1983

An evaluation of the broadband direction finding capabilities of array signal processing techniques

Larry Donald Bacon
Iowa State University

INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106

Bacon, Larry Donald

AN EVALUATION OF THE BROADBAND DIRECTION FINDING CAPABILITIES OF ARRAY SIGNAL PROCESSING TECHNIQUES

Iowa State University PH.D. 1983

An evaluation of the broadband direction finding capabilities of array signal processing techniques

by

Larry Donald Bacon

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1983

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

a_i = (where i is an integer) filter model coefficients

a_n = array element weighting coefficients

A = complex number which specifies the amplitude and phase of a wave at the origin

A(S) = complex number which specifies the amplitude and phase of a wave at the origin with direction specified by S

$\bar{A}$ = vector of MLM filter weights

A_0 = a constant

AF (·) = array factor of linear phased array

b_ℓ = (where ℓ is an integer) coefficients of moving average (MA) model

c = velocity of light

C = cos θ

$C_x(\cdot)$ = time-space covariance function of array data

d = spacing between array elements

E = column vector of exponentials
= $[1 \quad \exp(j\ 2\pi f_0 \Delta t) \quad \cdots \quad \exp(j2\pi[p-1]f_0\Delta t)]^T$

$\bar{E}$ = electric field intensity vector

E_y = the y-component of the electric field intensity vector

E[·] = the expectation operator

f = cyclic frequency

f_0 = operating frequency

f(t) = a function in the time domain

F	=	program variable equal to radio frequency of incoming plane waves
$\bar{H}$	=	magnetic field intensity vector
H(z)	=	system transfer function
j	=	square root of -1
k	=	propagation constant of medium -- the increase of phase difference per unit distance in the direction of propagation
$\bar{k}$	=	vector propagation constant -- a vector with magnitude k pointing in the direction opposite to the direction of propagation
K	=	ratio of physical length of array L to spatial modulation envelope length Z_d at the point of resolution
K(·)	=	ratio of physical length of array L to spatial modulation envelope length Z_d as a function of the arguments
ℓ	=	an integer
L	=	total physical length of array
LA	=	program variable equal to the order M of the autoregressive model
M	=	order of autoregressive model
n	=	an integer
N	=	number of samples
N	=	number of elements in array
P_0	=	a constant equal to the final prediction error power
P(f)	=	power spectrum as a function of frequency f
P(S)	=	$\lambda CA(S)$, a power spectrum in the spatial domain
$\bar{r}$	=	space-lag variable in correlation function

$\bar{R}_{xx}$ = signal correlation matrix

S = $\sin \theta$

$S(\omega)$ = a power spectrum

$S_x(f,\bar{\nu})$ = average power of plane wave component at frequency f and wavenumber ν

t = time variable

T = inter-sample spacing

T_d = period of time domain modulation envelope

v = phase velocity of wave

V = a volume in space

$V(\cdot)$ = complex voltage as a function of argument

W = $1/\Delta z$, a spatial frequency

$x(n)$ = a sequence

$x(t,\bar{z})$ = signal at point $\bar{z}$ in array

X = complex vector of array element voltages

$X(z)$ = z-transform of $x(n)$

$\tilde{X}(f,\bar{\nu})$ = complex phasor giving amplitude and phase of plane wave of frequency f and wave number ν

Z_d = length of spatial modulation envelope

$z(n)$ = location along the z-axis of the nth array element

$Z\{\cdot\}$ = z-transform operator

α = element to element phase shift

β = propagation constant of medium -- the increase of phase difference per unit distance in the direction of propagation

Δ = $\frac{\theta_2-\theta_1}{2}$, one-half of the difference between the bearings of two incoming plane waves

Δt	=	time interval between successive samples
Δz	=	spacing between array elements
ε_n	=	a noise sequence
η	=	characteristic impedance of medium
θ	=	bearing of signal source with respect to line of array
θ_m	=	bearing of desired maximum of phased array
θ_0	=	look angle - the angle halfway between the angles of arrival two incident plane waves
θ_1	=	bearing with respect to the line of the array of the incoming plane wave closest to 0°
θ_2	=	bearing with respect to the line of the array of the incoming plane wave farthest from 0°
λ	=	wavelength of incoming plane waves
$\bar{\nu}$	=	$2\pi\bar{k}$, vector wave number
ν_z	=	component of $\bar{\nu}$ in z-direction
σ^2	=	variance of spatial frequency estimate $\frac{2\pi}{\lambda}\cos\theta$
τ	=	time-lag variable in correlation function
Φ	=	angle with respect to real axis made by zeros in complex plane
ψ	=	$\frac{2\pi\Delta z}{\lambda}\cos\theta + \alpha$, composite variable used in array factor expression
ω	=	radian frequency
ω_d	=	$\omega_1 - \omega_2$, a difference radian frequency
ω_{sd}	=	$\frac{2\pi}{\lambda}(\cos\theta_1 - \cos\theta_2)$, spatial difference frequency
ω_1	=	radian frequency of first signal
ω_2	=	radian frequency of second signal
^{H}	=	indicates Hermitian transpose

^{T}	=	indicates matrix transpose
*	=	indicates complex conjugation
$\hat{}$	=	indicates function is an estimate
$\underline{\ }$	=	indicates symbol is a vector or matrix

I. INTRODUCTION

A. The Problem and the Scope of the Paper

Receiving array technology -- the use of an array of sensors to accurately, flexibly, and economically determine the characteristics of the field which surrounds it -- has broad applications in many areas, including radar, underwater acoustics, radio direction finding, and geophysics. A major problem facing all of these applications is the poor resolving capability of physically small receiving arrays, along with ambiguities due to sidelobe responses. This has led to the search for array signal processing techniques which exhibit higher angular resolution capability and low ambiguity.

To understand the reasons for these problems, we begin with a brief discussion of the basic ideas of array theory. Full details can be found in a text on antenna fundamentals such as Weeks (1968).

The array geometry which we will consider is shown in Figure 1.1. The elements of the array (which could be hydrophones, seismometers, or radio antennas) lie in a line along the z-axis. This line is known as the line of the array. Element number zero is located at $z = 0$. Assuming the location of element n is $z = z(n)$, the distance between elements is $z(n) - z(n-1) = \Delta z$. The angle θ is the polar angle measured from the line of the array. L is defined to be the total length of the array and is equal to $(N-1)\Delta z$.

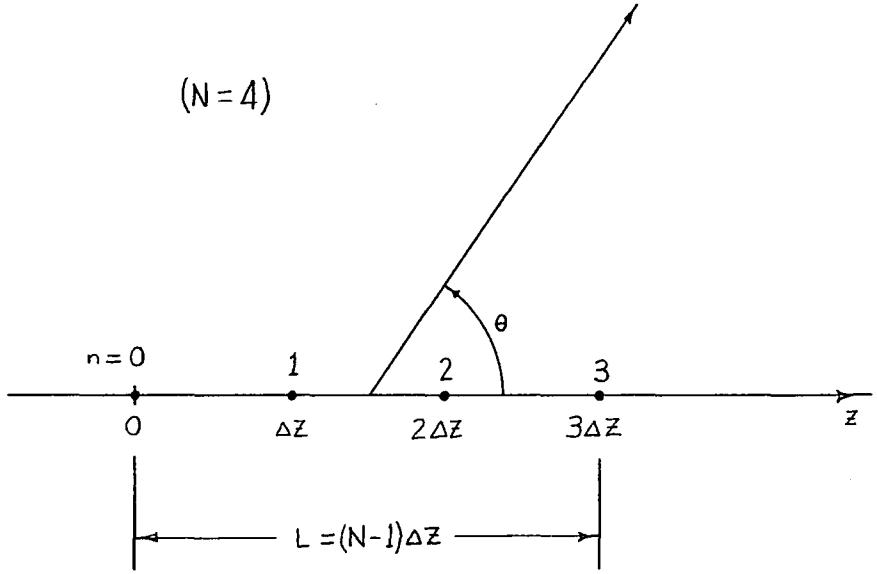


Figure 1.1. Basic array geometry

For simplicity, the elements are assumed to be isotropic, that is, they receive signals arriving from any direction equally well. Assuming that the array has N elements, and that each of these elements has the same gain, the response of the array (using conventional processing) as a function of the angle of arrival of the signal θ is controlled by the array factor $AF(\Psi)$. The magnitude of the array factor has the form (Weeks, 1968)

$$|AF(\psi)| = \frac{1}{N} \frac{\sin(1/2\ N\psi)}{\sin(1/2\ \psi)}, \quad \text{where}$$

$$\Psi = \frac{2\pi\Delta z}{\lambda} \cos\ \theta + \alpha$$

α = element to element phase shift
λ= wavelength of signal being received.

