac{\frac{1}{\sigma_j^2}\sum_{k=-N}^{N}(X_j - X_{j+k})e^{-\frac{(X_j-X_{j+k})^2}{2\sigma_j^2}}}{\sum_{h=-N,k\ne 0}^{N} e^{-\frac{(X_j-X_{j+h})^2}{2\sigma_j^2}}}$$

$$= \frac{1}{\sigma_j^2}\sum_{k=-N}^{N}(X_j - X_{j+K})w(j,k)\,,$$

where

$$w(j,k) = \frac{e^{-\frac{(X_j-X_{(j+K)}^2}{2\sigma_j^2}}}{\sum_{h=-N,k\ne 0}^{N} e^{-\frac{(X_j-X_{j+h})^2}{2\sigma_j^2}}}.$$

The probability density quotient involved in the second-order detector for a signal of unknown structure, given that the first-order detector has expected value 0, takes the form

$$\left[\frac{p''(X)}{p(X)}\right]_{X=X_j}$$

$$= \frac{\frac{1}{\sigma_j^4}\sum_{h=-N,k\ne 0}^{N}(X_j 0 X_{j+k})^2 k(j, j+k)}{\sum_{h=-N,k\ne 0}^{N} k(j, j+h)} - \frac{1}{\sigma_j^2}$$

$$= \frac{1}{\sigma_j^4}\sum_{k=-N}^{N}(X_j - X_{j+k})^2 w(j,k) - \frac{1}{\sigma_j^2}\,,$$

where $k(j, j+k) = e^{-\frac{(X_j - X_{j+k})^2}{2\sigma_j^2}}$ and with the weights $w(j,k)$ defined as they were for the first-order detector.

For undersea surveillance applications, it is often desirable to process the sample X_j with noise-only samples. However, the samples $\left\{X_{j+k} | -N \le k \le N, k \ne 0\right\}$ would in general contain signal and noise. When processing frequency domain data for narrowband signals, we can use samples $\left\{\tilde{X}_{j+k} | -N \le k \le N, k \ne 0\right\}$ from adjacent frequency and time bins to the one containing the sample to be processed. The resulting algorithms involve implicit modeling of the interference statistics for $X = A$ and $X = \Delta\theta$, respectively. Each sample $\tilde{X}_{j+k}$ defines a state with $\mu_{j+k} = \tilde{X}_{j+k}$ and σ_j^2. The probability density function of the samples in a state $j+k$ is the Gaussian kernel

$$p(X) = \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-\frac{(X - X_{j+k})^2}{2\sigma_j^2}}.$$

The weight $w(j,k)$ is the membership function for the sample $\tilde{X}_j$ in state $j+k$ viewed as a fuzzy set as described earlier and illustrated in figure 4. The weights $w(j,k)$ also have a simple algebraic interpretation for the first-order detector. Let $w(j,j) = 0$. Observe that $w(j,k) > 0$ for $k \ne j$ and that $\sum_{h=-N}^{N} w(j,h) = 1$. These conditions exhibit the calculation of the gain factor term as a filtering operation (Bond and Hui, to appear).

Implicit Gaussian Mixture Models. The general technique of probability density estimation by Gaussian kernels has to be modified to allow nonparametric estimation of Gaussian mixture models. As before, the statistics are estimated recursively. Suppose that z_j is to be transformed using interference statistics estimated from the samples $\left\{\tilde{z}_{j+k} | -N \le k \le N, k \ne 0\right\}$. Consider estimation of the probability density function by

$$p(|z|) = \frac{1}{2N}\left(\sum_{k=-N, k\ne 0}^{N} G_{j+k}(|z|)\right),$$

where $G_{j+k}(|z|) = \frac{1}{2\pi\sigma_{j+k}^2} e^{-\frac{|z|^2}{2\sigma_{j+k}^2}}$ with $\sigma_{j+k}^2 = |\tilde{z}_{j+k}|^2$.

Note that the above estimation is precisely the Gaussian kernel approximation where the parameters are the variances of the different states. We then arrive at the above form by observing that for z in state k, $|z|^2$ is an unbiased estimator of σ_k^2. To avoid numerical difficulties in the actual implementation, samples with energy less than a prescribed level are not used in building the implicit model.

The first-order detector, that is the detector for signal of known structure, for a nonparametric representation of the Gaussian mixture model is

$$\left[\sum_{k=-N, k\ne 0}^{N} \frac{\hat{G}_{j+k}(|z_j|)}{\sigma_{j+k}^2}\right] |z_j| \operatorname{Re}\left[z_j \frac{s_j^*}{|s_j|}\right],$$

while the second-order detector, that is the detector for signal of unknown structure, is

$$\sum_{k=-N, k\ne 0}^{0} \left[\frac{|z_j|^2}{\sigma_{j+k}^2}\right] |z_j| \operatorname{Re}\left[z_j \frac{s_j^*}{|s_j|}\right],$$

where in either case

$$\hat{G}_{j+k}(|z|) = \frac{G_{j+k}(|z|)}{\frac{1}{2N}\sum_{h=-N,k\neq 0}^{N} G_{j+h}(|z|)} .$$

The technique of probability density estimation that uses Gaussian kernels has to be modified slightly to allow nonparameteric estimation of Gaussian mixture model for symmetric phase differences. As for squard norms, the statistics are estimated recursively. Suppose that $\Delta\theta_j$ is to be transformed using interference statistics estimated from the samples $\left\{\tilde{\Delta\theta}_{j+k}| - N \leq k \leq N, k \neq 0\right\}$. Estimate the probability density function by

$$p(\Delta\theta) = \frac{1}{2N}\left(\sum_{k=-N,k\neq 0}^{N} G_{j+k}(\Delta\theta)\right)$$

where

$$G_{j+k}(\Delta\theta) = \frac{1}{\sqrt{2\pi}\,\sigma_{j+k}} e^{-\frac{\Delta\theta^2}{2\sigma_{j+k}^2}}$$

with

$$\sigma_{j+k}^2 = \tilde{\Delta\theta}^2 .$$

To avoid numerical difficulties, samples with symmetric phase difference less than a prescribed value are not used in building the implicit model for the symmetric phase difference Gaussian mixture model.

The first-order detector, that is the detector for signal of known structure, for this nonparametric representation of the Gaussian mixture model is

$$\left[\sum_{k=-N,k\neq 0}^{N} \frac{\hat{G}_{j+k}(\Delta\theta_j)}{\sigma_{j+k}^2}\right]\Delta\theta\,\mathrm{Re}\left[iz_j\frac{s_j^*}{|s_j|}\right],$$

while the second-order detector, that is the detector for signal of unknown structure, is

$$\sum_{k=-N,k\neq 0}^{N}\left[\frac{\Delta\theta_j^2}{\sigma_{j+k}^4} - \frac{2}{\sigma_{j+k}^2}\right]\hat{G}_{j+k}(\Delta\theta_j) ,$$

where in either case

$$\hat{G}_{j+k}(\Delta\theta) = \frac{G_{j+k}(\Delta\theta)}{\frac{1}{2N}\sum_{h=-N,k\neq 0} G_{j+h}(\Delta\theta)} .$$

The probability density function for a Gaussian mixture of amplitudes or of symmetric phase differences is modeled for sample z_j as a 2N-state model with the states all equally likely so that $p_k = \frac{1}{2N}$ for $-N \leq k \leq N, l \neq 0$. The k-th kernel $G_{j+k}(|z_j|)$ or $G_{j+k}(\Delta\theta_j)$ is the probability density function for the k-th state. For the amplitude case, this probability density function can arise from zero-mean Gaussian probability density functions of equal variance for the real and imaginary components of the sample. In this context, the variance of these components is simply $|\tilde{z}_{j+k}|^2$. Furthermore, if we assume that $P(\tilde{z}|k) = G_{j+k}(\tilde{z})$ then $w(j,k)$ becomes the probability that $|z_j|$ belongs to the k-th state. For the sym-

metric phase difference case, the component variance is $\tilde{\Delta}\theta_{j+k}^2$. In either case, the nonparametric estimation of a Gaussian mixture model leads to a fuzzy set model for the resulting detectors as for the kernel algorithms.

Useful information on the way to nonparametrically estimate a Gaussian mixture model is obtained by considering

$$\lim_{N \to \infty} \frac{1}{2N} \sum_{k=-N, k \neq 0} G_{j+k}(|z|)$$

when the samples $\tilde{z}_{j+k}$ are independent and identically distributed with probability density function

$$p(z) = \frac{1}{2\pi} \sum_{h=1}^{S} \frac{p_k}{\sigma_k^2} e^{-\frac{|z|^2}{2\sigma_k^2}}.$$

Then

$$\begin{aligned}
&\lim_{N \to \infty} \frac{1}{2N} \sum_{k=-N, k \neq 0} G_{j+k}(|z|) \\
&= \frac{1}{4\pi^2} \int \int_{R^2} \frac{e^{-|z|^2}}{|w|^2} \sum_{k=1}^{S} \frac{p_k}{\sigma_k^2} e^{-\frac{|w|^2}{2\sigma_k^2}} dw_1 dw_2 \\
&= \frac{1}{2\pi} \int_0^\infty \left(\frac{1}{r} e^{-\frac{|z|^2}{2r^2}} \right) \left(\sum_{k=1}^{S} \frac{p_k}{\sigma_k^2} e^{-\frac{r^2}{2}\sigma_k^2} \right) dr
\end{aligned}$$

Using the substitution $u = r^2$ and interchanging the order of integration and summation, we obtain

$$\begin{aligned}
&= \frac{1}{4\pi} \sum_{k=1}^{S} \frac{p_k}{\sigma_k^2} \int_0^\infty \frac{1}{u} e^{-\left(\frac{|z|^2}{2u} + \frac{u}{2\sigma_k^2} \right)} du \\
&= \frac{1}{2\pi} \sum_{k=1}^{S} \frac{p_k}{\sigma_k^2} K_o\left(\frac{|z|}{\sigma_k} \right).
\end{aligned}$$

where $K_o\left(\frac{|z|}{\sigma_k}\right)$ is the modified Bessel function of the second-kind order 0 (Gradshyteyna nd Ryshik, 1980). In an analogous manner,

$$\lim_{N \to \infty} \frac{1}{2N} \sum_{k=-N, k \neq 0}^{N} G_{j+k}(\Delta\theta) = \frac{1}{2\pi} \sum_{h=1}^{S} \frac{p_k}{\sigma_k} K_o\left(\frac{|\Delta\theta|}{\sigma_k} \right)$$

for

$$p(\Delta\theta) = \frac{1}{\sqrt{2\pi}} \sum_{h=1}^{S} \frac{p_k}{\sigma_k} e^{-\frac{\Delta\theta^2}{2\sigma_k^2}}.$$

In either case, the use of the nonparametreic techniques transforms $e^{-\frac{x^2}{2\sigma^2}}$ into $K_o\left(\frac{|x|}{\sigma}\right)$. Figure 5 shows the relationship between these two functions. Both the Bessel function and Gaussian function are symmetric and so figure 5 shows the functions for non-negative real numbers. In addition, they have been chosen to have unit area for the whole real axis. Since the Bessel function tends to infinity as the argument tends to 0, we need to modify the states for the samples with low values to obtain practical algorithms. The simplest approach is to rank order the samples from lowest to highest norm and not use the lowest 10% of the samples to construct the implicit model.

Processing Gain for Multistate Model Detectors

In this section, we obtain upper bounds and estimates for the square of the ratios of the deflection of adaptive locally optimum detectors to the traditional detectors. We call this squared ratio the processing gain for the first-order adaptive locally optimum processing algorithm relative to the traditional processing algorithm and ratio of deflections processing gain for second-order detectors. We obtain upper bounds for processing gain for noncentral mixture models of amplitude and symmetric phase differences for signals of known and unknown structure and Gaussian mixture models of amplitudes for signals of known and unknown structure. We have also obtained estimates of processing gain for the Gaussian mixture models of amplitude and conducted simulations to validate the bounds and estimates. These results relate achievable processing gain to mixture model parameters and provide a framework for assessing the potential processing gains achievable from the use of adaptive locally optimum processing instead of traditional processing. Also, note that it is unnecessary to obtain Gaussian mixture model symmetric phase difference processing gain bounds and estimates because this case is not of interest for ocean basin surveillance.

Processing Gain for a Noncentral Mixture Model First-Order Detector. Let the probability density function of the amplitudes of the noise be

$$p(A) \cong \frac{1}{\sqrt{2\pi}\,\sigma_{bn}} \sum_{k=1}^{S} p_k e^{-\frac{(A-\mu_k)^2}{2\sigma_{bn}^2}}, \; A \geq 0 .$$

We assume that $\mu_k >> \sigma_{bn}$ for $1 \leq k \leq S$. It then follows that $\int_0^\infty p(A)dA \cong \int_{-\infty}^{\int} p(A)dA$. We also assume that the phase is uniformly distributed. In addition, let the signal be of known structure, that is $s = (a, b) = a + ib$ and $\frac{s}{|s|}$ is known.

For the classical correlator,

$$c(x, y) = \frac{ax + by}{|s|} ,$$

and we have

$$E_n(c(x, y)) = 0 \text{ and } E_{s+n}(c(x, y)) = |s| .$$

The variance contains a broadband noise component σ_{bn}^2, the sum of the variances of the inphase and quadrature components of the background noise, and a component contributed by the interferer with different discrete values of amplitude. Under the assumption that the interferer and noise components are independent and radially symmetric, we have $\sigma_x^2 = \sigma_y^2 = \sigma^2$ so that

$$\begin{aligned} \sigma_n^2(c(x, y)) &= \frac{a^2}{|s|^2}\sigma_x^2 + \frac{b^2}{|s|^2}\sigma_y^2 \\ &= \sigma^2 \cong \sigma_{bn}^2 + \sum_{k=1}^{S} p_k \mu_k^2 . \end{aligned}$$

This follows from an easy computation using the approximate density function for *p(A)*.

Therefore, the deflection for the classical correlator is

$$\delta(c(x, y)) = \frac{|s|}{\sigma}$$

and its square is simply the signal-to-noise ratio.

To obtain an upper bound on the processing gain for the first-order adaptive locally optimum processing algorithm, we suppose that each sample is assigned to the correct state. This is equivalent to assuming that the adaptive locally optimum processing is effective in completely removing the interference.

Given the probability density function of amplitudes

$$p(A) \cong \frac{1}{\sqrt{2\pi}\,\sigma}\sum_{j=1}^{S} p_j e^{-\frac{(A-\mu_j)^2}{2\sigma^2}},$$

subject to the constraint that a sample's amplitude must be non-negative. The gain factor for the optimal first-order detector is

$$g_1(A) = \frac{1}{p(A)}\frac{1}{\sqrt{2\pi}\,\sigma^3}\sum_{j=1}^{S}(A-\mu_j)p_j e^{-\frac{(A-\mu_j)^2}{2\sigma^2}} + \frac{1}{A}.$$

The $\frac{1}{A}$ term in the gain factor can be neglected under the assumption of A. Indeed, practical experience with algorithms implemented for communication systems has indicated that excellent performance can be achieved by use of the modified gain factor

$$\hat{g}_1(A) = \frac{1}{p(A)}\frac{1}{\sqrt{2\pi}\,\sigma^3}\sum_{j=1}^{S}(A-\mu_j)p_j e^{-\frac{(A-\mu_j)^2}{2\sigma^2}} + \frac{1}{A}.$$

whose performance should be quite similar to that of the optimal detector when the small signal hypothesis holds for all of the discrete values of the deterministic interferer amplitudes, the condition required for the model being analyzed to be applicable. We now proceed to obtain the processing gain for the first-order detector $\hat{D}_1(x,y) = \hat{g}_1(A)\left(\mathrm{Re}\left[\frac{z}{|z|}\frac{s^*}{|s|}\right]\right)$ associated with this modified gain factor.

Under the assumption that each sample is assigned to the correct state:

$$\begin{aligned}
E_{n+u+s}(\hat{D}_1(x,y)) &= E_{n+s}(\hat{D}_1(x,y)) \\
&= E_{n+s}\left[\hat{g}_1(A)\left(\mathrm{Re}\left[\frac{z}{|z|}\frac{s^*}{|s|}\right]\right)\right] \\
&= E_{n+s}\left[\mathrm{Re}\frac{\hat{g}_1(A)z}{|z|}\frac{s^*}{|s|}\right] \\
&= e_{n+s}\left([proj_u(s) + proj_u(n)]\frac{s^*}{|s|}\right) \\
&= E_{n+s}\left([proj_u(s)]\frac{s^*}{|s|}\right)
\end{aligned}$$

from the earlier discussion of a one-state model for amplitudes. Under the added assumption that the phase of the signal relative to the interferer is uniformly distributed, the expected value of the projection is $\frac{1}{2}|s|$. Also, $\sigma_n^2(\hat{D}_1(x,y)) = \frac{1}{2}\sigma_{bn}^2$, so that $\delta^2(\hat{D}_1(x,y)) = \frac{1}{2}\frac{|s|^2}{\sigma_{bn}^2}$. It follows that the processing gain, under the assumptions that each sample is assigned to the correct state and the relative phase of signal to interference has uniform distribution, is

$$\frac{1}{2}\frac{\sigma_{bn}^2 + \sum_{k=1}^{S} p_k\mu_k^2}{\sigma_{bn}^2}.$$

Figures 6a and 6b present processing gain upper bound contours for a two-state noncentral mixture model first-order detector for $0 \le p_L \le 1$, $0 \le \mu_H \le 20$, $\mu_L = 1$, and $\sigma_{bn}^1 = 0.5$ and $\sigma_{bn}^2 = 2$, respectively. The curves show that the processing gain is a strong function of the level of the background noise and a weak function of the low-state probability.

Consider the first-order detector for symmetric phase differences modeled by a noncentral mixture model. Given the probability density function

$$q(\Delta\theta) \cong \sum_{k=1}^{S} q_k \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(\Delta\theta-\lambda_k)^2}{2\sigma^2}} .$$

The approximation is only good when the phase contribution of the Gaussian noise ψ to the received signal phase is small, that is

$$\int_{-\pi}^{\pi} q(\Delta\theta)d\theta \cong \int_{-\infty}^{\infty} q(\Delta\theta)d\theta .$$

For this density the first-order detector is

$$D_1(x,y) = h_1(\Delta\theta)\left(\mathrm{Re}\left[\frac{iz}{|z|}\frac{s^*}{|s|}\right]\right)$$

with

$$h_1(\Delta\theta) = \frac{1}{q(\Delta\theta)\sqrt{2\pi}\,\sigma^3}\sum_{j=1}^{S}(\Delta\theta - \lambda_j)p_j e^{-\frac{(\Delta\theta-\lambda_j)^2}{2\sigma^2}} .$$

Under the assumption that each sample is assigned to the correct state as for the amplitude case,

$$\begin{aligned} E_{n+u+s}(D_1(x,y)) &= E_{n+s}(D_1(x,y)) \\ &= E_{n+s}\left[h_1(\Delta\theta)(Re\left[\frac{iz}{|z|}\frac{s^*}{|s|}\right]\right) \\ &= E_{n+s}\left[\mathrm{Re}\frac{h_1(\Delta\theta)iz}{|z|}\frac{s^*}{|s|}\right] \\ &= E_{n+s}\left([proj_{iu}(\Delta\theta) + proj_{iu}(\Delta n)]\frac{s^*}{|s|}\right) \\ &= E_{n+s}\left([proj_{iu}(\Delta\theta)]\frac{s^*}{|s|}\right) . \end{aligned}$$

To evaluate this expectation, we assume that the signal sample j is uncorrelated with signal samples j–1 and j+1, and that the relative phase of the signal to the interferer is uniform. Under the former assumption

$$E_{n+s}\left([proj_{iu}(\Delta s)]\frac{s^*}{|s|}\right) = E_{n+s}\left([proj_{iu}(s)]\frac{s^*}{|s|}\right) ,$$

and under the latter assumption, as for the amplitude case,

$$E_{n+s}([proj_u(s)]\frac{s^*}{|s|}) = \frac{1}{2}|s| .$$

However, under the assumption of independence of the successive samples of the broadband noise component, the variance of the broadband noise component increases due to the adjacent noise terms so that

$$\sigma_n(D_1(x,y)) = \frac{3}{4}\sigma_{bn}^2 .$$

It follows that the deflection squared for the first-order detector of symmetric phase differences is

$$\frac{1}{3}\frac{|s|^2}{\sigma_{bn}^2} .$$

There is no traditional phase processing so we use the same traditional processing deflection value for phase as for amplitudes. Then the processing gain for symmetric phase differences depends only on the amplitude model as defined for the amplitude case considered in this section. The processing gain for the first-order detector of symmetric phase difference is then

$$\frac{1}{3}\frac{\sigma_{bn}^2 + \sum_{k=1}^{S} p_k \mu_k^2}{\sigma_{bn}^2} .$$

Note that due to the assumption that each sample is assigned to the correct state, the structure of the symmetric phase model plays no role in the achievable processing gain for the symmetric phase difference processing gain.

Processing Gain for a Noncentral Mixture Model Second-Order Detector. Suppose that the probability density function for the noncentral mixture model for amplitudes is

$$p(A) \cong \frac{1}{\sqrt{2\pi}\,\sigma_{bn}} \sum_{k=1}^{S} p_k e^{-\frac{(A-\mu_k)^2}{2\sigma_{bn}^2}} , A \geq 0 .$$

For the traditional processing

$$z \rightarrow T_2(z) = \frac{|z|^2}{2\sigma^2} - 1$$

with σ^2 the total variance of interferer-plus-Gaussian-noise real and imaginary components of the baseband samples. Note that the interferer component takes on discrete values of magnitude $\mu_1, \mu_2, ..., \mu_s$. It follows that the variance of the real components samples in state k is $\frac{\mu_k^2}{2\pi}\int_{-\pi}^{\pi} \cos^2\psi d\psi = \frac{\mu_k^2}{2}$ and the variance of the interference is given by $\frac{1}{2}\sum_{k=1}^{S} p_k \mu_k^2$. Thus

$$\begin{aligned}
\sigma_n^2(T_2(z)) &= \frac{1}{4\sigma^4} E[(u + proj_u n)^4] \\
&= \frac{1}{4\sigma^4} E[u^4 + 6u^2 proj_u^2(n) + proj_u^4(n)] \\
&= \frac{1}{4\sigma^4}\left[\sum_{k=1}^{S} p_k \mu_k^4 + 6\left(\sum_{k=1}^{S} p_k \mu_k^2\right)\sigma_{bn}^2 + 3\sigma_{bn}^4\right] .
\end{aligned}$$

It follows that the deflection for the traditional detector is

$$\frac{|s|^2}{\sqrt{\sum_{k=1}^{S} p_k \mu_k^4 + 6\left(\sum_{k=1}^{S} p_k \mu_k^2\right)\sigma_{bn}^2 + 3\sigma_{bn}^4}} .$$

To obtain an approximate upper bound on the second-order detector, we assume that each sample is assigned to the state containing its discrete component of interference so that the deflection for the multistate detector is the same as for the one-state detector, that is

$$\frac{1}{2\sqrt{3}}\frac{|s|^2}{\sigma_{bn}^2} .$$

It follows that an approximate upper bound for the processing gain for a second-order detector when the interference-plus-noise amplitudes are described by a noncentral mixture model is

$$\frac{1}{2\sqrt{3}\sigma_{bn}^2}\sqrt{\sum_{k=1}^{S} p_k\mu_k^4 + t\left(\sum_{k=1}^{S} p_k\mu_k su[2\right)\sigma_{bn}^2 + 3\sigma_{bn}^4}\ .$$

Figures 7a and 7b show the above processing gain upper bound contours for a two state noncentral mixture model for $0 \le p_L \le 1$, $0 \le \mu_H \le 20$, $\mu_L = 1$, and $\sigma_{bn}^1 = 0.5$ and $\sigma_{bn}^2 = 2$, respectively.

We have been unable to establish even an approximate upper bound for a noncentral distribution of symmetric phase differences second-order detector. The difficulty is that the argument used to obtain the only results we have for the single state model is inconsistent with the assumption that the symmetric phase differences are described by a multistate model. A further complication is that it is unreasonable to suppose for a multistate symmetric phase difference model that the amplitudes of the interference samples described by the model are independent of the state containing the symmetric phase difference. However, we have included the model, because in practice the nonparametric implementation for the symmetric phase difference algorithms has proven effective in communications (Bond and Schmidt, to appear) and further investigation of beamformer output statistics may reveal its applicability to ocean basin surveillance. This completes our treatment of noncentral mixture models of amplitudes and symmetric phase differences.

Processing Gain for a Gaussian MIxture Model First-Order Detector. Let the probability density function of the noise be

$$p(x,y) = \sum_{k=1}^{S} p_k \frac{1}{2\pi\sigma_k^2} e^{-\frac{x^2+y^2}{2\sigma_k^2}}$$

and let the signal of known structure be $s = (a,b) = a + ib$.

First, consider the classical correlator:

$$c(x,y) = \frac{ax + by}{|s|}\ .$$

We have

1. $E_n(c(x,y)) = 0$
2. $E_{s+n}(x(x,y)) = |s|$
3. $\sigma_n^2(c(x,y)) = \frac{a^2}{|s|^2}\sigma^2 + \frac{b^2}{|s|^2}\sigma^2 = \sigma^2$.

Therefore, the deflection for the classical correlator

$$\delta(c(x,y)) = \frac{|s|}{\sigma}$$

and its square is simply the signal-to-noise ratio. Furthermore, the performance does not depend on the details of the Gaussian mixture model. Indeed, the computations only used the fact that the noise had zero-mean and variance σ^2.

Recall that the deflection of the first-order detector satisfies:

$$\delta^2(D_1) = \int\int_{R^2} D_1^2(x,y)p(x,y)dxdy\ .$$

For Gaussian mixture noise,

$$D_1(x,y) = \frac{\sum_{k=1}^{S}\left(\frac{ax+by}{\sigma_k^2}\right)p_k\frac{1}{2\pi\sigma_k^2}e^{-\frac{x^2+y^2}{2\sigma_k^2}}}{\sum_{k=1}^{S}p_k\frac{1}{2\pi\sigma_k^2}e^{-\frac{x^2+y^2}{P}2\sigma_k^2}} .$$

By the Cauchy-Schwartz inequality:

$$D_1^2(x,y) \le \frac{\sum_{k=1}^{S}\left(\frac{ax+by}{\sigma_k^2}\right)p_k\frac{1}{2\pi\sigma_k^2}e^{-\frac{x^2+y^2}{2\sigma_k^2}}}{\sum_{k=1}^{S}p_k\frac{1}{2\pi\sigma_k^2}e^{-\frac{x^2+y^2}{P}2\sigma_k^2}} .$$

Therefore,

$$\begin{aligned}\delta^2(D_1) &\le \sum_{k=1}^{S}\int\int_{R^2}\left(\frac{ax+by}{\sigma_k^2}\right)^2 p_k\frac{1}{2\pi\sigma_k^2}e^{-\frac{x^2+y^2}{2\sigma_k^2}}dxdy \\ &= \sum_{k=1}^{S}\frac{p_k}{\sigma_k^4}(a^2\sigma_k^2+b^2\sigma_k^2) = \sum_{k=1}^{S}\frac{p_k}{\sigma_k^2}|s|^2 .\end{aligned}$$

We can obtain an approximation of σ_D^2 in the following way. Instead of applying the Cauchy-Schwartz inequality, use the fact that p_k is an unbiased estimator of $P(k|z)$ to obtain an approximation. Expand the product

$$D_1^2(x,y) = \left[\sum_{k=1}^{S}\left(\frac{ax+by}{\sigma_k^2}\right)p(k|z)\right]^2$$

and note that the cross-terms are expected to be small compared to the square-terms. Therefore, it is reasonable to assume that

$$\begin{aligned}D_1^2 &\cong \sum_{k=1}^{S}\left(\frac{ax+by}{\sigma_k^2}\right)^2 p^2(k|z) \\ &\cong \sum_{k=1}^{S}\left(\frac{ax+by}{\sigma_k^2}\right)^2 p_k p(k|z) .\end{aligned}$$

With this approximation,

$$\begin{aligned}\delta^2(D_1 &\cong \int_{-\infty}^{\infty}\sum_{k=1}^{S}\left(\frac{ax+by}{\sigma_k^2}\right)^2 p_k^2\frac{p(z|k)}{p(z)}p(z)dz \\ &= \sum_{k=1}^{S}\frac{p_k^2}{\sigma_k^2}|s|^2\end{aligned}$$

and so the deflection of the first-order detector is approximately

$$|s|\sqrt{\sum_{k=1}^{S}\frac{p_k^2}{\sigma_k^2}} .$$

It follows that an upper bound on the processing gain provided by the square of the ratio of the deflections of the first-order detector and the correlator is

$$g_1(p_1, \sigma_1, ..., p_s, \sigma_s) = \sum_{k=1}^{S} p_k \frac{\sigma^2}{\sigma_k^2} .$$

Note that

$$1 = \left(\sum_{k=1}^{S} p_k \right)^2 = \left(\sum_{k=1}^{S} p_k \frac{1}{\sigma_k \sigma_k} \right)^2$$
$$\leq \left(\sum_{k=1}^{S} p_k \frac{1}{\sigma_k^2} \right) \left(\sum_{k=1}^{S} p_k \sigma_k^2 \right) = g_1(p_1, \sigma_1, ..., p_s, \sigma_s) .$$

The above formula can be used to obtain a global processing gain bound for a multistate model in terms of a two-state model. The processing gain global bound for any multistate model first-order detector for specified interference variance and specified highest state to lowest state variance (a measure of the dynamic range of interference power) is provided by the two-state model with these parameters. The proof is presented in appendix A. Thus, it is natural to investigate two-state Gaussian mixture model first-order detectors to assess how much processing gain is available. Additional evidence of the central role of multistate models with two or three states is provided by an application of the upper bound obtained here to Middleton Class A noise first-order detectors. This discussion can be found in appendix B.

Figure 8 shows the upper bound for processing gain $g_1(p_1, \sigma_1^2, p_2, \sigma_2^2)$ in decibels as a function of p_1 and $\rho = \frac{\sigma_2^2}{\sigma_1^2}$. Results are presented for ρ with $1 \leq \rho \leq 20$, for which the greatest processing gain is 7 dB.

The lower bound of 1 is sharp is restraints are not placed on the state probabilities (either explicitly or implicitly). Consider a multistate model of $N > 2$ states. It is then possible to increase the probability of any internal state (not the lowest or highest variance state) while leaving all of the state variances equal. Let K be the index of the internal state whose probability is to be chosen as arbitrarily high. Let $\{\hat{p}_k | 1 \leq k \leq N, k \neq K\}$ be any set of positive real numbers whose sum is 1 such that $\sum_{k=1, k \neq K}^{N} \hat{p}_k \sigma_k^2 = \sigma_K^2$. Let $p_K = p$ and $p_k = (1 - p)\hat{p}_k$ for $k \neq K$.

Then

$$\sum_{k=1}^{N} p_k = 1$$

and

$$\sigma^2 = \sum_{k=1}^{N} p_k \sigma_k^2$$
$$= (1 - p) \sum_{k=1, k \neq K}^{N} \hat{p}_k \sigma_k^2 + p_k \sigma_K^2 = \sigma_K^2$$

so that the overall variance is equal to that of the k-th state and independent of p. Then

$$\lim_{p \to 1} \sum_{k=1}^{N} p_k \left(\frac{\sigma^2}{\sigma_k^2} \right)$$
$$= \lim_{p \to 1} \left\{ \sum_{k=1, k \neq K} (1 - p) \hat{p}_k \left(\frac{\sigma su[2}{\sigma_k^2} \right) + p \left(\frac{\sigma^2}{\sigma_K^2} \right) \right\} = 1 .$$

Using the estimate obtained earlier for the square of the deflection, we obtain the following estimate for the processing gain for the detector:

$$\sum_{k=1}^{s} p_k^2 \frac{\sigma_n^2}{\sigma_k^2} .$$

Figure 9 presents processing gain contours for a two-state Gaussian mixture model first-order detector.

The contours are for low-state probability $0 \leq p_L \leq 1$ and high-state-to-low-state variance ratios $1 \leq \frac{\sigma_H^2}{\sigma_L^2} \leq 20$.

Figure 10 presents the differences between the processing upper bounds and processing gain estimates for a two-state mixture model first-order detector. These contours estimate the losses in processing gain due to use of probabilistic assignment of samples to the states of the model, in other words, losses encountered because the first-order detector assigns the samples to fuzzy sets. These losses might be avoided through exploiting temporal correlations between successive samples. Figure 10 suggests that there is a small penalty paid for treating the interference samples as independent even when they are correlated.

Processing Gain for a Gaussian Mixture Model Second-Order Detector. Observe that

$$E_{n+s}[D(z)] = \int_{-\infty}^{\infty} \frac{p_n''(z)}{p_n(z)} p_{n+s}(z) dz$$
$$= \int_{-\infty}^{\infty} \frac{p_n''(z)}{p_n(z)} \left[\sum_{k=1}^{\infty} \frac{1}{k!} p_n^k(z) \int_{-\infty}^{\infty} w^k p_s(w) dw \right] dz ,$$

where we use convolutional expansion for the distribution of signal plus interference. The last expression equals $\frac{\sigma_s^2}{2} \int_{-\infty}^{\infty} \left[\frac{p_n''(z)}{p_n(z)} \right]^2 p_n(z) dz +$ higher order terms in the moments of s, $\cong \frac{\sigma_s^2}{2} \sigma_D^2$ by neglecting the higher order terms in the moments of s. It follows that $\frac{E_{n+2}(D(z))}{\sigma_D} \cong \frac{\sigma_s^2 \sigma_D}{2}$.

To obtain a bound on the variance of σ_D^2, observe that

$$\left[\sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right) \frac{1}{\sigma_k^2} p(k|z) \right]^2$$
$$\leq \sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right)^2 \frac{1}{\sigma_k^2} p(k|z)$$

by the Cauchy-Schwartz inequality (Halmos, 1957). Note that equality would hold if all the $p(k|z)$ were either 0 or 1, that is, each sample was assigned to the appropriate state. Finally,

$$\begin{aligned}\sigma_D^2 &\leq \int_{-\infty}^{\infty} \sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right)^2 \frac{1}{\sigma_k^4} p_k \frac{p(z|k)}{p(z)} p(z) dz \\ &= 4 \sum_{k=1}^{S} \frac{p_k}{\sigma_k^2} .\end{aligned}$$

Therefore, an upper bound for the deflection of the adaptive locally optimum processing quantity is given by

$$\sigma_s^2 \sqrt{\sum_{k=1}^{S} \frac{p_k}{\sigma_k^4}} .$$

In addition, the upper bound obtained is expected to be fairly sharp since it was obtained by dropping high-order terms in the magnitude of the signal and by supposing that the state assignment process yields the same results as a deterministic state assignment process without errors. Both of these assumptions underlie the derivation of the adaptive locally processing algorithm under discussion.

We can also obtain an approximation of σ_D^2 in the following way. Instead of applying the Cauchy-Schwartz inequality, we use the fact that p_k is an unbiased estimator of $p(k|z)$ to obtain an approximation. Expand the project $\left[\sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma k^2}{\sigma_k^2} \right) \frac{1}{\sigma_k^2} p(k|z) \right]^2$, and note that the cross-terms are expected to be small compared to the square-terms. Therefore, it is reasonable to assume that

$$\begin{aligned}&\left[\sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right) \frac{1}{\sigma_k^2} p(k|z) \right]^2 \\ &\cong \sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right)^2 \frac{1}{\sigma_k^4} p^2(k|z) \\ &\cong \sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right)^2 \frac{1}{\sigma_k^4} p_k p(k|z) .\end{aligned}$$

With the above approximation,

$$\begin{aligned}\sigma_D^2 &\cong \int_{-\infty}^{\infty} \sum_{k=1}^{S} \left(\frac{|z|^2 - 2\sigma_k^2}{\sigma_k^2} \right)^2 \frac{1}{\sigma_k^4} p_k^2 \frac{p(z|k)}{[(z)} p(z) dz \\ &= 4 \sum_{k=1}^{S} \frac{p_k^2}{\sigma_k^4}\end{aligned}$$

and therefore the deflection is approximately

$$\sigma_s^2 \sqrt{\sum_{k=1}^{S} \frac{p_k^2}{\sigma_k^4}} .$$

The numerator of the deflection for the traditional detection quantity when the noise is described by a Gaussian mixture model is given by

$$E_{s+n}[T(z)] = \frac{\sigma_s^2}{\sigma_n^2} .$$

The variance of the traditional detection quantity for noise only

$$\sigma_T^2 = E\left[\left(\frac{|z|^2}{\sigma_n^2} - 1\right)^2\right]$$
$$\sum_{k=1}^{S} \frac{p_k}{\sigma_k^2} \int_0^\infty \left(\frac{A^2}{2\sigma_n^2} - 1\right)^2 e^{-\frac{A^2}{2\sigma_k^2}} A dA ,$$

where the double integral in x and y is replaced by a double integral in polar coordinates and θ (for which the integrand is a constant) is integrated out. This integral can be evaluated using integration by parts to obtain

$$\sigma_T^2 = 2 \sum_{k=1}^{S} p_k \frac{\sigma_k^4}{\sigma_n^4} - 1 .$$

Therefore the deflection for the traditional detection quantity is

$$\frac{\sigma_s^2}{\sigma_n^2} \frac{1}{\sqrt{2 \sum_{k=1}^{S} p_k \frac{\sigma_k^4}{\sigma_n^4} - 1}} .$$

The upper bound for processing gain, the ratio of the deflections, can be written in terms of model parameters alone:

$$s_2(p_1, \sigma_1^2, ..., p_s, \sigma_s^2)$$
$$= \sqrt{\sum_{k=1}^{S} p_k \left(\frac{\sigma_n^4}{\sigma_k^4}\right)} \left(2 \sum_{k=1}^{S} p_k \frac{\sigma_k^4}{\sigma_n^4} - 1\right) .$$

The above formula can be used to obtain a global processing gain bound for a multistate model in terms of a two-state model. The processing gain global bound for any multistate model second-order detector for specified interference variance and specified highest-state-to-lowest-state variance is provided by the two-state model with these parameters. The proof is similar to that of the same result for first-order detectors and is presented in appendix A. Thus it is natural to investigate two-state Gaussian mixture model second-order detectors to assess how much processing gain is available. Additional evidence of the central role of multistate models with two or three states is provided by an application of the upper bound obtained here to Middleton Class A noise model second-order detectors. This discussion can be found in appendix B.

The lower bound of 1 is sharp if restraints are not placed on the state probabilities (either explicitly or implicitly). Consider a multistate model of $N > 2$ states. It is then possible to increase the probability of any internal state (not the lowest or highest variance state) while leaving all the state variances equal. Let K be the index of the internal state whose probability is to be chosen as arbitrarily high. Let $\{\hat{p}_k | 1 \le k \le N, k \ne K\}$ be any set of positive real numbers whose sum is 1 such that $\sum_{k=1,k!+K}^{N} \hat{p}_k \sigma_k^2 = \sigma_k^2$.

Let $p_K = p$ and $p_k = (1 - p)\hat{p}_k$ for $k \neq K$. Then $\sum_{k=1}^{N} p_k = 1$ and $\sigma^2 = \sum_{k=1}^{N} p_k \sigma_k^2 = (1–p)$ $\sum_{k=1, k \neq K}^{N} \hat{p}_k \sigma_K^2 + p_K \sigma_{K^2 = \sigma_K^2}$ so that the overall variance is equal to that of the K-th state and independent of p. Then

$$\lim_{p \to 1} \sum_{k=1}^{N} p_k \left(\frac{\sigma^2}{\sigma_k^2} \right)^2$$
$$= \lim_{p \to 1} \left\{ \sum_{k=1, k \neq K}^{N} (1 - p)\hat{p}_k \left(\frac{\sigma^2}{\sigma_k^2} \right)^2 + p \left(\frac{\sigma^2}{\sigma_k^2} \right)^2 \right\} = 1$$

and

$$\lim_{p \to 1} \sum_{k=1}^{N} p_k \left(\frac{\sigma_k^2}{\sigma^2} \right)^2$$
$$= \lim_{p \to 1} 2 \left\{ \sum_{k=1.k \neq K}^{N} (1 - p)\hat{p}_k \left(\frac{\sigma_k^2}{\sigma^2} \right)^2 + p \left(\frac{\sigma_K^2}{\sigma^2} \right)^2 \right\} - 1 = 1 .$$

Using the estimate obtained for the deflection of the second-order detector, we can obtain an estimate of the processing gain for the second-order detector as a function of mixture model parameters. This estimate is given by

$$\sqrt{\left\{ \sum_{k=1}^{S} p_k^2 \frac{\sigma_m^4}{\sigma_k^4} \right) \left(2 \sum_{k=1}^{S} p_k \frac{\sigma_k^4}{\sigma_n^4} - 1 \right)} .$$

Figure 11 shows the upper bound for processing gain for the second-order detector for a two-state Gaussian mixture model in decibels as a function of p_L and $\frac{\sigma_H^2}{\sigma_L^2}$ for $0 \leq p_L \leq 1$ and $\leq \frac{\sigma_H^2}{\sigma_L^2} \leq 20$. For this range of parameters the upper bound for processing gain is 12 dB for the second-order detector. Figure 12 shows the processing gain estimates for the second-order detector for the same ranges of parameter values as used to obtain the results presented for the processing gain upper bounds. Figure 13 presents contours of teh difference between the processing gain upper bounds and the processing gain estimates. The differences shown in figure 13 are estimates of the losses due to assignment of samples to states defining fuzzy sets.

Figures 14 and 15 present processing gain estimates and processing gain bounds for a three-state model, where $p_L = 0.25$, $p_M = 0.5$, $p_H = 0.25$, $\sigma_L = 1$, σ_M varies from 2 to 8, and σ_H varies from 8 to 44. This example was chosen to illustrate the processing gain for a class of three-state models shown earlier to arise from modal interference at beamformer outputs. The figures show that there is substantial gain in the region of middle-state and high-state variances considered and that processing gain is mainly a function of the high-state variance. This is because the deflection for the second-order detector is almost constant for the range of middle-state and high-state variances considered and the deflection of the traditional detector is heavily influenced by the high-state variance. The difference between the estimate and the upper bound for the processing gain varies from 2.8 to 3 dB.

Performance Comparisons Using Deflection

During the theoretical development of adaptive locally optimum processing algorithms suitable for processing interferer Gaussian mixture models for signals of unknown structure, we have obtained upper bounds and estimates of processing gain and identified algorithms based on parametric and implicit models of interference. In this subsection, the deflections obtained through simulations of the detectors are compared to the analytical results. The simulations address the performance of the second-order detectors at the sample level. In the next subsection, we address second order line detector performance. We concentrated on establishing results for second-order detectors for Gaussian mixture models because they have proven applicability to undersea surveillance.

The signal and interferer models for the simulations are baseband models. The simulation baseband signal model is a 10 Hz sinusoid with unit amplitude:

$$s = \cos 20\pi\, t + i \sin 20\pi\, t\,.$$

The baseband sample rate is 128 samples per second.

The expected values of the real and imaginary components of the baseband signal for the second-order detector are zero and their variances are one-half for the majority of the simulations. This signal level was chosen to illustrate the performance of the second-order detector when the validity of the Taylor's expansion about zero signal, which is used to derive the detector, begins to be questionable.

For each simulation run the statistics are assumed to be described by a two-state Gaussian mixture model. The low-state variance $\sigma_L^2 = 1$, while the other mixture model parameters, p_L and σ_H^2 are varied. A basic unit for a simulation run is a trial consisting of 512 samples of the interference. For each trial, the average deflection is calculated for the samples without signal and the same samples with signal. Statistics were collected for a minimum of 15 trials to allow for the calculation of the mean and standard deviation of the average deflection for a simulation run.

Simulations were conducted for the traditional detector and three second-order detector cases:

1. Gaussian mixture model with exact parameters,
2. Gaussian mixture model with predetermined errors in the parameters, and
3. Gaussian mixture model represented by an implicit mixture model.

The average deflections values are compared with the analytically obtained upper bound on deflection and the analytically obtained estimate of deflection as functions of Gaussian mixture model parameters. The relationship between single sample detection and false alarm and line detection is discussed in the next subsection: probability of line detection versus probability of false alarm is related to the average deflection. These results relate average deflection as a measure of performance to traditional descriptions of line detector performance by probability of detection versus probability of false alarm for a given input signal-to-noise ratio.

Deflection Bounds and Deflections for Known Parameters. Earlier we obtained an upper bound for the second-order detector deflection as a function of mixture state parameters. This upper bound provides the value of deflection for a clairvoyant detector, that is for a detector with each sample assigned to the correct state. An estimate of the deflection as a function of mixture state parameters was also obtained. Here, we address the accuracy of the estimate of deflection. Detailed results are presented for two families of mixture models: (1) $\sigma_L^2 = 1$, $\sigma_H^2 = 4$, and p_L = 0.1, 0.2, 0.3, 0.4, and

0.5 and (2) $\sigma_L^2 = 1$, $\sigma_H^2 = 16$, and p_L = 0.1, 0.2, 0.3, 0.4, and 0.5. The mean and standard deviations of the deflection values for 15 trials are presented in figures 16 and 18 with corresponding processing gains in decibels presented in figures 17 and 19.

The traditional detector curve would be smooth in figures 16 and 18 except that the curve is obtained by simulation. The upper bounds and estimates of deflection are calculated from the formulas derived earlier. The simulation data for the two-state Gaussian mixture model second-order detector are shown by the curves which are not labeled in figures 16 and 18. The average deflections nobtained are presented with vertical line segments connecting the average plus and minus one standard deviation of the 15 trials run to obtain the average value of deflection. Figures 17 and 19 present the processing gains in decibels obtained as the difference is decibels of the deflections for the second-order detector and the traditional detector.

There is very little processing gain available when $\sigma_L^2 = 1$ and $\sigma_H^2 = 4$ as indicated by figure 17. The estimate seems to be a little optimistic, estimating a processing gain from less than 0.5 dB to about 2 dB, while the simulations indicate from 0 dB to about 1.25 dB. These simulations indicate the adaptive locally optimum processing should not be used for a two-state model with high-state to low-state variance ratio below 4. The estimate is remarkably accurate for the case of $\sigma_L^2 = 1$ and $\sigma_H^2 = 16$ with the estimates and simulation processing gains differing by less than 0.5 dB. For this case the processing gain varies from about 3 to 8 dB.

Figures 20, 21, and 22 present the average deflection values, processing gain, and fuzzy set losses, respectively, for $\sigma_L^2 = 1$, p_L = 0.1, 0.2, 0.3, 0.4, and 0.5 and $\sigma_H^2 = 4$, 9, 16, 25, 36, 49, and 64. The different line types for the three figures correspond to the same cases. In figure 21, the processing gain is an increasing function of high-state variance since the low-state variance is fixed. Note that figure 20 exhibits deflection as a function of mainly low-state probability and low-state variance. This is because the adaptive locally optimum processing algorithms emphasize the samples estimated to be in the low states over those estimated to be in the high states. The fuzzy set loss is also a function for the simulations under discussion of the low-state probability, decreasing in general from about 4 dB for a low state probability of 0.10 to about 2 dB for a low state probability of 0.50. Figures 21 and 22 indicate that most of the processing gain can be achieved by modeling the interference samples as independent.

Dependency of Processing Gain for Gaussian Mixture Models on Model Parameter Errors. Simulations were conducted to determine the dependency of average deflection for the two-state Gaussian mixture model on parameter errors. Simulations were conducted for true two-state Gaussian mixture models given by p_L = 0.1, and $\sigma_L^2 = 1$ and $\sigma_H^2 = 6$, 10, and 16. The error sensitivity analysis was performed by supposing that the low-state probability and low-state variance were in error, with the overall variance for the model accurate. The low-state probability was varied from 0.1 to 0.8 in increments of 0.1 and the low-state variance from 0.5 to 1.75 in increments of 0.25. The contours are generated using the contour plot routine of MATLAB. Thus the high-state variance is determined from the supposed low-state probability and low-state variance. Figures 23, 24, and 25 present processing gain loss contours as a function of parameter error for the three high-state variance cases. The figures show that the processing gain achieved by using a two-state Gaussian mixture model is not sensitive to low-state variance and low-state probability errors. The processing gain is nearly flat near the true state. Indeed, in figure 23, possibly due to statistical variations in the estimation of the processing gains, which are very sensitive to the estimates of deflection for the traditional detector, the maximum processing gain was not achieved for the model with exact parameters, but rather slightly more gain (0.1 dB) was realized for a number of cases wth higher low-state probabilities than

that for the true parameters within the 0-dB contour. Generally speaking, there is more sensitivity to the low-state variance. In particular, it appears to be better to overestimate the low-state variance than to underestimate it.

The lack of sensitivity of the processing gain to low-state probability suggests that modeling the noise statistics on adjacent frequency bins to that containing the signal for an interferer occupying four or more adjacent bins should provide sufficient accuracy to support effective adaptive locally optimum processing. The modal propagation effects described in an earlier section leads to periods of low-state variance, the low variance values during these periods should be nearly the same for adjacent frequency bins, while it might be expected that the durations and frequency of occurrence of the periods might vary more, which would lead to the low-state probabilities varying from frequency bin to frequency bin.

Dependency of Processing Gain for Gaussian Mixture Model on Implicit Model States. For the implicit models, the number of states is the same as the number of samples used to construct the model. The initial trials for implicit models constructed by using 16, 32, and 64 samples were conducted for p_L = 0.1, 0.2, 0.3, 0.4, and 0.5, $\sigma_L^2 = 1$, and $\sigma_H^2 = 16$. The simulation approach assures that the samples used to model the interference are independent samples. The appropriate way to interpret the implicit model results is that the number of samples correspond to the number of independent samples of the interference for an implementation of the algorithm. For each trial, new samples were selected to construct the implicit model used to obtain results for the 512 samples drawn from the Gaussian mixture model distribution. Figures 26, 28, and 30 summarize the deflection results, while figures 27, 29, and 31 summarize the corresponding processing gain results. It is clear from an examination of these figures that there is some loss of performance for using a model constructed from 16 samples, but very little loss from using a model constructed from 32 samples in comparison with 64 samples. Thus an implicit model can be built using about one-fourth the number of samples needed to estimate model parameters using the EM algorithm. The modeling loss for 32 samples is less than 1 dB. Thus the implicit modeling approach provides an alternative to the EM algorithm. It is clear that the implicit model can be used to model interference with rapidly changing statistics.

Figure 32 presents average deflections with standard deviations of the 15 trials used to obtain the averages for implicit models using 32 samples for a low-state variance of 1 and high-state variances of 4. Figure 33 presents the processing gains corresponding to the deflections presented in figure 32. Figure 33 indicates that implicit modeling should not be used when the ratio of high-state to low-state variance is 4 or less.

Performance contours as a function of true two-state Gaussian mixture model parameters were obtained for the second-order detector using for implicit interference modeling by 32 samples. Figures 34, 35, and 36 present the average deflection values, processing gain, and modeling losses for 32 samples, for $\sigma_L^2 = 1$, p_L = 0.1, 0.2, 0.3, 0.4, and 0.5, and $\sigma_H^2 = 4$, 9, 16 25, 36, 49, and 64. The different line types for the three figures correspond to the same cases. As a result, the different cases can be inferred by examining figure 35 from the fact that performance increases with increasing high-state variance.

The implicit model deflection curves shown in figure 34 are closely clustered indicating that the performance depends mainly on the low-state parameters. The results presented in figure 34 suggest that the process gain penalty shown in figure 35 for implicit modeling by 32 samples results from the fact that the model membership function is a Bessel function rather than a Gaussian distribution.

Observe from figure 35 that use of the implicit model leads to gains except for the case of high-state variance 4 already discussed. Figure 36 indicates that provided the probability of the low state is at least 0.2 (and by inference, but not demonstrated, less than 0.8) the implicit modeling loss is between 1 and 2 dB. We tried constructing implicit models by including repeats of the state with the threshold variance (the 10 percentile of the sample norms) for each sample less than equal to the norm. For every case, this modeling approach performed 1 dB or worse than the case for which results have been presented.

Implementation Issues and Summary

Adaptive locally optimum second-order detectors are a natural extension of the noise equalization algorithm traditionally used to process beamformed ocean surveillance data. Our analysis suggests that adaptive locally optimum processing can significantly improve the detection of signals masked by ship-generated interfering lines.

The main unresolved issue for application of locally optimum processing to ocean basin surveillance is the nature of the received narrowband signal and interference signals. Simulations and hydrophone MDA data indicate that the received narrowband signals could be better described by a multi-state Gaussian mixture model than a one-state Gaussian mixture model. The scatter plots of the hydrophone MDA data indicate that the interference is better modeled by a Gaussian mixture model than by a noncentral mixture model. Likewise it may be assumed that the signal is normally of unknown structure. Given the present state of knowledge, only the second-order detectors for Gaussian mixture models for amplitudes have established relevance to ocean basin surveillance. However, conceptual considerations indicate that first-order and second-order detectors for noncentral mixture models of amplitudes and symmetric phase differences and first-order detectors for Gaussian mixture models for amplitudes should be used to complement the second-order Gaussian mixture model amplitude processing.

Additional processing of beamformed data containing real signals and interference should be carried out to determine whether the best general approach to implementing first-order and second-order detectors for Gaussian mixture models should use the model approach or real-time parametric or nonparametric estimates of the interference. The robustness of the two-state detector indicated that most of the processing gain for a second-order detector could be obtained for approximate models. This suggestts that a reasonable number of models might suffice as predetermined candidate models for the received signal plus interference complex sample amplitudes. Therefore, the modeling approach provides another implementation option.

The results obtained on the characterization of interference statistics indicated more complexity of the received signal statistics than captured by a two-state or three-state Gaussian mixture model. These results leave open the possibility that the nonparametric approach, which does not require *a priori* knowledge of the number of the states of the model, may provide better performance than the parametric approach. At the present time, our results suggest that either approach provides considerable processing gain over traditional processing when narrowband interference masks a signal of interest.

The overall scheme for the selection of samples and the processing of the beamformer output data needs to be decided before implementing the adaptive locally optimum processing. In particular, the detectors could be implemented using block recursive with or without overlap processing or sample recursive schemes. The samples to model the interference to allow the adaptive locally optimum

processing of a given sample should probably be drawn in such a way from adjacent frequency and temporal bins that the desired signal can occupy two adjacent bins. The total number of samples needed to model the interference when the EM algorithm is used to extract model parameters should be around 100 samples, while 32 samples suffice to model the interference implicitly.

The detectors in this subsection were characterized for independent samples. Averaging these detectors leads to line detectors, whose performance is briefly addressed in the next subsection. Observe that the line detector for a first-order detector involves coherent integration, while that for the second-order detector involves noncoherent integration. This is another reason why the application of first-order detectors to ocean basin surveillance should be pursued further.

LINE DETECTION AND CLASSIFICATION

Introduction

In the last section, we established that nonlinear processing of beamformed data could be viewed as a generalization of the NSE algorithm to improve detection of signals masked by narrowband interference described by multistate Gaussian mixture models. The nonlinear processing treats the frequency domain beamformer output samples as independent. In this subsection, we briefly discuss spatial cell combining techniques suitable for the detection of narrowband signals (lines). A theory of such techniques tailored to adaptive locally optimum processed matched field beamformer outputs has yet to be developed. At the present time, the noise statistics of the background after locally optimum processing based on Gaussian mixture models have not been characterized for either first- or second-order detectors. It is probably reasonable to treat this background noise as Gaussian so that the extensive work of Bar-Sahlom (1988, 1990) provides a starting point for such an investigation.

Line detection and classification for undersea surveillance systems is often performed by eye integration of NSE processed data presented on LOFARGRAMs as described in the introduction and background subsections. The operators are alerted to lines of interest by periodic printouts listing lines of interest by frequency and bearing. We suggest a generalization of this approach for high gain arrays to classify lines automatically detected by line tracking techniques such as adaptive line enhancement.

A classical LOFARGRAM is a three-variable presentation, Fourier coefficient magnitude (intensity of a pixel), frequency (abscissa0, and time (ordinate). The fourth variable, bearing, is indicated by the beam output whose output is displayed on the LOFARGRAM. A matched field beamformer output is a function of these variables as well as range from the array and depth of the source. In addition, due to increased resolution in bearing, it is no longer reasonable to base detection on the display of data from a single beam. Color displays with chrominance and intensity allow the presentation of one additional variable than a monochromatic LOFARGRAM. The challenge is to reduce the number of displayed variables to allow an operator to confirm automatic detections and classify detected lines.

One approach is to use adaptive line enhancement techniques to automatically detect potential lines and then to structure the database to allow display of the tracked line in a color LOFARGRAM-like display with depth suppressed time intensity for a given frequency (magnitude of Fourier coefficient), chrominance (depth), bearing (abscissa), and range (ordinate). Chrominance would be used to combine information for a certain number of depth bins around the average depth of the line being displayed. Mouse selection of a sequence of adjacent bins could be used to activate temporal histories of any of the displayed parameters of the line.

Display technology is rapidly advancing and so more sophisticated display technology may soon be available to allow the operator to assess the characteristics of a line detected by adaptive line enhancement processing than described in the previous paragraph. To pursue this subject further, we have concentrated our attention on processing that might be used to automatically detect lines. We begin this discussion by describing the relationship between the detectors discussed in the last section and line detectors. These results provide an upper bound on the line detection performance that can be achieved using interference tracking algorithms.

Upper Bounds for Line Detectors

Upper bounds for line detectors are obtained by averaging deflection. In a more general context, the results characterize performance when the signal is tracked perfectly. Given a detctor $D : R^2 \rightarrow R$, recall that the line detector for a sequence of complex samples $z_1 = x_1 + iy_1, ..., z_N = x_N + iy_N$ is $\frac{1}{N}\sum_{k=1}^{N} D(x_k, y_k)$. For a given sequence of samples without signal, and a threshold T, a probability of detection and probability of false alarm are defined for both the detector and the line detector in a natural way. Let $M >> N$ be the number of samples in each sequence. Let M_S be the number of samples with signal for which $D(x, y) \geq T$ and let F_S be the number of samples without signal for which $D(x, y) \geq T$. Let M_{SS} be the number of subsequences for which the line detector is calculated. Let M_L be the number of these subsequences of samples with signal for which $\frac{1}{N}\sum_{k=1}^{N} D(x_{j+k}, y_{j+k}) \geq T$ and let F_L be the number of these subsequences of samples without signal for which $\frac{1}{N}\sum_{k=1}^{N} D(x_{j+k}, y_{j+k}) \geq T$. Then the estimates of sample probability of detection and sample probability of false alarm are $PS_D = \frac{M_s}{M}$ and $PS_{FA} = \frac{F_s}{M}$, respectively, and the estimates of sample probability of line detection and line probability of false alarm are $PL_D = \frac{M_L}{M_{ss}}$ and $PL_{FA} = \frac{F_L}{M_{ss}}$, respectively.

Figures 37 and 38 are scatter plots of the probability of sample detection for a probability of false alarm of 10% against the deflection for the traditional detector and for the two-state Gaussian mixture model second-order detector for low-state variance 1 and high–state variance 6 and 16. The data presented is the same data used to generate the contour plots presented in figures 23 and 24 of the previous section. Figures 39, 40, 41, and 42 are scatter plots of the probability of line detection for a false alarm of 10% against the deflection for the traditional line detector for the two-state Gaussian mixture model second-order line detector for low-state variance 1 and high-state variance 6, 10, 16, and pooled for these cases, respectively. The line detection statistics are obtained from the sample statistics by averaging 128 samples. In each plot, the traditional detector results are presented by asterisks and the adaptive locally optimum second-order detector by crosses, except for figure 42 where “+” is high-state variance 6 data, “x” is high-state variance 10 data, and “o” is high-state variance 16 data. For every case the low-state variance is 1 and the signal variance 0.5 and data are presented for all the different model parameter choices to model the underlying two-state Gaussian mixture model.

Some correlation between probability of detection and deflection is shown in figure 37. For this case, the adaptive locally optimum processing sometimes performed worse than the traditional detector and sometimes about 1 to 2 dB better. Figure 38 shows that by the time the high state to low state variance has reached 16 to 1, the probability of detection and deflection have become more highly correlated and the data points for the adaptive locally optimum processing are separated from a cluster of data points for the traditional detector.

Figures 39 through 42 show how highly correlated probability of line detection and deflection are for the data sets under discussion—the correlation is independent of parameter error. These figures justify using deflection as a criterion for the development of the detectors. They also clearly indicate the performance improvements derivable from use of adaptive locally optimum processing errors even in the presence of large modeling errors.

A series of simulations was conducted to generate receiver operation characteristic (ROC) curves and soft decision gains for two-state Gaussian mixture model second-order detectors (Stein, Bond, and Zeidler, 1993). The simulations were conducted to establish the achievable performance gains using adaptive locally optimum processing techniques. The simulations investigated the performance of the algorithms using a parametreic description of the interference. The parameters were obtained using the EM technique. (See appendix B, “Gaussian Mixture Models for Acoustic Inference,” for a description of how the EM algorithm was used to provide parameter estimates.) The interference was assumed to be described independent samples from a stationary Gaussian mixture model. In the context of the last subsection, if the use of the EM algorithm to estimate noise-only statistics from adjacent frequency bins to the bin containing the signal yields a model whose parameters are approximately correct, the performance obtained should be within 1 to 2 dB of that estimated by the simulations. The simulations also address cases in which the signal energy is comparable or exceeds the low-state variance. For these cases, the derivation of the algorithms as implementations of optimal detectors no longer apply. However, the algorithms can be quite effective if modified to compensate for the presence of the signal in the processed samples.

The ROC curves were generated in the following way. A two-state Gaussian mixture model was chosen. Each run consisted of 10,240 independent trials. For each independent trial, 130 independent samples were generated for the assumed two-state Gaussian mixture model and the model parameters estimated using the Estimate and Maximize (EM) algorithm to obtain a detector for the samples for a signal plus noise and another 130 independent samples were generated and the model parameters estimated using the EM algorithm to obtain a detector for the samples for noise alone.

Probability of detection results for different thresholds were obtained for four detectors:

1. the traditional detector,
2. a detector with the processing based on noise-only statistics (the detector obtained in the limit as the signal goes to zero),
3. a detector obtained by adjusting state membership functions for the presence of signal, and
4. a detector obtained by adjusting state membership functions by locally scaling the estimated noise variance.

The traditional detector is an energy detector. The detector with processing based on noise only assigns the individual signal-plus-interference samples to the low and high states of the model based on the norm of the interference. This detector cannot be achieved in practice. It is included to provide an upper bound for an adaptive locally optimum second-order detector when the interference is described by a two-state Gaussian mixture model and the samples have independent interferer components. The third detector includes an adjustment for the presence of signal. The variance of the signal is estimated from the difference of the variance of the samples processed (with signal added) from the variance of the noise, which is known by assumption. The variance of the signal can then be subtracted from the norms of the samples to adjust the low-state and high-state membership functions. The fourth detector was obtained by scaling the norms for the samples with signal so taht their norms

provide the same variance as that for the noise only. Both of these detectors would reduce to the adaptive locally optimum detector when the signal energy is small compared with the low-state variance.

Probability of detection results were obtained for each of the four detectors for a signal that was present in one frequency bin for the first 65 samples and then present in the adjacent bins for the next 65 samples. This feature was incorporated in the simulations to provide some indication of the robustness of the techniques when the duration of the signal in any one frequency bin is unknown. The detectors 3 and 4 were implemented based on the estimates obtained using all 130 samples in each of these bins. Probability of false alarm estimates were obtained for each of the thresholds used to generate probability of detection using the noise only samples.

ROC curves were generated for various two-state Gaussian mixture models. Figure 43 presents the curves for $p_L = 0.5$, $\frac{\sigma_H^2}{\sigma_L^2} = 10$, and (a) $\frac{\sigma_s^2}{\sigma_L^2} = 2$, (b) $\frac{\sigma_s^2}{\sigma_L^2} = 1$, (c) $\frac{\sigma_s^2}{\sigma_L^2} = 0.5$, and (d) $\frac{\sigma_s^2}{\sigma_L^2} = 0.25$, respectively; figure 44 presents the curves for $p_L = 0.25$, $\frac{\sigma_H^2}{\sigma_L^2} = 10$, and (a) $\frac{\sigma_s^2}{\sigma_L^2} = 2$, (b) $\frac{\sigma_s^2}{\sigma_L^2} = 1$, (c) $\frac{\sigma_s^2}{\sigma_L^2} = 0.5$, and (d) $\frac{\sigma_s^2}{\sigma_L^2} = 0.25$, respectively; figure 43 presents the curves for $p_L = 0.1$, $\frac{\sigma_H^2}{\sigma_L^2} = 4.2$, and (a) $\frac{\sigma_s^2}{\sigma_L^2} = 7.3$, (b) $\frac{\sigma_s^2}{\sigma_L^2} = 3.6$, (c) $\frac{\sigma_s^2}{\sigma_L^2} = 1.8$, and (d) $\frac{\sigma_s^2}{\sigma_L^2} = 0.9$, respectively. The processing gain relative to the traditional detector is independent of the signal energy. Modest processing gains of 4.2, 2.5, and 1 dB are predicted by the estimated ratio of deflections obtained in the last subsection for the cases presented in figures 43, 44, and 45, respectively. Even modest processing gains lead to clearly discernible improvement of probability of detection for the adaptive locally optimum detectors over the traditional detector for a given probability of false alarm as illustrated by figures 43, 44, and 45.

For each case, the processing gain increases as the signal-level to low-state variance decreases, that is as the processing becomes more optimal. The figures indicate the inherent robustness of the second-order detector. It still provides significant processing gain under signal to interference conditions for which the derivation of the adaptive locally optimum detector breaks down. This is not always a feature of adaptive locally optimum processing techniques. In particular, for applications when coherent detection of the reconstructed signal is required, it is definitely not the case: signal distortion becomes manifest for a ratio of interference to signal of 2 to 1 (Bond and Hui, to appear).

Target Tracking

It is necessary to combine energies in successive temporal beamformer spatial cells to detect signals from a target whose depth is changing, whose range from the receiving array, and whose frequency may be changing from one cell to an adjacent cell. The tracking of signals exploits the continuity of the signal term in the beamformer output. In the last subsection, techniques were described to exploit interference and the techniques developed ignored temporal correlation of the interference. In general, the correlation among the signal component of the sample cannot be ignored. Ignoring correlation of the interference among samples described by a Gaussian mixture model causes a small proportional performance loss. In contrast, the correlation of signal in samples usually needs to be exploited for detection and often necessary for classification.

There are two general approaches to combining energy. One approach is to use database browsing techniques to identify potential detections and through operator interaction provide the capability to the operator to obtain an estimate of the probability that the energies in the operator-designated track

would have occurred due to noise alone. A technique similar to this has proved effective for associating correlation peaks over time in interarray processing. The other general approach is to process the beamformed data with automatic detection and tracking algorithms and use the algorithms to alert the operator to cases of interest. In this approach, the operator then examines the data using various display options to confirm its interest.

The development of optimal tracking algorithms for Gaussian mixture model first- or second-order detectors remains to be done. At the present time, the properties of the noise at the outputs of the detectors have not been characterized. Tracking techniques using Markov models of the transitional probabilities relating the next spatial cell containing the signal to a finite number of previous spatial cells containing the signal could be investigated.

A preliminary analysis of the tracking problem has been conducted. The results are presented in appendix C. In this analysis, likelihood ratios are assigned to tracks made up of segments over which samples can be coherently combined and models movement from one segment to the next using random walks constrained to result in a net movement in some direction. The movement in the chosen direction is analyzed through use of a one-dimensional model. The one-dimensional model results indicate positive cell signal-to-noise ratios for the segments before tracking provides gain over single-segment detection. In a general context, the algorithms moderately improve detection over those provided by the individual segments and allows for the tracking of targets from segment to segment, which aids in the classification of detected targets.

Summary

A variety of traditional and recently developed information processing techniques have applicability to processing the beamformed output of large arrays of hydrophones. No one technique is the best for all of the scenarios which may be of interest. The signal of interest may be narrowband, broadband, or broadband with associated narrowband signals, the interference can be narrowband or broadband.

Adaptive locally optimum processing can be used to improve detection of signals masked by interference. The adaptive locally optimum processing should be used between a time domain beamformer and spectral analysis. This allows adaptive locally optimum processing techniques to be used to cancel narrowband interferers with slowly varying frequency and amplitude, whenever they exceed the background by 6 dB or more. In addition, the Gaussian mixture model first-order detector can be used to handle nonstationary interference. These techniques can be used whether the desired signal is broadband or narrowband. Also, the search for narrowband lines with or without associated broadband features, can be based on use of traditional spectral analysis or recently developed time-frequency domain analysis techniques.

The adaptive locally optimum processing can also be used after a frequency domain beamformer. A general theory of these techniques has been presented. In particular, we have shown the Gaussian mixture model second-order detectors to be applicable to signals received with constructive and destructive propagation mode interactions.

The processing of beamformer outputs by different detection algorithms provides different information about the signals present, narrowband and broadband. In general, it may be helpful to an informed operator to provide the capability to present displays of the various detection algorithms side by side, and possibly superimposed using different colors for the displays being overlaid.

The output of a matched field beamformer consists of outputs for three spatial dimensions, a time dimension, and a frequency dimension. For high-spatial and/or high-frequency resolution beamformer outputs, it will usually be necessary to combine energy from different spatial cells over time to allow for detection of weak signals. The algorithms to best accomplish this need to be developed. The traditional techinque of eye integration may still have a central role in the multidimensional case. The operator could be provided with database browse software, which would allow for the display of beamformed output for any surface defined in combined spatial, temporal, and frequency space.

Automated classification algorithms could be processing the outputs of the beamformer after detection processing, to identify signals of interest. The operator could examine the thus identified signals of potential interest using the browse feature.

The information processing analysis has revealed a central role for adaptive locally optimum processing in ocean basin surveillance. The techniques can substantially improve the detection of weak signals masked by other signals.

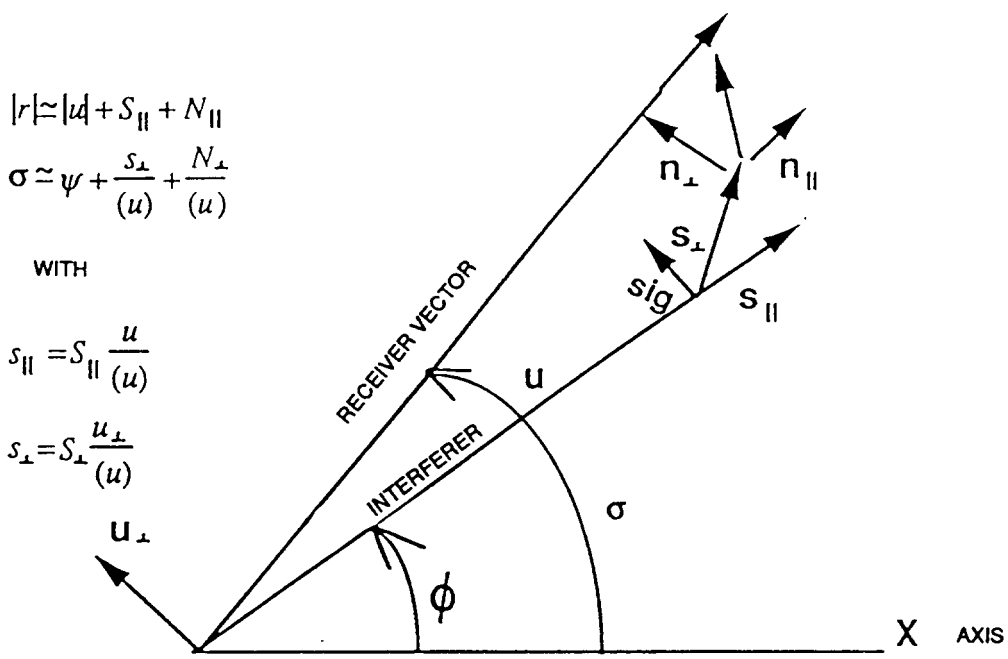


Figure 1. Signal and noise representations in coordinate system defined by interferer.

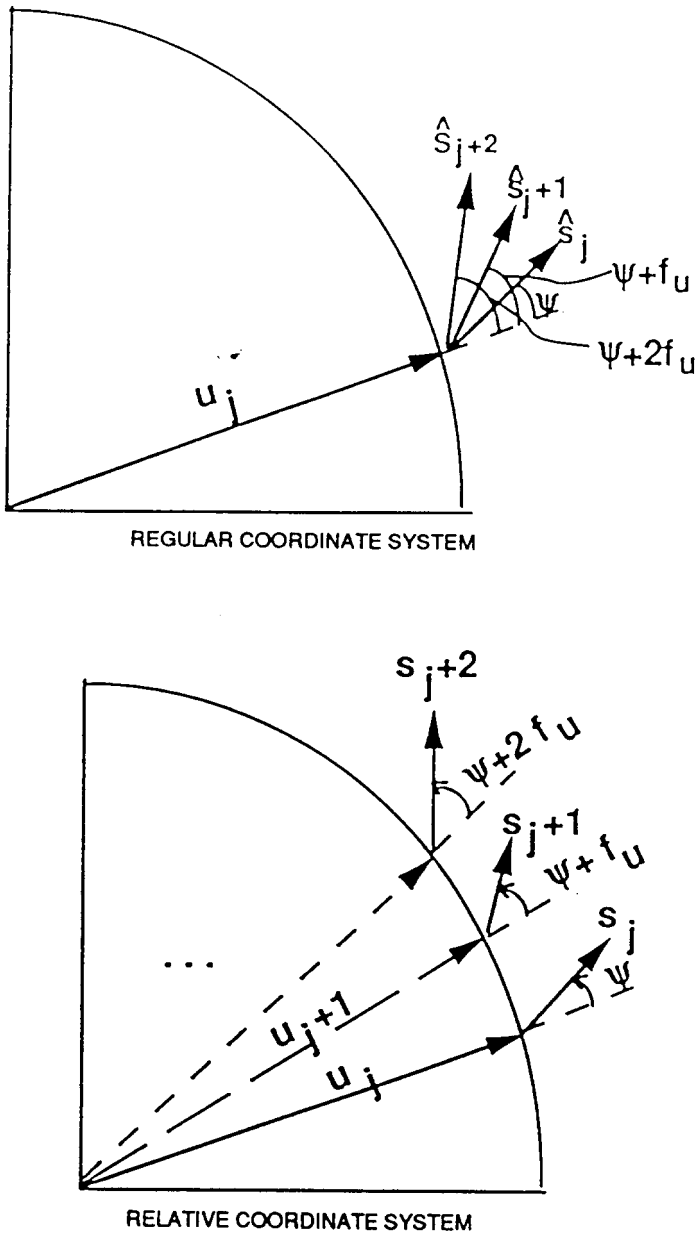
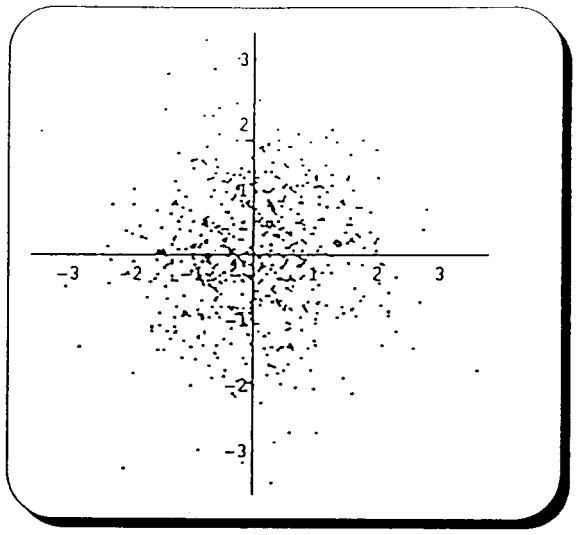
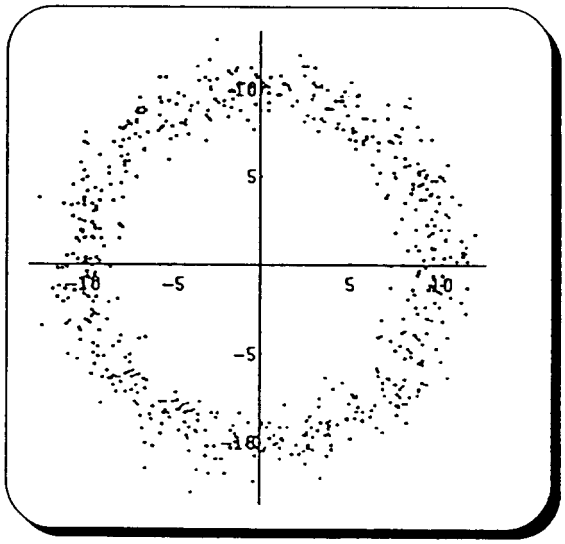


Figure 2. Geometry of describing samples of narrowband signal with greater frequency than narrowband interferer.

Amplitude of CW = 10, Noise: zero mean, unit variance for both components



Noise only



Noise + CW

Figure 3. Complex sample scatter plots for mixture models.

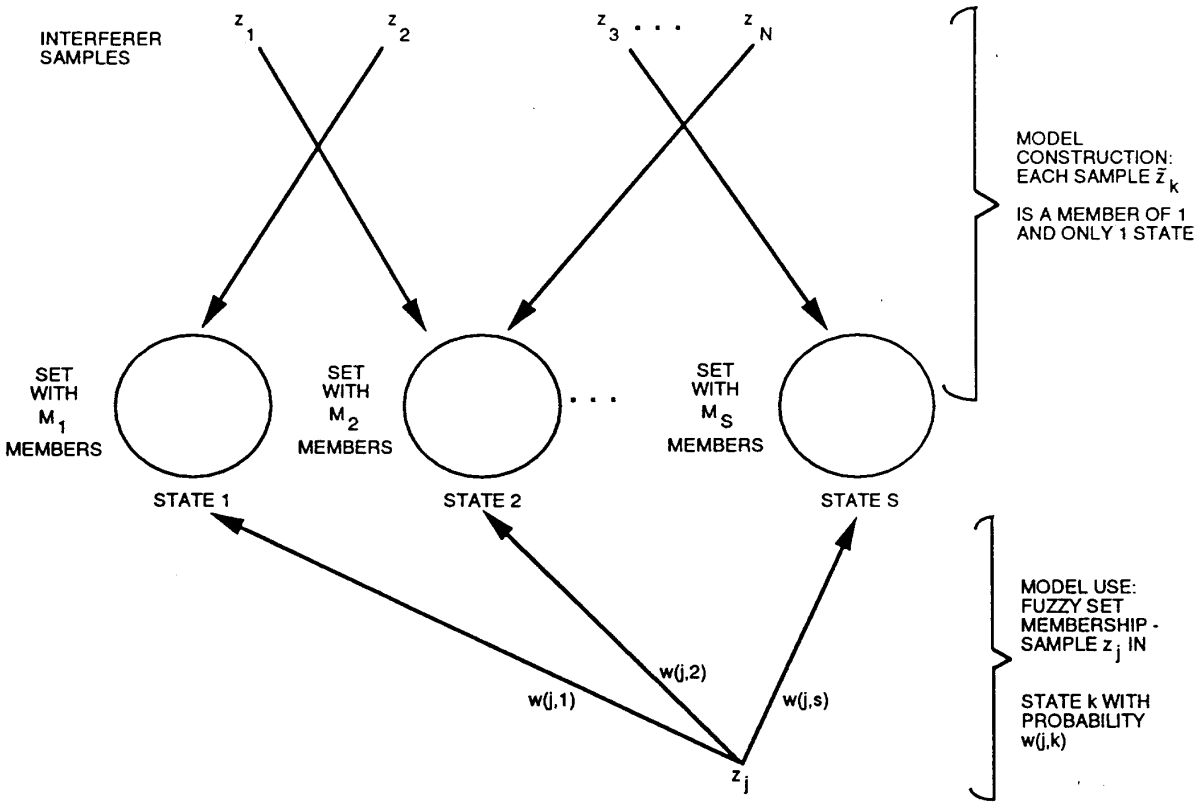


Figure 4. Fuzzy set interpretation of mixture model construction process.