This function has the form sketched in Figure 1.2. It is a periodic function with period 2π. The large response near $\psi = 0$ is called the main lobe. The other, smaller lobes are the sidelobes.

The resolution of the array is closely related to the width of the main lobe, that is, the distance between the two points of zero response, or nulls, nearest $\psi = 0$. This is sometimes called the beamwidth between first nulls (BWFN). A commonly used measure of an

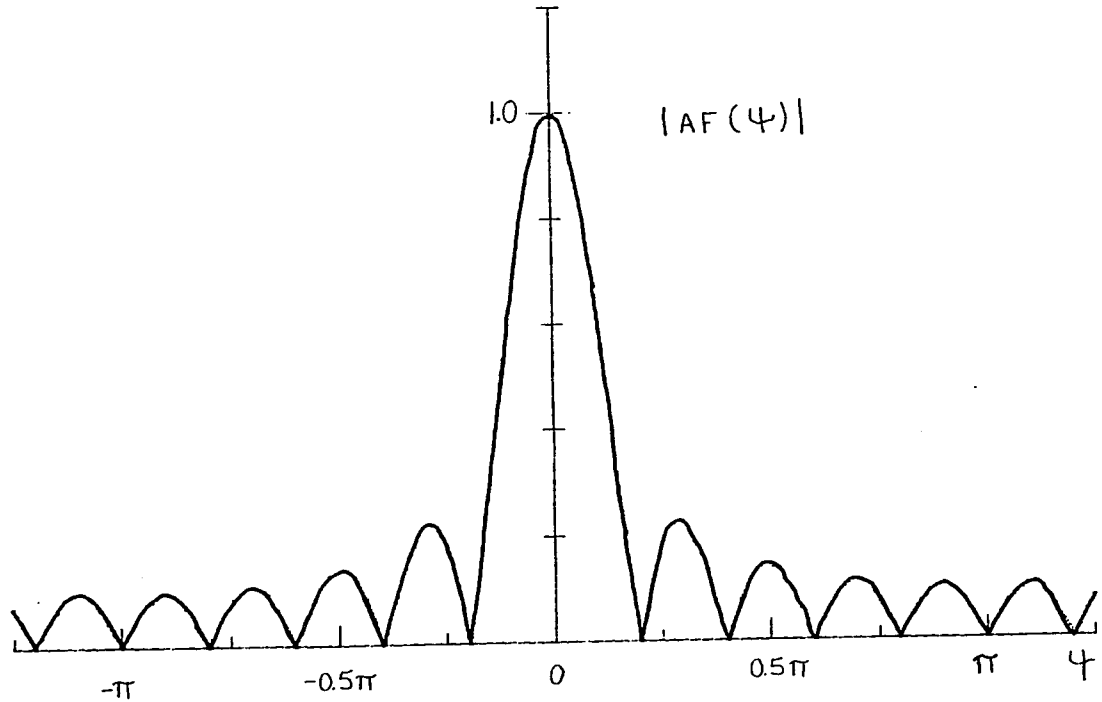


Figure 1.2. Array factor magnitude for N = 10

array's resolution is the Rayleigh criterion, which states that two sources are resolvable if their angular separation is at least 1/2 BWFN (Kraus, 1966). This limit simply means that the array can look at one source with full gain while the second source is in the null, and, thus, not interfering.

The BWFN depends upon the argument of the sine function in the numerator of AF. Notice that (with $\alpha = 0$)

$$\frac{1}{2} N \psi = \frac{1}{2} N \left(\frac{2\pi\Delta z}{\lambda} \cos\theta\right) = \pi \frac{(N\Delta z)}{\lambda} \cos\theta$$

Thus, the longer the array is in terms of wavelengths, the narrower the BWFN, and, hence, the better the resolution of the array. This result holds in general, even when $\alpha \neq 0$ or when each element does not have the same gain (nonuniform excitation). Conversely, if the array is small when compared to the wavelength of the received signal, the resolution is poor.

Notice, also, that the sidelobe responses are relatively high. For a uniform linear array, the first sidelobe is only about 13 dB down from the main lobe. Thus, a strong signal in one of the sidelobes could appear to be coming in from the direction of the main lobe. Ambiguity in pinpointing the signal's direction of arrival is the result.

The resolution problem brought about by small array size is particularly severe when it is desired to perform highly precise

direction finding over a broad range of frequencies. Clearly, the resolution of an array can be very good at the high frequency end of the operating band, where L/λ is large. As the sensors tune toward the low end, however, the size of the array in terms of wavelengths decreases. This, of course, changes the characteristics of the pattern and can dramatically reduce the resolution.

In an effort to minimize this problem, several new array signal processing techniques have been developed. These techniques seek to make the best possible use of the information available from the array. The motivation behind this approach is the fact that by processing the element signals in the optimum manner, the array should maintain acceptable resolution over the broadest frequency range possible for the given array size.

The objective of this study is to determine the broadband direction finding capabilities of array signal processing algorithms. The investigation will address the question, "Is it possible to use signal processing array techniques to achieve satisfactory broadband direction finding performance with a given array size, and, if so, what level of performance can be expected?" Satisfactory broadband performance means that the array has (1) high resolution -- the ability to distinguish sources which are physically close in angular separation, (2) low ambiguity -- low side lobe or spurious responses, and (3) broad bandwidth -- the ability to maintain this resolution and low ambiguity over a broad range of frequencies.

This paper will cover the two phases of the research, first, the investigation and comparison of array signal processing algorithms available, and, second, computer simulation of array performance using the most promising algorithms. Also included is the necessary background for understanding the operation of the algorithms.

Section I B is a survey of the literature pertinent to this investigation. Chapter II describes the relationship between array beamforming and spectral analysis. Array beamforming with these algorithms is basically a process of spectrum estimation, and this chapter describes the transition from time-series spectral analysis to beamforming. Chapter III outlines several of these high resolution spectral analysis algorithms, and discusses their advantages and disadvantages in the present application.

Chapter IV and V detail the second phase of the work, where the most promising of the techniques are simulated by computer and evaluated as to their broadband direction finding performance.

Finally, Chapter VI presents the conclusions which can be drawn from this study and suggests the directions that further research might take.

B. Survey of Relevant Literature

The literature of spectrum estimation techniques, the heart of these high resolution beamforming algorithms, is extensive. Only a few particularly helpful references will be cited here, with others

mentioned in the sections describing each specific method.

Kay and Marple (1981), in their excellent tutorial, describe, develop, and compare most of the techniques considered here. Their comparison of algorithm performance with a data set consisting of both narrow and broadband processes is particularly helpful. They include many references for more detailed study.

Modern Spectrum Analysis, an IEEE Press collection of selected reprints edited by D. G. Childers (1978), contains many of the most important papers written during the development of nonlinear spectral analysis techniques. Another excellent source describing these techniques is Nonlinear Methods of Spectral Analysis, edited by S. Haykin (1979). The chapter by McDonough (1979) directly addresses the problem of processing spatial array data. His earlier paper (McDonough, 1974) and a paper by Barnard (1982) also deal with this topic. A slightly different, but related, approach is taken by Gabriel (1980), who compares certain adaptive array algorithms with these high resolution spectral analysis techniques.