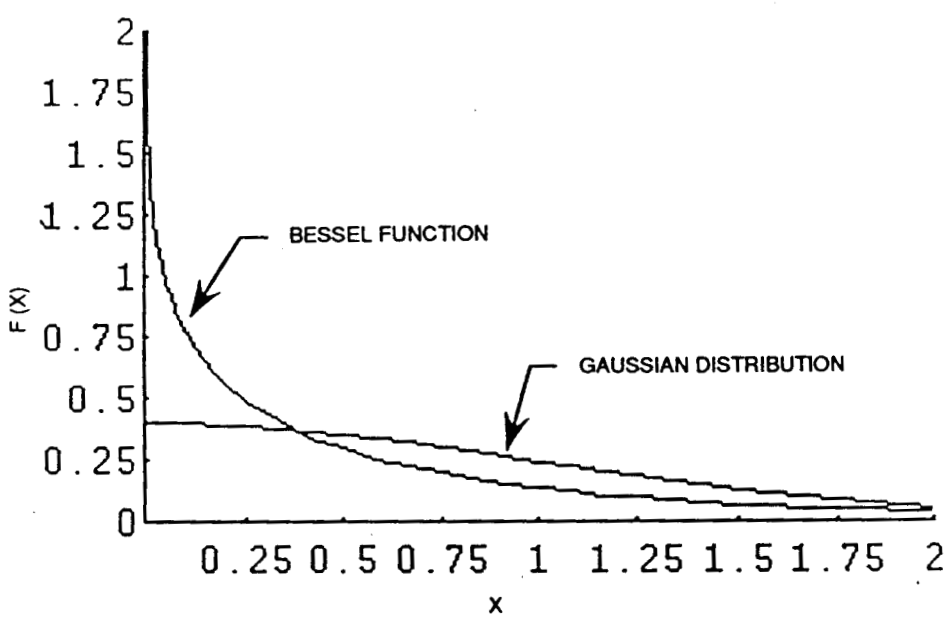


Figure 5. Normalized Bessel function K_0 and the Gaussian distribution.

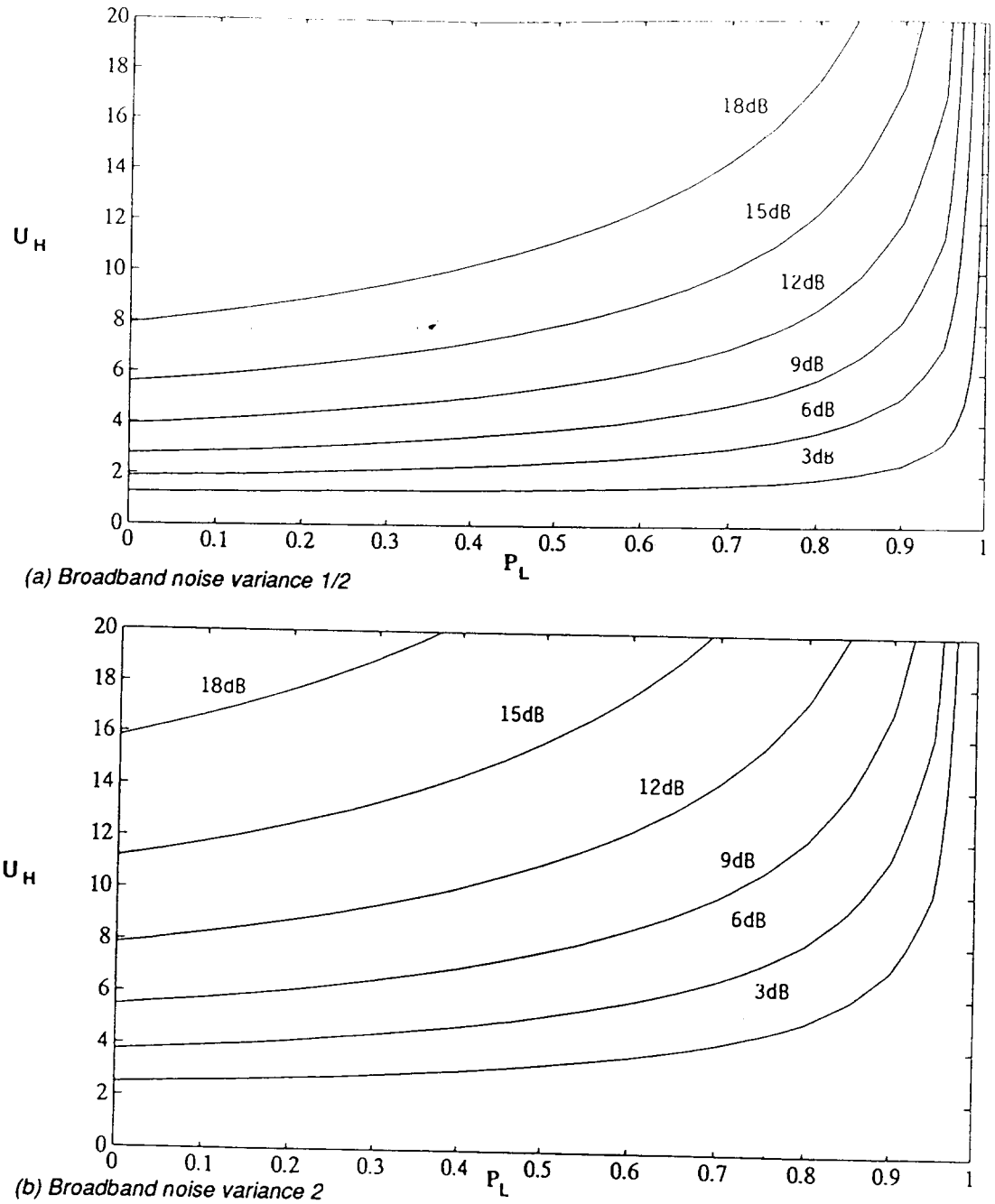
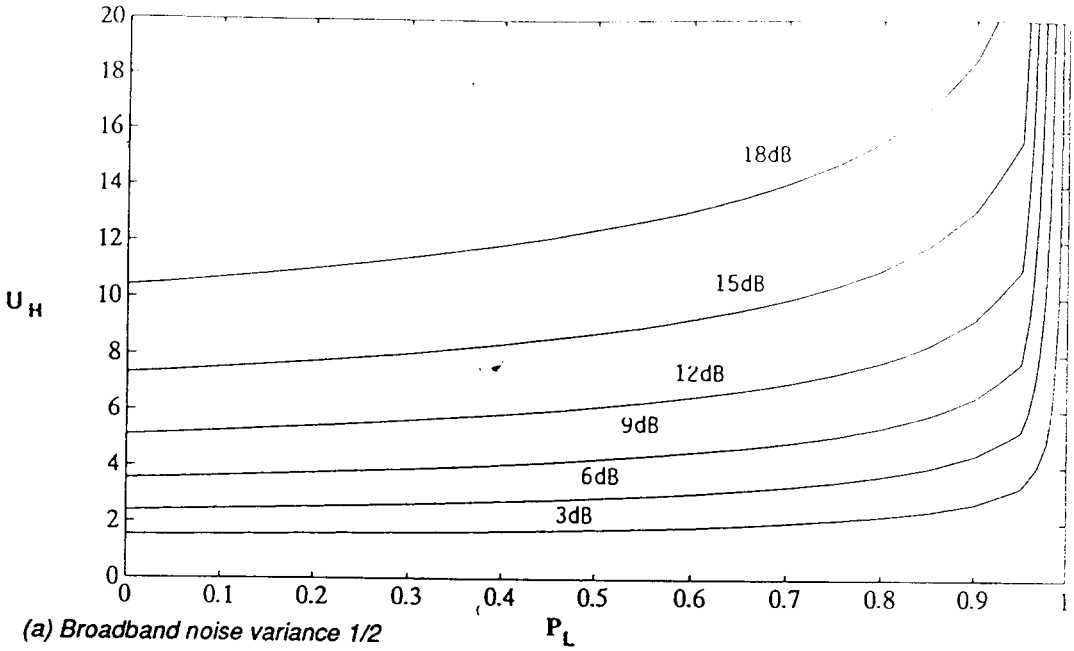
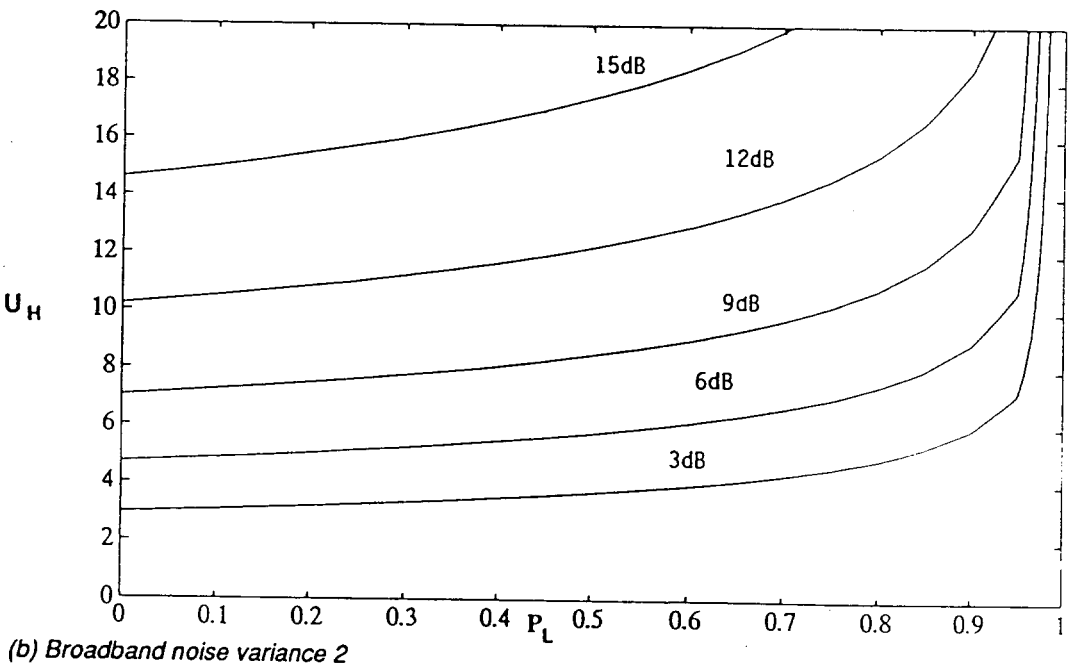


Figure 6. Processing gain upper bound contours for noncentral mixture model first-order detector.



(a) Broadband noise variance 1/2



(b) Broadband noise variance 2

Figure 7. Processing gain upper bound contours for noncentral mixture model second-order detector.

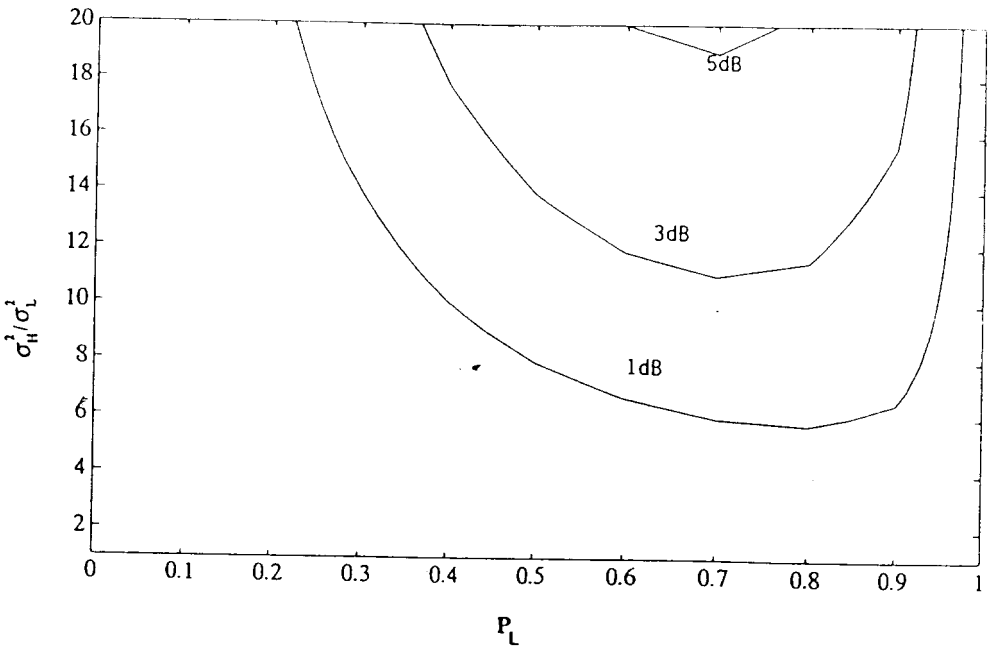


Figure 8. Processing gain upper bound contours for a two-state Gaussian mixture model first-order detector.

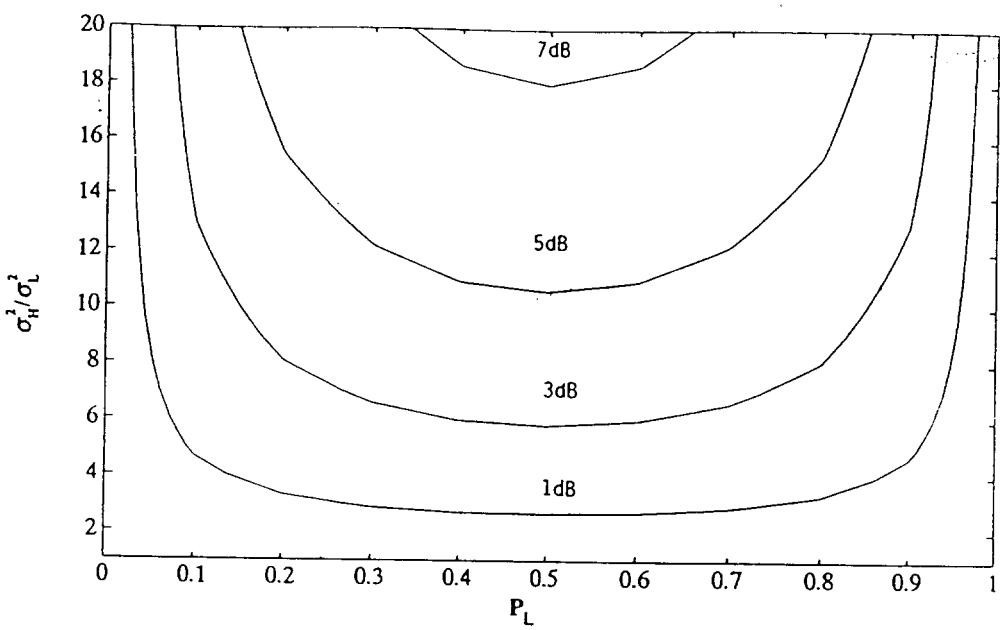


Figure 9. Processing gain contours for a two-state Gaussian mixture model first-order detector.

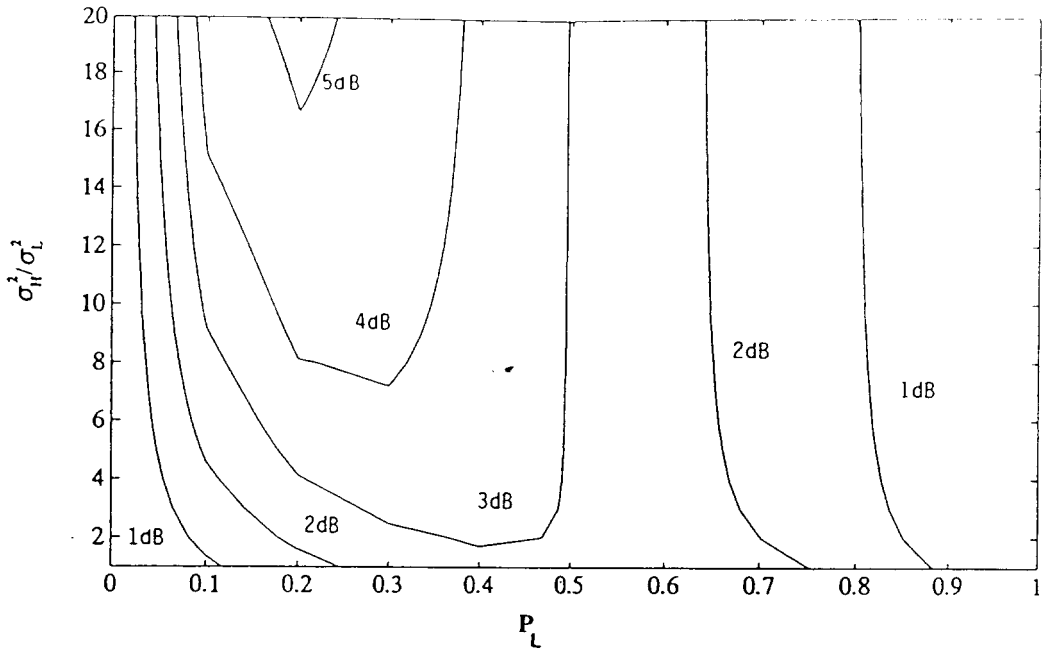


Figure 10. Processing gain losses due to fuzzy set membership for a two-state Gaussian mixture model first-order detector.

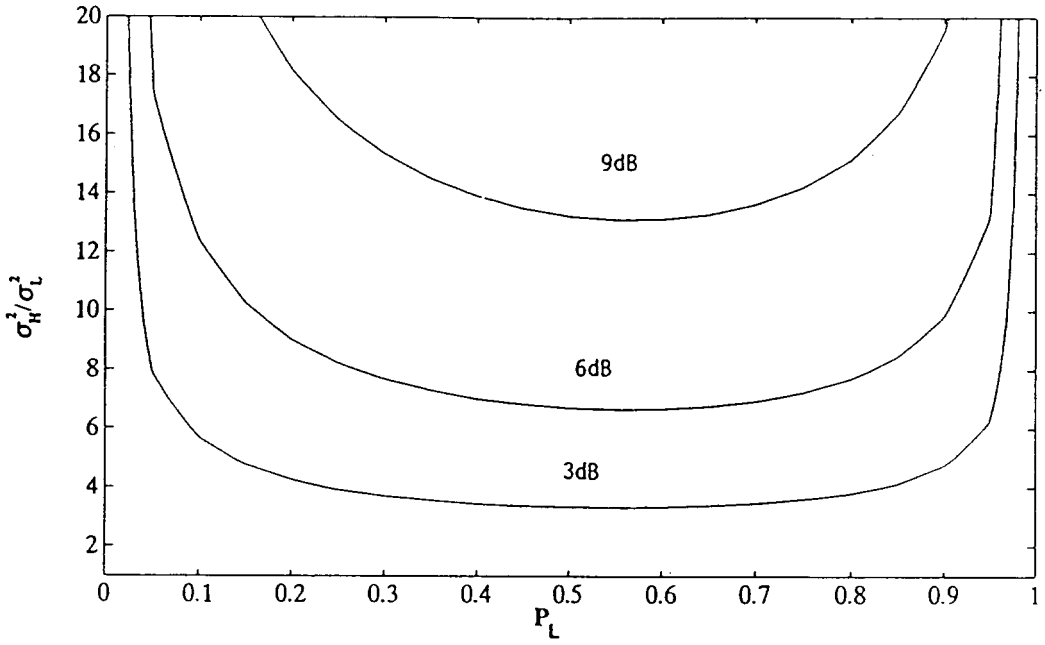


Figure 11. Processing gain upper bound contours for a two-state Gaussian mixture model second-order detector.

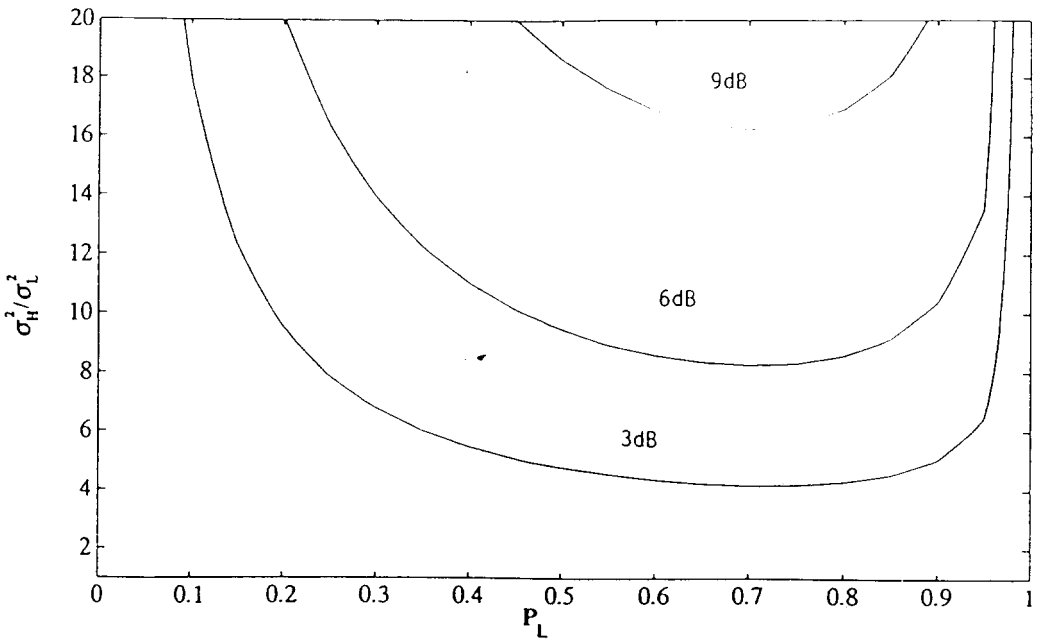


Figure 12. Processing gain contours for a two-state Gaussian mixture model second-order detector.

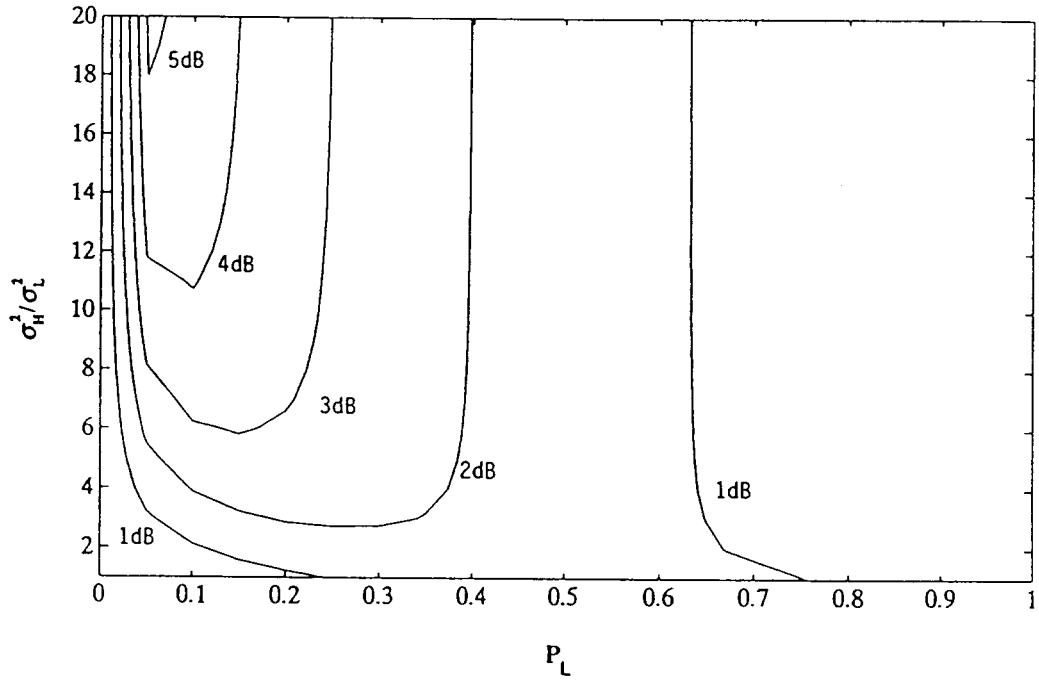


Figure 13. Processing gain losses due to fuzzy set membership for a two-state Gaussian mixture model second-order detector.

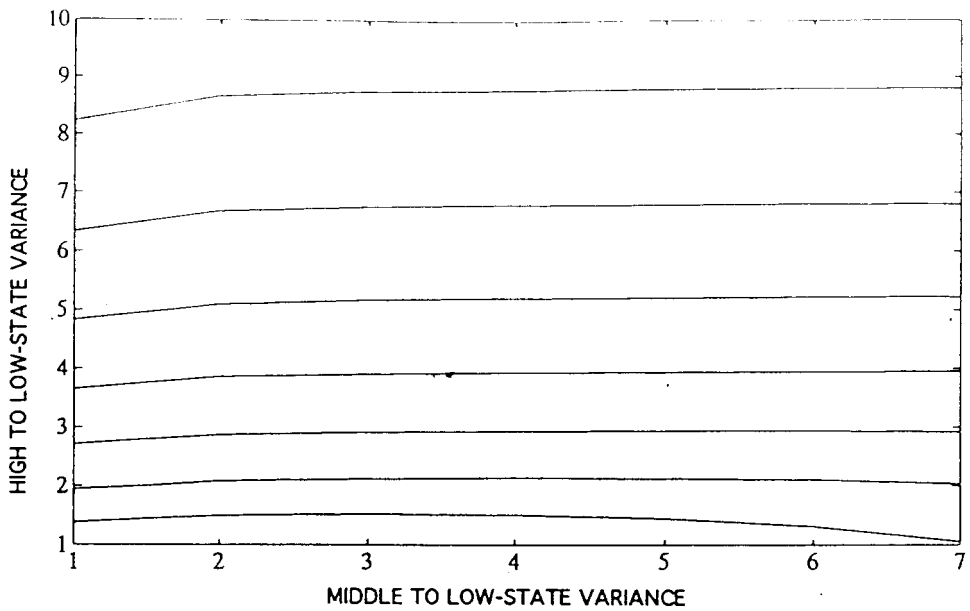


Figure 14. Processing gain estimates for a class of three-state Gaussian mixture model second-order detectors.

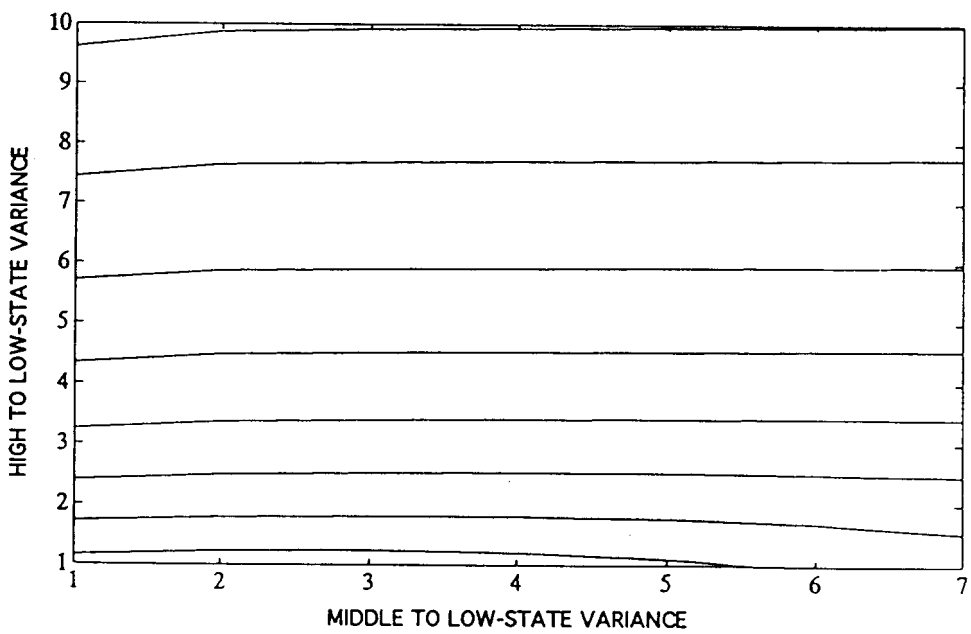


Figure 15. Processing gain upper bounds for a class of three-state Gaussian mixture model second-order detectors.

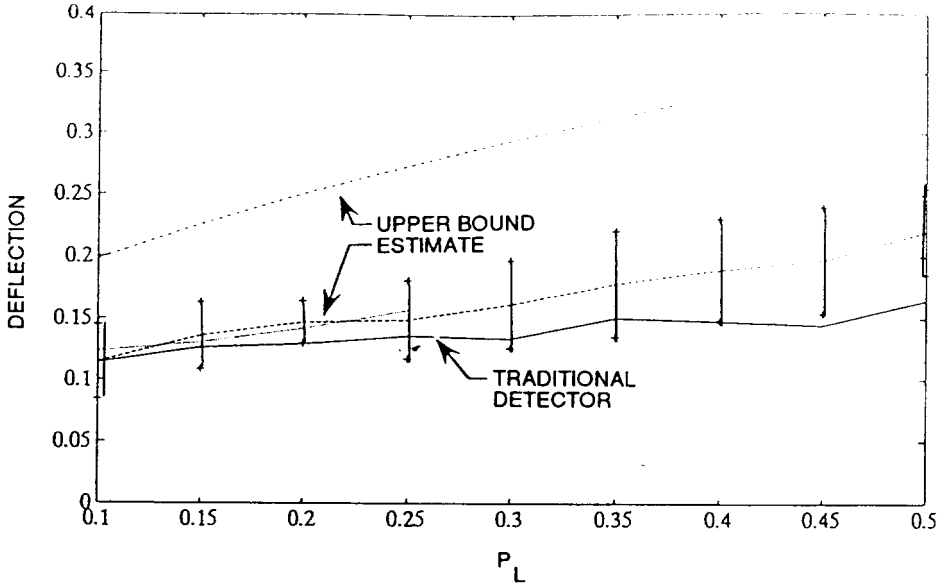


Figure 16. Deflection comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 4$.

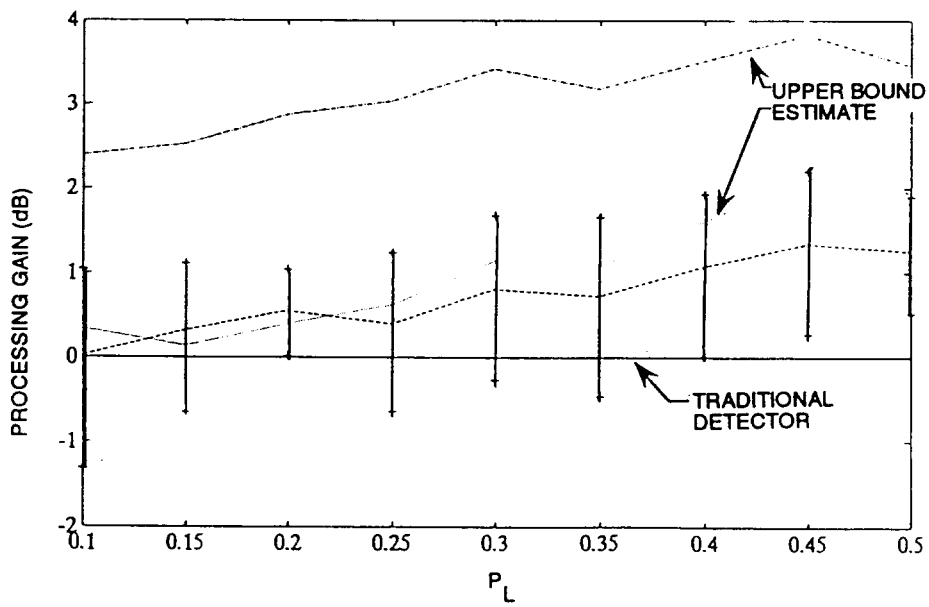


Figure 17. Processing gain comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 4$.

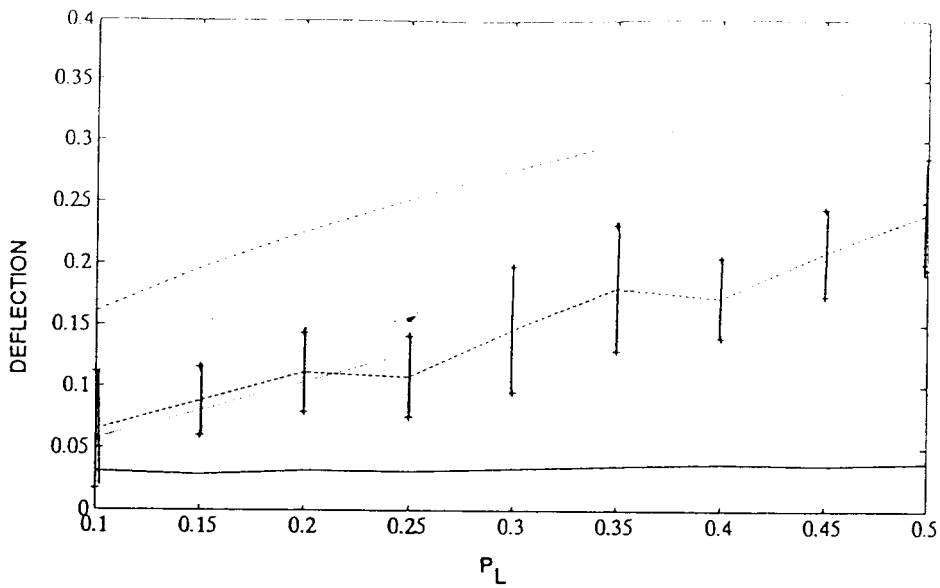


Figure 18. Deflection comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$.

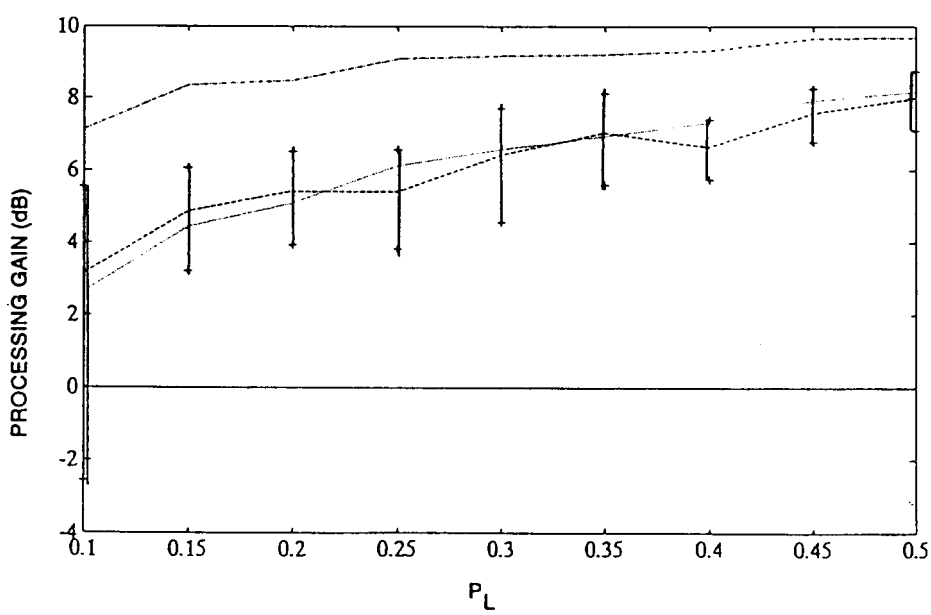


Figure 19. Processing gain comparison for two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$.

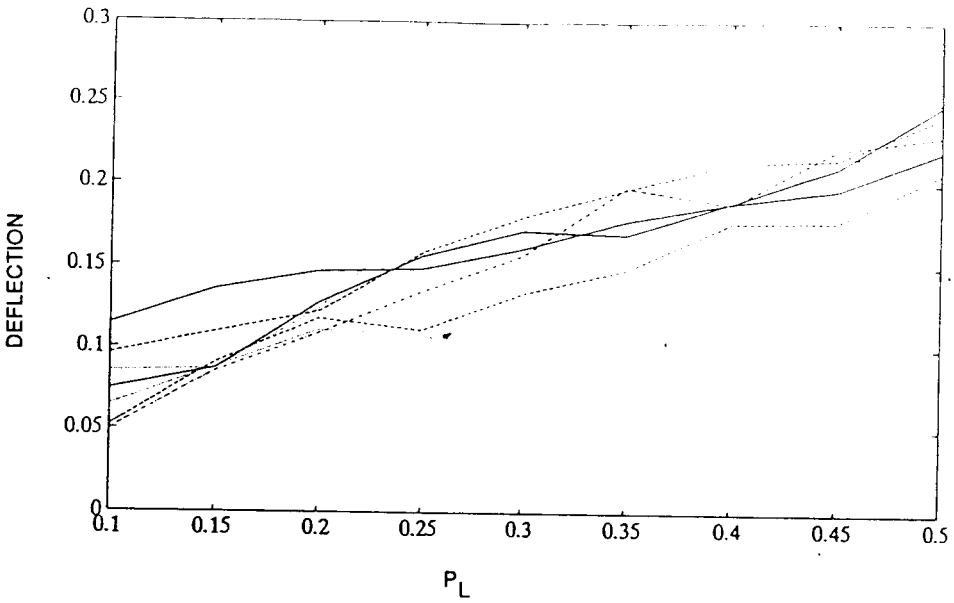


Figure 20. Deflection for the two-state Gaussian mixture model second-order detector.

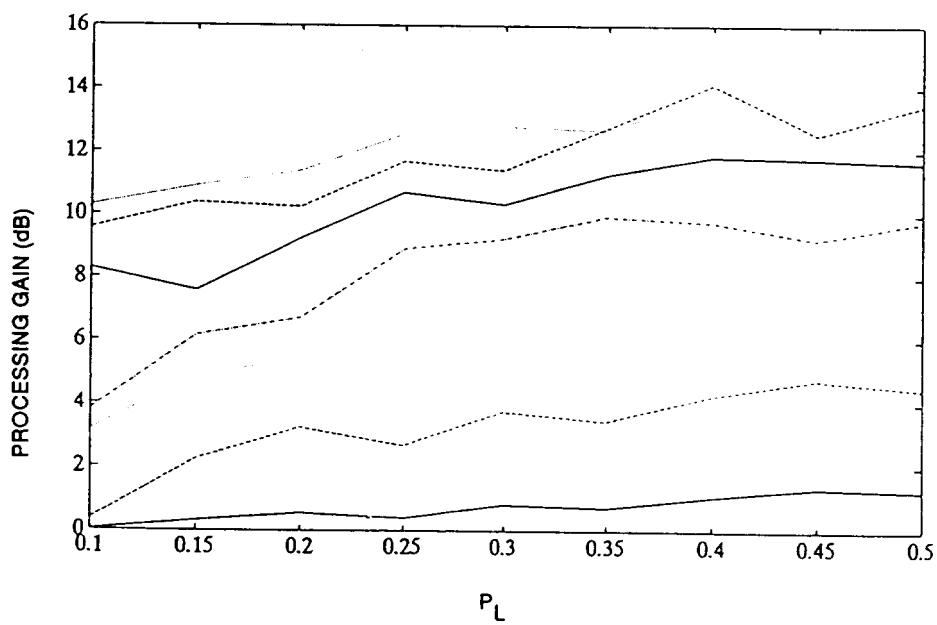


Figure 21. Processing gain for the two-state Gaussian mixtur model second-order detector.

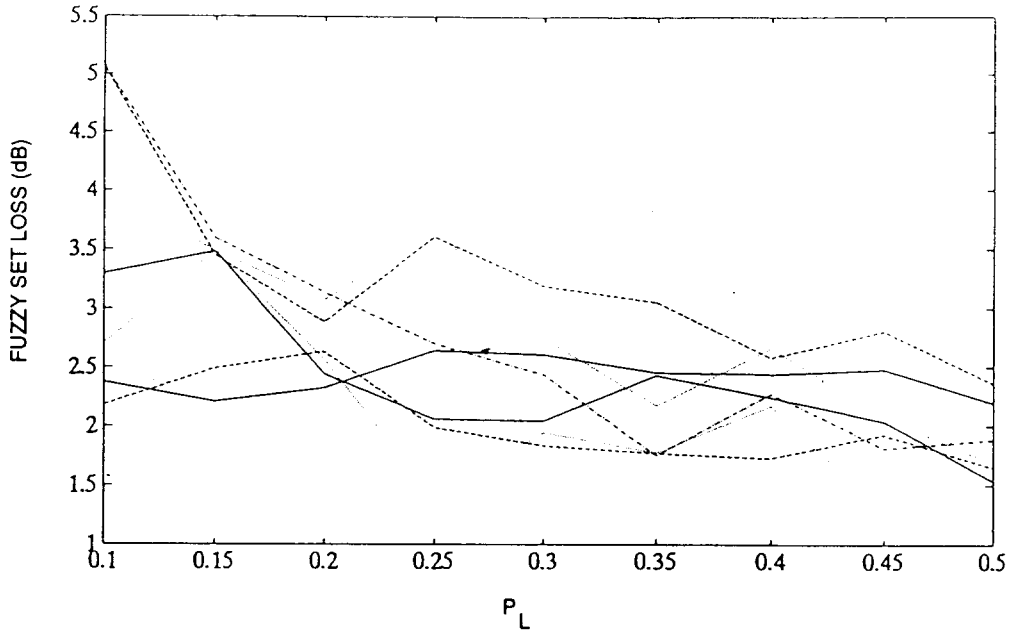


Figure 22. Fuzzy set loss for the two-state Gaussian mixture model second-order detector.

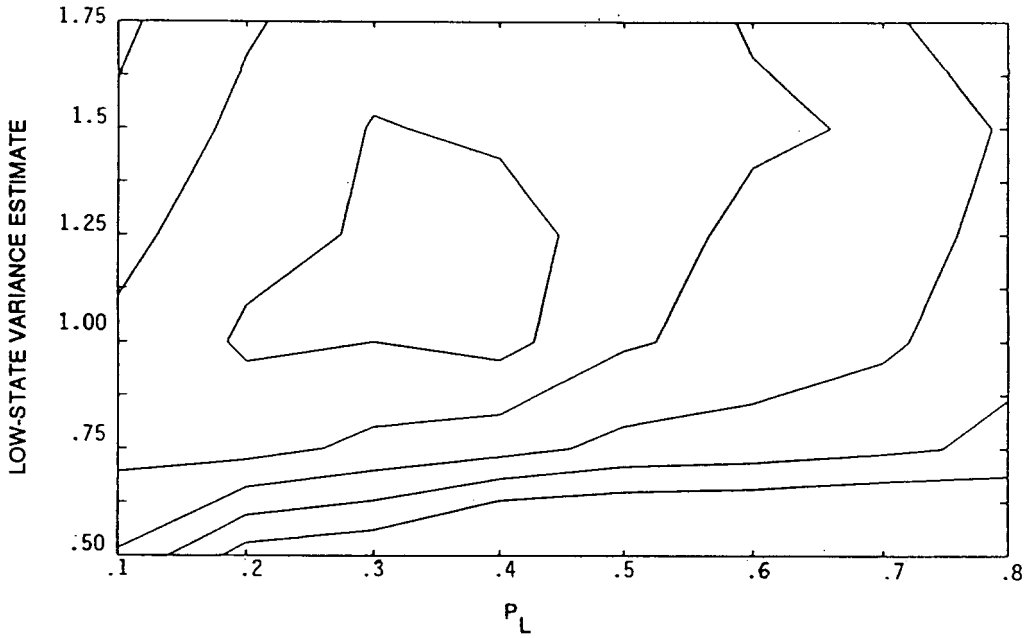


Figure 23. Parameter sensitivity for a two-state Gaussian mixture model for $\sigma_H^2 = 6$.

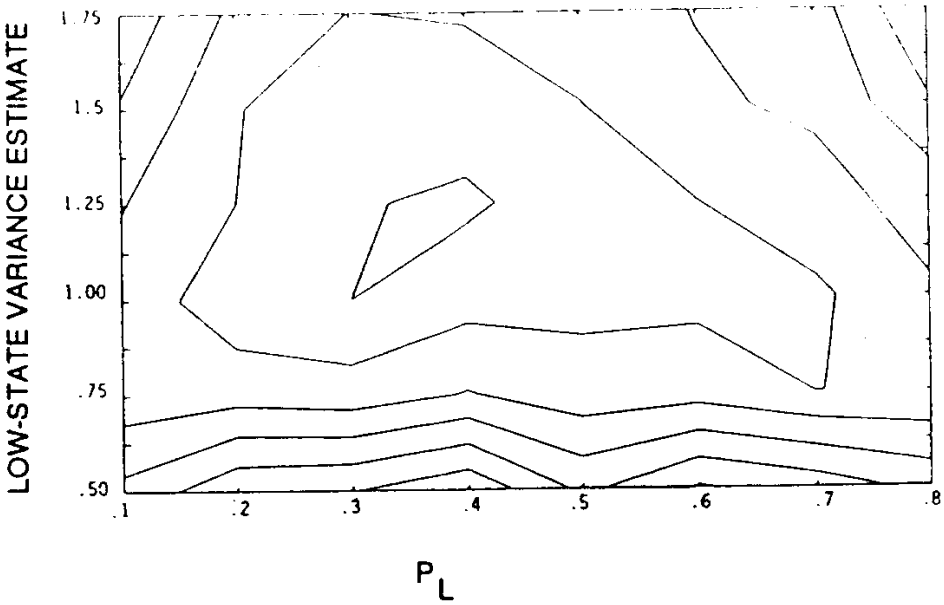


Figure 24. Parameter sensitivity for a two-state Gaussian mixture model for $\sigma_H^2 = 10$.

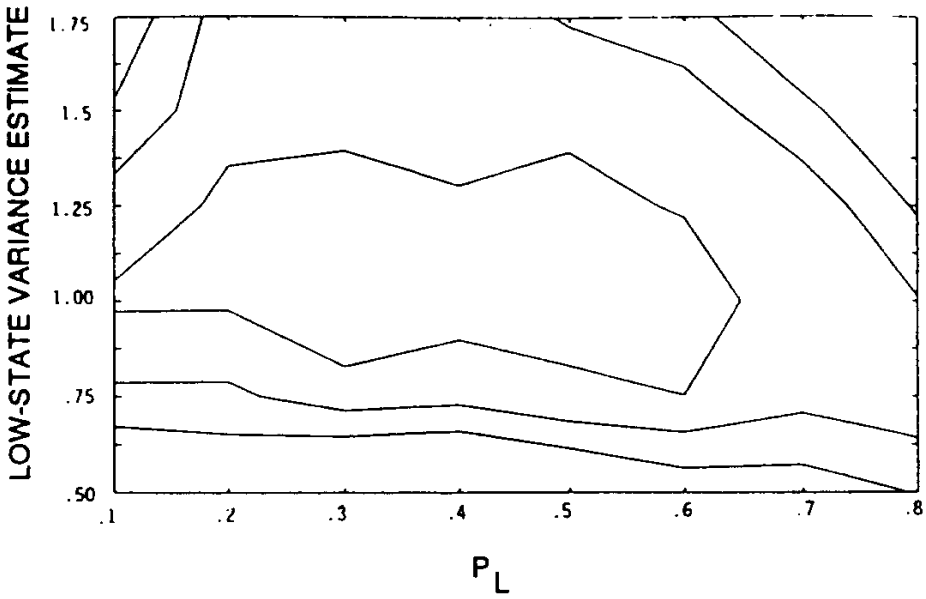


Figure 25. Parameter sensitivity for a two-state Gaussian mixture model for $\sigma_H^2 = 16$.

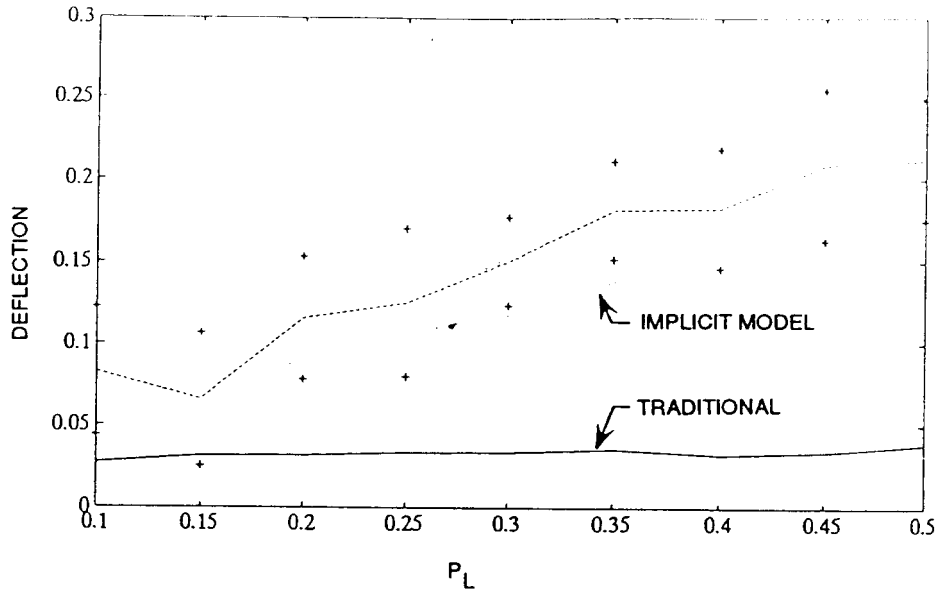


Figure 26. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 16 samples and high-state variance 16.

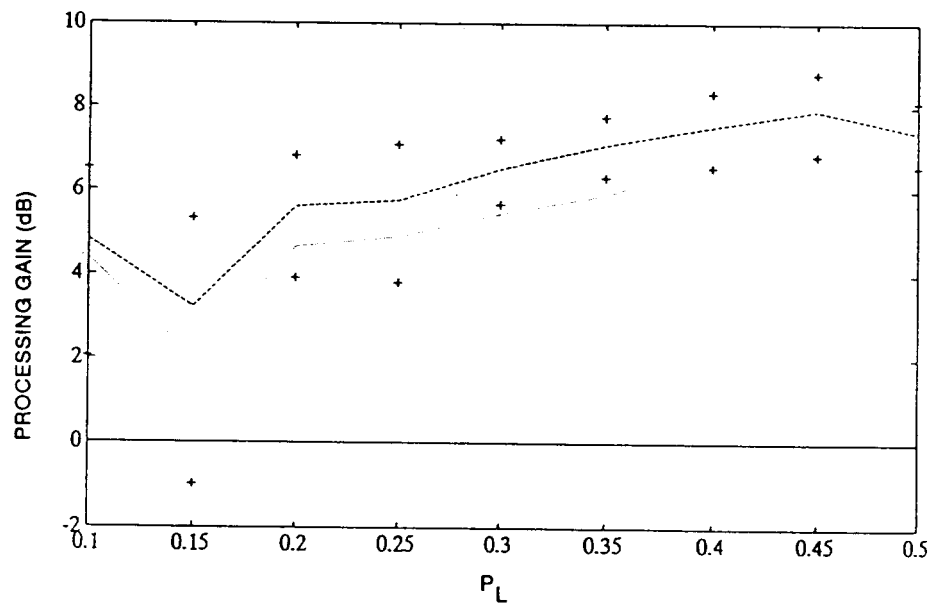


Figure 27. Implicit model processing gain for a two-state Gaussian mixture

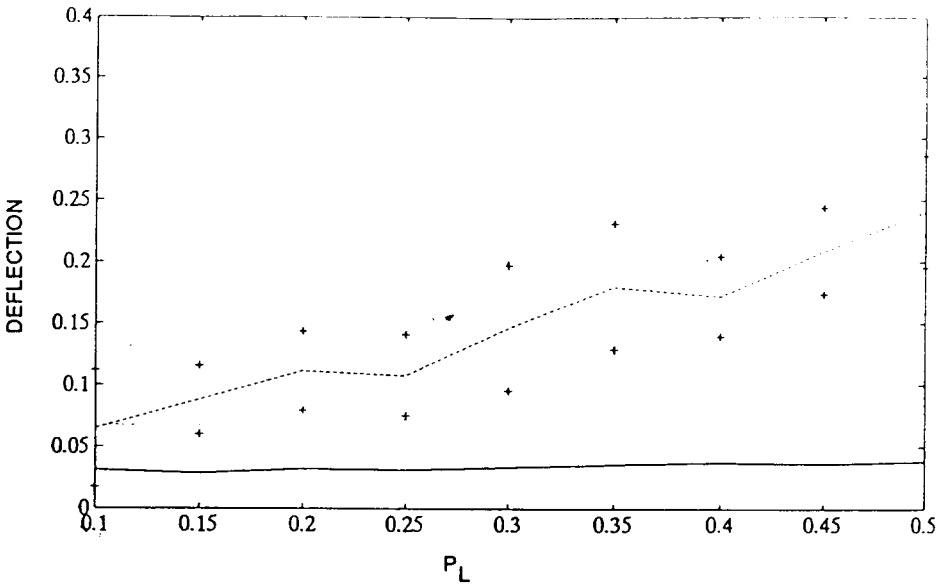


Figure 28. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 16.

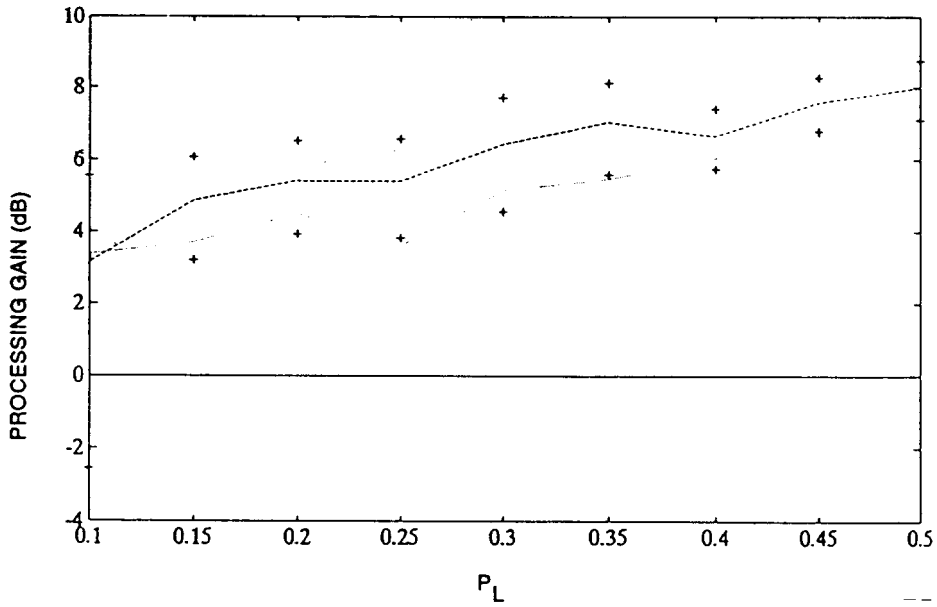


Figure 29. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 16.

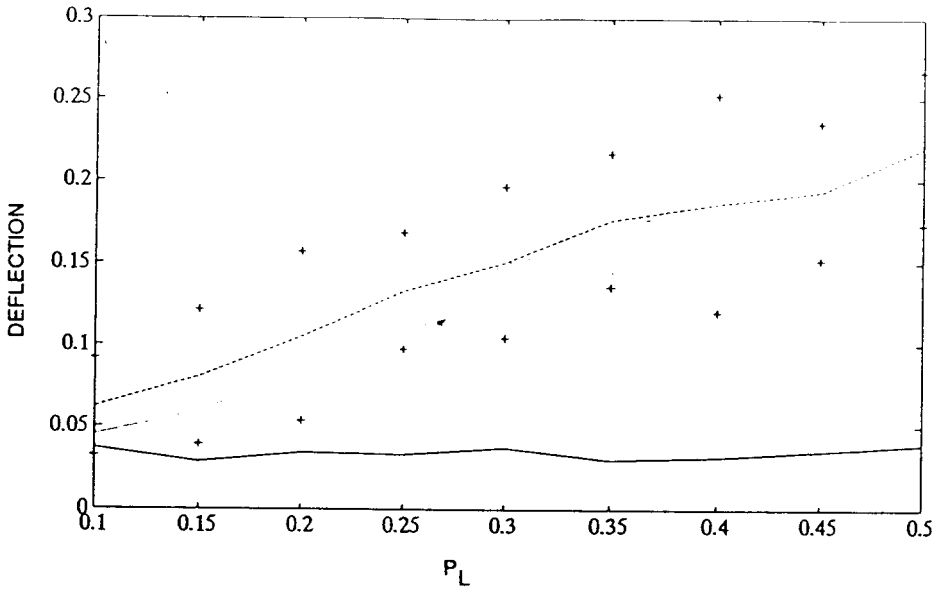


Figure 30. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 64 samples and high-state variance 16.

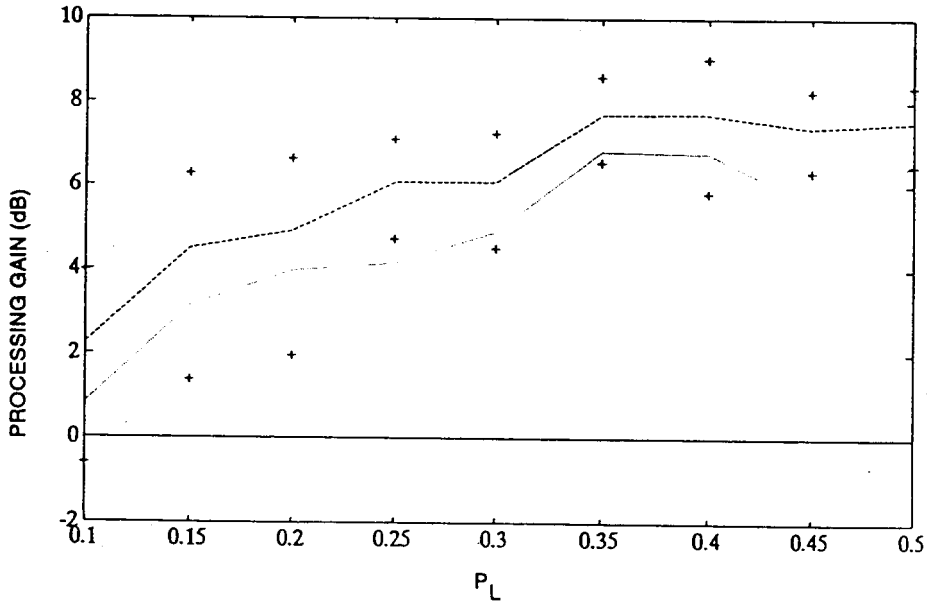


Figure 31. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 64 samples and high-state variance 16.

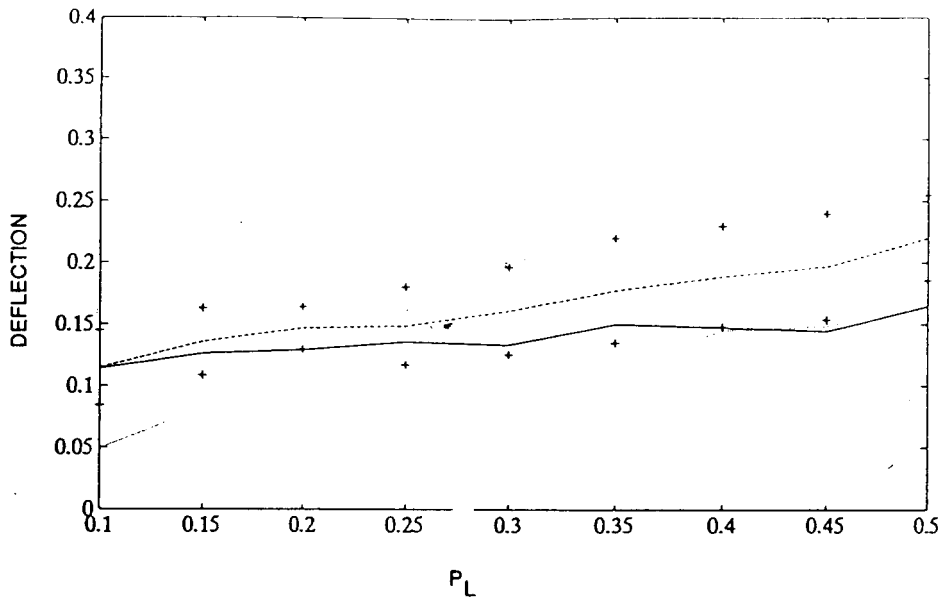


Figure 32. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 4.

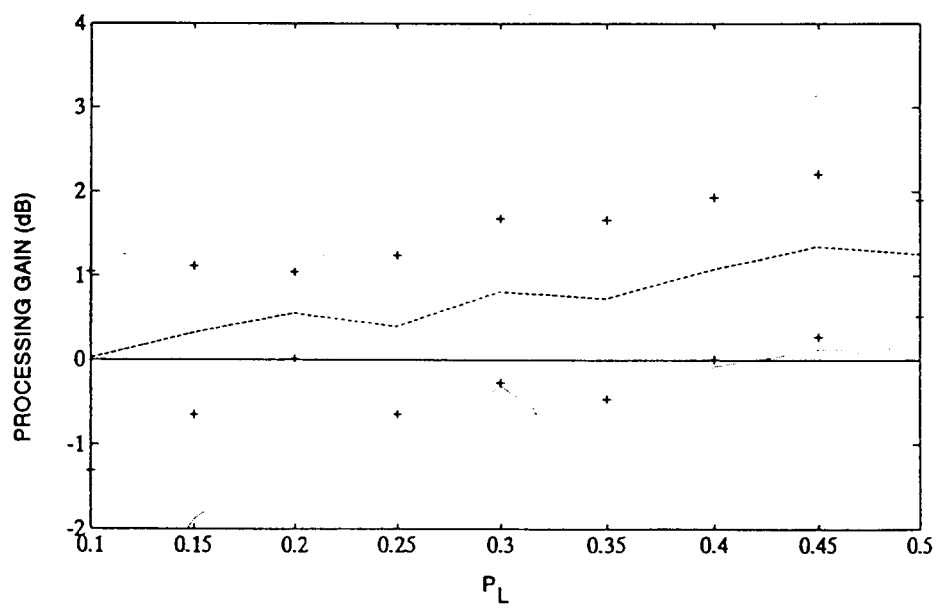


Figure 33. Implicit model processing gain for a two-state Gaussian mixture model second-order detector for 32 samples and high-state variance 4.

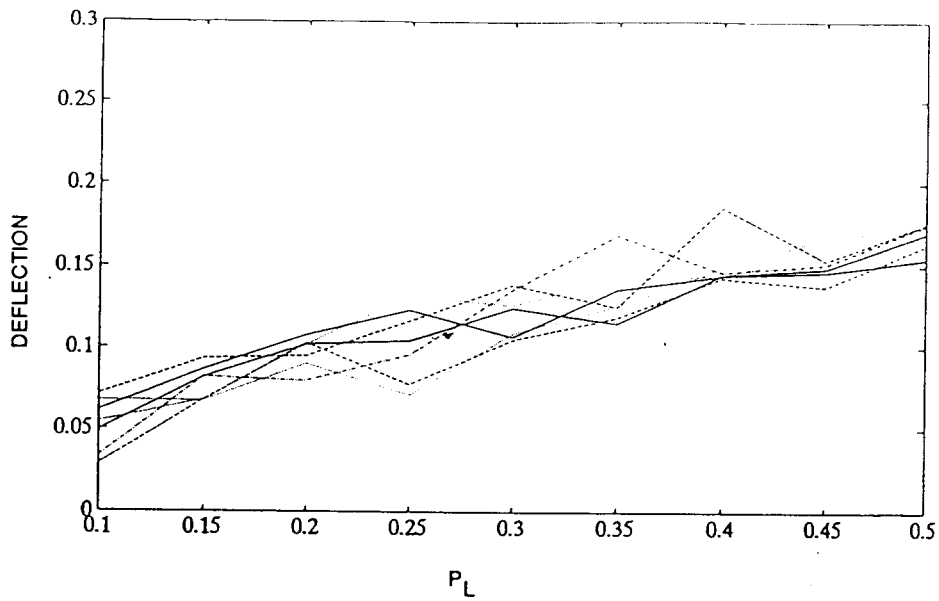


Figure 34. Implicit model deflection values for a two-state Gaussian mixture model second-order detector for 64 samples and high-state variance 16.

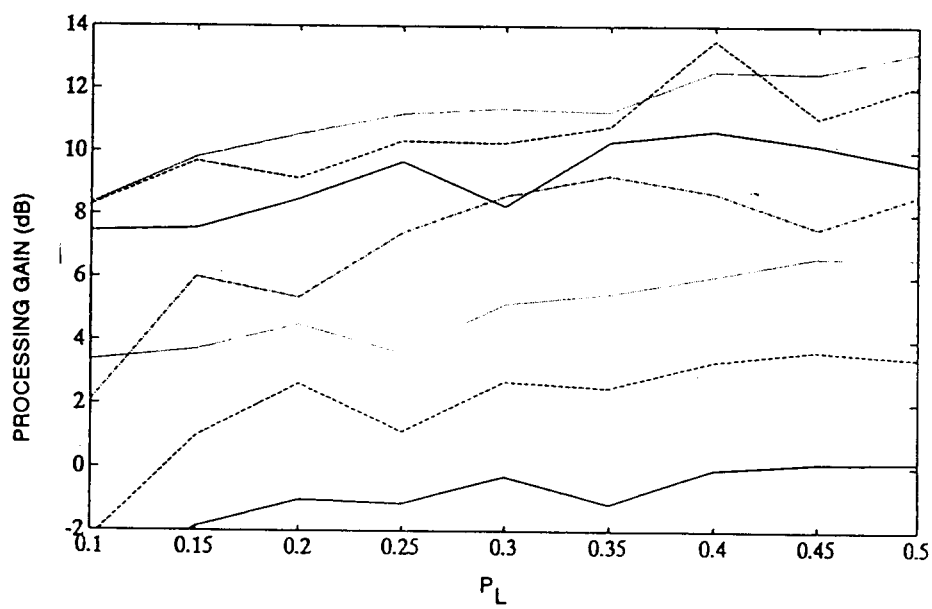


Figure 35. Implicit model processing gains for a two-state Gaussian mixture model second-order detector.

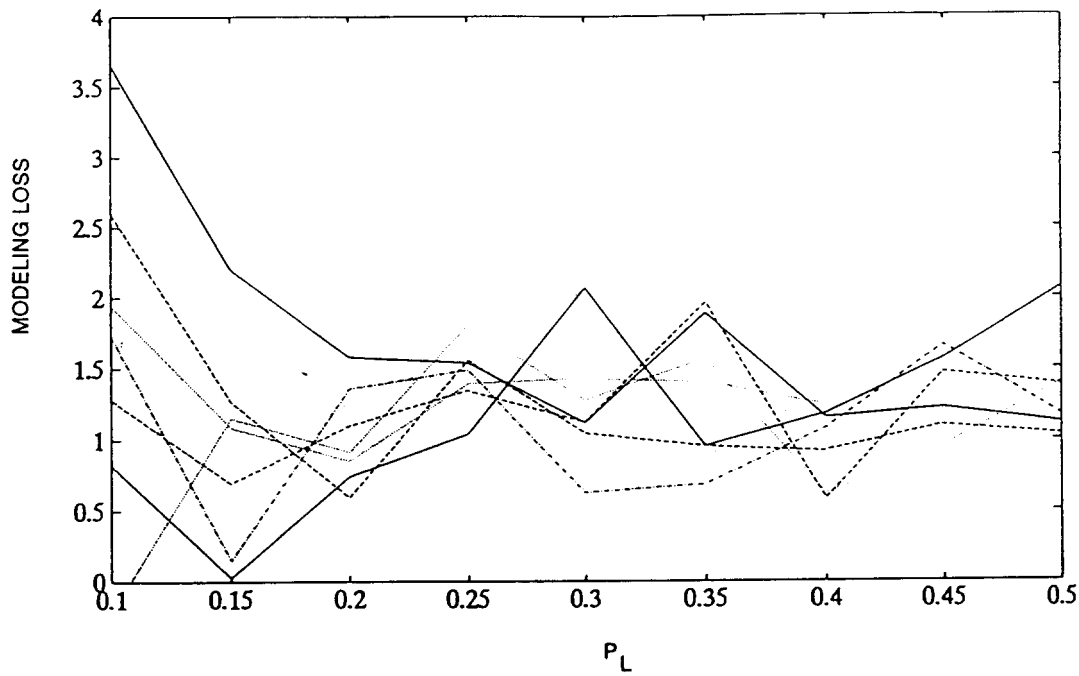


Figure 36. Implicit modeling loss for a two-state Gaussian mixture model second-order detector.

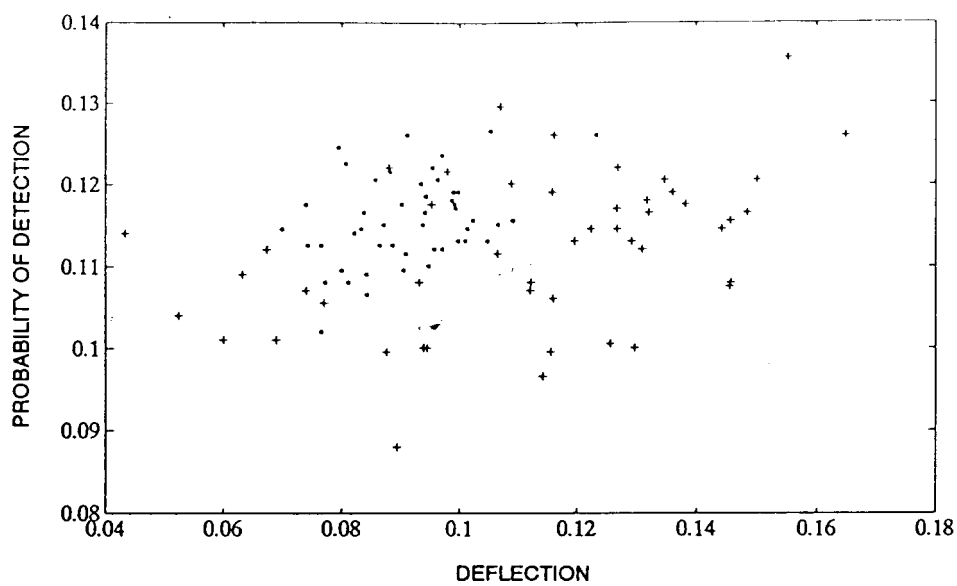


Figure 37. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 6$.

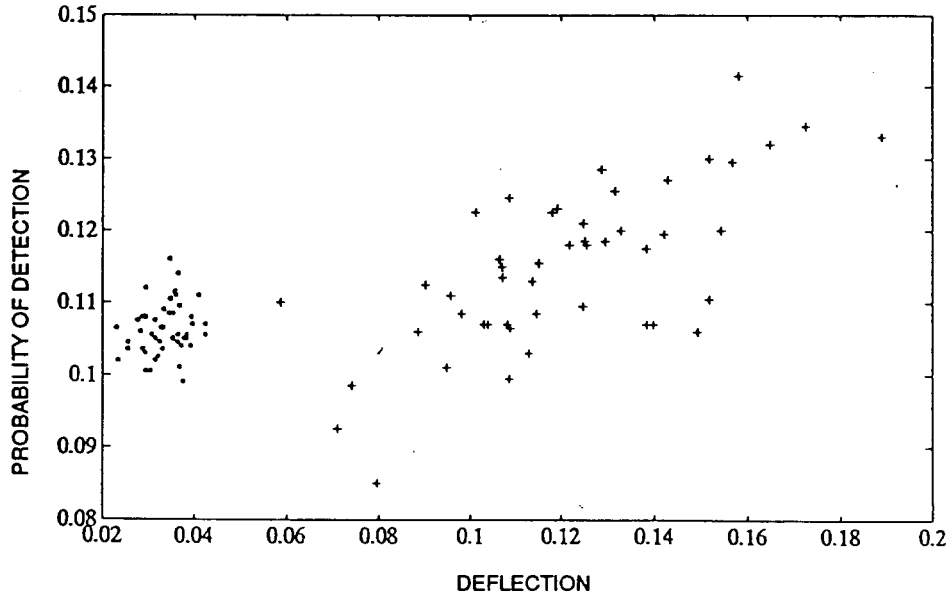


Figure 38. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$.

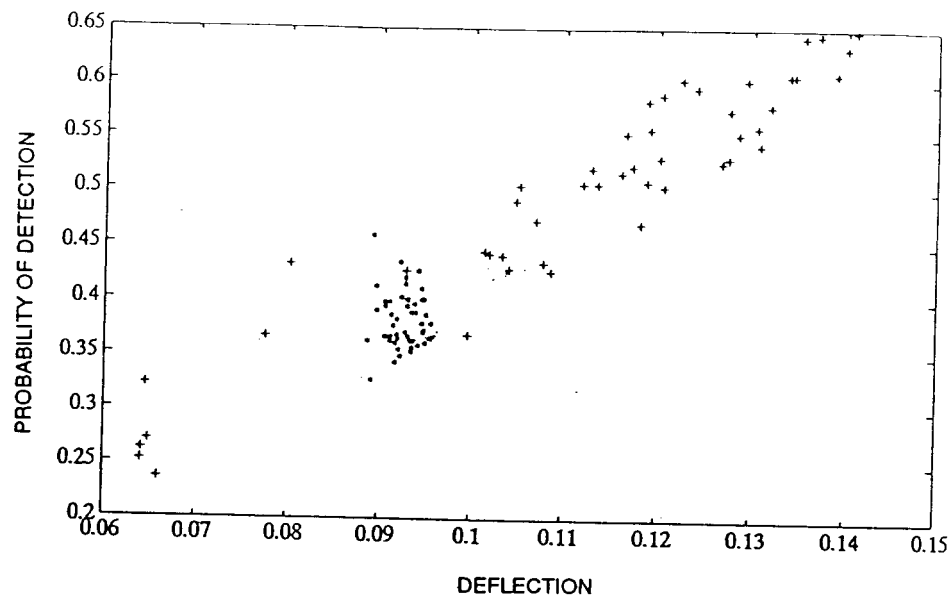


Figure 39. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 6$.

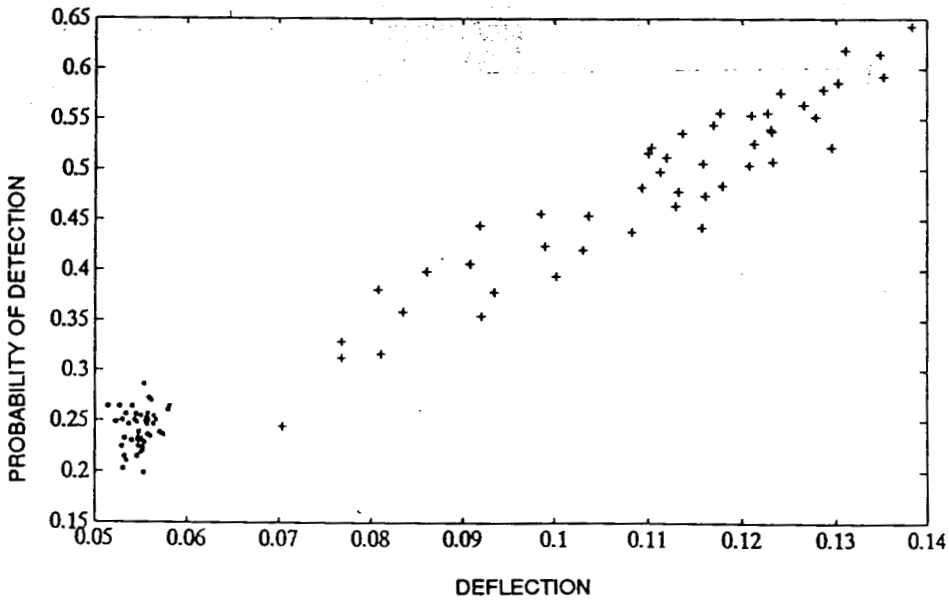


Figure 40. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 10$.

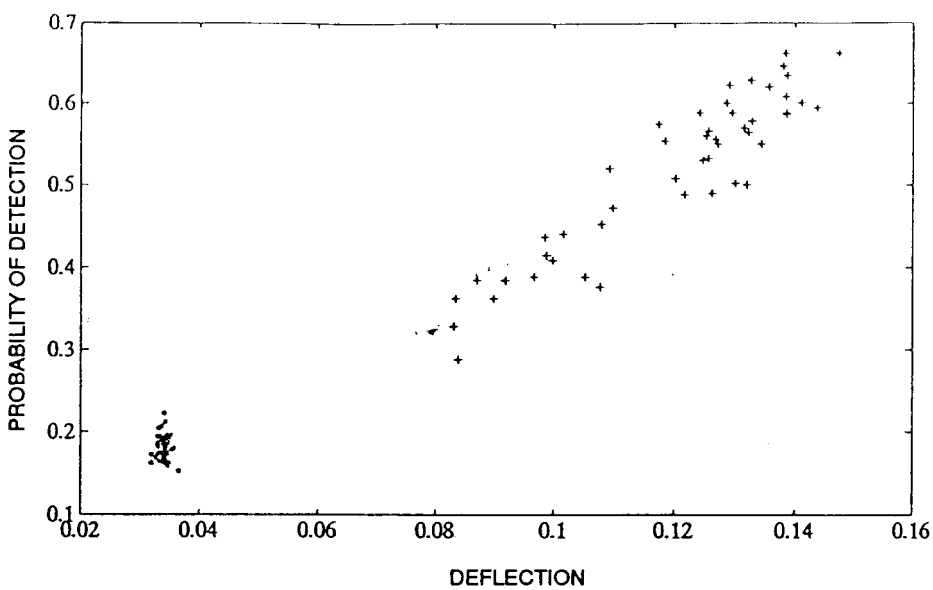


Figure 41. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 16$.