II. DIRECTION FINDING AND SPECTRAL ANALYSIS

A. Beamforming and its Links with Spectral Analysis

The purpose of this section is to point out the reasonableness of the view that array beamforming can be implemented through spectral estimation techniques. The idea that the two problems are closely related should have strong intuitive appeal, because many array concepts point to it. For example, one of the first principles taught in array theory is that the excitation distribution across the aperture of an antenna and the resulting pattern in the far field are Fourier transforms of each other (Weeks, 1968). Consider also the Butler multiple beamforming network. Its diagram is basically the flow diagram of a Fast Fourier transform, making the Butler array an analog implementation of the FFT (Skolnik, 1980).

Barnard (1982) points out the link between beamforming and spectral analysis in his discussion of conventional beamforming techniques. All conventional arrays use one of two methods to form their patterns:

1. Time shift and sum. The signal striking the array from the desired direction is enhanced by inserting the appropriate time delay between adjacent elements.

2. Phase shift and sum. Here, the signal from the desired direction is enhanced in the frequency domain by inserting the appropriate phase shift between adjacent elements.

Barnard emphasizes that both of these methods are equivalent to a direct multidimensional Fourier transform from time and space to

frequency and wavenumber. The resulting frequency-wavenumber spectrum contains information concerning the power as a function of frequency and the vector velocities (directions of arrival) of propagating waves (Capon, 1979).

The following section should confirm the reasonableness of the equivalence of beamforming and spectral analysis and place it on firm mathematical footing.

B. The Angular Spectrum of Plane Waves

The radiation field from which the array is to derive this frequency-wavenumber information is usually assumed to be the superposition of an infinite number of sinusoidal waves, of various amplitudes and phases, coming from a continuous distribution of far-distant sources. These waves sweep across the array from their individual directions as plane wave fronts (McDonough, 1979). Just as Fourier analysis of a time series decomposes the data into complex exponentials, frequency-wavenumber analysis decomposes the field into these plane waves (Lacoss, 1976). Many years ago, Booker and Clemmow (1950) provided a rigorous basis for these assumptions, and tied it to spectral analysis:

> The field at all points in front of a plane aperture of any distribution may be regarded as arising from an aggregate of plane waves traveling in various directions. The amplitude and phase of the waves, as a function of their direction of travel, constitutes an angular spectrum, and this angular spectrum, appropriately

> expressed, is, without approximation, the Fourier transform of the aperture distribution (Booker and Clemmow, 1950).

Reviewing their discussion of this point will illustrate why the distribution of sources in the far field can be computed from a knowledge of the spatial frequencies of the resultant signal along the line of the array.

Consider the situation shown in Fig. 2.1. The region $x > 0$ is homogeneous, with propagation constant k and characteristic impedance η. For a plane wave in the medium, k is the increase of phase difference per unit distance in the direction of propagation. The characteristic impedance η is the ratio of electric to magnetic field intensities.

Booker and Clemmow made the following simplifying assumptions in their analysis:

a) All fields oscillate sinusoidally, and have wavelength λ in the medium.

b) All fields are two dimensional, with no variation in the z-direction.

c) $\bar{H}$ is parallel to z, and $\bar{E}$ is parallel to the x-y plane.

If we maintain an $\bar{E}$ field in the plane $x=0$, this field will propagate into the region $x > 0$. Thus, we can think of the plane $x=0$ as the aperture of a planar array. We specify in the aperture plane the

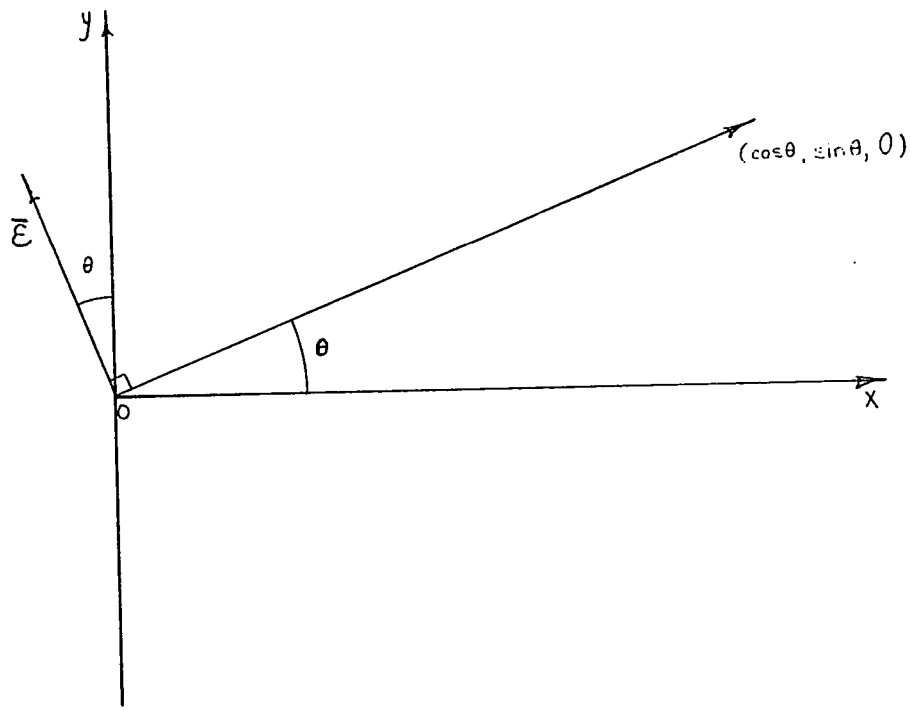


Figure 2.1. Angular spectrum of plane waves - geometry

y-component of $\bar{E}$ or the z-component of $\bar{H}$ as functions of y. The fields are completely specified by one or the other.

Suppose that the field propagating in the region $x > 0$ is a plane wave with direction vector $(\cos\theta, \sin\theta, 0)$. What field distribution in the aperture plane would be necessary to generate such a wave?

Following Booker and Clemmow, let $C = \cos\theta$, $S = \sin\theta$.
Then $\bar{E}$ and $\bar{H}$ can be expressed as

$$\bar{E}(x,y) = A(-S, C, 0) \exp[-jk(Cx+Sy)] \tag{2.1}$$

$$\bar{H}(x,y) = \frac{A}{\eta}(0,0,1) \exp[-jk(Cx+Sy)] \tag{2.2}$$

Setting $x=0$ to get the aperture distribution required, and looking at the y-component,

$$E_y(0,y) = AC \exp(-jkSy) \tag{2.3}$$

Notice that this is a wave traveling over the aperture of the array with a propagation constant kS. Any wave of this form produces a plane wave in the medium in the direction θ corresponding to $\sin\theta = S$. If we looked at the signal along the y-axis at any given moment, the distribution of the field would be a sinusoidal function of y. The spatial frequency of this sinusoid, in cycles per unit distance, depends upon kS, which, in turn, depends upon the angle of departure θ. If we

turned the situation around, we could compute the angle of arrival θ of a plane wave striking the array from a knowledge of the spatial frequency of the signal along y.

It is possible for kS to be greater than k, leading to an imaginary θ. This situation corresponds to an evanescent (nonpropagating) wave hugging the aperture. This wave would be the source of reactive terms in the near field.

Suppose that the aperture distribution were not quite as simple as the single sinusoidal traveling wave. By Fourier analysis, a complex distribution could be broken down into an infinite sum of sinusoidal waves, each having different propagation constants along the y-axis. Each of these components will give rise to a plane wave propagating in a direction θ (or to a reactive field).

Each wave has, in general, a different amplitude and phase, with the combination of them all forming what Booker and Clemmow called an angular spectrum of plane waves.

The constant A is a complex number which determines the amplitude and phase of a wave at the origin. It will be different, in general, for each wave. Thus, it will be a function of S. Define the function

$$P(S) = \lambda CA(S). \tag{2.4}$$

Substituting Eq. 2.4 into Eqs. 2.1, 2.2 and 2.3, and summing the contribution of all of the waves in the angular spectrum, that is, integrating over S from $-\infty$ to ∞, we obtain

$$\bar{E}\ (x,y) = \frac{1}{\lambda} \int_{-\infty}^{\infty} P(S)\ (-S,C,0)\ \exp\ [-jk(Cx+Sy)]\ \frac{dS}{C} \tag{2.5}$$

$$\bar{H}(x,y) = \frac{1}{\eta\lambda} \int_{-\infty}^{\infty} P(S)\ (0,0,1)\ \exp\ [-jk(Cx+Sy)]\ \frac{dS}{C} \tag{2.6}$$

produced by the aperture distribution

$$E_y(0,y) = \frac{1}{\lambda} \int_{-\infty}^{\infty} P(S)\ \exp\ (-jkSy)\ dS \tag{2.7}$$

Note that $\frac{dS}{C}$ is simply $d\theta$

Compare equation 2.7 with

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega)\ \exp\ (j\omega t)\ d\omega, \tag{2.8}$$

the inverse Fourier transform in the time domain. If we make the following substitutions:

Time Domain	Spatial Domain
t	-y
f(t)	$E_y(0,y)$
ω	kS
$S(\omega)$	P(S)

we see that Eq. 2.7 and Eq. 2.8 are of the same form.

The expression for the forward Fourier transform in the time domain is

$$S(\omega) = \int_{-\infty}^{\infty} f(t) \exp(-j\omega t)\, dt$$

Substituting the appropriate spatial variables, we obtain

$$P(S) = \int_{-\infty}^{\infty} E_y(0,y) \exp(jkSy)\, dy$$

Thus, the aperture distribution of the field and the angular spectrum of plane waves form a Fourier transform pair in the spatial domain.

The major objective of direction finding with an array is to determine the power and angle of arrival of each of the components in the angular spectrum. Clearly, in this simplified case, this information can be determined from spectral analysis in the spatial domain of the signal in the aperture.

Direct application of these ideas to beamforming, however, could run into several major problems. First, physical signals will always contain random components, due to receiver and ambient noise sources, fluctuations in the propagation medium, etc. This situation may require a probabilistic, or stochastic, description of the field. Just as conventional time-domain analysis must be modified when dealing with stochastic signals, so must conventional array spatial processing be modified when dealing with stochastic fields. Second, notice from the limits on the integral that to compute the angular spectrum exactly, a knowledge of $E_y(0,y)$ is required for all y. This would require an array of infinite extent unless one knew, as in the transmitting case, what E_y were outside the dimensions of the array. A finite amount of data will have great impact upon the attainable resolution, just as it does in time-domain processing. These problems will be discussed in the following section.

C. Frequency-Wavenumber Analysis of Array Data

The first problem mentioned in the preceding section, that of processing stochastic fields, can be handled in a manner analogous to the procedure for time-series analysis. This method uses the Wiener-Khinchine relation, which links the correlation function and the spectral density. The problem of limited data is particularly troublesome in the spatial domain, where obtaining more data points can be very difficult. This second problem has been the impetus for the

great deal of interest in applying high resolution spectral analysis to array beamforming.

1. Correlation processing of array data

McDonough (1974, 1979) has dealt thoroughly with the analysis of spatial array data using correlation processing. Following his development, we assume

$$x(t,\bar{z}) = \text{signal at point } \bar{z} \text{ in array.}$$

As discussed above, we consider $x(t,\bar{z})$ to be the superposition of infinitely many plane waves of the form, $\exp(j2\pi(ft + \bar{\nu}\cdot\bar{z}))$, where $\bar{\nu}$ is defined to be the vector wave number. ($2\pi\bar{\nu} = \bar{k}$, where $\bar{k}$ is the vector propagation constant.)

The wave propagation is in the direction of $-\bar{\nu}$, with phase velocity

$$v = f/|\bar{\nu}|.$$

As we did before, we must sum up the contributions of waves of every frequency and wavenumber, so that

$$x(t,\bar{z}) = \int_{-\infty}^{\infty} \int_{V} \bar{X}(f,\bar{\nu}) \exp\left[j2\pi(ft + \bar{\nu}\cdot\bar{z})\right] d\bar{\nu}\, df.$$

V refers to all wavenumber space.

$\tilde{X}(f,\bar{\nu})$ is a complex phasor representing the amplitude and phase of the plane wave component in the frequency and wavenumber ranges (f, f + df) and $(\bar{\nu},\bar{\nu} + d\bar{\nu})$. $\tilde{X}(f,\bar{\nu})$ is taken to be an independent random variable at each f and $\bar{\nu}$.

We want to find the average power of each travelling wave component

$$S_x(f,\bar{\nu}) = E\ [|\tilde{X}(f,\bar{\nu})|^2] ,$$

where E[•] is the expectation operator, by processing those values of $x(t,\bar{z})$ which are available from the array.

Motivated by the fact that the correlation function and the spectral density are related by Fourier transform in time-series analysis, consider the time-space covariance function of the array data:

$$C_x(\tau,\bar{r}) = E\ [x(t+\tau,\ \bar{z}+\bar{r})\ x^*(t,\bar{z})]$$

We assume that the field possesses wide sense stationarity, so that C_x depends only upon time lag τ and spatial offset $\bar{r}$, not on the values of t and $\bar{z}$.

Using the integral expression for $x(t,\bar{z})$, and remembering that the elemental sources were assumed to be independent, we find that

$$C_x(\tau,\bar{r}) = \int_{-\infty}^{\infty} \int_V S_x(f,\bar{\nu})\ \exp\ [j2\pi(f\tau + \bar{\nu}\cdot\bar{r})]\ d\bar{\nu}\ df$$

This can be recognized to be a multidimensional inverse Fourier transform, so that the frequency-wavenumber spectrum we are looking for is given by the forward transform

$$S_x(f,\bar{\nu}) = \int_{-\infty}^{\infty} \int_V C_x(\tau,\bar{r}) \exp[-j(ft+\bar{\nu}\,\bar{r})]\, d\bar{r}\, d\tau,$$

where V now refers to $\bar{r}$ space.

The strategy for frequency-wavenumber analysis, then, is simple: estimate the space-time covariance function of the data and then compute its spectrum (McDonough, 1979).

If there is an upper bound on the frequencies striking the array, then $|\bar{\nu}|$ also has an upper bound, and $x(t,\bar{z})$ is band limited in the space variable $\bar{z}$. In that case, $C_x(\tau,\bar{r})$ is limited in $\bar{r}$. Using sampling theory, this means that the spectrum can be computed from a sufficient number of equally spaced samples of $C_x(\tau,\bar{r})$ sampled in $\bar{r}$ (calculated from the signal values at the element of a discrete array), using any spectrum estimation technique which can be used with time series. Thus, each time-domain spectral analysis procedure is also an array direction finding algorithm. High resolution spectral estimation techniques will yield high resolution direction finding algorithms (McDonough, 1974), and it is with these techniques that we will mainly be concerned.

2. The resolution problem

Our goal is to estimate the frequency-wavenumber spectrum

$$S_x(f,\bar{\nu}) = \int_{-\infty}^{\infty} \int_V C_x(\tau,\bar{r}) \exp\,[-j2\pi(f+\bar{\nu}\,\bar{r})]\, d\bar{r}\, d\tau$$

from values of the signal $x(t,\bar{z})$ at a finite number of points in space (the values of $\bar{z}$ where array elements are located). These points are also contained within a finite region of space; physical arrays cannot be infinitely large. Working in three dimensions for generality, assume that the array is contained in a cube with $-L<x<L$, $-L<y<L$, and $-L<z<L$. What this means is that we do not have the estimates available for the cross covariance function $C_x(\tau,\bar{r})$ over <u>all</u> values of $\bar{r}$, although the expression above requires them for all $\bar{r}$. A similar problem exists with the values of time delay τ. We can gather data only for a finite period of time T. Thus, the time and space lags available are $-2T<\tau<2T$, $-2L<r_i<2L$ (where $r_i = x$, $i=1$, $= y$, $i=2$, $= z$, $i=3$), and the <u>estimate</u> of the frequency-wavenumber spectrum is

$$\hat{S}_x(f,\bar{\nu}) = \int_{-2T<\tau<2T} \int_{-2L<r_i<2L} \hat{C}_x(\tau,\bar{r}) \exp\,[-j2\pi(f\tau+\bar{\nu}\cdot\bar{r})]\, d\bar{r}\, d\tau$$

(where the $\hat{}$ over the function indicates that it is an estimate)

Whereas the original expression gave the frequency-wavenumber spectrum exactly, with perfect resolution of adjacent sources, this

estimate cannot provide such performance. How well it does depends upon the resolution properties of the spectral analysis technique used.