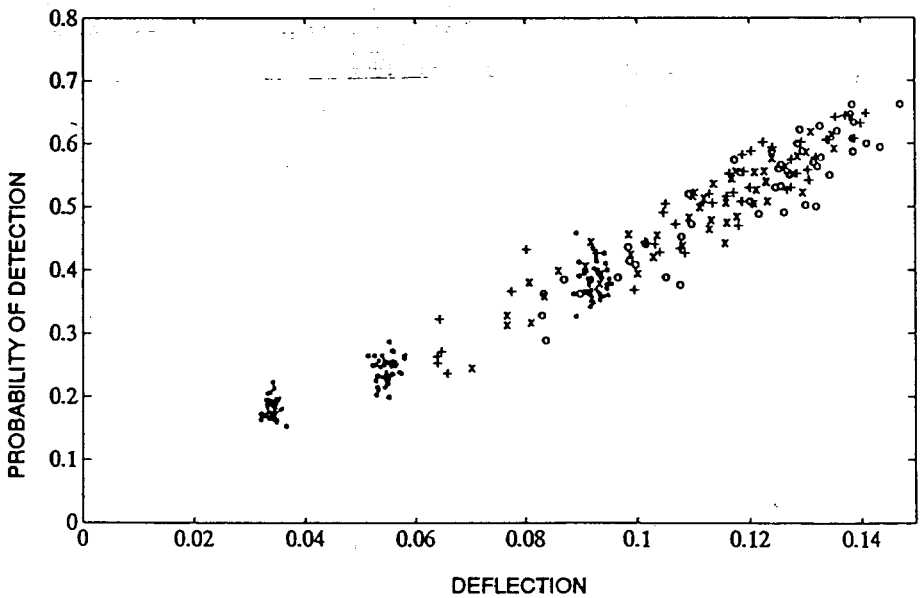
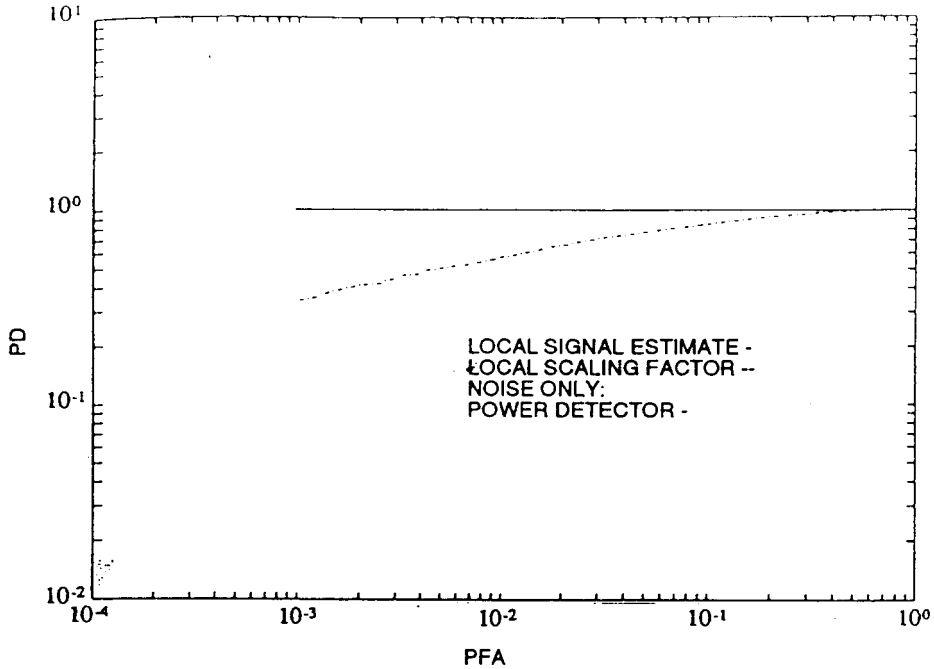
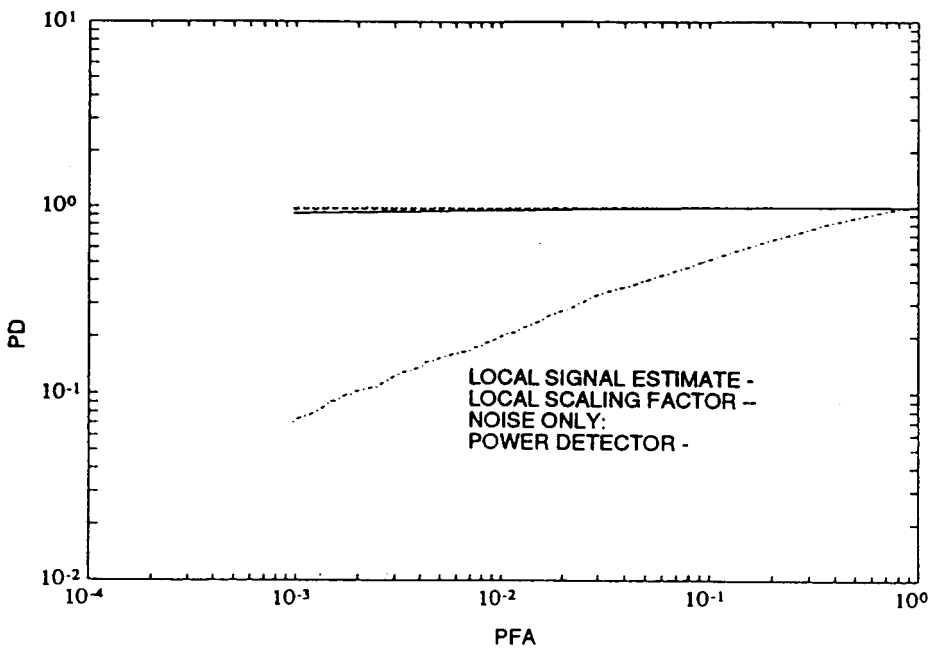


Figure 42. Probability of detection versus deflection for the two-state Gaussian mixture model second-order detector for $\sigma_H^2 = 6,\ 10,$ and 16.

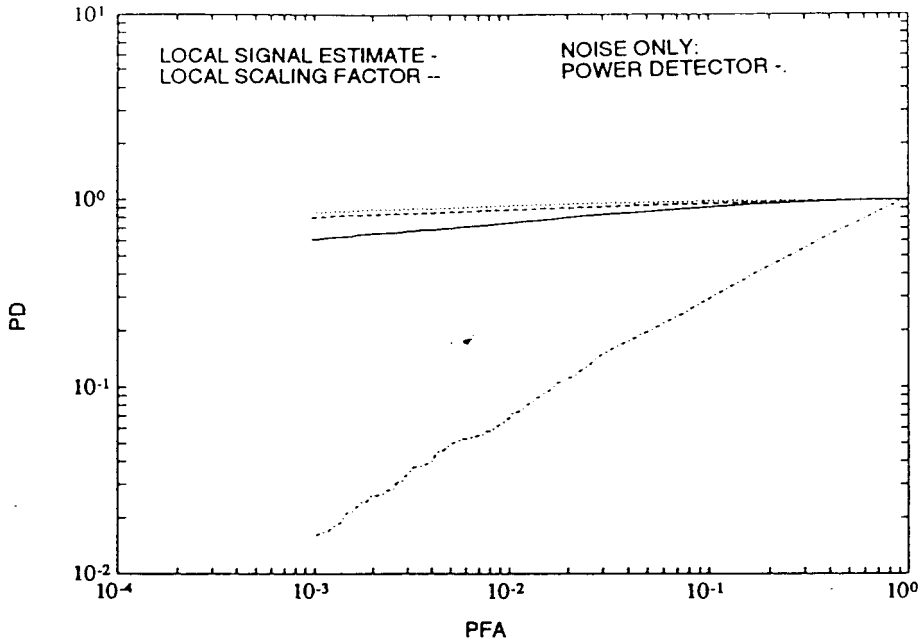


(a) Signal variance to low-state variance 2

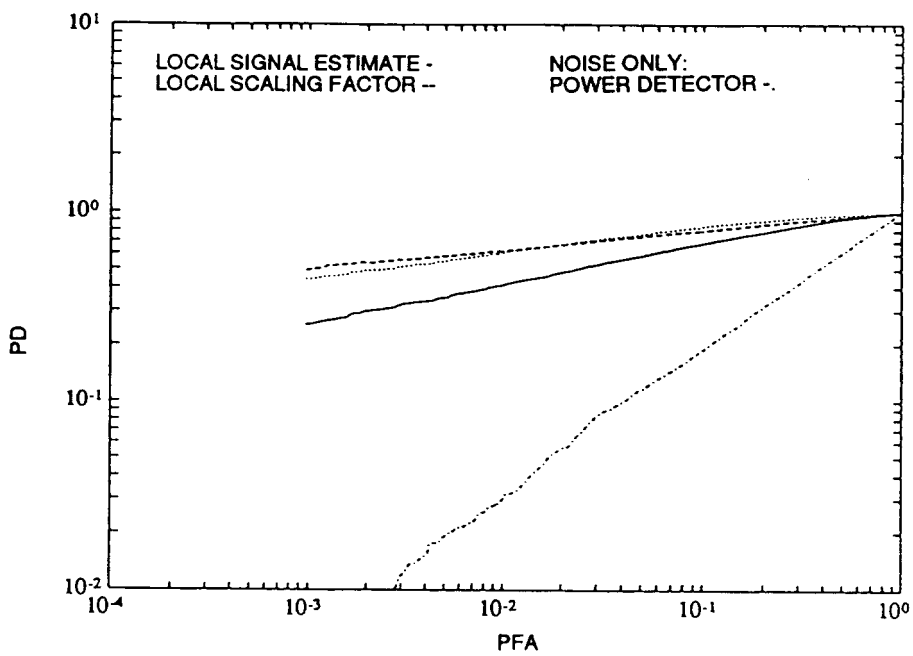


(b) Signal variance to low-state variance 1

Figure 43. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.5$ and $\sigma_H^2/\sigma_L^2 = 10$.

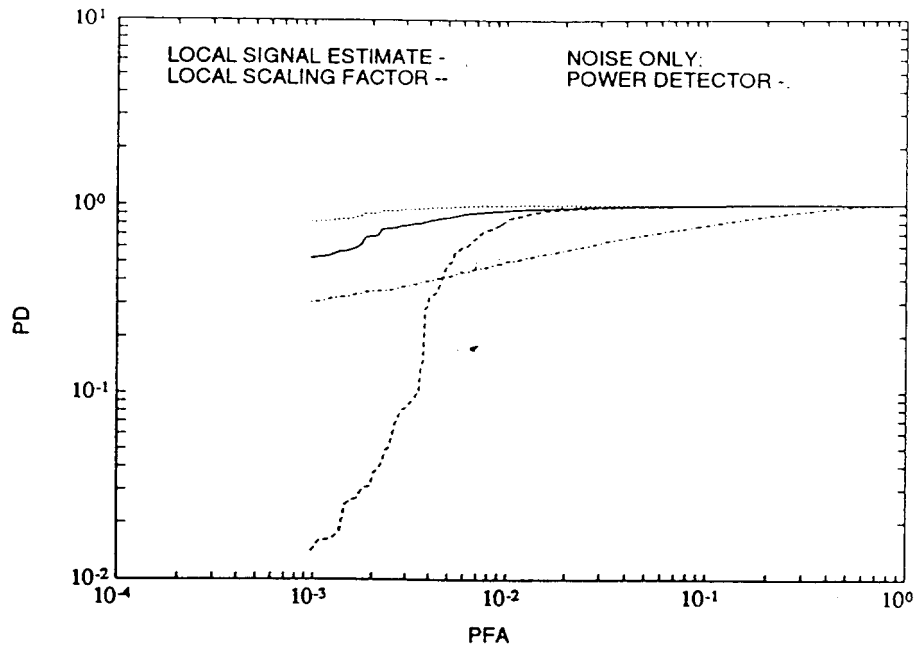


(c) Signal variance to low-state variance 0.5

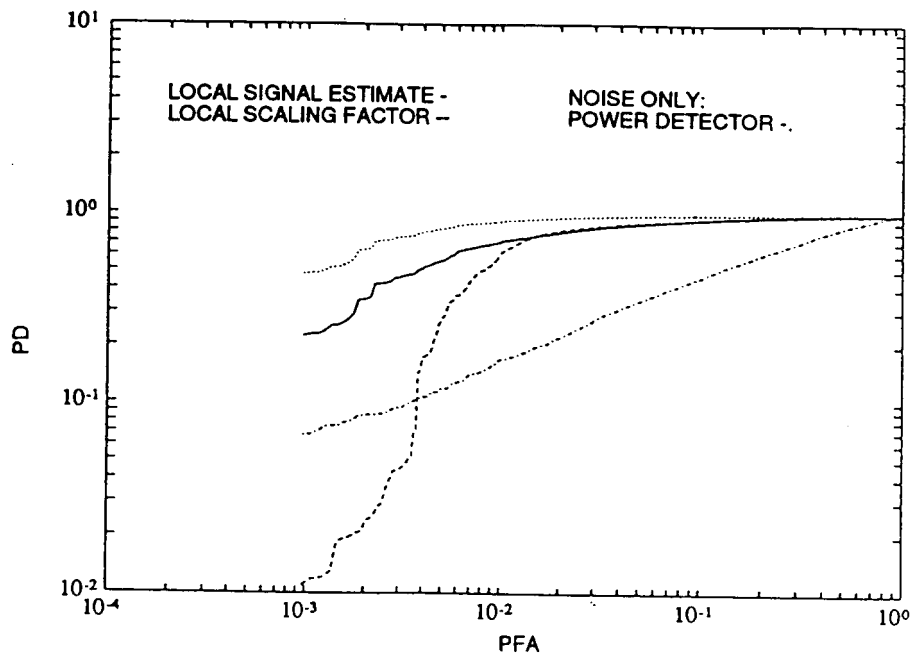


(d) Signal variance to low-state variance 0.25

Figure 43. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.5$ and $\sigma_H^2/\sigma_L^2 = 10$.

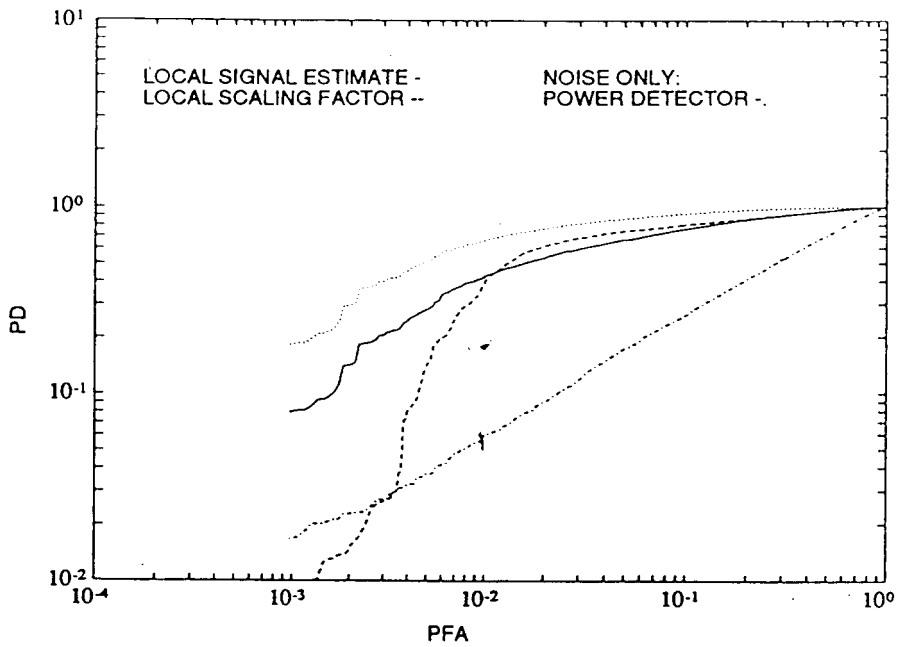


(a) Signal variance to low-state variance 2

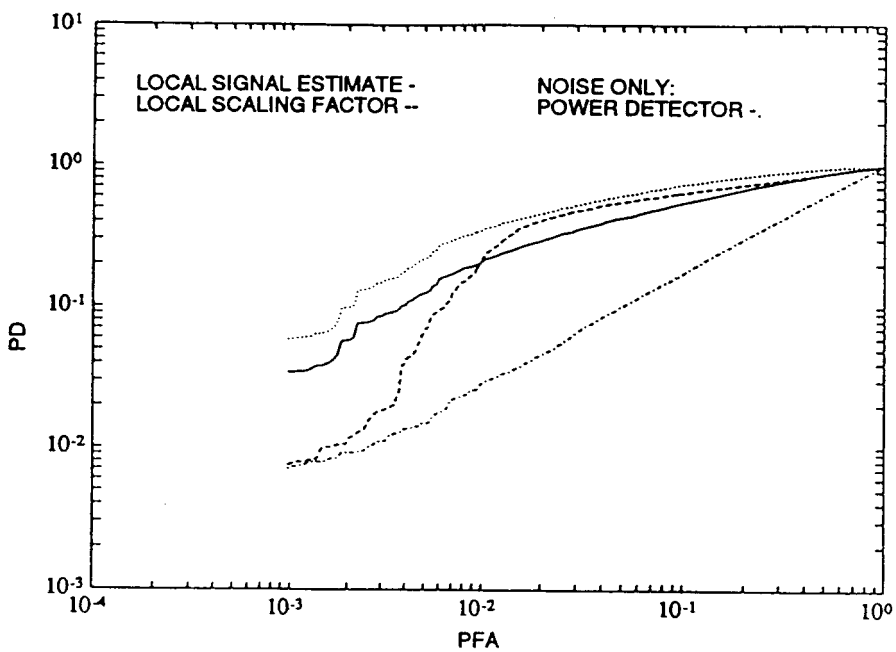


(b) Signal variance to low-state variance 1

Figure 44. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.25$ and $\sigma_H^2/\sigma_L^2 = 10$.

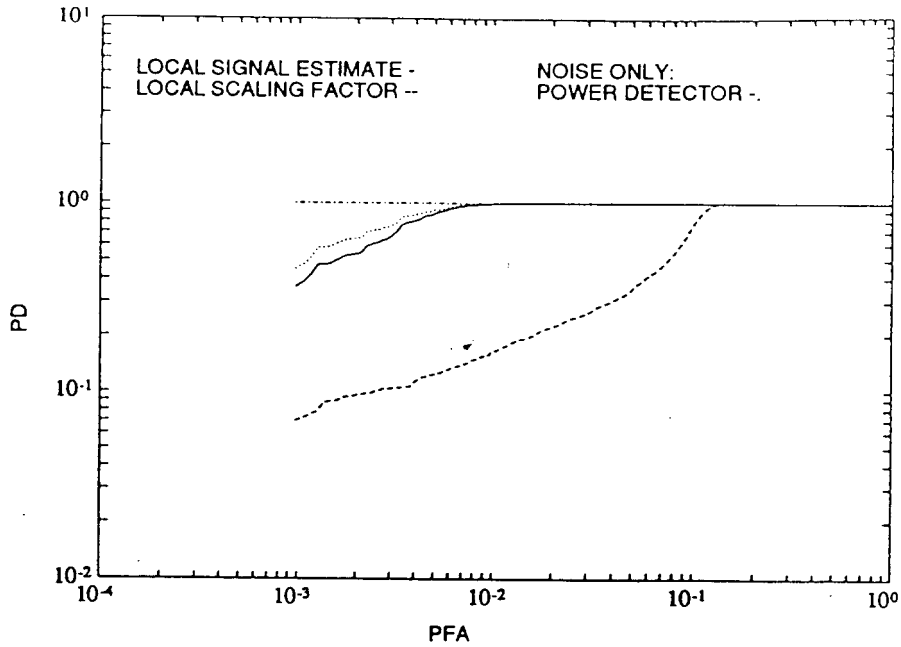


(c) Signal variance to low-state variance 0.5

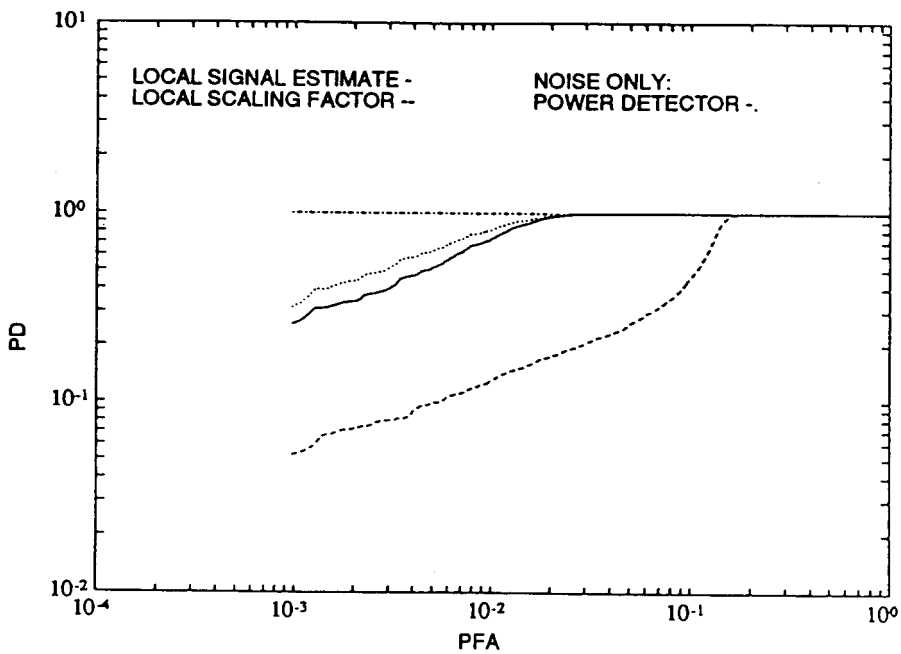


(d) Signal variance to low-state variance 0.25

Figure 44. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.25$ and $\sigma_H^2/\sigma_L^2 = 10$.

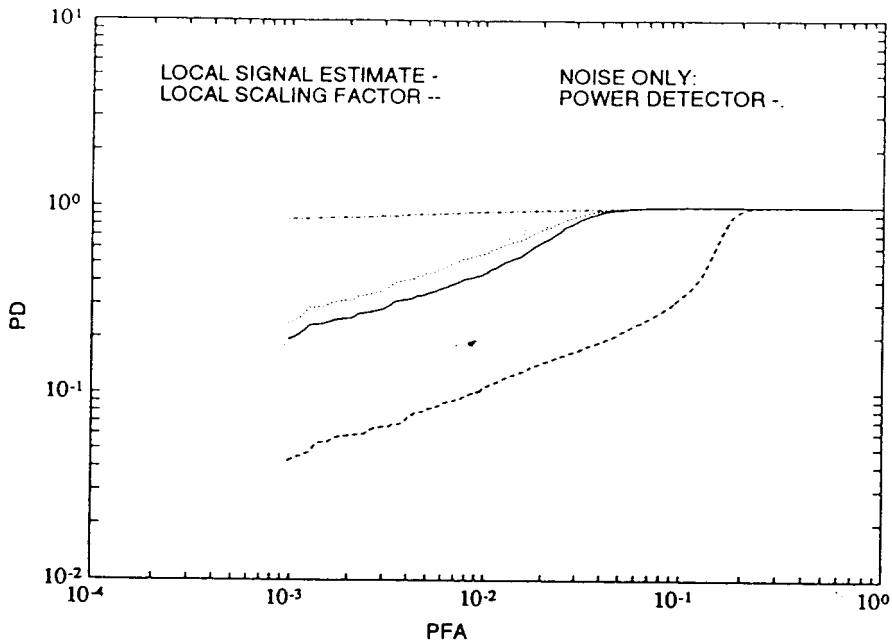


(a) Signal variance to low-state variance 2

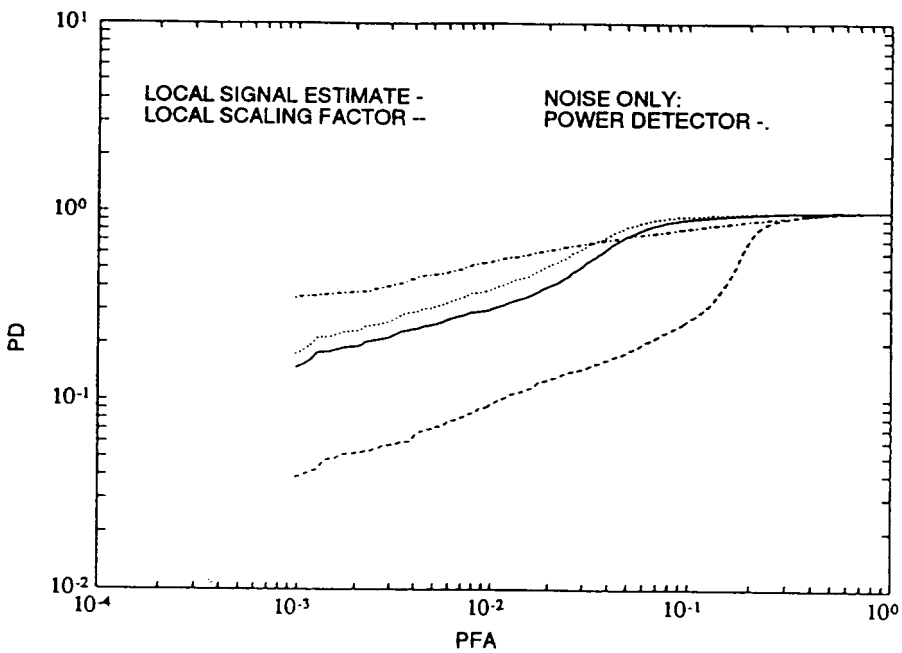


(b) Signal variance to low-state variance 1

Figure 45. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.1$ and $\sigma_H^2/\sigma_L^2 = 10$.



(c) Signal variance to low-state variance 0.5



(d) Signal variance to low-state variance 0.25

Figure 45. Receiver operating curves for a two-state Gaussian mixture model second-order line detector for $P_L = 0.1$ and $\sigma_H^2/\sigma_L^2 = 10$.

REFERENCES

Arnstein, D. A. 4–7 November 1991. “Smart AGC: A New Anti-Jam Device for Military Satellite Systems,” IEEE Mil. Comm. Conf. Record, McLean, VA, Paper no. 30.7.

Arnstein, D., C. Pike, and G. Estep. 11–14 October 1992. “On-Board AJ Enhancement Using Adaptive Non-Linear Processing: Practical Aspects of ‘Smart AGC’ Implementation,” IEEE Mil. Comm. Conf. Record, San Diego, CA, Paper no. 7.5.

Baker, C. R. and A. G. Gualtierotti. “Likelihood Ratio Detection of Stochastic Signals,” to appear in Advances in Statistical SIgnal Processing, Volume 2: Signal Detection, H. V. Poor and J. B. Thomas, eds., JAI Press, London, UK.

Beckman, P. 1962. “The Amplitude Probability Distribution of Atmospheric Radio Noise,” Institute of Radio Engineering and Electronics, Czechoslovak Academy of Sciences, 26.

Beckman, P. 1964. “Amplitude Probability Distribution of Atmospheric Radio Noise,” Radio Science, 68D, pp. 723–73.

Barniv, Y. January 1985. “Dynamic Programming Solution for Detecting Dim Moving Targets,” IEEE Transactions on Aerospace and Electronic Systms, vol. 21, no. 1, pp. 144–156.

Bar-Shalom, Y., T. Fortman. 1988. *Tracking and Data Association*, Academic Press, Boston, MA.

Bar-Shalom, Y. (Editor). 1990. *Multitarget-Multisensor Tracking*, Artech House, Norwood, MA.

Berry, L. A. 1981. “Understanding Middleton’s Canonical Formula for Class A Noise,” IEEE Trans. EMC 23(4), pp. 337–344.

Blachman, N. M. April 1964. “Bandpass Non-Linearities,” IEEE Trans. Inform. Theory, vol. IT–10, no. 2, pp. 162–164.

Blachman, N. M. 26–29 June 1992. “Biased vth-Law Rectifier Signal-to-Interference-Ratio Improvement,” Prov. 16th Biennial Conf. Comm., Queen’s University, Kingston, Ontario, Canada.

Blachman, N. M. July 1971. “Detectors, Bandpass Non-Linearities, and Their Optimization: Inversion of the Chebyshev Transform,” IEEE Transactions Inform. Theory, vol. IT–17, no. 4, pp. 398–404.

Blachman, N. M. 15 July 1971. “Signal/Noise-Ratio Performance of Optimum Nonlinearities,” Electron. Letter, vol. 17, no. 4, pp. 398–395.

Blachman, N. M. 1982. *Noise and Its Effect on Communication*, 2nd ed. Krieger, Malabar, FL.

Blostein, S. D. and T. S. Huang. 1991. “Detecting Small, Moving Objects in Image Sequences Using Sequential Hypothesis Testing,” IEEE Transactions on Signal Processing, vol. 39, no. 7, pp. 1611–1629.

Bond, J. 4–7 November 1991. “A Geometric Interpretation of Adaptive Locally Processing,” MILCOM 91, IEEE, para. 3.2.1.

Bond, J. 11–14 October 1992. “A Relationship between Adaptive Locally Processing and Adaptive Wiener Filtering,” MILCOM 92, classified volume.

Bond, J. and S. Hui. “A Geometric Theory of Locally Optimum Processing,” to appear in an Institute of Mathematics book on Signal Processing.

Bond, J. and H. Schmidt. “Adaptive Locally Optimum Processing for Moderately Spread SATCOM AJ Modes,” to appear in MILCOM 93 conference proceedings.

Bouvet, M. and S. C. Schwartz. 1988. “Underwater Noises: Statistical Modeling, Detection, and Normalization,” J. Acoust. Soc. Am. vol. 83, no. 3, pp. 1023–1031.

Bouvet, M. and S. C. Schwartz. 1989. “Comparison of Adaptive and Robust Receivers,” IEEE Transactions ASSP, vol. 37, no. 5.

Broman, V. December 1992. “Bay R Contact Management Software, Theoretical Extensions and Implementations and Performance Issues,” NCCOSC RDT&E Division TR 1561.

Chen, Y. May 1989. “On Suboptimal Detection of 3-Dimensional Moving Targets,” IEEE Transactions on Aerospace and Electronic Systems, vol. 25, no. 3, pp. 343–350.

Cowart, A. E., W. E. Snyder, and W. H. Ruedger. 1983. “The Detection of Unresolved Targets Using the Hough Transform,” Computer Vision, Graphics, Image Processing, vol. 21, pp. 222–238.

Gibbons, G. D., J. A. Musselman, J. D. Pehoushek, and J. R. Russel. 1987. “Data Fusion in the SES: A Sonar Expert System,” Technical Proceedings 1987 Tri-Service Data Fusion Symposium, vol. 1, pp. 357–364.

Gradshteyn, I. S., and I. Ryzhik. 1980. *Table of Integrals, Series, and Products*. Corrected and enlarged edition. Academic Press, New York, NY., Sec. 3.471, vol. 12, p. 340.

Halmos, P. 1957. *Introduction to Hilbert Space*, 2nd edition, Chelsea, New York, NY.

Higbie, J. 23–26 October 1988. “Adaptive Nonlinear Suppression of Interference,” MILCOM 88, IEEE, P. 23.3.1.

Kassam, S. A. and H. V. Poor. 1985. “Robust Techniques for Signal Processing: A Survey,” Prox. IEEE vol. 73, no. 3, pp. 433–481.

Kessler, K. M., M. R. Sutphen, T. J. Holmes, and B. K. McQuiston. November 1988. “Dynamic Programming Algorithm (DPA) Analyses,” Final Report for Titan Systems, Inc.

Klir, J. and T. Folger. 1988. *Fuzzy Sets, Uncertainty and Information*. Prentice–Hall, Englewood Cliffs, NJ.

Lee, H., C. Chung, and K. Chang. 1991. “Maneuvering Target Tracking in a Cluttered Environment from Noisy Images by an Adaptive Probabilistic Data Association Filter,” International Journal of System Science, vol. 22, no. 1, pp. 133–149.

Middelton, D. 1960. *Introduction to Statistical Communications Theory.* McGraw–Hill, New York, NY.

Middleton, D. 1966. “Canonically Optimum Threshold Detection,” Trans. Inform Theory, vol. 12, no. 2, pp. 230–243.

Middelton, D. 1967. “A Statistically Theory of Reverberation and Similar First-Order Scattered Fields,” Trans. Inform. Theory, vol. 13, no. 3, pp. 372–393.

Middelton, D. 1977. “Statistical Physical Models of Electromagnetic Interference,” Trans. on EMC, vol. 19, no. 3, pp. 106–127.

Middleton, D. 1983. “Canonical and Quasi-Canonical Probability Models of Class A Interference,” IEEE Transactions on EMC, vol. 25, no. 2, pp. 76–106.

Middleton, D. May 1984. “Multiple-Element Threshold Signal Detection of Underwater Acoustic Signals in Non-Gaussian Interference Environments,” NOSC CR 23.

Middleton, D. 1987. “Channel Modeling and Threshold Signal Processing in Underwater Acoustics: An Analytical Overview,” IEEE J. Oceanic Eng. OE-12, 1, pp. 4–28.

Middleton, D. 12–14 March 1991. “Threshold Detection and Estimation in Correlated Interference,” 9th Intl. Zurich Symposium on EMC: “EMC ′91,” Switzerland.

Middleton, D. and A. Spaulding. May 1983. “Optimum Reception in Non-Gaussian Electromagnetic Interference Environments: II Optimum and Suboptimum Threshold Signal Detection in Class A and B Noise,” NTIA Report 83–120.

Middleton, D. and A. Spaulding. May 1986. “A Tutorial Review of Elements of Weak Signal Detection in Non-Gaussian EMI Environments,” NTIA Report 86–184.

Mohanty, N. C. September 1981. “Computer Tracking of Moving Point Targets in Space,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 3, no. 5, pp. 606–611.

Murphy, J. M., P. B. Tilley, and F. M. Torre. March 1990. “Adaptive Nonlinear Coherent Processor Design,” vol. I and II, Rome Air Development Center.

Nagarajan, V., R. N. Sharma, and M. R. Chidambara. September 1984. “An Algorithm for Tracking a Maneuvering Target in Clutter,” IEEE Transactions on Aerospace and Electronic Systems, vol. 20, no. 5, pp. 560–573.

Nitzberg, R. January 1972. “Constant False Alarm Rate Signal Processors for Several Types of Interferences,” IEEE Transactions on Aerospace & Electronic Systems, AES–8, pp. 27–34.

Nitzberg, R. 1992. *Adaptive Signal Processing for Radar*, Artech House, Boston, MA.

Oppenheim, A. V., and R. W. Shafter. 1989. *Discrete-Time Signal Processing*, Prentice–Hall, Englewood Cliffs, NJ.

Papoulis, A. 1977. *Signal Analysis*, McGraw Hill, New York, NY.

Poor, H. V. 1988. “An Introduction to Signal Detection and Estimation,” Springer Verlag, New York, NY.

Porat, B., and B. Friedlander. April 1990. “A Frequency Domain Algorithm for Multiframe Detection and Estimation of Dim Targets,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 4, pp. 398–401.

Pryor, C. M. 2 August 1971. “Calculation of the Minimum Detectable Signal for Practical Spectrum Analyzers,” Naval Ordnance Laboratory Technical Report, 71–92.

Rudin. 1973. *Functional Analysis*, McGraw–Hill Book Company, New York, NY.

Scholz, J. B. September 1991. “Trial Results of a Novel Bit Error Rate Monitor Scheme Over a Skywave HF Link,” IEEE Letters, vol. 27, no. 19, pp. 1730–1732.

Scholz, J. B. April 1992. “Real-Time Performance Estimation and Optimization of Digital Communications Links,” Ph.D. thesis, Univ. of Adelaide, Australia.

Scholz, J. B. and T. C. Giles. October 1990. “A Revolutionary High Performance Passive Real-Time Link Evaluator for Fast Adaptive Communications,” Comm. 90 Conf., Melbourne, Aust., pp. 186–191.

Scholz, J. B. and T. C. Giles. September 1992. “Generalized Maximum A Posterior Probability (MAP) Estimation in Communications Systdems,” in Proceedings of the First Intl Conf. on DSP for Comm. Systems, Univ. of Warwick, UK, 7–9.

Shensa, M. J. and V. Broman. 1985. “The Sequential Processing of Report Detections in Multitarget Data Association,” Proc. 1985 American Control Conference, pp. 1026–1031.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, UK.

Stein, D. W., J. W. Bond, and J. R. Zeidler. February 1993. “Adaptive Locally Optimal Detection of Narrowband Signals in the Presence of Narrowband Interference,” NCCOSC RDT&E TR 1566.

Struzinski, W. A. 14 February 1978. “Analysis of Existing Detection and Classification Systems Investigated for Use in the All Target Classifier (ATC) (U),” NUSC TD 5777 (Secret NOFORN).

Wei, P., J. Zeidler, and W. H. Ku. April 1992. “Target Detection from Image Sequences Using Pixel-Based Decision Criterion,” SPIE Proceedings, vol. 1698.

Wei, P., J. Zeidler, and W. H. Ku. May 1993. “Adaptive Interference Suppression for CDMA Based Personal Communications Networks,” Proc. of the IEEE International Conference on Communications.

Wei, P., J. Zeidler, and W. H. Ku. 1993. “Analysis of the Track-Before-Detect Approach to Target Detection using Pixel Statistics,” to appear in IEEE Transactions on Aerospace and Electronic Systems.

Whalen, A. D. 1971. *Detection of Signals in Noise*, Academic Press, New York, NY.

Wolfram Research, Inc. 1993. *Mathematica*, Version 2.2. Wolfram Research, Inc., Champaign, IL.

APPENDIX A

GLOBAL PROCESSING GAIN BOUNDS FOR MULTISTATE GAUSSIAN MIXTURE MODELS

In this appendix, we obtain global bounds for the processing gain for multistate Gaussian mixture model first-order and second-order detectors. The bounds are obtained by relating the processing gain of any three or more state Gaussian mixture model to the processing gain of a constructed Gaussian mixture model with fewer states.

We would expect the processing gain achievable for a multistate model for the first-order detector to depend on the low-state and high-state variances and the overall noise variance. We would like to obtain an upper bound for the processing gain subject to $\sigma_K^2 = \sigma_1^2$, $\sigma_H^2 = \sigma_S^2$ and $\sum_{k=1}^{s} p_k \sigma_k^2 = \sigma^2$. Moving energy from a middle state to outer states should make the distribution less Gaussian-like and increase processing gain, while moving energy from the outer states to a middle state should decrease processing gain. It is rather easy to make this idea precise, as we now proceed to do. This suggests that the maximal processing gain is achieved when the interference is modeled by a two-state model with the energy in the lowest and highest states, for which processing gain is determined by the probability of the lower state and the ratio of the variances of the states, that is, the dynamic range of the interference and the percentage of time with little interference. This, in turn, provides a model-free criteria for when it is worthwhile to use adaptive locally optimum processing first-order detectors rather than traditional coherent detection processing.

We can make precise the argument outlined in the previous paragraph by focusing our attention on any three states of the N > 2 state model. Suppose the states have probabilities p_L, p_M, and p_H with variances $\sigma_L^2 < \sigma_M^2 < \sigma_H^2$. Let $\hat{p} = p_L + p_M + p_H$ and $\hat{\sigma}^2 = p_L\sigma_L^2 + p_M\sigma_M^2 + p_H\sigma_H^2$. Then, introduce the normalized probabilities $\hat{p}_L = \frac{p_L}{\hat{p}}$, $\hat{p}_M = \frac{p_M}{\hat{p}}$, and $\hat{p}_H = \frac{p_H}{\hat{p}}$ and the normalized variances $a = \frac{\hat{p}}{\hat{\sigma}^2}\sigma_L^2$, $b = \frac{\hat{p}}{\hat{\sigma}^2}\sigma_M^2$, and $c = \frac{\hat{p}}{\hat{\sigma}^2}\sigma_H^2$. The conditions on the normalized probabilities define a probability region in three-dimensional space as indicated in figure A-1. The probability region lies in the plane passing through the three points (1, 0, 0), (0, 1, 0), and (0, 0, 1) and within the triangle with these points as vertices. The constant variance restraint, $a\hat{p}_L + b\hat{p}_M + c\hat{p}_H = 1$, defines a plane passing through $(\frac{1}{a}, 0, 0)$, $(0, \frac{1}{b}, 0)$, and $(0, 0, \frac{1}{c})$. Figure A-1 shows the case when b > 1 and the intersection of the constant variance plane and the probability region is a line intersecting sides B and C of the probability region triangle; when b $\leq$ 1, their intersection can be parametrized by the low-state normalized probability by using, in succession, two substitutions: (1) $\hat{p}_M = 1 - \hat{p}_L - \hat{p}_H$ and (2) $\hat{p}_H(c–b) = (1–b)–\hat{p}_L(a–b)$, which is simply rewriting the sum of the state variances times their normalized probabilities. Equation (2) can be rewritten as $\hat{p}_L(b - a) = \hat{p}_H(c - b) + (b - 1) \geq b - 1$ because the normalized probabilities are non-negative. This leads to $\hat{p}_L \geq 0$ for the geometry of case b $\leq$ 1 and $\hat{p}_L \geq \frac{b-1}{b-a}$ for the geometry of case b > 1, The processing gain $g_1(p)$ is a simple function of the low-state probability restricted to the line of intersection for both geometries. It is described by a line with positive slope.