The performance of conventional linear methods of spectral analysis, such as the Blackman-Tukey autocorrelation approach or the averaged periodogram technique, depends very heavily on the length of the data sample (in both space and time). It is well-known that the resolution of these linear methods is roughly equal to 1/T in the frequency domain. Similarly, the resolution in wavenumber will be proportional to 1/L.

Implicit with these techniques is a windowing of the data. This problem will be discussed in more detail in the description of maximum entropy spectral analysis, but the basic consequence of windowing is a reduction of resolution. Kay and Marple (1981) state that data windowing is the fundamental factor in determining the frequency resolution of the periodogram. It influences the resolution so strongly because the spectral estimate is the convolution of the true spectrum and the transform of the window function.

If the power in the actual spectrum is concentrated in a narrow band, the convolution with the spectrum of the window function will spread the power out into adjacent frequencies. This is called leakage. The problem is particularly bad with short data sets, since then the window function must necessarily be narrow. The narrow window has a broad spectrum which can spread power over a broad frequency (or wavenumber) range during the convolution.

Another aspect of this same problem concerns sidelobes. If the data are merely truncated, the window function implicit in the analysis is rectangular. The spectrum of a rectangular function has many large sidelobes, half of which are negative. Clearly, in the convolution process a strong adjacent signal in a sidelobe could completely mask a weaker signal in the main lobe. It is even possible to get an estimate of negative power from signals in the negative sidelobes. Sidelobes can be reduced by "smoothing" or tapering the window function appropriately, but this always broadens the main lobe of the window transform, leading to a reduction of the resolution of the spectral estimate.

Kay and Marple (1981) point out a misconception concerning the resolution of the periodogram approach. Quite often, zeros are appended to the end of a data record prior to transforming to give an apparently longer sample. The effect of this procedure is to interpolate between the values which would be obtained from a non-zero-padded transform. It results in a smoother spectrum, but does not improve the resolution of the technique.

The actual resolution of linear processing methods can only be improved with more data. Returning to the array signal processing problem, it is apparent that improving the resolution in frequency requires only that the equipment spend more time gathering data. Improving the wavenumber resolution, the ability to distinguish sources which are close together in angle, however, would require a larger array. Often, given system constraints, this is not possible.

In estimating the frequency-wavenumber spectrum, then, the assumption is that T is large but L is not. That is, for each frequency we desire to look at, many more oscillation periods occur in time T than there are wavelengths across L. Conventional processing will suffice for the time domain but not the spatial domain. McDonough (1979) commented:

> ...the trade-off between data processing effort as opposed to increased data collection heavily favors the former, in the spatial domain, and nonlinear spectral estimation methods have been of great interest in array processing for that reason.

The nonlinear methods of spectral estimation are especially useful when good resolution of spectral peaks is desired (Pisarenko, 1973). This is exactly the desired result in direction finding with an array. Gabriel (1980) explains that this additional spatial information is obtained over the results of a conventional array because the array degrees of freedom are being used in a more effective, data adaptive manner.

Chapter III will discuss several of these nonlinear techniques and compare their performance in array applications.

D. Direction Finding with a Linear Array

In all of the discussion that follows, the array under consideration will be a discrete, equally spaced, linear array of isotropic receiving elements. Because it is symmetric, such an array is unable to distinguish sources with the same angle of incidence with respect to the line of the array. The linear array is therefore blind

to differences in polar angle in each plane perpendicular to the array axis (Barnard, 1982).

Another way of saying this is that the array is sensitive only to the wavenumber component which lies along the line of the array. This component is a function of both wavelength and angle of arrival.

Assume that the array lies along the z-axis. From Figure 2.2, we can see that the component of wavenumber along the array is

$$\nu_z = \frac{1}{\lambda} \cos\theta = \frac{f}{v} \cos\theta,$$

where f is the frequency of the signal and v is the velocity of propagation in the medium. Solving for θ,

$$\theta = \cos^{-1} \left(\frac{(v)(\nu_z)}{f}\right).$$

Returning for a moment to the discussion in Section II C 1., recall that

$$S_x(f,\bar{\nu}) = \int_{-\infty}^{\infty} \int_V C_x(\tau,\bar{r}) \exp\left[-j2\pi(f\tau+\bar{\nu}\cdot\bar{r})\right] d\bar{r}d\tau.$$

Let

$$P(f,\bar{r}) \triangleq \int_{-\infty}^{\infty} C_x(\tau,\bar{r}) \exp\left[-j2\pi f\tau\right] d\tau$$

Then

$$S(f,\bar{\nu}) = \int_V P(f,\bar{r}) \exp\left[-j2\pi\bar{\nu}\cdot\bar{r}\right] d\bar{r},$$

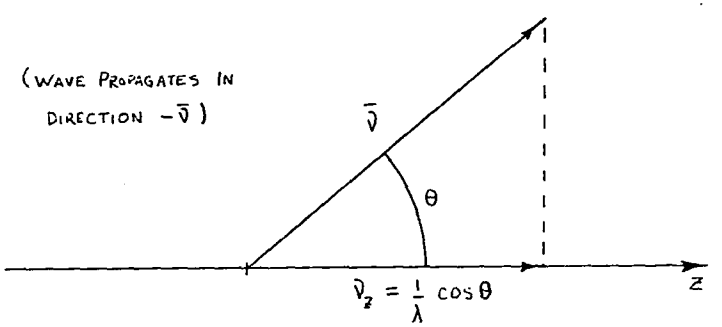


Figure 2.2. Component of wavenumber along array

which for a linear array along the z-axis becomes

$$S(f,\nu_z) = \int_{-2L}^{2L} P(f,r) \exp\,[-j2\pi\nu_z r]dr$$

$$= \int_{-2L}^{2L} P(f,r) \exp\,[-j2\pi(\frac{1}{\lambda}\cos\theta)r]dr$$

If we draw an analogy between this wavenumber spectrum and time-series analysis, it is clear that ν_z plays the part of frequency, and element location plays the part of the time-sampling instant. $P(f,r)$ is the function whose spectrum is being computed. If the interelement spacing Δz is no more than $\lambda/2$ at the highest signal frequency of interest, all energy in the spectrum $S(f,\nu_z)$ lies in the wavenumber range $[-W/2, W/2]$, where $W = {}^{1}/_{\Delta z}$ (Barnard, 1982). This is the spatial equivalent of the sampling theorem. Once the wavenumber spectrum is computed the results can be plotted so that

$$\theta = \cos^{-1}(\frac{v\nu_z}{f}),$$

giving the power of each plane wave component as a function of angle of arrival θ.

With this background of the extension of spectral analysis algorithms to array direction finding, we are now ready to investigate

several high resolution spectral analysis techniques. The hope is to find techniques with resolution sufficiently high that they will allow useful array direction finding performance over a much broader range of operating frequencies than conventional methods.

III. HIGH RESOLUTION SPECTRAL ANALYSIS ALGORITHMS

The past fifteen years have seen the development of several new spectral analysis algorithms, along with the reconsideration of some old ones. These spectral density estimators, called nonlinear because their design is data dependent (Haykin, 1979), can, under the proper conditions, exhibit much higher resolution than conventional techniques. The resolution advantages of two of these methods in particular, the maximum likelihood method (MLM) and the maximum entropy method (MEM), are most significant when processing short data sets (Gabriel, 1980). This implies potential for improved resolution for array antennas with few elements. Unfortunately, when conditions depart from ideal, the performance of these algorithms can degrade significantly.

In this section, we will consider the advantages and disadvantages of several of these new high resolution spectral estimators. The goal is to determine which of these techniques, if any, could provide improved broadband array direction finding performance.

A. The Maximum Likelihood Method

MLM was originally developed by J. Capon (1969) as a tool for analyzing the data from a very large seismic array. Thus, the original use of the technique was for frequency-wavenumber analysis. Capon (1979) states that the wavenumber resolution achievable by MLM is much greater than conventional methods and is limited primarily by the signal-to-noise ratio.

Gabriel (1980), in his description of a type of adaptive array whose operation is equivalent to an MLM spectral estimation, listed several advantages of this technique:

1. It permits calibration and measurement of the relative source strength of a single signal. Other methods (MEM in particular) do not give a clear or accurate estimate of signal strength.

2. A pseudolinear superposition holds at the peaks, if the sources can be resolved, giving true relative source strengths among several signals. Again, some other methods cannot do this.

3. The output of the filter is a real, physical signal. Steered to a particular source, one can monitor it at full array gain while rejecting all others. Other techniques do not provide physical signals. (A null, for example, can give a very good indication of angle of arrival, but the signal coming in at that angle cannot be monitored at full gain.)

4. The residual background ripple is low and well behaved.

5. The elements do not need to be equally spaced or near $\lambda/2$ apart. (Efficient solution of MEM requires equal spacing.)

The basic idea of MLM is that the power spectral density (PSD) is estimated by measuring the power out of a set of narrow band filters. The coefficients of the filters are chosen so that the response at the frequency of interest is unity and the output variance is minimized. The filters adjust to reject power from other frequencies not near the frequency of interest in an optimal, adaptive manner.

The conventional Fourier transform methods can also be considered to be a bank of filters, in this case, at harmonically related frequencies. The difference between the transform methods and MLM is that the shape of the MLM filter changes, in general, for each frequency to optimally reject out of band signals. The shape of the transform method filter is fixed (Kay and Marple, 1981).

The MLM filters are finite impulse response (FIR) with p filter weights a_i, $i = 0,1,\cdots,p-1$.

$$\bar{A} = [a_0, a_1, a_2, \cdots, a_{p-1}]^T.$$

The derivation of the method involves minimizing the output variance (output power)

$$\sigma^2 = \bar{A}^H \bar{R}_{xx} \bar{A},$$

where $\bar{R}_{xx}$ is the signal correlation matrix, subject to the unity frequency response condition at f_0

$$\bar{E}^H \bar{A} = 1.$$

Here, the H indicates Hermitian transpose, and

$$E = [1 \quad \exp(j2\pi f_0 \Delta t) \quad \cdots \quad \exp(j2\pi[p-1]f_0 \Delta t)]^T$$

The spectral estimate, $P_{ML}(f_0)$, corresponding to this optimal set of filter coefficients can be shown to be

$$P_{ML}(f_0) = \frac{\Delta t}{\bar{E}^H \bar{R}_{xx}^{-1} \bar{E}}$$

Note the dependence upon the data through $\bar{R}_{xx}^{-1}$.

Despite the advantages of MLM cited above, the method has several serious disadvantages which have discouraged its use. The first problem with the method is that, although it can have more resolution than Blackman-Tukey (BT) estimates or the periodogram, it has less resolution than the autoregressive estimate (AR or MEM) (Kay and Marple, 1981).

Burg (1972) very neatly explained this when he showed analytically that the MLM is actually an average of MEM spectra of order one to p:

$$\frac{1}{P_{ML}(f)} = \frac{1}{p} \sum_{M=1}^{P} \frac{1}{P_{MEM}(f)}$$

The low order, low resolution MEM spectra combine with the high order, high resolution MEM spectra to produce a smoothed, lower resolution result. Its variance, however, is less than the MEM estimate (Kay and Marple, 1981).

A second problem which could lead to doubts about its use is that the inverse Fourier transform of the MLM spectral estimate does not give

back the original autocorrelation function. The inverse transform of an MEM spectrum, on the other hand, does (Kay and Marple, 1981).

The most serious problem with this method (and the other nonlinear spectral estimators as well), however, is that the resolution may deteriorate very seriously if the actual data structure does not fit our assumptions about it. MLM appears to be particularly sensitive to this. Its resolution can actually become worse than the conventional techniques if the incoming waves depart from their assumed planar form (Seligson, 1970). Evans (1979) and White (1979) both found that the performance of MLM was quite poor when there is correlated interference. Evans concluded that MEM was able to offer substantially better performance than MLM and conventional beam sum techniques.

For these reasons, MLM will not be pursued further in this work. The next section will provide background for a more promising technique: the maximum entropy method.

B. The Maximum Entropy Method -- Autoregressive Techniques

1. Basic ideas

The maximum entropy method (MEM) was introduced in 1967 by J. P. Burg (1978) for the processing of geophysical data.

MEM differs from conventional methods of spectral analysis in a very fundamental way. In conventional spectral analysis, values of autocorrelation lags for which no data are available are assumed to be zero. Those values which are known are usually tapered smoothly to zero

to reduce the sidelobes of the data window function. The effect of this is to seriously reduce the resolution. Burg's idea was to extend the autocorrelation function (thus, increase the resolution) by estimating its values from a knowledge of the previous lags.

To ensure that this procedure does not bias the resulting spectrum, the extension must be done in such a way that it adds no information to the process. The idea is to choose the spectrum which corresponds to the most random time series whose autocorrelation agrees with all of the known values. This condition corresponds to the concept of maximum entropy as it is used in information theory (Haykin and Kesler, 1979), hence the name maximum entropy method.

Like MLM, MEM can be thought of as a digital filter which adjusts itself to be least disturbed by power at frequencies different from the one to which it is tuned (Haykin and Kesler, 1978).

The maximum entropy method has several attributes which make it well-suited for high resolution direction finding over a broad frequency range. First, it has greater (often much greater) resolution than conventional techniques. A second, particularly important advantage is that it is very well-suited to the analysis of short data records. This is highly desirable when dealing with arrays which have a small number of elements. The computational load is of the same order of magnitude as conventional methods (Haykin and Kesler, 1979). And, because there is no windowing of data, there are no sidelobes. Thus, the ambiguity of the array should be low (Burg, 1978).

2. Implicit models

To better understand why MEM has these advantages, let us first examine the differences between the assumptions of this method and the conventional Blackman-Tukey (BT) and periodogram (FFT) approaches.

All spectrum estimators make some sort of assumptions about the data being analyzed. These assumptions imply a particular model for the input data. According to Kay and Marple (1981), the differences in the performance of different spectral estimators are often due to differences in how well the models assumed by each estimator describe the process being analyzed.

Conventional FFT analysis implies a Fourier series model. In other words, the input data are assumed to be the weighted sum of a set of harmonically related sinusoids. The resulting periodogram is a least-squares fit of the data to this model.

Note that any noise present in the input signal is not accounted for explicitly in the model. It must also be represented by the harmonic sinusoids. This is the reason for the need to average over an ensemble of FFTs to obtain a stable estimator (Kay and Marple, 1981).

The most important implication of this underlying model is that the input data are periodic. Whether this assumption is significant or not depends, of course, on the problem, but for short arrays and noisy signals, it is clearly not correct.

The other commonly used conventional spectrum estimator is the Blackman-Tukey autocorrelation method. This approach makes use of the

fact that the autocorrelation function (AC) and the power spectral density function form a Fourier transform pair. The idea is to calculate an approximate autocorrelation function from the data available and then Fourier transform it to get the power spectrum. The autocorrelation function, as discussed earlier, can be only approximated due to the finite extent of the data. The unavailable lag values are assumed to be zero, and the resulting truncation, which implies a rectangular AC window function, introduces severe sidelobe problems. To reduce these difficulties, the known AC values are modified by a window which tapers smoothly to zero at the ends of the known lag interval.

When considered carefully, both the FFT and BT methods strongly violate what Ables (1974) calls the First Principle of Data Reduction:

> The result of any transformation imposed on the experimental data shall incorporate and be consistent with all relevant data and be maximally non-committal with regard to unavailable data.

The FFT assumes that the input data repeat periodically -- which is certainly not maximally noncommittal. The BT approach violates both aspects: unavailable lags are assumed to be zero, and known lag values are modified in a very ad hoc fashion by a fixed window. The price paid for these violations of the First Principle is less than maximum resolution.

According to Ables (1974), MEM attempts to set up new rules for determining the spectrum which abide by the First Principle as closely as possible.