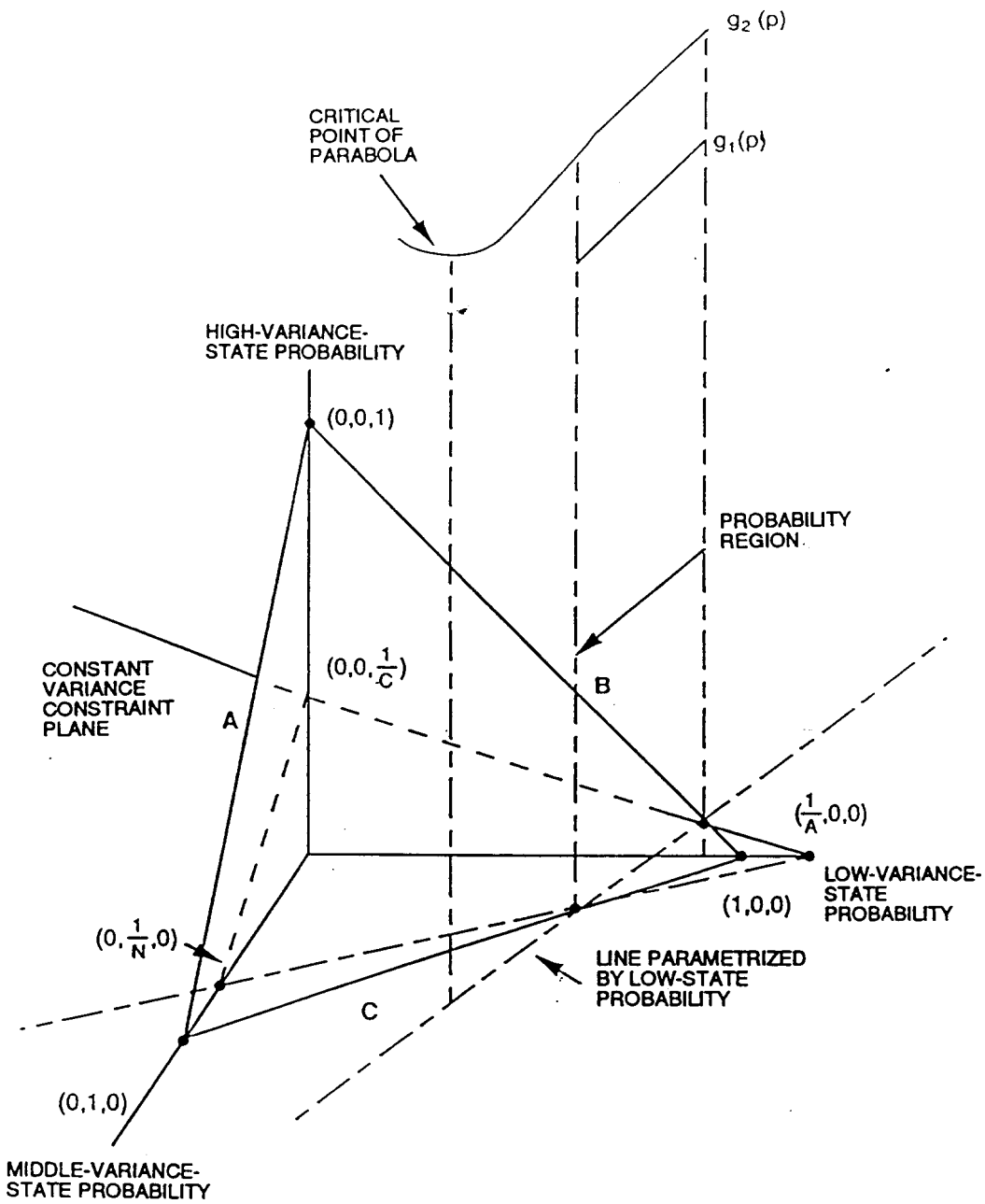


Figure A-1. Processing gain parametrized by low-state probability.

Rewrite the upper bound for processing gain as follows

$$\sum_{k=1}^{S} p_k \frac{\sigma^2}{\sigma_k^2} = \hat{p}^2 \frac{\sigma^2}{\hat{\sigma}^2}\left(\hat{p}_L \frac{1}{a} + \hat{p}_M \frac{1}{b} + \hat{p}_H \frac{1}{c}\right) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma^2}{\sigma_k^2}\right)$$

$$= \hat{p}^2 \frac{\sigma^2}{\hat{\sigma}^2}\left(\hat{p}_L\left(\frac{1}{a} - \frac{1}{b}\right) + \frac{1}{b} + \hat{p}_H\left(\frac{1}{c} - \frac{1}{b}\right)\right) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma^2}{\sigma_k^2}\right)$$

$$= \hat{p}^2 \frac{\sigma^2}{\hat{\sigma}^2} \frac{1}{abc}(\hat{p}_L c(b - a) + ac + \hat{p}_H a(b - c) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma^2}{\sigma_k^2}\right)$$

$$= \hat{p}^2 \frac{\sigma^2}{\hat{\sigma}^2} \frac{1}{abc}(\hat{p}_L[c(b - a) + (a - b)a] + ac + (b - 1)a + \sum_{\text{remaining states}} p_k \left(\frac{\sigma^2}{\sigma_k^2}\right)$$

and we are done because $c(b–a) + (a–b)a = (b–a)(c–a) > 0$ by the choice of the states.

As for the first-order detector, we would expect the upper bound for the processing gain achievable for a multistate model second-order detector to depend on the low-state and high-state variances and the overall noise variances just as it did for the signal of known structure. Once again, we would like to obtain upper and lower bounds on $g_2(p_1, \alpha_1^2, ..., p_N, \alpha_s^2)$ subject to $\alpha_1^2 = \alpha_{L:}^2$, $\alpha_N^2 = \alpha_H^2$, and $\sum_{k=1}^{S} p_k \sigma_k^2 = \sigma^2$. As before, moving energy from a middle state to the outer states should lead to a less Gaussian-like distribution and thus should increase processing gain, while moving energy from the outer states to a middle state leads to a more Gaussian-like distribution and thus decreases the processing gain. For the case of a signal of known structure, the upper bound for processing gain is a linear function of positive slope as a function of the low-state probability given the constraints. For the case of a signal of unknown structure, the square of the upper bound for processing gain $g_2(p)$ turns out to be a parabolic function opening upwards as a function of the low-state probability (see figure A-1).

As before, we focus our attention on any three states of the N > 2 state model. Suppose the states have probabilities p_L, p_M, and p_H with variances $\sigma_L^1 < \sigma_M^2 < \sigma_H^2$. Let

$$\hat{p} = p_L + p_M + p_H, \hat{\sigma}^2 = p_L\sigma_L^2 + p_M\sigma_M^2 + p_H\sigma_H^2, \hat{p}_L = \frac{p_L}{\hat{p}}, \hat{p}_M = \frac{p_M}{\hat{p}}, \hat{p}_H = \frac{p_H}{\hat{p}},$$

$$a = \frac{\hat{p}}{\sigma^2}\sigma_L^2, b = \frac{\hat{p}}{\hat{\sigma}^2}\sigma_M^2, \text{and } c = \frac{\hat{p}}{\hat{\sigma}^2}\sigma_H^2.$$

The conditions on the normalized probabilities define a probability region in three space as indicated in figure A-1. Figure A-1 shows the case when b > 1, and the intersection of the constant variance plane and the probability region is a line intersecting sides B and C of the probability region triangle; when $b \leq 1$, their intersection would result in a line intersecting sides A and B of the probability region triangle. The lines of intersection can be parametrized by the low-state normalized probability by using in succession two substitutions: (1) $\hat{p}_M = 1 - \hat{p}_L - \hat{p}_H$ and (2) $\hat{p}_H(c - b) = (1 - b) - \hat{p}_L(a - b)$. Equation (2) can be rewritten as $\hat{p}_L(b–a) = \hat{p}_H(c–b) + (b–1) \geq b–1$ because the normalized probabilities are non-negative. This leads to $\hat{p}_L \geq 0$ for the geometry of case b $\leq$ 1 and $\hat{p}_L \geq \frac{b-1}{b-a}$ for the geometry of case b > 1.

For the first factor of the square of the upper bound for processing gain

$$\begin{aligned}
\sum_{k=1}^{S} p_k \frac{\sigma_n^4}{\sigma_k^4} &= \hat{p}^3 \frac{\sigma_N^4}{\hat{\sigma}^4}\left(\hat{p}_L \frac{1}{a^2} + \hat{p}_M \frac{1}{b^2} + \hat{p}_H \frac{1}{c^2}\right) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma_n^4}{\sigma_k^4}\right) \\
&= \hat{p}^3 \frac{\sigma_n^4}{\hat{\sigma}^4}\left(\hat{p}_L\left(\frac{1}{a^2} - \frac{1}{b^2}\right) + \frac{1}{b^2} + \hat{p}_H\left(\frac{1}{c^2} - \frac{1}{b^2}\right)\right) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma_n^4}{\sigma_k^4}\right) \\
&= \hat{p}^3 \frac{\sigma_n^4}{\hat{\sigma}^4} \frac{1}{a^2 b^2 c^2}(\hat{p}_L c^2(b^2 - a^2) + a^2 c^2 + \hat{p}_H a^2(b^2 - c^2)) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma_n^4}{\sigma_k^4}\right) \\
&= \hat{p}^3 \frac{\sigma_n^4}{\hat{\sigma}^4} \frac{1}{a^2 b^2 c^2}(\hat{p}_L[c(b - a)(c - a)(bc + ba + ac) + + a^2 c^2 + a^2(b + c)(b - 1) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma_n^4}{\sigma_k^4}\right) \\
&= \hat{p}^3 \frac{\sigma_n^4}{\hat{\sigma}^4} \frac{1}{a^2 b^2 c^2}[p(bc + ba + ac) + a^2 c^2 + a^2(b + c)(b - 1) + D]\,,
\end{aligned}$$

where $p = (b - a)(c - a)\hat{p}_L$ and $D = \left(\hat{p}^3 \frac{\sigma_n^4}{\hat{\sigma}^4} \frac{1}{a^2 b^2 c^2}\right)^{-1} \left(\sum_{\text{remaining states}} p_k \left(\frac{\sigma_n^4}{\sigma_k^4}\right)\right)$.

For the second factor of the square of the upper bound for processing gain

$$\begin{aligned}
2\sum_{k=1}^{S} p_k \frac{\sigma_k^4}{\sigma_n^4} - 1 &= 2\left[\frac{\sigma^4}{\hat{p}\sigma_n^4}\left(\hat{p}_L a^2 + \hat{p}_M b^2 + \hat{p}_H c^2\right) + \sum_{\text{remaining states}} p_k \left(\frac{\sigma_k^4}{\sigma_n^4}\right)\right] - 1 \\
&= 2\left[\frac{\sigma^4}{\hat{p}\sigma_n^4}\left\{\hat{p}_L(a^2 - b^2) + b^2 + \hat{p}_H(c^2 - b^2)\right\}\right. + \sum_{\text{remaining states}} p_k \left(\frac{\sigma_k^4}{\sigma_n^4}\right) \\
&= \frac{\hat{\sigma}^4}{\hat{p}\sigma_n^4}[2p + b + c - bc + E - F]\,,
\end{aligned}$$

where p is as above, $E = 2\left(\frac{\hat{\sigma}^4}{\hat{p}\sigma_n^4}\right)^{-1} \sum_{\text{remaining states}} p_k \left(\frac{\sigma_k^4}{\sigma_n^4}\right)$, and $F = \left(\frac{\hat{\sigma}^4}{\hat{p}\sigma_n^4}\right)^{-1}$.

In preparation for calculations to follow, we need $2 + E - F \geq 0$. Note that

$$2\left(\frac{\hat{\sigma}^4}{\hat{p}\sigma_n^4}\right) + 2\sum_{\text{remaining states}} p_k \left(\frac{\sigma_k^4}{\sigma_n^4}\right) - 1 \geq 2\sum_{k=1}^{N} p_k \left(\frac{\sigma_k^4}{\sigma_n^4}\right) - 1 \geq 1 \text{ because } \hat{p} \leq 1 \text{ and}$$

$$\sum p_k \sigma_k^4 \geq \left(\sum p_k \sigma_k^2\right)^2 = \sigma_n^4 \text{ by the Cauchy–Schwartz inequality .}$$

The coefficients of the normalized low-state probability are obviously positive for both factors of the upper bound for processing gain so that the upper bound for the processing gain squared is a parabola opening upwards. It is then clear that the maximum of the upper bound for processing gain as a function of state probabilities occurs when one of the probabilities is zero. It follows that an upper bound for an N state model is less than that for an N-1 state model, and thus by induction for a two-state model. However, more is true, namely, the prabola is opening up over the allowable values of p so that the upper bound for processing gain is everywhere an increasing function of the low-state probability for $N > 2$. We proceed to prove this.

Suppose that $b \leq 1$. To show that the critical point of the parabola occurs when $p \leq 0$, it suffices to show that the coefficient of p in the product of the two factors is positive because the coefficient of p^2

is positive. To conclude this, observe that the coefficient of p is a quadratic polynomial in c with the coefficient of c^2 given by $2a^2 + 2(1-b)(b+a) > 0$ and that of c given by $(b+a)[2b + E - F + 2(b-1)a]$. We proceed by showing that the critical point of the parabola in c is ≤ 1 and showing that the coefficient $p > 0$ when $c = 1$. That the critical point is as desired follows from $4a^2 + 4(1-b)(b+a) \leq (b+a)\{-2b - E + F + 2(1-b)a\}$ using $a < b < 1$ and $2 + E - F \geq 0$. For $c = 1$, the coefficient of p is positive follows from $(2 + E - F)(a + b + ab) + 2a^2b^2 + 2D > 0$, which can easily be established by directly evaluating the coefficient of p in the product of the factors before expressing the coefficient as a quadratic in c.

Suppose that b > 1. In this case, it is convenient to introduce $\tilde{p}$ defined by $p = \tilde{p} + (b-1)(c-a)$ so that the lower bound on p can be replaced by $\tilde{p} \geq 0$. To show that the critical point of the parabola occurs when $\tilde{p} \leq 0$, it suffices to show that the coefficient of $\tilde{p}$ in the product of the two factors is positive because, as before, the coefficient of $\tilde{p}^2$ (which is the same as that of p^2) is positive. To do this we observe that the coefficient of $\tilde{p}$ is a quadratic in c with the coefficient of c^2 given by $2a^2 + 2(b-1)(b+a) > 0$ and the coefficient of c given by $(b + a)[2(b + a - ab) + E - F]$. Since the coefficient of c^2 is positive, it suffices to show that the critical point for the parabola in c is $\leq b$ and showing that the coefficient of p is positive when $c = b$. That the critical point is as desired follows from $2b\{2a^2 + 2(b-1)(b+a)\} \geq -(b+a)[2(b + a - ab) + E - F]$, which is equivalent to $4ba^2 + (4-2a)(b-1) + (b+a)(2b + E - F) \geq 0$, which is clearly true because $2b + E - F \geq 2 + E - F \geq 0$. It is easy by direct calculation to evaluate the coefficient of $\tilde{p}$ when $c = b$ and show that it is positive by using the same inequality as in the previous sentence.

The argument presented in the last paragraph shows that given any N-state model with $N > 2$, it is always possible to construct a two-state model $\alpha_1^2 = \alpha_L^2$, $\alpha_N^2 = \alpha_H^2$, and $\sum_{k=1}^{2} \hat{p}_k \sigma_k^2 = \sigma^2$ such that $g_2(p_a \alpha_1^2, ..., p_N, \alpha_N^2) < g_2(\hat{p}_1, \alpha_1^2, \hat{p}_2, \alpha_2^2)$. A global bound for any N-state mixture model of two or more states is obtained by observing that the constraints completely define the two-state model obtained by the construction process. The probabilities of the two states can be calculated from $p\sigma_1^2 + (1-p)\sigma_2^2 = \sigma^2$. Solving this equation for p and $1 - p$ yields $p = \frac{\sigma^2 - \sigma_H^2}{\sigma_L^2 - \sigma_H^2}$ and $1 - p = \frac{\sigma^2 - \sigma_L^2}{\sigma_H^2 - \sigma_L^2}$. Then $p\left(\frac{\sigma^2}{\sigma_L^2}\right)^2 + (1-p)\left(\frac{\sigma^2}{\sigma_H^2}\right)^2 = 1 - [p + (1-p)] + \left[p\left(\frac{\sigma^2}{\sigma_L^2}\right)^2 + (1-p)\left(\frac{\sigma^2}{\sigma_H^2}\right)^2\right] = 1 + \left[\frac{\sigma^2 - \sigma_H^2}{\sigma_L^2 - \sigma_L^2}\right]\left(\left(\frac{\sigma^2}{\sigma_L^2}\right)^2 - 1\right) + \left[\frac{\sigma^2 - \sigma_L^2}{\sigma_H^2 - \sigma_L^2}\right]\left(\left(\frac{\sigma^2}{\sigma_H^2}\right)^2 - 1\right), = 1 + \left(\frac{1}{\sigma_L^4 \sigma_H^4}\right)(\sigma^2 - \sigma_L^2)(\sigma_H^2 - \sigma^2)\sigma^2(\sigma_H^2 + \sigma_L^2)$ provides a formula for the first factor, while by easy algebra the second factor can be written as $2[p(\frac{\sigma_L^2}{\sigma^2})^2 + 1(1 = p)(\frac{\sigma_H^2}{\sigma^2})^2] - 1 = 1 + \frac{2}{\sigma^4}(\sigma^2 - \sigma_L^2)(\sigma_H^2 - \sigma^2)$.

These expressions exhibit the upper bound for processing gain for a nontrivial two-state model as positive. Letting $a = \frac{\sigma_L^2}{\sigma^2}$ and $b = \frac{\sigma_H^2}{\sigma^2}$, it follows that

$$g(a, b) = \frac{1}{ab}\sqrt{[a^2b^2 + (1-a)(b-1)(a+b)][1 + 2(1-a)(b-1)]}\ .$$

APPENDIX B

PROCESSING GAIN BOUNDS FOR THE MIDDLETON CLASS A NOISE MODEL

In this section, we derive processing gain bounds for the Middleton Class A noise model. To make this appendix self contained, we recall a few basic facts about the Middleton class A noise model. Middleton's class A noise model is a Gaussian mixture model with an infinite number of terms. Middleton has also introduced noise models of class B and C. These models describe noise for which the individual noise sources are received at certain times and not at others. A noise impulse is categorized as class A if it produces a decaying response from a receiver, the case for undersea surveillance applications, and class B if it produces ringing of the receiver, and class C if sometimes it causes ringing and other times not, often the case for communications receivers.

Middleton (1977, 1983) and Middleton and Spaulding (1983) derive univariate probability density functions for class A and class B noise, by supposing the noise consists of Gaussian background noise and interference from discrete sources which are Poisson-distributed in space. (References in this appendix are listed at the end of the body.) The density function that they obtained for class A noise is given by $p(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=1}^{\infty} \frac{1}{\sigma_m} e^{-A} \frac{A^m}{m!} e^{-\frac{x^2}{2\sigma_m^2}}$. .

We assume that $\sigma_0^2 = \sigma_{bn}^2$ is the variance of the background noise. This probability density function is formulated for real quantities and we wish to use the noise model to describe complex samples. To do this, we introduce the spherically invariant probability density function $p(|z|^2) = \frac{1}{2\pi} \sum_{m=1}^{\infty} \frac{1}{\sigma_m^2}$ $e^{-\frac{|x|^2}{2\sigma_m^2}}$ and interpret the parameters of the model as applying to the norms of the samples in place of the real quantities of the original formulation. In other words, we use the Middleton class A noise model to describe received interference power.

We provide a physical interpretation of the Middleton class A noise model by assuming that each sample belongs to a single state. The states for the Middleton class A noise model are then determined by the number m of discrete interference sources active at any given time. In the equation for the probability of being in the m-th state is $p_m(A) = e^{-A} \frac{A^m}{m!}$. This is the probability that interference generated by m discrete sources is received. The $m = 0$ state describes the Gaussian background noise level when the background noise does not contain interference from any of the discrete sources. The expected number of discrete sources for which interference is received is A. Middleton calls A the overlap index, which he defines as $A = rD$, where r is the expected number of sources whose interference is received per unit, and D is the expected duration of such a reception. For this model, the background noise has variance σ_0^2 and the expected received nosie plus interference power for the m-th state is given by $\sigma_m^2 = \sigma_0^2 + \frac{m\Omega}{A}$ with $\frac{\Omega}{A}$ the expected power received from each interferer. It follows that the overall noise variance for the model is $\sigma^2 = \sigma_0^2 + \Omega$.

The processing gain upper bounds obtained for Gaussian mixture models for signals of known and unknown structure can be extended to Middleton's class A model by an easy limit argument. Let

$\frac{\sigma^2}{\sigma_m^2} = \frac{\sigma_0^2 + \Omega}{\sigma_0^2 + \frac{m\Omega}{A}} = \frac{\Gamma + 1}{\Gamma + \frac{m}{A}} = R_m\ (A,\ \Gamma\ )$. Then the processing gain upper bound for the Middleton class A noise model and signal of known structure is $g_1(A,\Gamma) = \sum_{m=0}^{\infty} p_m\ (A) R_m\ (A,\ \Gamma\ )$, while the processing gain upper bound for the Middleton class A noise model and signal of unknown structure is

$$g_2(A,\Gamma) = \sqrt{\left[\sum_{m=0}^{\infty} p_m(A) R_m^2(A,\Gamma)\right]\left[2\sum_{m=0}^{\infty}\frac{p_m(A)}{R_m^2(A, GAMMA)} - 1\right]}.$$

We now show that as A and Γ tend to infinity, both the first-order and second-order gains tend to 1. In other words, there is virtually no gain over classical processing when A or Γ is large. We first consider the case of $g_1(A,\Gamma)$ and show that $\lim_{A \to \infty} g_1(A,\Gamma) = 1$ and $\lim_{\Gamma \to \infty} g_1(A, GAMMA) = 1$.

To prove the first limit, we use Chebychev's Inequality (Guttman, Wilks, and Hunter, 1971, p. 94) and some basic properties of the Poisson distribution (Guttman et al., 1971, p. 116). Recall that Chebychev's Inequality states that for any probability density function p with mean μ and stanadrd deviation σ, and any $k > 0$, $p(|X - \mu| > k\sigma) \le \frac{1}{k^2}$. The Poisson distribution with parameter A is given by $p(m) = e^{-A}\frac{A^m}{m!}$, m = 0, 1, 2, ... The mean and the variance of the Poisson distribution both have value A. Let $\lambda(A)$ be a positive function of A such that $\lim_{A \to \infty} \frac{\lambda(A)}{\sqrt{A}} = 0$ and $\lim_{A \to \infty} \frac{1}{\lambda^2(A)}$. Examples of such $\lambda(A)$'s are $A^{\frac{1}{3}}$ and $\frac{\sqrt{A}}{\ln A}$. Then using Chebychev's Inequality, $\sum_{|m - A \ngtr \lambda(A)\sqrt{A}} e^{-A}\frac{A^m}{m!} \le \frac{1}{\lambda^2(A)}$. Recall that $g_1(A,\Gamma) = \sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m + A\Gamma}\right)$. Therefore, when $\Gamma > 0$, and using the fact that $\frac{(1+\Gamma)}{m + A\Gamma}$ is a decreasing function of m,

$$\left(1 - \frac{1}{\lambda^2(A)}\right)\frac{(1+\Gamma)A}{A + \lambda(A)\sqrt{A} + A\Gamma} \le g_1(A,\Gamma) \le \left(\frac{1+\Gamma)A}{A - \lambda(A)\sqrt{A} + A\Gamma}\right) + \frac{1}{\lambda^2(A)}\frac{(1+\Gamma)A}{A\Gamma}\ .$$

Now observe from the defining properties of $\lambda(A)$ that the first and last terms of the above inequality tend to 1 as $A \to \infty$ to conclude that $\lim_{A \to \infty} g_1(A,\Gamma) = 1$. Note that in evaluating $\lim_{A \to ing} g_1(A,\Gamma)$, the limit cannot be taken inside the sum since $\lim_{A \to \infty} e^{-A}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m + A\Gamma} = 0$ for each m and the sum would then be 0.

We next prove that $\lim_{\Gamma \to \infty} g_1(A,\Gamma) = 1$. Observe that $\lim_{A \to \infty} \frac{(1+\Gamma)A}{m + A\Gamma} = 1$ and $\frac{(1+\Gamma)A}{m + A\Gamma} = 1 - \frac{m - A}{m + A\Gamma}$. From the first equality, observe that $\lim_{\Gamma \to \infty} g_1(A,\Gamma) = 1$ is true if we formally interchange the sum and the limit. However, as we noted previously, one must be careful that the operation of interchanging the sum and the limit is justified. From the second equality, observe that $\frac{(1+\Gamma)A}{m + A\Gamma}$ is an increasing function of Γ when $m >$ A. Therefore,

$$\lim_{\Gamma \to \infty} g_1(A,\Gamma) = \lim_{\Gamma \to \infty} \sum_{m \le A} e^{-A}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m + A\Gamma} + \lim_{\Gamma \to \infty} \sum_{m > A} e^{-A}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m + A\Gamma} = \sum_{m \le A} e^{-A}\frac{A^m}{m!} + \sum_{m > A} e^{-A}\frac{A^m}{m!} = 1.$$

Note that the first limit on the right converges since there are only a finite number of terms in the sum

and the second limit converges by the Monotone Convergence Theorem (Halmos, 1957, 103, p. 112, Theorem B).

Consider the gain for the second-order detector. Recall that

$$g_2(A,\Gamma) = \sqrt{\left[\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m+A\Gamma}\right)^2\right]\left[2\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\left(\frac{m+A\Gamma}{1+\Gamma)A}\right)^2 - 1\right]}\,.$$

From the properties of the Poisson distribution, $\sum_{m=1}^{\infty} e^{-A}\frac{A^m}{m!}m = A$ and $\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}(m-A)^2 = A$. Therefore, $\sum_{m=1}^{\infty} e^{-A}\frac{A^m}{m!}m^2 = A^2 + A$. Then

$$\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\left(\frac{m+A\Gamma}{(1+\Gamma)A}\right)^2 = \frac{1}{(1+\Gamma)^2}\left[\frac{1}{A^2}\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}m^2 + (2\Gamma+\Gamma^2)\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\right]$$
$$= \frac{1}{(1+\Gamma)^2}\left[\frac{1}{A^2}(A^2+A) + 2\Gamma + \Gamma^2\right] - 1 + \frac{1}{(1+\Gamma)^2 A}\,.$$

Thus,

$$g_2(A,\Gamma) = \sqrt{\left[\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m+A\Gamma}\right)^2\right]\left[1 + \frac{2}{(1+\Gamma)^2 A}\right]}\,.$$

Clearly, the second term inside the radical tends to 1 as A or $\Gamma \to \infty$. The first term inside the radical can be shown to tend to 1 as A or $\Gamma \to \infty$ in exactly the same way as for $g_1(A,\Gamma)$ and we omit the details.

We now determine the numbers of terms needed to obtain good approximations of $g_1(A,\Gamma)$ and $g_2(A,\Gamma)$. The result that we obtain states that for any fixed A and any Γ, using $2A + 1$ terms in the sum will give very good approximations.

We first consider $g_1(A,\Gamma)$. Recall that Stirling's formula states that

$$\sqrt{2\pi n}\left(\frac{n}{3}\right)^n < n! < \left(\frac{12n}{12n-1}\right)\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \text{ for } n = 1, 2, \ldots$$

Let $A > 0$ be given. Let $N = A$ if A is an integer and let $N = [A] + 1$ otherwise. We have for $m \geq N$ that $\frac{A^m}{m!} = \frac{A^N}{N!}A^{m-N}\frac{N1}{m!} \leq \frac{A^N}{N!}N^{m-N}\left(\frac{12N}{12N-1}\right)\frac{\sqrt{2\pi N}\left(\frac{N}{e}\right)^N}{\sqrt{2\pi m}\left(\frac{m}{e}\right)^m} = \frac{A^N}{N!}\left(\frac{12N}{12N-1}\right)\frac{\sqrt{N}}{\sqrt{m}}\left(\frac{eN}{m}\right)\frac{1}{e^n}$. When $m = 2N$, $\left(\frac{e^N}{m}\right)^m\frac{1}{e^N} = \left(\frac{\sqrt{e}}{2}\right)^{2N}$ and when $m \geq 2N$, $\frac{\left(\frac{eN}{m+1}\right)^{m+1}}{\left(\frac{eN}{m}\right)^m} = \frac{eN}{m}\left(\frac{m}{m+1}\right)^{m+1} \leq \frac{eN}{m}\frac{1}{e} = \frac{N}{m} \leq \frac{1}{2}$. Therefore, for $m \geq 2N$, $\frac{A^m}{m!} \leq \frac{A^N}{N!}\left(\frac{12N}{12N-1}\right)\frac{\sqrt{N}}{\sqrt{m}}\left(\frac{\sqrt{e}}{2}\right)^{m-2N}$.

Hence

$$\sum_{m=2N+1}^{\infty} \frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma} \le \frac{A^N}{N!}\frac{(1+\Gamma)A}{N+A\Gamma}\left(\frac{12N}{12N-1}\right)\sqrt{N}\left(\frac{\sqrt{e}}{2}\right)^{2N} \sum_{m=2N+1}^{\infty} \frac{1}{\sqrt{m}}\left(\frac{1}{2}\right)^{m-2N}$$

$$\le \frac{A^N}{N!}\frac{(1+\Gamma)A}{N+A\Gamma}\left(\frac{12N}{12N-1}\right)\left(\frac{\sqrt{e}}{2}\right)^{2N}\frac{1}{\sqrt{2}}$$

and

$$\left|\sum_{m=0}^{\infty} \frac{A^m}{m!}\frac{!+\Gamma)A}{m+A\Gamma} - \sum_{m=0}^{2N} \frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}\right| = \left|\sum_{m=2N+1}^{\infty} \frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}\right|$$

$$\le \left|\frac{A^N}{N!}\frac{(1+\Gamma)A}{m+A\Gamma}\left(\frac{12N}{12N-1}\right)\left(\frac{\sqrt{e}}{2}\right)^{2N}\frac{1}{\sqrt{2}}\right| \le \left|\left(\frac{12N}{12N-1}\right)\left(\frac{\sqrt{e}}{2}\right)^{2N}\frac{1}{\sqrt{2}}\sum_{m=0}^{2N}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}\right|.$$

Therefore,

$$\left|1 - \frac{\sum_{m=0}^{\infty}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}}{\sum_{m=0}^{2N}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}}\right| \le \left|\frac{1}{\sqrt{2}}\left(\frac{12N}{12N-1}\right)\left(\frac{\sqrt{e}}{2}\right)^{2N}\right| \text{ and by the Mean Value Theorem}$$

$$\left|10\log\frac{g_1(A,\Gamma)}{e^{-A}\sum_{m=0}^{2N}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}}\right| = \left|10\log\left[\frac{\sum_{m=0}^{\infty}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}}{\sum_{m=0}^{2N}\frac{A^m}{m!}\frac{(1+\Gamma)A}{m+A\Gamma}}\right]\right| \le \frac{10}{\ln(10}\left|\frac{1}{\sqrt{2}\left(\frac{12N}{12N-1}\right)\left(\frac{\sqrt{e}}{2}\right)^{2N}}\right|.$$

For example, a 31-term approximation is within 0.01 dB for $0 \le A \le 15$, and for 60 terms, the approximation is within 0.00003 dB for $0 \le A \le 30$. Note that the above upper bound is very conservative since we have shown that the tail of the series is much less than the N-th term alone.

Consider $g_2(A,\Gamma)$. Recall that $g_2(A,\Gamma) = \sqrt{\left[\sum_{m=0}^{\infty} e^{-A}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m+A\Gamma}\right)^2\right]\left[1+\frac{2}{(1+\Gamma)^2A}\right\}}$ . .

Using the exact same computation as before,

$$\left|10\log\frac{g_2(A,\Gamma)}{\sqrt{\sum_{m=0}^{2N} e^{-A}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m+A\Gamma}\right)^2\left[1+\frac{2}{(1+\Gamma)^2A}\right]}}\right|$$

$$\left|5\log\left[\frac{\sum_{m=0}^{\infty}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m+A\Gamma}\right)^2}{\sum_{m=0}^{2N}\frac{A^m}{m!}\left(\frac{(1+\Gamma)A}{m+A\Gamma}\right)^2}\right]\right| \le \frac{5}{\ln(10)}\left|\frac{1}{\sqrt{2}}\left(\frac{12N}{12N-1}\right)\left(\frac{\sqrt{e}}{2}\right)^{2N}\right|$$

Figures B-1 and B-2 present processing gain of first-order and second-order detectors for the Middleton class A noise model. Each figure contains a three-dimensional plot of the gain in decibels and a contour plot of the gain in decibels. The plots were generated based on the above considerations by using the first 31 terms of the infinite sums in $g_1(A, \Gamma)$ and $g_2(A, \Gamma)$ for $1 \leq A \leq 16$ and $0.01 \leq \Gamma \leq 0.3$. The plots indicate significant gain for small A and low background noise as expected.

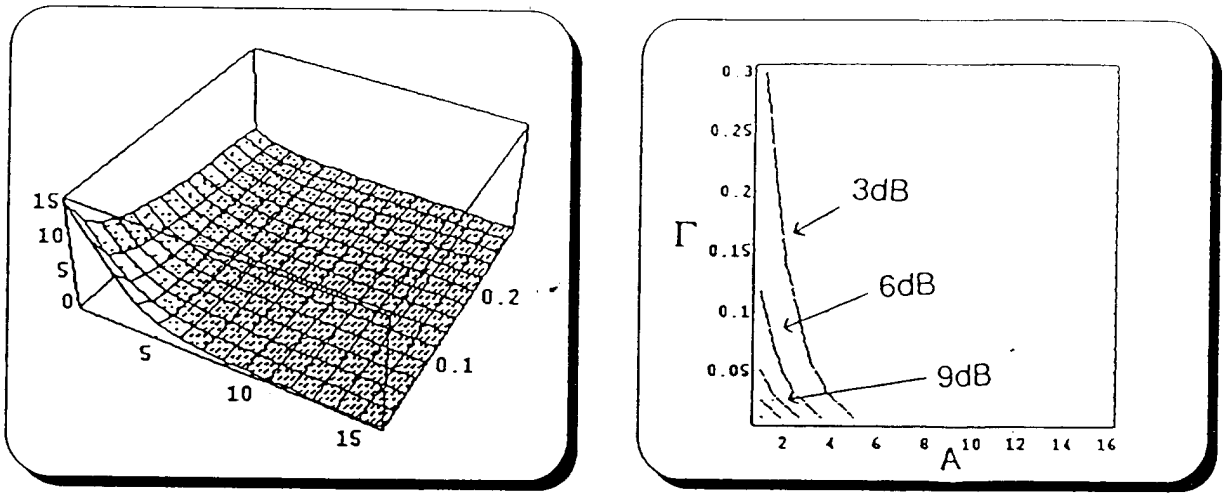


Figure B-1. Processing gain of a first-order detector for Middleton class A noise model.

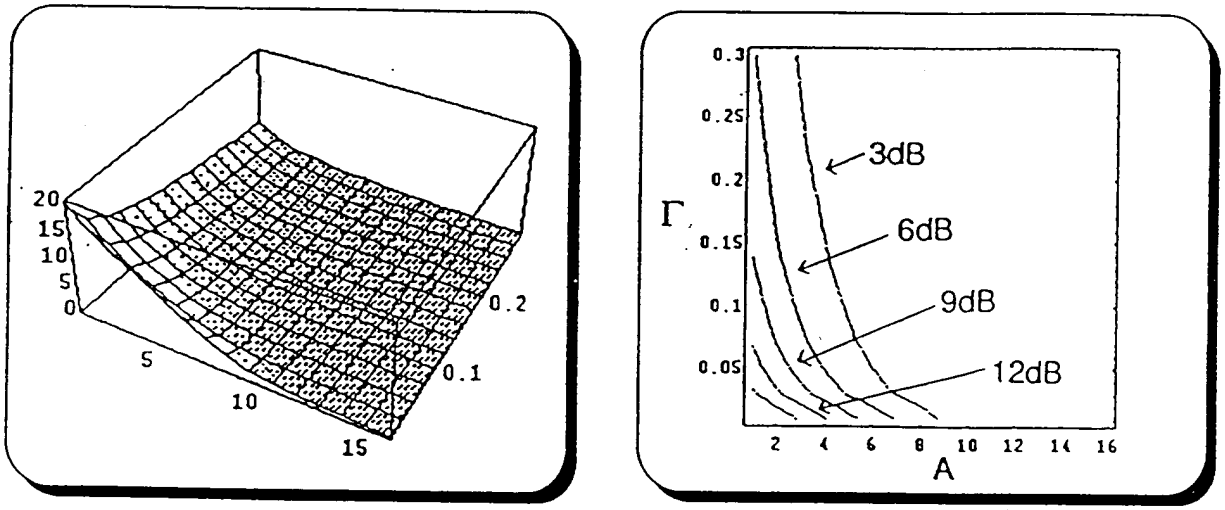


Figure B-2. Processing gain of a second-order detector for Middleton class A noise model.

A natural two-state model to approximate the Middleton class A noise model has low state variance equal to σ_0^2 and low-state probability e^{-A}. The high-state variance is then determined by enforcing the condition that $\sigma^2 = \sigma_0^2 + \Omega = \sigma_0^2 e^{-A} + \sigma_H^2(1 - e^{-A})$, which implies that $\sigma_H^2 = \dfrac{\sigma_0^2(1 - e^{-A}) + \Omega}{(1 - e^{-A})}$.

Note that for small A, $1 - e^{-A} \cong A$ and the above σ_H^2 is close to the high state variance obtained for the two-state approximation of the Middleton class A noise model described in appendix A to the companion report "Gaussian Mixture Models for Acoustic Interference."

For this two-state model and signal of known structure, an upper bound for the processing gain is $g_1(A,\Gamma) = e^{-A}\left(\frac{1+\Gamma}{\Gamma}\right) + (1-e^{-A})^2\left(\frac{1+\Gamma}{1+\Gamma(1-e^{-A})}\right)$. .