The model implied by MEM allows more realistic assumptions about the unavailable data than that they are zero. MEM retains all of the known lag values without modification. It estimates the values of the unknown autocorrelation lags (Burg, 1978). Thus, windows and the distortion they introduce are eliminated.

Van den Bos (1971) has shown that the model of the input process implied by MEM is an autoregressive (AR) model. An autoregressive model for a sequence x has the form

$$x(n) = a_1\, x(n-1) + a_2\, x(n-2) + \cdots + a_M\, x(n-M) + \varepsilon_n,$$

where x(n) is the current value of the sequence, x(n-1) is the previous value, etc., the a's are the coefficients of the AR model, and ε_n is a noise term.

The name autoregressive comes from the fact that a linear model which relates a dependent variable x(n) to independent variables x(n-1), x(n-2),$\cdots$ x(n-M) is known as a regression model (Robinson, 1979). Since, in this case, the sequence x is regressed on itself, the model is called autoregressive.

MEM performs a least squares fitting of the coefficients of the AR model (a_1, a_2, $\cdots$, a_M) to the input process (Van den Bos, 1971).

3. Linear prediction and spectra

It is probably not immediately apparent why the parameters of the AR model of a sequence should be related to the maximum entropy spectrum. The derivation of the maximum entropy spectrum from information theory concepts is mathematically complicated and will not be repeated here. (For details, see Ulrych and Bishop (1975), or Haykin (1979).) Once the connection between this spectrum and that of an autoregressive process is accepted, however, it is not difficult to see how the spectrum of the input sequence can be computed from those parameters. We will look at this relationship and at the computation of the parameters from the viewpoint of linear prediction.

Suppose we have a sequence x(n) to which we wish to fit an AR model. That is, we want to find the coefficients $a_1, a_2, \cdots, a_M$ in the equation

$$x(n) = a_1 x(n-1) + a_2\, x(n-2) + a_3\, x(n-3) + \cdots + a_M\, x(n-M) + \varepsilon_n$$

such that the noise sequence, which we now consider to be an error sequence, is a minimum. In other words, we compute an estimate

$$x(n) = a_1 x(n-1) + a_2\, x(n-2) + \cdots + a_M\, x(n-M),$$

where $x(n) - x(n) = \varepsilon(n)$.

This process is called linear prediction, since we predict the value of x(n) from a linear combination of past values x(n-1), x(n-2), ···, x(n-M). ε_n is the prediction error to be minimized by proper choice of the a's.

This procedure is illustrated in Figure 3.1. The case of M=4 is shown. N, the length of the sequence, is 11. Here, members of a sequence x(n) are being predicted by multiplying the value of the sequence at a point by the coefficient immediately below it. The M products are then summed to produce an estimate of the next number in the sequence. In general, both the sequence and the coefficients can be complex numbers. The predicition filter is then advanced one element in the sequence and the next value is predicted.

A total of N-M prediction equations can be written without extending the prediction filter beyond the ends of the data sequence:

$$\hat{x}(n) = \sum_{k=1}^{M} a_k x(n-k) \qquad M \leq n \leq N$$

The filter may then be reversed, its coefficients conjugated, and then run backwards over the data to yield N-M more prediction equations (Ulrych and Clayton, 1976). Prediction error equations are written from these 2(N-M) prediction equations. Efficient algorithms exist for finding the solution which minimizes the prediction error (Ulrych and Bishop, 1975; Marple, 1980).

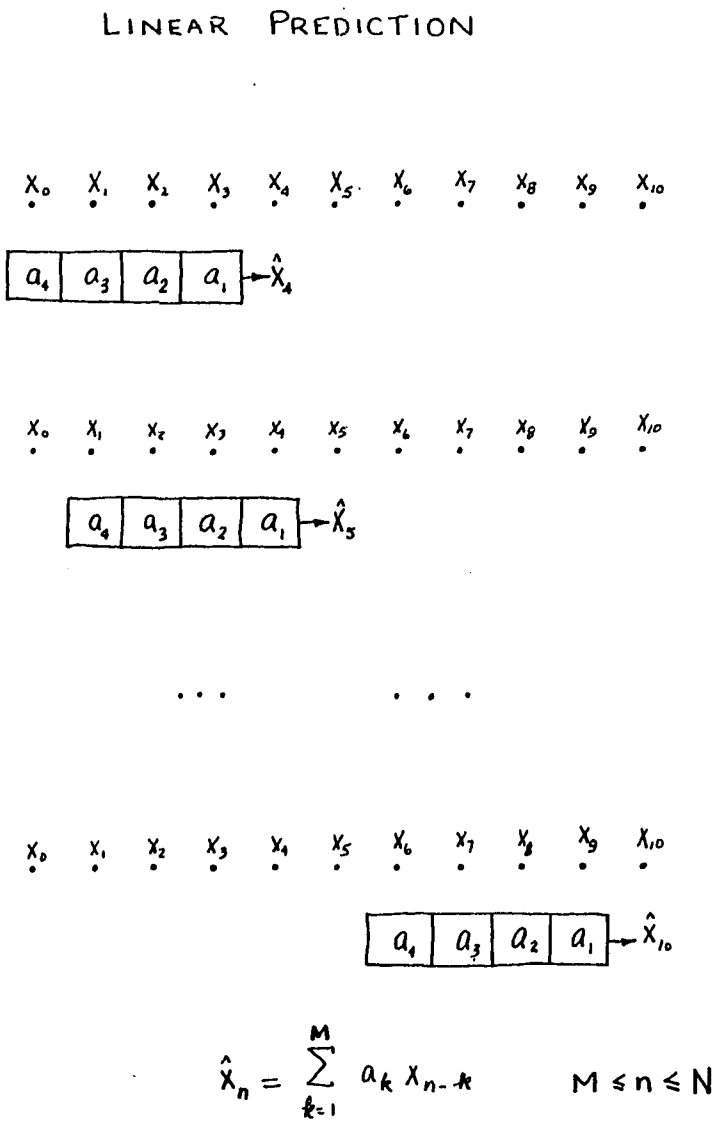


Figure 3.1. Linear prediction M = 4, N = 11

An implementation of a prediction error filter is shown in Figure 3.2. The blocks with z^{-1} inside represent unit time delays. This filter has a transfer function H(z) which we will define as the z-transform of the output sequence divided by the z-transform of the input sequence:

$$H(z) = \frac{Z\{\varepsilon(n)\}}{X(z)}$$

If the input sequence x is accurately represented by the autoregressive model, this prediction error filter will remove all of the predictable components of the input spectrum. The output sequence ε will then be completely random (unpredictable). The spectrum of a completely random sequence is uniform, or white. (The prediction error filter is often called a whitening filter for this reason.) Thus, the output spectrum is $Z\{\varepsilon(n)\} = P_0$, a constant.

Since

$$X(z) = \frac{Z\{\varepsilon(n)\}}{H(z)} = \frac{P_0}{H(z)} = \frac{P_0}{1-\sum_{k=1}^{M} a_k z^{-k}}, \qquad (3.1)$$

all of the spectral information (to within a constant) is contained in the zeros of H(z). These zeros are determined by the model coefficients a_k.