For this two-state model and signal of unknown structure, an upper bound for the processing gain is

$$g_2(A,\Gamma) = \sqrt{f_1(A,\Gamma)(2f_2(A,\Gamma)-1)} \text{ with}$$

$$f_1(A,\Gamma) = 3^{-A}\left(\frac{1+\Gamma}{\Gamma}\right)^2 + (1-e^{-A})^3\left(\frac{1+\Gamma}{1+\Gamma(1-e^{-A})}\right)^2$$

and

$$f_2(A,\Gamma) = e^{-A}\left(\frac{\Gamma}{1+\Gamma}\right)^2 + \frac{1}{(1-e^{-A})}\left(\frac{1+\Gamma(1-e^{-A})}{1+\Gamma}\right)^2 .$$

A natural three-state model to approximate the Middleton class A noise model has low-state, middle-state, and high-state probabilities and variances

$$e^{-A}, \sigma_0^2, Ae^{-A}, \sigma_0^2 + \frac{\Omega}{A}, \text{ and } 1 - e^{-A} - Ae^{-A}, \sigma_0^2 + \frac{1-e^{-A}}{1-e^{-A}-Ae^{-A}}\Omega .$$

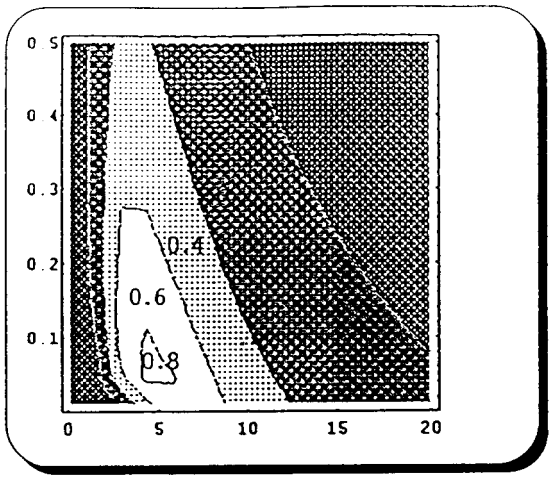
For this three-state model and signal of known structure, an upper bound for the processing gain is $h_2(A,\Gamma) = e^{-A}\left(\frac{1+\Gamma}{\Gamma}\right) + A^3e^{-A}\left(\frac{1+\Gamma}{1+A\Gamma}\right) + (1-Ae^{-A}-e^{-A})^2\left(\frac{1+\Gamma}{(1-e^{-A})+\Gamma(1-Ae^{-A}-e^{-A})}\right)$.

For this three-state model and signal of unknown structure, an upper bound for the processing gain is $h_2(A,\Gamma) = \sqrt{f_1(A,\Gamma)(2f_2(A,\Gamma)-1)}$ with

$$f_1(A,\Gamma) = e^{-A}\left(\frac{1+\Gamma}{\Gamma}\right)^2 + A^3e^{-A}\left(\frac{1+\Gamma}{1+A\Gamma}\right)^2 + \left(1-Ae^{-A}-e^{-A}\right)^3\left(\frac{(1+\Gamma}{(1-e^{-A})+\Gamma(1-Ae^{-A}-e^{-A})}\right)^2 \text{ and}$$

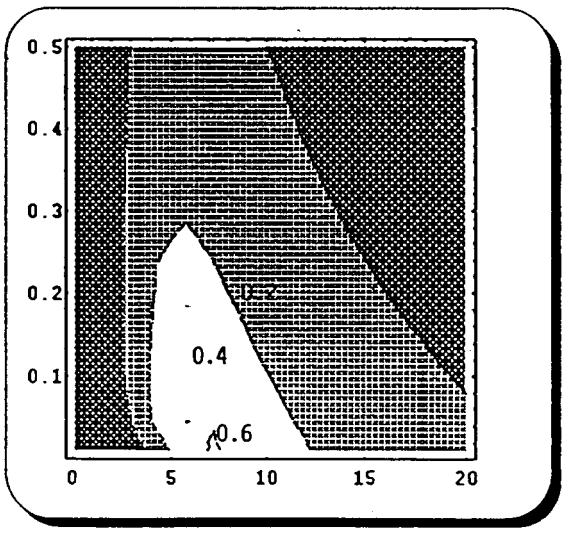
$$f_2(A,\Gamma) = e^{-A}\left(\frac{\Gamma}{1+\Gamma}\right)^2 + A^{-1}e^{-A}\left(\frac{1+A\Gamma}{1+\Gamma}\right)^2 + (1-Ae^{-A}-e^{-A})^{-1}\left\{\frac{(1-e^{-A})+\Gamma(1-Ae^{-A}-e^{-A})}{1+\Gamma}\right)^2 . .$$

Figures B-3 and B-4 compare the processing gain bounds obtained for the first-order and second-order detectors, respectively, for the two-state and three-state approximations to the Middleton Class A noise model and the 31-term approximation of the Middleton Class A noise model. The comparisons are contour plots for abscissa A and ordinate Γ satisfying $0.1 \le A \le 20$ and $0.01 \le \Gamma \le 0.5$. The contours are for differences between the processing gain bounds in decibels. For the range of mixture model parameters surveyed, the upper bounds were very close for a first-order detector, less than 0.8 dB for the two-state model and 0.6 dB for the three-state model, and a comparison of the plots indicates further that they never differed by more than 0.2 dB gain. The same comparison for the second-order detectors, indicate slightly larger maximum differences for the two-state and three-state models processing gain bounds than exhibited for the first-order detectors, a maximum of 1.4 dB for a two-state model, a maximum of 0.8 dB for a three-state model, and a gain of up to 0.6 dB for using a three-state model instead of a two-state model. To the extent that the upper bounds for processing gain estimate the achievable processing gains, these results support placing emphasis on fitting actual data with two-state and three-state mixture models.



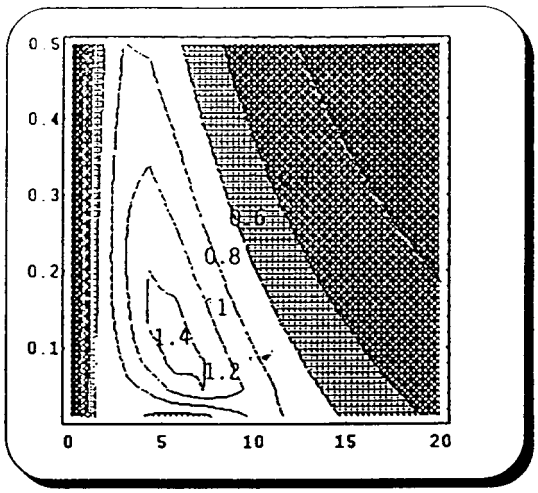
(a) vs a two-state

Note: The probability and variance of the high state in each finite state model are adjusted to keep the overall variance constant. Range of plots:
$0.01 \leq \Gamma \leq 0.5$, $0.1 \leq A \leq 20$.



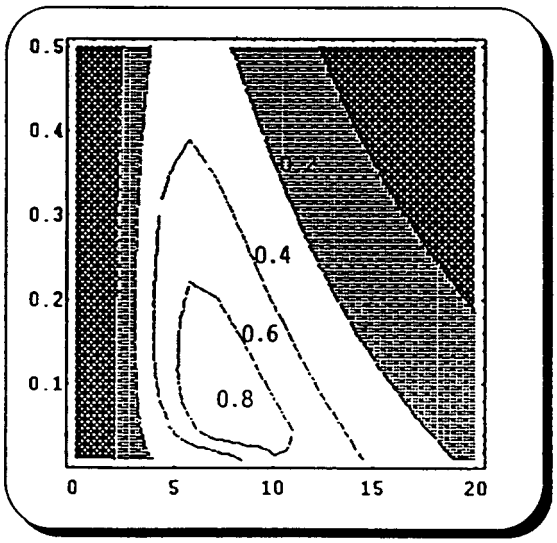
(b) vs a three-state

Figure B-3. Comparison of processing gains of the Middleton class A first-order detector and finite state approximations.



(a) vs a two-state

Note: The probability and variance of the high state in each finite state model are adjusted to keep the overall variance constant. Range of plots:
$0.01 \leq \Gamma \leq 0.5$, $0.1 \leq A \leq 20$.



(b) vs a three-state

Figure B-4. Comparison of processing gains of the Middleton class A second-order detector and finite state approximations.

APPENDIX C

PERFORMANCE LOSS DUE TO TARGET MOTION

INTRODUCTION

Detecting quiet targets in a surveillance system typically requires processing that integrates beamformer outputs over many samples in order to accumulate sufficient signal energy to detect. This integration often requires tracking for large aperture arrays with fine spatial resolution, because a moving target could travel through multiple spatial cells during the integration time required for detection. If the target's velocity is unknown, as it generally is before detection, the integration time has to be limited to the length of time a target is likely to remain in a single cell unless tracking algorithms or eye integration is used.

Figure C-1 illustrates the need for combining spatial energy for different spatial cells to achieve detection in a scenario similar to some modeled by the HGI simulator. Figure C-1 displays the cumulative probability distribution for the number of spatial cell boundaries crossed by a randomly chosen target track over one-half hour. The target's initial position has a uniform random distribution on a disk of radius 1000 km, with the spatial cells being 1-degree by 2-km annular sectors, and the velocity distribution is uniform on a disk of radius 28 km/hr. Depth and frequency are fixed and known. Note that in this example very few targets remain in one cell, and about half of them cross at least three cell boundaries. This example suggests the need for combining energy for multiple spatial cells for large surveillance arrays employing matched field beamforming or similar beamforming.

The purpose of this appendix is to present analysis and simulation results showing the extent to which unknown target motion degrades achievable detection performance in large surveillance arrays.

One approach to combining energy in different spatial cells is track-before-detect. References to examples appear in the next few paragraphs. In general, a tracking detector hypothesizes various possible target motions and evaluates a detection statistic conditioned on each target motion under consideration. When the statistic falls in a critical region, both the existence of the target and its velocity are reported. Since the number of possible target motions is huge, tracking detector implementations require a great deal of processing power. Algorithms based on similar signal and noise models are mainly distinguished by the shortcuts and approximations required to reduce the processing loads on the computers used to implement the algorithms.

Tracking detector implementations can be grouped in two broad categories: track initiation techniques and probability map techniques. The former compute detection statistics for potential tracks, which are initiated by energy peaks crossing a low threshold and are terminated when the track fails some kind of M-out-of-N rule designed to monitor persistence of signal power. Examples include Blostein and Huang (1991), Bruton and Bartley (1986), Nagarajan, Sharma, and Chidambara (1984), Shensa and Broman (1985), and Broman (1992). (References for this appendix are listed at the end of the body of the report.) These track initiation approaches are appropriate for low dimensional problems, but they may not scale up to the dimensions of HGI beamformer outputs. They generally employ complex decision logic and require a good deal of tuning and adjustment of error-handling heuristics in order to achieve good performance. See for examples Gibbons et al. (1987) and Struzinski (1978) and the articles referred to therein.

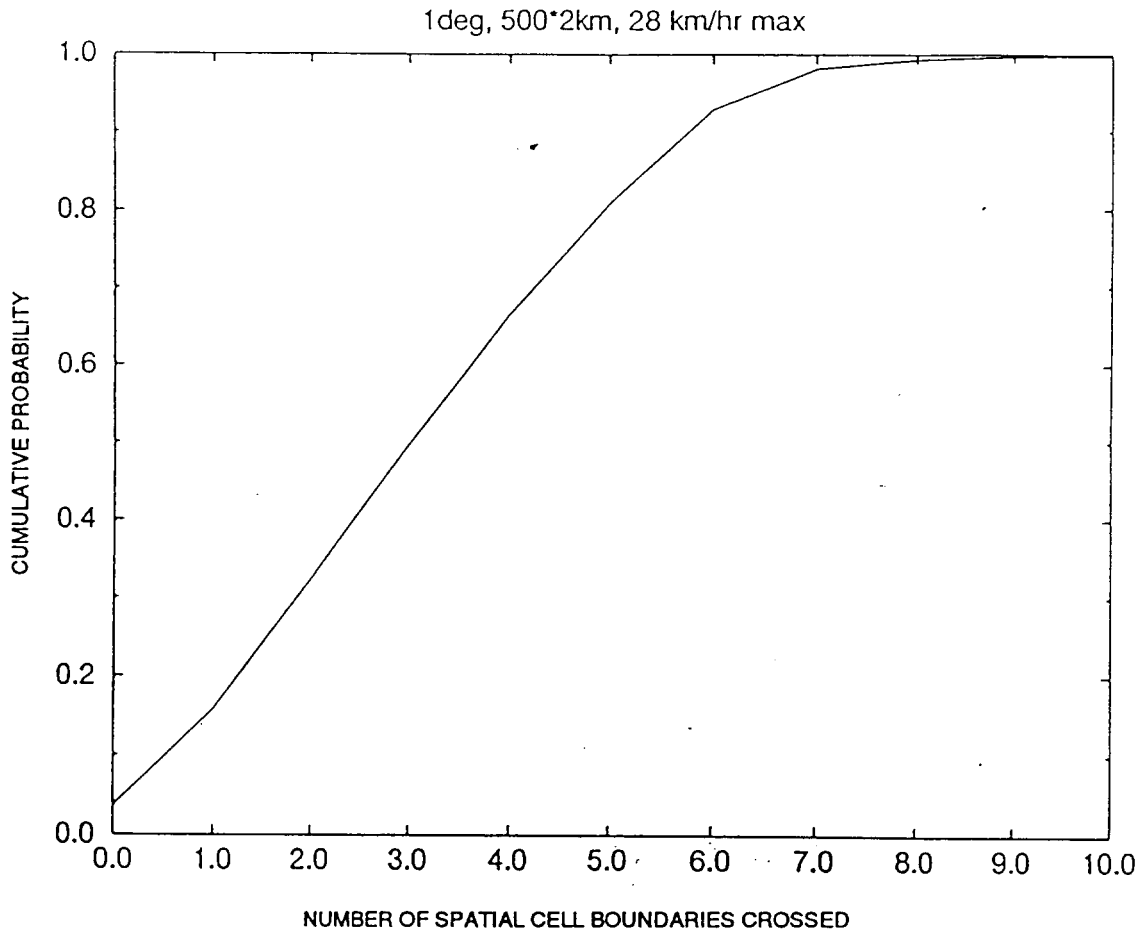


Figure C-1. Cumulative distribution of cell-boundary crossings.

Probability map techniques from detection statistics for a large number of designated possible target motions, without picking and choosing the motions based on observed signal energies. This approach uses simpler logic, but can require more processing power than track initiation techniques. On the other hand, the processing required is highly parallelizable because of its simplicity. The dynamic programming implementations in Kessler et al. (1988) and Barniv (1985), the bounded hop pixel statistic in Wei, Zeidler, and Ku (1992, 1993), and the random walk approach introduced in this appendix, are examples of probability map schemes. Other comparable approaches are found in Mohanty (1981), Chen (1989), Cowart, Snyder, and Ruedger (1983) and Porat and Friedlander (1990). Most pattern-matching and interactive approaches are also in this category, but are not considered because their performance depends on the details of the beamformer.

THE STATE SPACE

We assume that the inputs to the tracking detector are beamformed, matched field processors, Fourier transformed and normalized by some kind of noise spectrum equalizer or by an adaptive locally optimum processing algorithm. Accordingly, at each time step the detector receives an array of spatial cells that may consist of noise or signal-plus-noise measurements. The spatial cells are parametrized by frequency, discrete target range from the array, bearing, and depth. Generally, the particular coor-

dinate system used for this state space depends on the beamforming, but a state space having dimension of at least 4 is likely for an HGI array.

The question arises whether the dimension, or the precise form of the coordinate system of the state space is important for evaluating the effect of target motion on detection? Intuition suggests that the principal factor influencing the performance of most tracking detectors is the amount of noise introduced, instead of signal plus noise, due to the inclusion of measurements from state space cells not containing the target. Uncertainties about the true target path bring in more noise-only measurements in a higher dimensional space than in a lower dimensional one because each spatial cell has more neighbors. We suppose that the driving factor is the number of extraneous noise-only measurements introduced by path uncertainties, and only secondarily the dimension of the space in which tracking is performed. Accordingly, we focus on the number of cells involved, and for simplicity's sake perform simulations in one dimension.

SIGNAL MODEL

The model chosen for the narrowband signal and the noise for our simulations is chosen to illuminate essential characteristics of the track-before-detect problem. The results obtained are not expected to depend strongly on model details. In particular, the structure of the tracking detection algorithm presented is applicable to measurements for which a likelihood ratio can be computed. In particular, we suppose that the measured signal and noise is sampled and reported at intervals of T/N seconds, so that for each spatial cell, a window of T seconds of data can be input into an N-point real Fourier transform.

For simplicity's sake, we assume first that signal energy from each target appears in only one state space bin at each time step. This requires that the frequency of the signal be centered in its Fourier transform frequency bin so as not to leak out, and that moving targets jump discretely from one spatial cell to the next in between the T-second windows. Granted that real targets move continuously between spatial cells, it will become clear later that failures of this assumption are only a minor problem for the random walk approach introduced in this appendix. Of course, they could be more serious for other approaches that do not average results from neighboring spatial cells. The time interval T must be short enough to justify the assumption of a motionless target in the spatial processing and the assumptions of linear phase and constant amplitude of the signal during each T-second coherent processing window. Therefore the analysis applies to those cases when an adaptive locally optimum first-order detector is used to process the beamformer outputs and not to the cases when an adaptive locally optimum second-order detector is used to process the beamformer outputs.

On the basis of this no-leakage assumption, we can now describe in standard terms a detection problem for each spatial cell, each frequency bin, and each T-second interval. Assume the spectrum equalized noise samples $g(t)$ at $t = nT/N$ for $n = 0, .., N - 1$, have zero mean, unit variance, and independent Gaussian distributions. We postulate a random sinusoidal signal $s(t) = ar \cos(2\pi ft + 2\pi\theta)$, where $f = k/T$ for an integer k strictly between 0 and $N/2$, the phase offset θ is uniformly distributed on [0,1]. The amplitude factor r is Rayleigh distributed with parameter $\sigma = 1$, and the factor a is known. Then we measure $z(t) = g(t)$ or $g(t) + s(t)$ under hypothesis H_0 or H_1, respectively.

Modeling the phase offset and the amplitude as constant for exactly T seconds and then changing instantaneously to new random values for the next T seconds is an approximation that is analytically tractable, and it allows the phase and amplitude to vary randomly. This is the weakest assumption taht can be made on moving targets and has consequences that we wish to investigate.

In analyses like that in Pryor (1971), processing gain from incoherent averaging is computed based on an assumption of constant, known signal power. As long as the detector is only thresholding measured energy, this common assumption is inconsequential, but for a tracking detector, the randomness of the signal power is important.

Based on the assumptions introduced above, the likelihood functions for the measurement N-vector z is $P(z|H_0) = (2\pi)^{-\frac{1}{2}N} e^{-\frac{1}{2}\sum_{n=0}^{N-1} z(nT/N)^2}$. Recalling that $P(\theta) = 1$ and $P(r) = re^{-\frac{1}{2}r^2}$, we find that $P(z, r, \theta|H_1) = (2\pi)^{-\frac{1}{2}N} e^{-\frac{1}{2}\sum_{n=0}^{N-1} [z(nT/N) - ar\cos(2\pi ft + 2\pi\theta)]^2} re^{-\frac{1}{2}r^2}$. The marginal density function for z can be computed from these equations directly, with some difficulty. Simplicity is gained by transforming variables as follows: $x = r\cos(2\pi\theta), y = r\sin(2\pi\theta), dxdy = 2\pi drd\theta$. Then $P(z, x, y|H_1) = (2\pi)^{-\frac{1}{2}(N+2)}$ $e^{-\frac{1}{2}\sum_{n=0}^{N-1} [z_n - ax\cos(2\pi kn/N) + ay\sin(2\pi kn/N)]^2} e^{-\frac{1}{2}(x^2+y^2)} = P(z|H_0)(2\pi)^{-1} e^{-\frac{1}{2}G}$, where we expand the exponent G as follows. Defining N-vectors c_k and s_k, whose n-th components are $\cos(2\pi kn/N)$ and $\sin(2\pi in/N)$, respectively:

$$
\begin{aligned}
G &= \| z - axc_k + ays_k \|^2 - \| z \|^2 + x^2 + y^2 \\
&= x^2 - 2axc_k \cdot z + a^2x^2 \| c_k \|^2 + y^2 + 2ays_k \cdot z + a^2y^2 \| s_k \|^2 \\
&= (1 + a^2 \| c_k \|^2)\left[x - \frac{ac_k codtz}{1 + a^2 \| c_k \|^2}\right]^2 - \frac{a^2(c_k \cdot z)^2}{1 + a^2 \| c_k \|^2} \\
&\quad + (1 + a^2 \| s_k \|^2)\left[y + \frac{as_k \cdot z}{1 + a^2 \| s_k \|^2}\right]^2 - \frac{a^2(s_k \cdot z)^2}{1 + a^2 \| s_k \|^2} .
\end{aligned}
$$

In this form it is easy to integrate out the x and y dependence to obtain the likelihood ratio:

$$
\begin{aligned}
l(z) &= \int\int \frac{P(z, x, y|H_1)}{P(z|H_0(} dxdy = \frac{1}{2\pi}\int\int e^{-\frac{1}{2}G} dxdy \\
&= (1 + a^2 \| c_k \|^2)^{-1/2}(1 + a^2 \| s_k \|^2)^{-1/2} e^{\frac{1}{2}\frac{a^2(c_k \cdot z)^2}{1 + a^2\|c_k\|^2} + \frac{1}{2}\frac{a^2(s_k \cdot z)^2}{1 + a^2\|s_k\|^2}} .
\end{aligned}
$$

Recall that $\| c_k \|^2 = \| s_k \|^2 = \frac{N}{2}$ and that the k-th coefficient of the discrete Fourier transform of the sequence z is $\hat{z}(k) = z \cdot (c_k - is_k)$. For notational simplicity, let $\zeta(a, N) = \frac{1}{2}\frac{a^2}{1 + a^2\frac{N}{2}}$.

The above expression for *l(z)* can then be written $l(z) = \frac{2}{a^2}\zeta(a, N)e^{\zeta(a,N)|\hat{z}(k)|^2}$. Since a and N are fixed, we see that the likelihood ratio is a monotone function of the power in the k-th frequency bin.

Next, we determine the distribution of $|\hat{z}(k)|^2$ under H_0 and H_1, and compute the expected signal-to-noise ratio of the detector inputs. Under H_1, the signal plus noise case, the measurement vector $z = g + axc_k - asy_k$ consists of N zero mean Gaussian distributions, so any linear functional of z has a zero-mean Gaussian distribution. In particular, $\frac{z \cdot c_k}{\| c_k \|} = \frac{g \cdot c_k}{\| c_k \|} + ax \| c_k \|$, $\frac{z \cdot s_k}{\| s_k \|} = \frac{g \cdot s_k}{\| s_k \|} - ay \| s_k \|$, $\mathrm{Var}\left(\frac{z \cdot c_k}{\| c_k \|}\right) = 1 + a^2 \| c_k \|^2 = \frac{a^2}{2\zeta(a, N)}$, and $\mathrm{Var}\left(\frac{z \cdot s_k}{\| s_k \|}\right) = 1 + a^2 \| s_k \|^2 = \frac{a^2}{2\zeta(a, N)}$.

Remembering that a Chi-squared random variable with 2 degrees of freedom is just an exponential random variable with mean 2, we conclude that $|\hat{z}(k)|^2$ is exponential with mean $N\left(1 + \frac{Na^2}{2}\right)$ under H_1.

We obtain the distribution for $|\hat{z}(k)|^2$ under H_0 by setting $a = 0$. This means the presence of a signal in $|\hat{z}(k)|^2$ changes the mean of its exponential distribution from N to $N\left(1 + \frac{Na^2}{2}\right)$.

The signal-to-noise ratio (per frequency bin, not per hertz), including the processing gain due to band pass filtering, is easy to compute. The signal power is $\frac{1}{2}a^2r^2$ with expected value a^2. Letting g denote the vector of N noise-only inputs, the expected signal power in the kth frequency bin is

$$\frac{1}{N}E\left[\left(\frac{g \cdot c_k}{\| c_k \|}\right)^2 + \left(\frac{g \cdot s_k}{\| s_k \|}\right)^2\right] = \frac{2}{N} .$$

Consequently, the ratio of expected signal power to expected noise power is $\frac{N}{a}a^2$.

KNOWN PATH DETECTION PERFORMANCE

A detector based on the Neyman-Pearson criteria employs the likelihood ratio as a test statistic: $l(z) = \frac{2}{a^2}\zeta(a, N)e^{\zeta(a,N)|\hat{z}(k)|^2}$, where a and N are assumed known *a priori*, or else are accurately estimated from other data.

Now, to obtain a detector for measurements like z, we set a threshold on $|\hat{z}(k)|^2$, or on $l(z)$, or on any other monotonic function of the likelihood ratio, in order to select the best track. But because of the tracking to be done below, we need to handle convex combinations of likelihood ratios, and the addition operations involved prevent our transforming the likelihood ratios nonlinearly, as we might otherwise do by adding logarithms and replacing products of likelihood ratios by sums of "energies." This is the reason why the tracking detector under discussion should not be thought of as just "summing energy over time" for an unknown target track.

Assuming a known target path, the likelihood ratio for M successive measurements taken along that path is $\lambda_M = \prod_{j=1}^{M} l_j(z)$. The random variable $\ln(\lambda_M) + M \ln(1 + \frac{Na^2}{2})$ is then equivalent to $\frac{Na^2}{2}$ times the sum of M exponential random variables with mean one, and it has a Γ distribution with parameters M and $2/(na^2)$. By reversing the transformation, the distribution for λ_M can be derived from the Gamma distribution.

We now describe the Neyman-Pearson detection performance of our algorithm for the known path case by means of these distributions. An important point arises because our surveillance system does not just process M measurements and then shut down, it continues with the next M. If a system designer doubles M, he not only doubles the amount of information available at each detection opportunity, but he also halves the number of detection opportunities per hour. Both effects are significant for system performance. All of our performance results are formulated in terms of the minimum detectable signal level associated with a given operating point (P_D, P_{FA}), for various path lengths M. In evaluating such a minimum detectable signal level, it is appropriate to try to fix the probability of false alarm or detection *per detection opportunity*, when the length of the detection opportunity itself is being varied? A fair assessment might be based on detections or false alarms *per hour* instead of *per opportunity.*

To help assure that equality of minimum detection signal thresholds implies equal detection performance on minimally detectable signals, we allow the operating point ($P_D(M)$, $P_{FA}(M)$) to depend

on M, and attempt to formulate an appropriate equivalence condition to relate operating points chosen for different values of M. Our criteria is that nominally equivalent operating points should generate approximately equal false-alarm rates per unit time. If, say, we are comparing detectors with path lenghts of M = 1, 2, 4, 8, or 16 points, then it seems appropriate to assume that the 16/M detection opportunities occurring in an interval of length 16 are combined with an "if-any" combiner, which declares a detection if any one or more of the 16/M subintervals declares a detection. In this case, identical detection statistics for each combined 16-point detector can be guaranteed by requiring that $[1–P_D(M)]^{16/m} = 1–P_D(16)$ and $[1–P_{FA}(M)]^{16/M} = 1–P_{FA}(16)$.

More generally, we require that $[1 - P_{D,FA}(M)]^{1/M}$ be invariant over M for equivalent operating points. Checking the consequences of this rule, we see that for small P_{FA}, the false-alarm rate per unit is approximately $\frac{P_{FA}(M)}{M} = \frac{1 - [1 - P_{FA}(1)]^M}{M} \approx \frac{P_{FA}(1)}{1}$ just as desired.

As a specific example, the following P_D settings are considered equivalent under this rule:

M	1	2	4	8	16
$P_D(M)$	0.159	0.293	0.5	0.75	0.875

Figure C-2 shows the strength of the minimum detectable signal level in decibels per bin, i.e., 10 log 10(c), for a known path detector achieving $P_D(1)$ = 1/2 (or equivalent), given various values of M and $P_{FA}(1)$ (or equivalent). From the figure we conclude that for fixed $P_{FA}(1)$, the minimum detectable signal level in decibels seems to be approximately a linear function of log(M), and that the level decreases about 0.5 to 1 dB per doubling of M. A similar graph, as shown in figure C-3, based on $P_D(4)$ = 1/2 and various $P_{FA}(4)$, looks almost identical, except for the level being shifted 4.2 dB lower.

Figure C-4, by contrast, plots the same minimum detectable signal levels against $P_{FA}(M)$ for a fixed $P_D(M)$. This shows the effect of increasing the path length M, without compensating for the change in the number of detection opportunities per hour. As expected, the curves are more widely spaced than in figures C-2 and C-3, illustrating the benefits expected from increasing M in a system that is shut down after the first and only detection opportunity.

Intuition suggests that the minimum detectable level of the tracking detector might be improved to an arbitrary level by increasing M. This turns out to be false, understood in terms of minimum detectable signal level for fixed ($P_D(1)$, $P_{FA}(1)$). An approximate limit as M increases without bound can be obtained by means of an asymptotic expansion (Prof. T. Sellke, Stanford, private communication): $\lim_{M \to \infty} c_M + 1 = \frac{W(P_{FA} - 1)}{W\left(\frac{P_D - 1}{e}\right)}$, where c_M is the minimum detectable level. The limit is in terms of the Lambert function $W(x)$, which is defined by $x = W(x)e^{W(x)}$ for $x > -1/e$. The limit found in this manner is plotted as the "M large" curve in Figure C-2, where "large" is interpreted to mean $M \geq 1/P_{FA}$. This curve has a horizontal asymptote of 5.2 dB for small P_{FA}. It is not known whether this limit is approached monotonically nor that the limit is bound for the actual performance, although, this is probably the case. If this were the case, even given perfect knowledge of the taret path there would be a limit to the performance gain obtainable by combining measurements over time.

Figure C-5 displays the thresholds on a Gamma distribution for $P_D(1)$ = 0.9 and $P_{FA}(1)$ = 0.5 and various M. The curves appear linear. If they were linear (and they are not), then the limiting MDS for

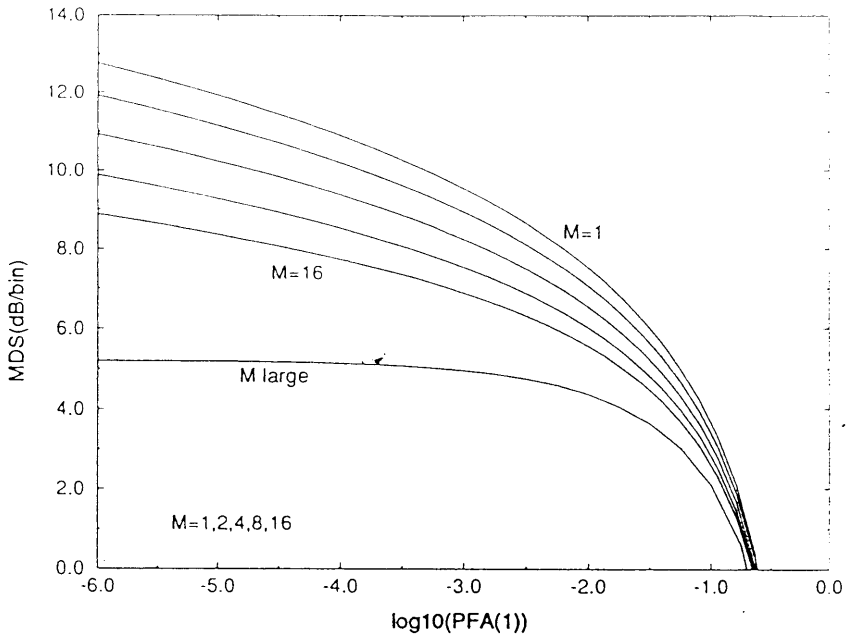


Figure C-2. Known path detector performance for PD(1) = 0.50.

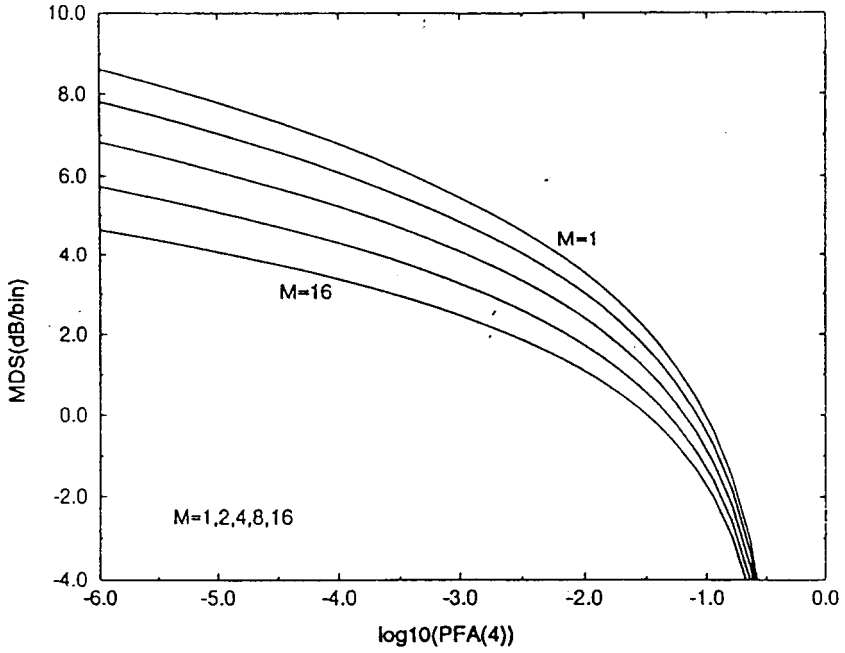


Figure C-3. Known path detector performance for PD(4) = 0.50.

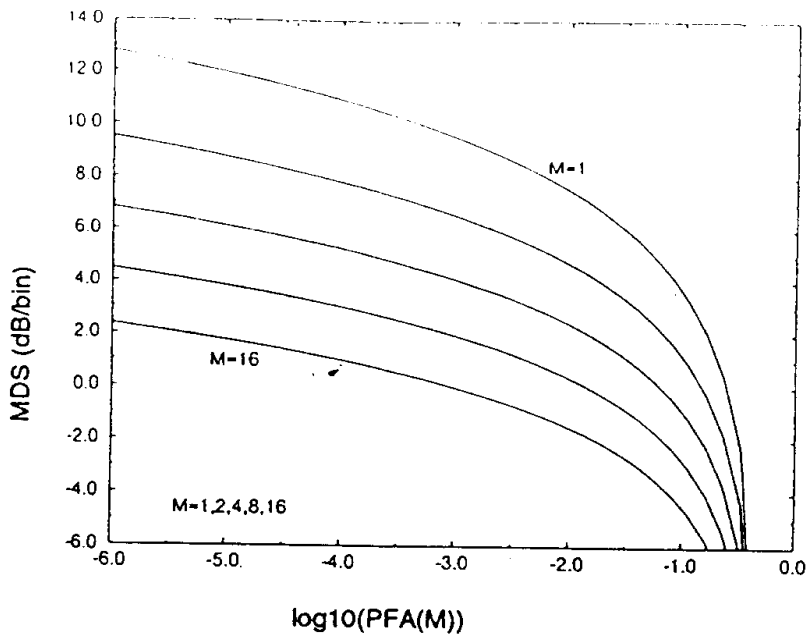


Figure C-4. Known path detector performance for PD(M) = 0.50.

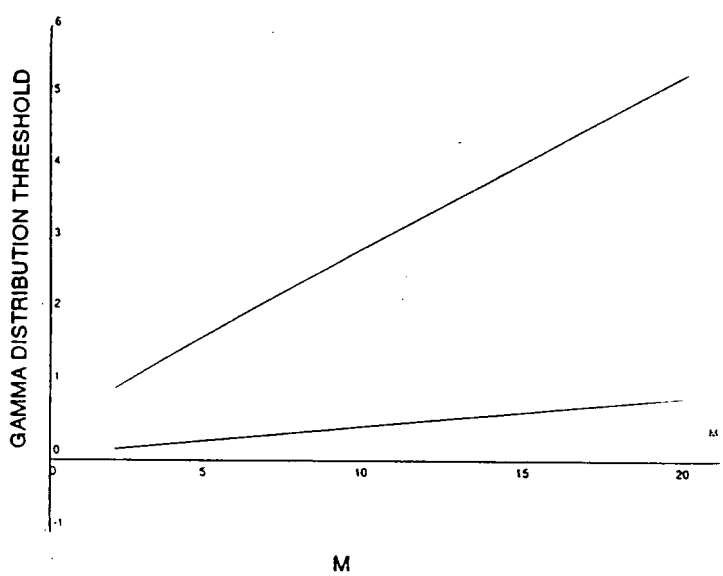


Figure C-5. Gamma distribution thresholds.

large M would be A the ratio of slopes minus one. Unfortunately, poor numerical conditioning prevents the computation of these curves for small P_{FA} or large M.

MOTION MODELS AND COMBINING LIKELIHOOD RATIOS OVER SPATIAL CELLS

Most target models treat velocity as approximately constant, estimating $2d$ parameters in a d-dimensional space, for position and velocity. Normally, either highly accurate measurements or some idea of target mission, plan, or tactics are required to estimate acceleration. Acceleration estimation may be appropriate in some aircraft and missile applications, but for undersea surveillance applications it is not considered necessary.

Wei's and Zeidler's (1992, 1993) bounded hop motion model is an example of a model estimating only d parameters. Target velocity is assumed bounded by an *a priori* maximum speed, but is otherwise free to change arbitrarily from time step to time step. This approach was previously investigated under the HGI project, and concerns arose about whether the performance was relevant, mainly because it id not seem to take any advantage of the consistency in target velocity that one would expect to see for submarines. Accumulation of likelihood ratio statistics under the assumption of known target velocity requires forming the product of the likelihood ratios of the measurements at each time step at the predicted position of the target. The following recursive formula computes the accumulated likelihood ratio at time step $M + 1$ and position x_{M+1}, from the individual likelihood ratios for the measurement at x_{M+1} and the accumulated likelihood ratio from the previous time step and the previous position: $L_{M+1}(x_{M+1}) = l_{M+1}(x_{M+1})L_M(x_M)$. If the velocity is not known, but only bounded by B spatial cells per time step, then Wei's and Zeidler's algorithm accumulates a pixel statistic (which does not seem to be a likelihood ratio for $M > 1$), by the following formula: $[L_{M+1}(x_{M+1}) = l_{M+1}(x_{M+1}) \max_{|x_M - x_{M+1}| \le B} L_M(x_M)$. Only one likelihood ratio is maintained for each spatial cell at each time step. Performance results quoted below include results from Wei's algorithm, which are treated as a baseline for comparison.

RANDOM WALK PATHS

A likelihood ratio statistic with better detection performance than Wei's pixel statistic can be based on a target motion model of a random walk on a d-dimensional grid, with independent identically distributed increments. The extra information used is consistency of target velocity over time. A random walk modeling ignorance of target velocity models the velocity by a distribution with mean zero and substantial variance; a random walk modeling fairly good knowledge of velocity models the velocity by a distribution with nonzero mean and small variance. The requirement that the target velocity distribution be discrete and not vary from time step to time step imposes a lower bound on the variance of the average velocity, but this bound decreases like $1/M$ and is not onerous.

Let the state space cells be points on a grid in a discrete vector space, $x_M \in Z^d$. Consider the random generation of the target path to occur in two stages: first, a starting point and a random walk are chosen, the walk only roughly determining the average velocity of the taret, and second, an instance of that random walk is chosen to generate the exact path taken by the target. Consider a random walk, W, starting at x_1, with increments, or steps, with density $P(\Delta x)$, which does not vary over time or space. The initial segments of the walk are random sequences of points written as $w_M = (x_1, x_2, ..., x_M)$, and the expected average velocity of any segment of the walk is $E(\Delta x)/T$.

We accumulate measurement likelihood ratios for a walk ending at X_{M+1} with a formula of the form $L_{M+1}(x_{M+1}, W) = l_{M+1}(x_{M+1}) \sum_{x_M} P(x_M|x_{M+1})L_M(x_M, W)$. We exclude the possibility of two targets with crossing paths, i.e., we assume that all the alternatives are exhausted by specifying hypothesis H_0 to mean that no targets exist along any path being considered and hypothesis $H_1(x_M)$ to mean that exactly one target exists, and it is traveling along a path through x_M (at the time considered). Let

$z^M(x_M)$ denote the measurement vector at the spatial cell x_M taken in the time window ending at time T_M. Then the likelihood ratio l_M obtained from a single measurement is

$$l_M(x_M) = \frac{P(z^M(x_M)|H_1(x_M))}{P(z^M(x_M)|H_0)} .$$

Let x_M denote either the point, or the event, $H_1(x_M)$. The event w_M is the conjunction of points, or events, $H_1(x_1)$, $H_1(x_2)$, ..., $H_1(x_M)$. The accumulated likelihood ratio is the sum over all paths which are instances of W:

$$L_{M+1}(x_{M+1}, W) = \frac{\sum_{x_M} P\left(z^{M+1}(x_{M+1}), z^M(x_M), ;;;, z^1(x_1), w_M | x_{M+1}\right)}{\prod_{m=1}^{M+1} P(z^m(x_m)|H_0)}$$

Expand and rearrange the numerator of this expression to obtain a recursive relation:

$$\sum_{w_M} P\left(z^{M+1}(x_{M+1}), z^M(x_M), ..., z^1(x_1), w_M | x_{M=1}\right)$$

$$= P\left(z^{M+1}(x_{M+1})|x_{M+1}\right) \sum_{w_M} P\left(z^M(x_M), ..., z^1(x_1), w_M | x_{M+1}\right)$$

$$= P\left(z^{M+1}(x_{M+1})|x_{M+1}\right) \sum_{x_M} P(x_M|x_{M+1})$$

$$\times P(z_M(x_M)|x_M) \sum_{w_{M-1}} P\left(z^{M-1}(x_{M-1}), ..., z^1(x_1), w_{M-1}|x_M\right) .$$

Divide the equations above by the denominator of $L_{M+1}(x_{M+1}, W)$, $\prod_{m=1}^{M} P(z^m(x_m)|H_0)$, and substitute in the transition probability $P(x_M|x_{M+1} = P(\Delta x = x_{M+1} - x_M)$ defined by the random walk to obtain

$$L_{M+1}(x_{M+1}, W) = l_{M+1}(x_{M+1}) \sum_{x_M} P(\Delta x = x_{M+1} - x_M) L_M(x_M, W) .$$

The initial values that start the recursion are $L_0(x_0, W) = 1$ for all x_0 and W. In a practical implementation, one would set $P(\Delta x) = 0$ for all but two or three increment values Δx, so that the sum over x_M would involve only two or three points. But, the presence of addition operations complicates any attempt to manipulate log likelihoods instead of likelihoods.

One can visualize the relative contribution of the likelihood ratios obtained over time from neighboring spatial cells by plotting the probability of the random walk passing through each of those cells. Figure C-6 is a surface plot of the probability at each time step, 1 through 15, of a random walk that ends up in cell 0 at time 16 and that has an expected velocity of minus one cell per time step. The points in the past history that contribute part of their likelihood to $L_{16}(x_{16}, X_{-1.0})$ are arranged in a narrow fan shape, the middle of the fan points in the direction of the expected average velocity of the random walk. The amount each point contributes to the accumulated likelihood ratio is greatest near the middle of the fan. This averaging of nearby likelihoods is the reason little difficulty occurs with our random walk algorithm when signal energy leaks out of the spatial cell containing the true target position.

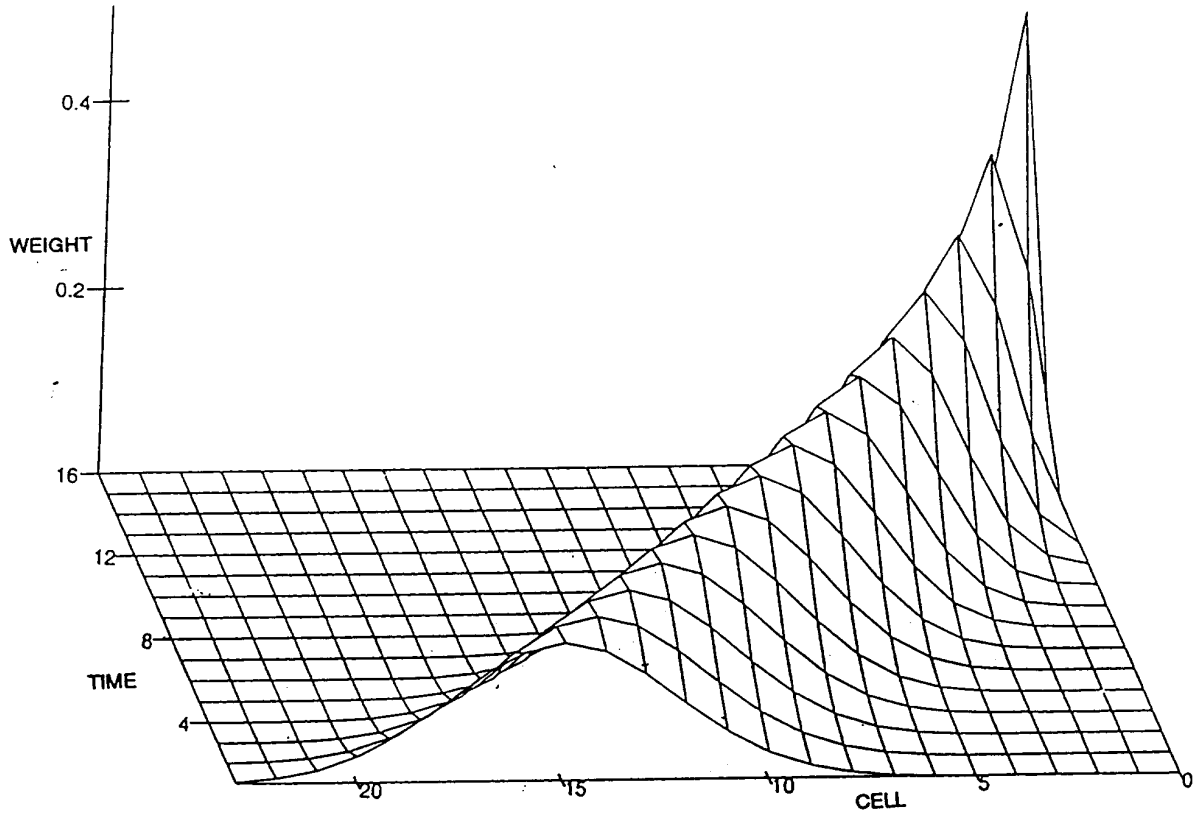
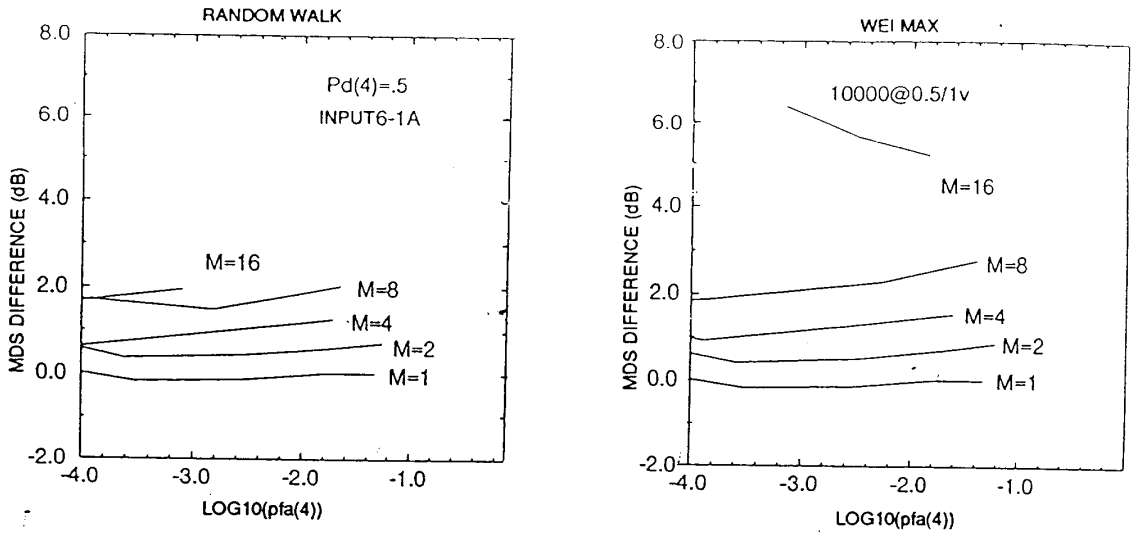


Figure C-6. Example of random walk spatial occupation probability.

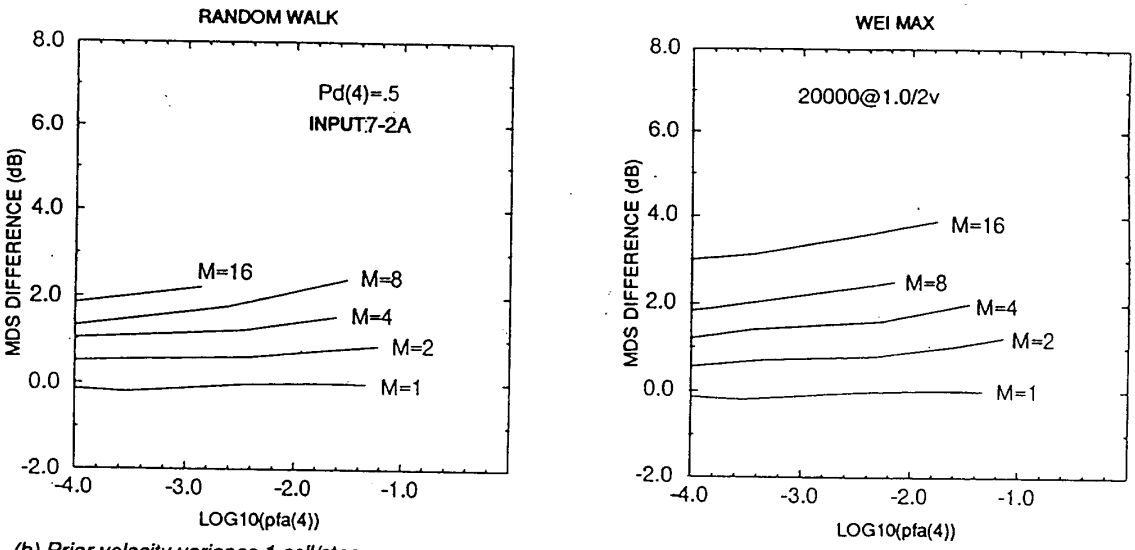
Serious attempts at deriving a probability distribution for $L_{M+1}(x_{M+1}, W)$, even under the simplest conditions, yielded unworkably complex formulas, so results from the unknown path case are obtained from Monte Carlo simulations. To conduct Monte Carlo simulations, we create an array of likelihoods, one for each spatial cell and each random walk going through that cell, and update the array recursively at each time step, by applying equation 1. In order to evaluate the likelihood of a target being in the spatial cell, regardless of how it got there, we compute a weighted sum of all the likelihoods for that spatial cell, the weights being the prior probability of each of the random walks going through that cell. The increments of each random walk have a density structured to produce the desired expected velocity and to satisfy continuity conditions. The density on the average velocity of the random walk is trinomial, but has a bell-shape. A prior Gaussian density on target velocity is approximated by a mixture of a small collection of these multinomials with the mixture weights the prior probability of the corresponding random walk. For comparisons with the random walk algorithm, the hop-bound B for Wei's algorithm is set relative to the prior target velocity distribution at the 2σ point.

SIGNAL GAIN DEGRADATION FOR UNKNOWN PATH

Figures C-7a, b, c, and d show the increase in minimum detectable signal level caused by lack of knowledge of the target path. This is computed from the difference in levels required in Monte Carlo simulations of tracking detection minus the analytically defined minimum detectable signal level for the known path case. Each figure shows random walk and Wei's algorithm results for M = 1, 2, 4, 8,

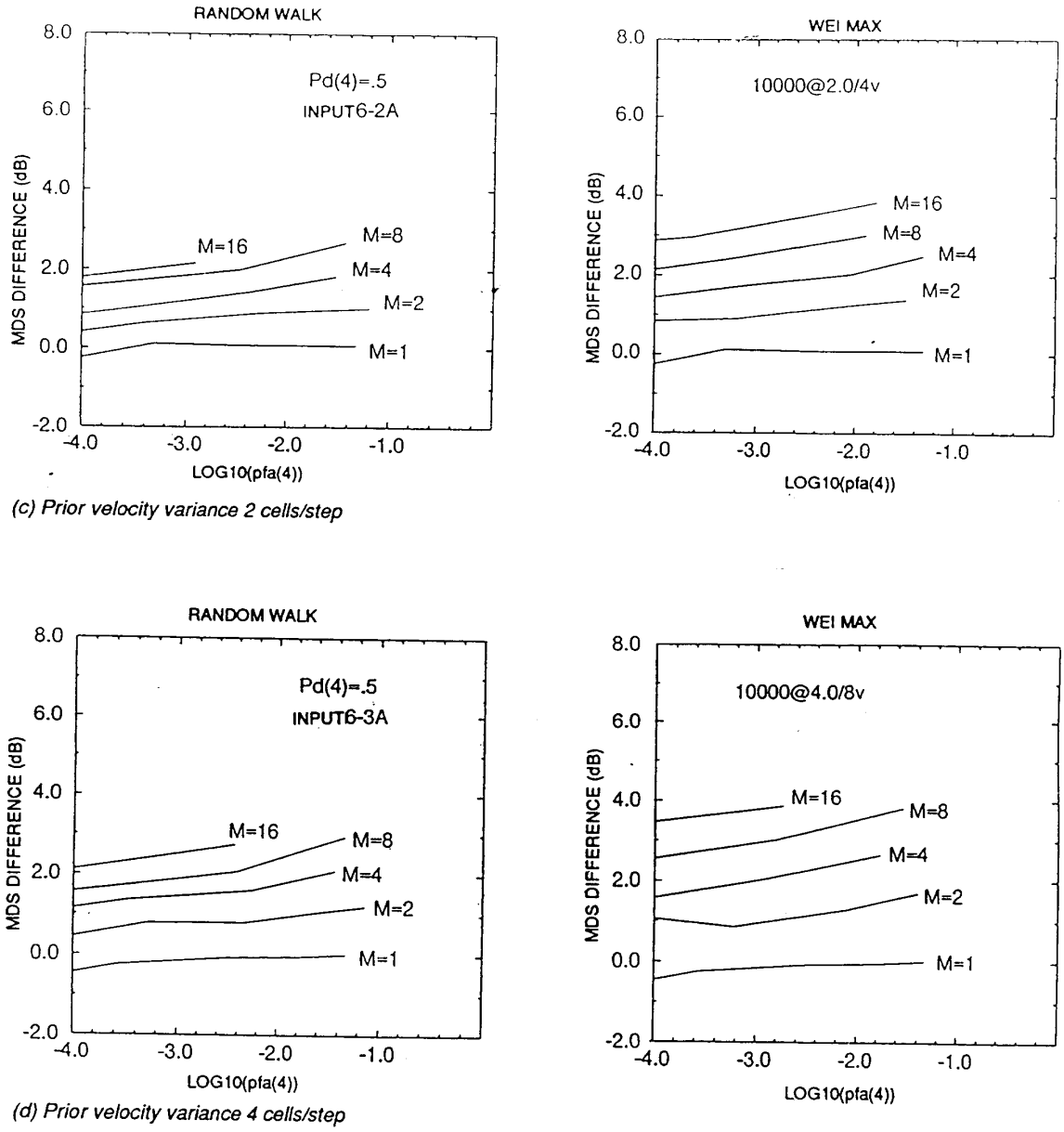


(a) Prior velocity variance 0.5 cell/step



(b) Prior velocity variance 1 cell/step

Figure C-7. Dependence of detector on prior velocity variance.



(c) Prior velocity variance 2 cells/step

(d) Prior velocity variance 4 cells/step

Figure C-7. Dependence of detector on prior velocity variance (Continued).

and 16. The P_D is fixed at $P_D(4) = 1/2$ and $\log(P_{FA}(4))$ is plotted on the abscissa, whlie level differences in decibels are plotted on the ordinate. Figures C-7a, b, c, and d differ from each other in the size of the variance of the prior velocity distribution. This determines the number of neighboring spatial cells a target might travel into during one time step of T seconds. The figures C-7a, b, c, and d present results for standard deviations of 0.5, 1, 2, and 4 cells per time step. The results look similar. The level differences depend weakly on P_{FA}, and increase by about 0.5 dB per doubling of M, for the random walk algorithm, and somewhat more for Wei's algorithm. Figures C-8 and C-9 illustrate the degradation caused by poor estimates of the prior velocity variance, which err by a factor of 4 and 1/4, respectively.

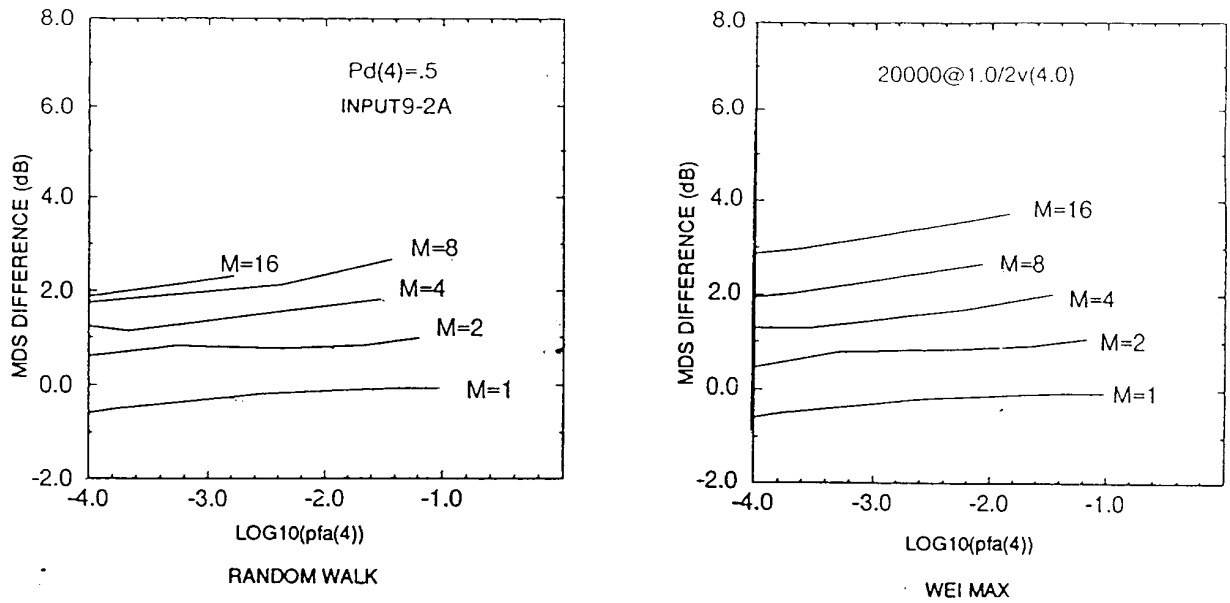


Figure C-8. Dependence of detector performance on low estimate of prior velocity variance.

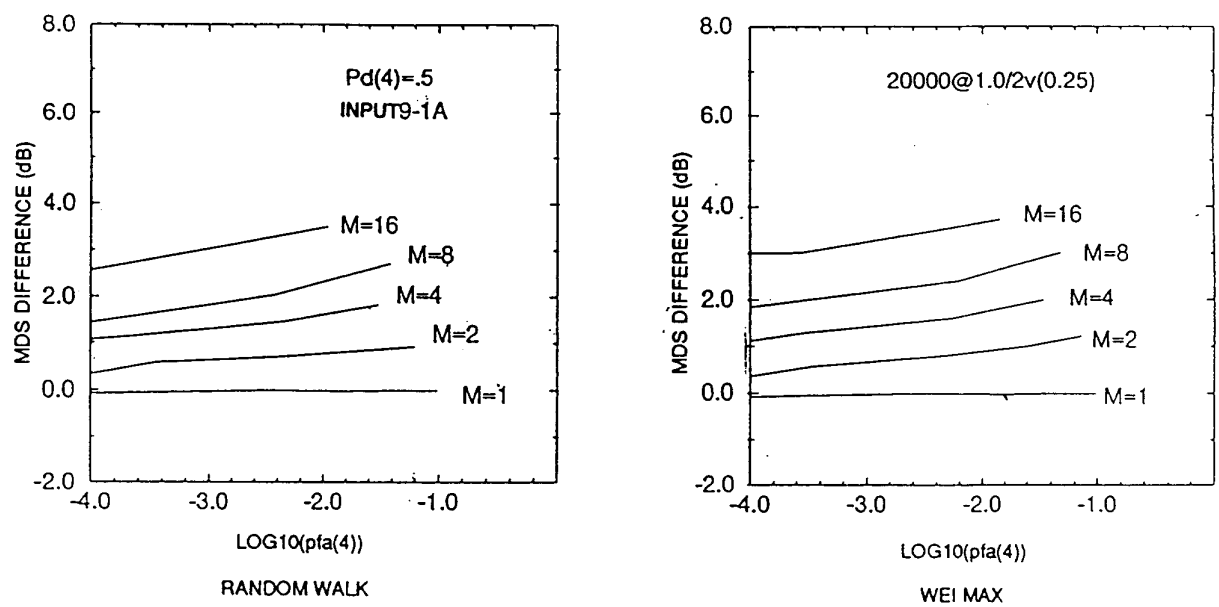
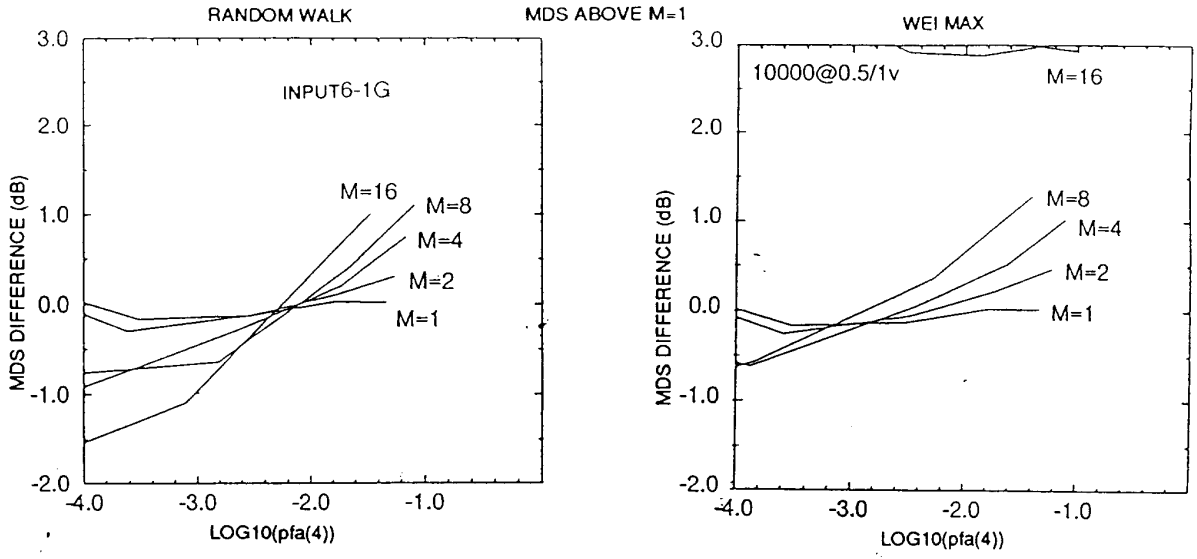


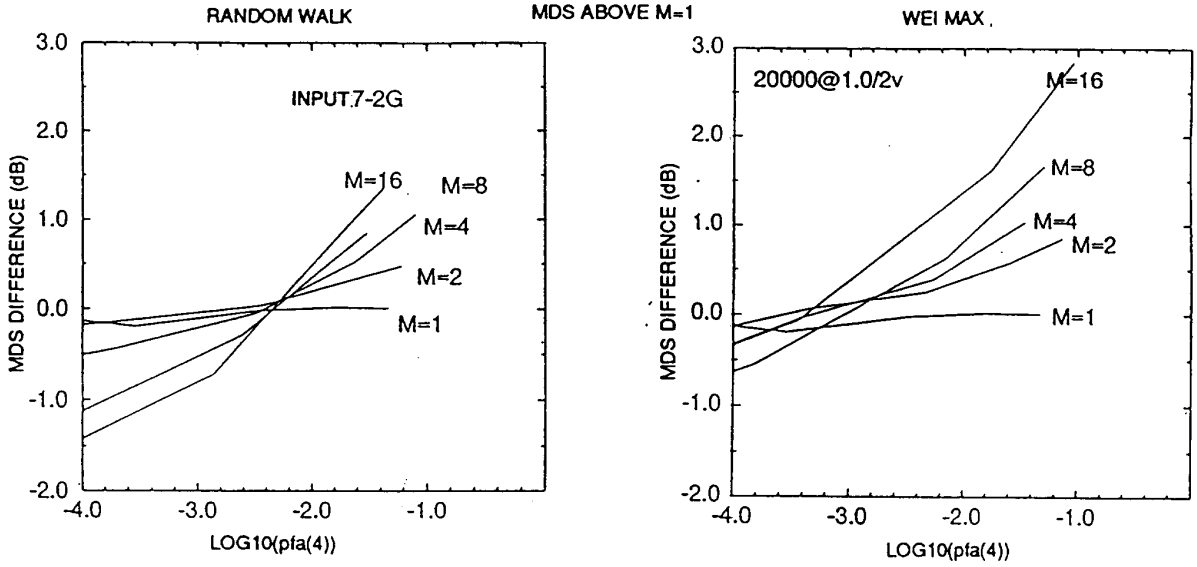
Figure C-9. Dependence of detector performance on high estimate of prior velocity variance.

SIGNAL GAIN FROM INCOHERENT INTEGRATION

Figures C-10a, b, c, and d illustrate the signal loss, or gain, i.e., the change in minimum detectable signal level, caused for an M greater than one. This was computed from the difference in level required in Monte Carlo simulations of tracking detection for various M less the analytically obtainable level for the case $M = 1$. Each figure presents random walk and Wei's algorithm results for $M = 1, 2, 4, 8$, and 16. The P_D is fixed at $P_D(4) = 1/2$ and $P_{FA}(4)$ is plotted on the abscissa, while

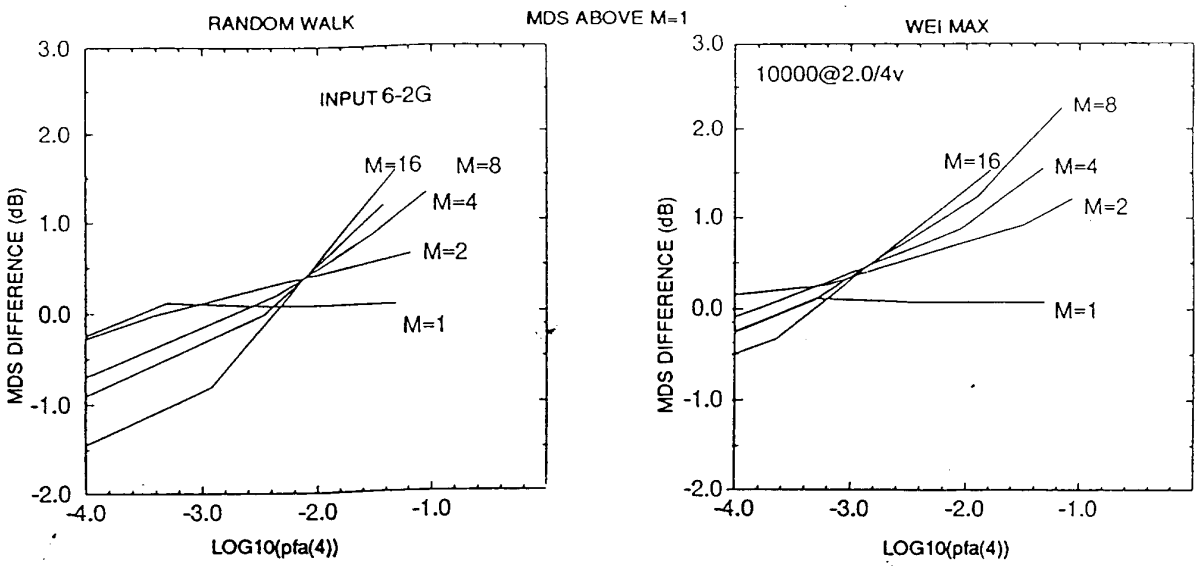


(a) Prior velocity variance 0.5 cell/step

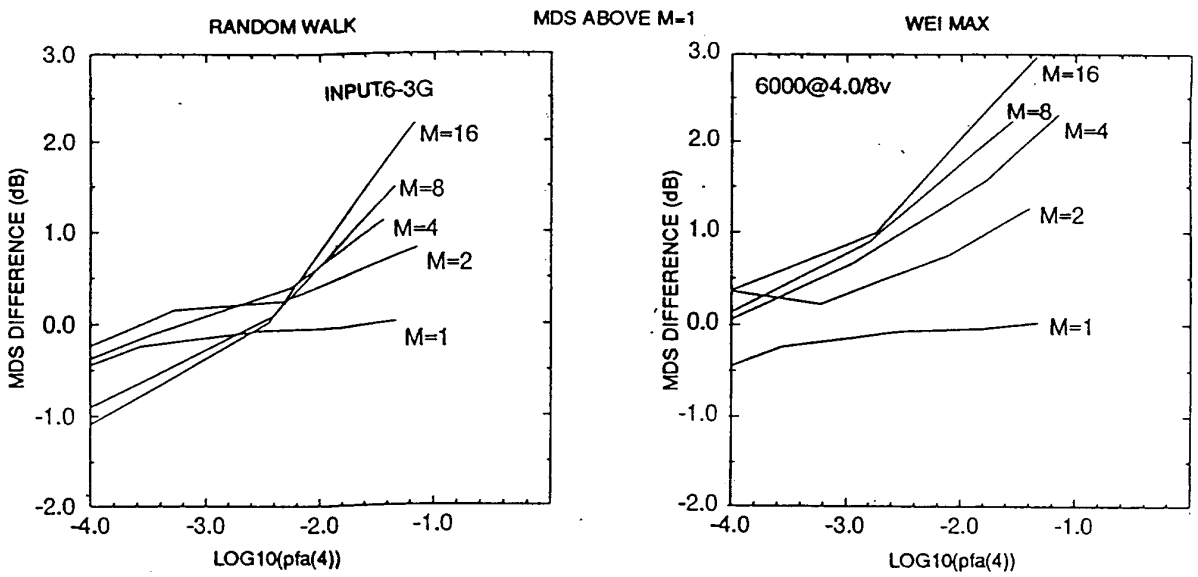


(b) Prior velocity variance 1 cell/step

Figure C-10. Dependence of detector performance on low estimate of prior velocity variance.



(c) Prior velocity variance 2 cells/step

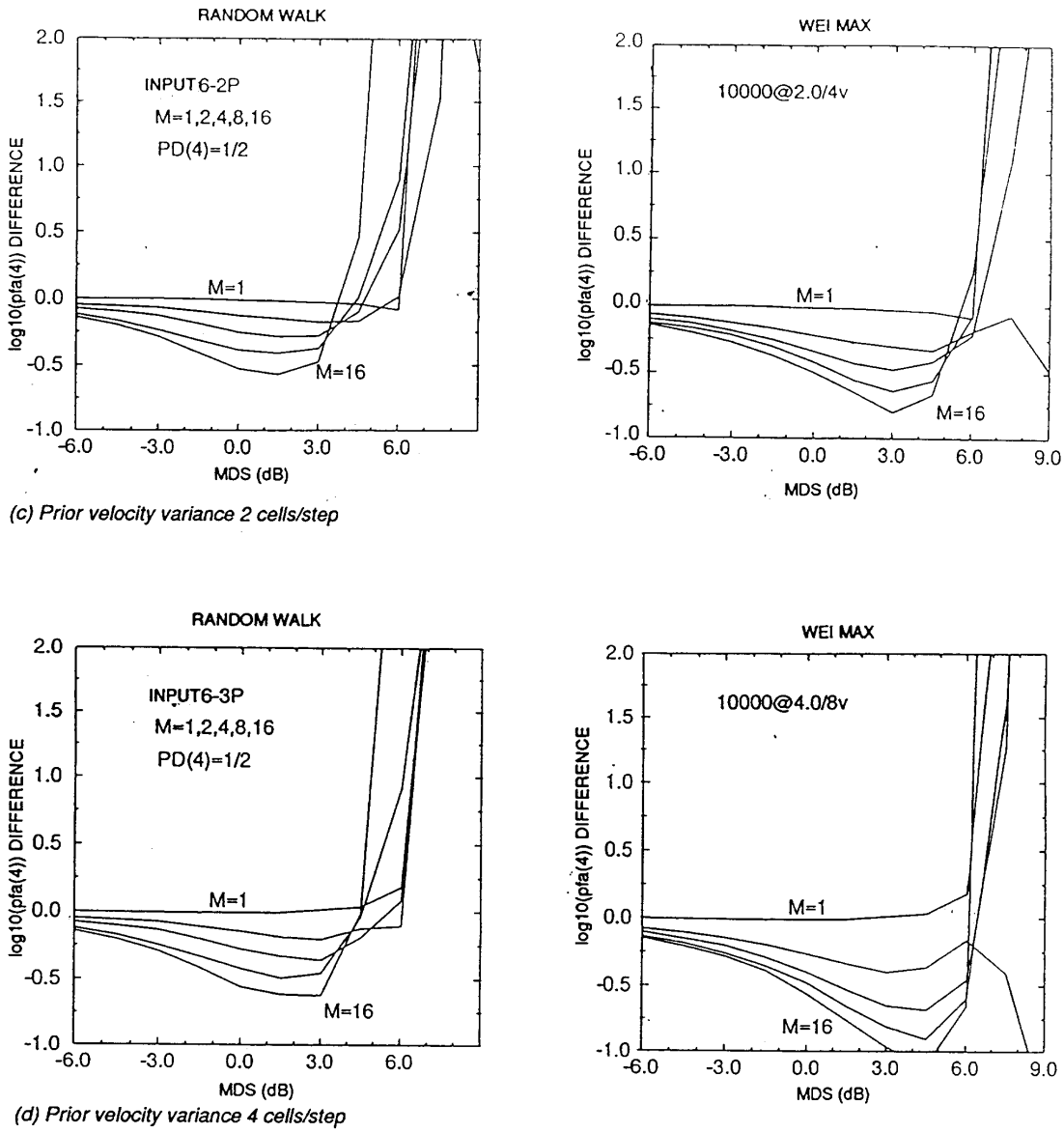


(d) Prior velocity variance 4 cells/step

Figure C-10. Dependence of detector performance on low estimate of prior velocity variance (Continued).

level differences in decibels are plotted on the ordinate. The results presented in figures C-10a, b, c, and d are for different variances of the prior velocity distribution, as for figures C-7a, b, c, and d. The crossing of the curves at $P_{FA}(4) = 10^{-2.2}$ for the random walk, and at $P_{FA}(4) = 10^{-3.3}$ for Wei's algorithm, show that weak signals (supporting only large P_{FA}'s) cause trouble for tracking detectors, and that for the weakest signals setting $M = 1$ appears optimal. The break even point being higher for the random walk algorithm is a sign that it is performing more strongly than Wei's algorithm.

A more useful description of the regime of operation in which tracking detection profits for $M > 1$ can be displayed in terms of the increase, or decrease, of $P_{FA}(4)$ caused by choosing a path length of M instead of one, given fixed $P_D(4)$ and various minimum detectable signal levels is illustrated in figures C-11a, b, c, and d. Observe that the break-even points are near 3.5 dB for the random walk algorithm, and 6 dB for Wei's algorithm. These break-even points bound the range signal-to-noise ratios for which tracking detection offers a potential benefit.



(c) Prior velocity variance 2 cells/step

(d) Prior velocity variance 4 cells/step

Figure C-11. Dependence of detector performance on path length.

HIGH P_D REGIME

The question arises whether the price of not knowing the target velocity is significant for strong signals, or high values of P_D, since the track ought to be evident from the data in such cases. If so, the tracking detector performance should approach that of the known-path case, given strong enough sig-

nals. Figure C-12 illustrates the change in minimum detectable signal levels from the known path case to the tracking detector case, for $P_D(4) = 0.95$ and various P_{FA}. The difference is not zero. The increase in levels for the random walk detector might be attributable to the fact that likelihood ratios for neighboring spatial cells are averaged together whether the signal is strong or weak, so that some noise-only measurements contribute to the detector. This seems unavoidable for the present algorithm implementation, because the variance of the velocity of the random walk has a lower bound imposed by the discretization of target motion.

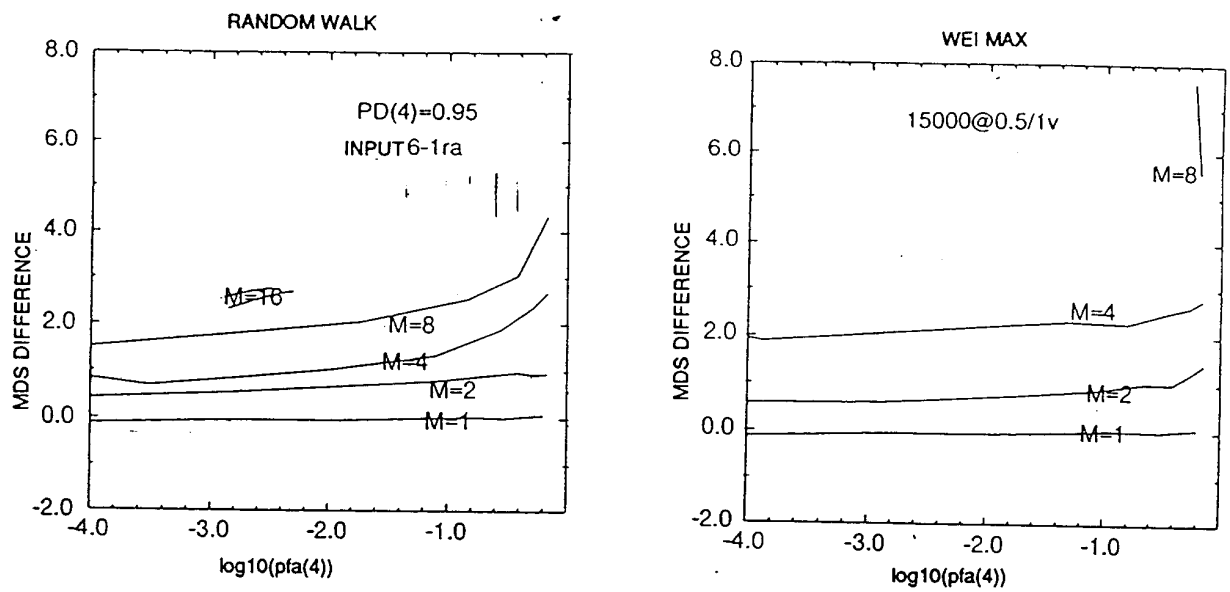


Figure C-12. Dependence of detector performance on knowledge of track.

SUMMARY

The processing load increases with increasing M depending on the specific tracking algorithm. For Wei's algorithm, the processing is proportional to M. For the random walk algorithm, there is an additional processing load, proportional to $M^{d/2}$, which depends on the number of random walks required to model possible target velocities. Chen (1989) examines the signal loss caused by finite discretization of the velocity in a probability map tracking detector. He shows that for $d = 2$ the number of velocity cells required to meet a loss constraint increases as M^1, which corresponds to our $M^{d/2}$ rate.

The minimum detectable signal level for the known path case seems bounded as $M \rightarrow \infty$, so that tracking detectors can gain onl a bounded advantage over an if-any combiner, but the bound may be large. In any case, the loss in level caused by doubling M, which is weakly dependent on the operating point, is a little less than 1 dB per doubling.

The signal loss caused by lack of knowledge of the target path is approximately $\frac{1}{2}\log_2(M)$ dB. This loss is also weakly dependent on the operating point. The break even point in signal-to-noise ratio, below which a tracking detector offers no benefit for detection, depends on the detection algorithm employed. Observe that for too weak a signal the signal cannot be identified among noise peaks that line up by coincidence.