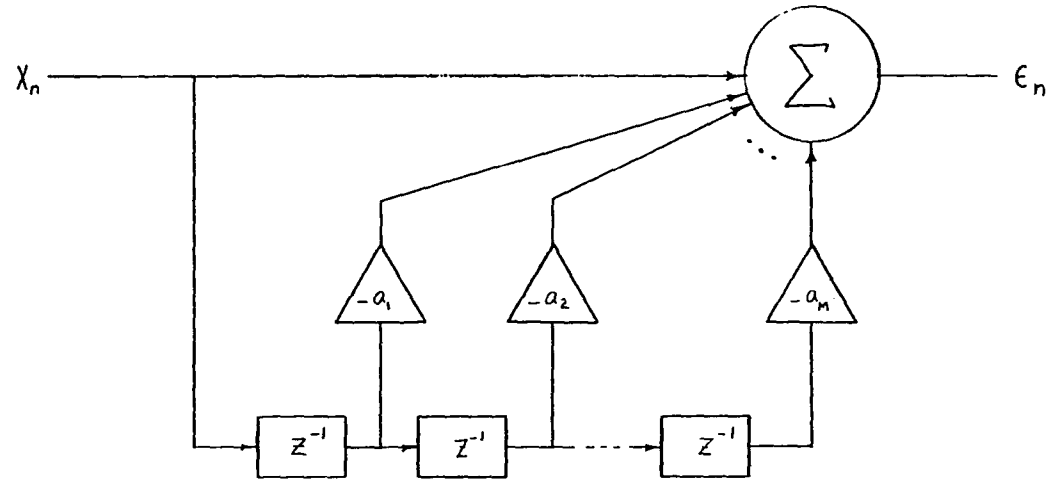


Figure 3.2. Prediction error filter block diagram

Thus, an algorithm which fits an autoregressive model to a sequence also gives the information needed to determine the maximum entropy spectrum of the sequence. We see that the maximum entropy method generates a unique filter based on information contained in the available data samples. The input spectrum is proportional to the reciprocal of the power response of this filter (Chen and Stegen, 1974).

The autoregressive model has M/2 pole pairs available to be placed in the z-plane to form the modelling filter. Since real frequencies fall on the unit circle in the z-plane, the resolution depends upon how near these poles are to the unit circle. This model can realize arbitrarily narrow passbands at M/2 different frequencies. Unlike the Fourier series model, these frequencies need not be harmonically related.

4. Applicability of the AR model

The AR model is a special case of the class of rational transfer function models. Many discrete processes, that is, processes whose values are defined only at discrete points in time or space, can be well-represented by models of this type (Kay and Marple, 1981). Both deterministic and random processes can often be approximated in this way.

The most general model of this type relates an input sequence $\varepsilon(n)$ to an output sequence $x(n)$ by the difference equation

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + \sum_{\ell=0}^{q} b_\ell \varepsilon(n-\ell).$$

Note that if all of the b's are zero except b_0, we have an AR model similar to the one discussed earlier. If all of the a's vanish, the output is a simple moving average of the past q+1 inputs. This yields what is known as a moving-average (MA) process. If both kinds of terms contribute, the process is known as autoregressive-moving-average (ARMA).

The transfer function H(z) relating the input and output sequences is rational:

$$H(z) = \frac{\sum_{m=0}^{q} b_m z^{-m}}{1 - \sum_{m=1}^{p} a_m z^{-m}}$$

Since the moving-average components contribute the zeros of this rational function, a moving-average process is also known as an <u>all-zero</u> process. Likewise, AR processes are called <u>all-pole</u> processes.

A theorem of statistics known as the Wold decomposition theorem says that a stationary ARMA or MA process can be represented by an AR process of possibly infinite order p (Kay and Marple, 1981). Modelling an ARMA process with a <u>finite</u> order AR model, the peaks in the spectrum are more accurately represented than the troughs, since the troughs correspond to the zeros of the MA part (Gersch and Sharpe, 1973). A higher order AR model must be used to compensate. If this is done,

however, nearly every time series found in real problems can be accurately represented by a finite AR model (Haykin and Kesler, 1979).

These facts are important for a very good reason. An AR model leads to linear equations, as we saw earlier. The other two models do not. This gives AR representations a very significant computational advantage.

Another advantage of the AR model has to do with pole-zero locations in the complex z-plane. If the order of a (p,q) ARMA process is overestimated, the model puts spurious zeros on the unit circle, where they can have tremendous impact on the resulting spectrum. The spurious zeros of the AR model tend to be further inside the unit circle where they have less of an effect (Ulrych and Clayton, 1976). We will have more to say about this in a later section.

Of course, how good the results of AR spectral analysis are depends upon whether the AR model is appropriate and upon the accuracy with which the data values are known (Frost, 1976). Since it employs an AR model, the maximum entropy method works best for AR processes. If the model does not fit the process being analyzed, the results are not as good.

As we saw in Chapter II, the estimation of the frequencies of sinusoids is particularly important in the array direction finding problem. A sinusoid can be considered to be a strong resonance of a very narrowband filter. Thus, at a high signal-to-noise ratio (SNR), an AR filter model is a good representation of the process. As we will

see, however, at a low SNR this model will give poor results, especially if the filter order is too small (Frost, 1976).

The role played by the model order is very important. If, in the ideal case, the order of the estimate is the same as the order of the actual process, the parameters are maximum likelihood estimates (Gerhardt, 1978). Using a model order which is too low broadens the resultant spectral peaks and lowers the resolution. If the MEM filter order is too high, the filter amplifies the effects of the noise and produces spurious spectral peaks of significant magnitudes (Chen and Stegen, 1974).

Since the maximum entropy method is equivalent to using an AR model to represent the input sequence, it is possible to apply the results of AR time series analysis for determining the appropriate model order to use (Ulrych and Bishop, 1975). Several criteria have been applied, including the Final Prediction Error (FPE) criterion, the Information Theoretic criterion (AIC), and the Autoregressive Transfer Function criterion (CAT). (See Haykin and Kesler, 1979.)

These criteria have been applied in time series analysis with some success. Ulrych, however, notes that the FPE criterion was of limited usefulness when dealing with short realizations of harmonic processes. He found that limiting the model order to between N/3 and N/2 gave the most consistent results (Ulrych and Clayton, 1976). Haykin and Kesler (1979) also reported that the filter order should be some percentage of the record length N. They mention that the optimum value usually lies in the range 0.05N to 0.2N.

5. Factors affecting the resolution of MEM

At a high signal-to-noise ratio (SNR), autoregressive techniques can place poles very close to the unit circle. This can lead to very high resolution. When a significant amount of noise is present, however, the pole locations must be moved away from the unit circle in order to model the data. As a result of this, resolution is degraded to that of the Fourier techniques (Frost, 1976). This degradation should be expected. Noise turns an AR process into an ARMA process, and the all-pole model implied by MEM is no longer valid (Kay and Marple, 1981).

Marple (1978) found empirically that the resolution of this technique is proportional to a power (approximately -0.31) of the signal-to-noise ratio expressed as a fraction, times the number of available autocorrelation lags. He states that at 20 dB SNR, MEM has roughly four times the resolution of an FFT, at 0 dB SNR it has twice the resolution, and at -10 dB the resolution is almost the same.

Chen and Stegen (1974) note that increasing both the number of data points and the model order improves the spectrum estimate of a noisy signal. This is reasonable, since a larger model order makes more poles available for representing the data. The poles representing the desired signal do not have to be displaced as far from their proper locations since there are poles available to account for the noise. Since the number of prediction equations is 2(N-M), however, increasing the model order M without also increasing the number of points N will reduce the

number of equations. This, in effect, reduces the averaging that the algorithm can perform in solving for the prediction filter coefficients. The variance of the frequency estimate is increased as a result. This point will be discussed further when we talk about improvements to MEM.

In the two sinusoid case, the relative phase between the signals also influenced the resolution. Zero phase difference at the beginning of the sampled interval was found to give the worst two sinusoid resolution. Marple (1978) explains this effect qualitatively by noting that the net transform of two or more windowed sinusoids is the result of the interference between the sidelobes of the $\frac{\sin f}{f}$ functions. This interaction is a function of the initial phase.

6. Special problems of MEM

MEM has several special problems which must be taken into account in its application. One of the most important of these problems is called line-splitting. Line-splitting occurs when a valid, single spectral peak splits into two closely spaced peaks. This effect is also quite phase and noise sensitive. It is worst when an odd multiple of a quarter cycle of the input signal is contained in the sampled interval, and when the initial phase is an odd multiple of 45 degrees. It is not as bad with other lengths and phases, but it can still contribute to shifts in the peak locations (Kesler and Haykin, 1978).

Another problem which plagues both AR and Fourier methods is known as frequency pulling. As two signals approach each other in frequency, the peaks remain at their proper locations until just before they merge. At that time, the peaks tend to pull in toward each other. The amount of frequency pulling is about the same for both techniques (Frost, 1976).

Chen and Stegen (1974) showed that the location of the spectral peak due to a single sinusoid is greatly influenced by the initial phase of the sinusoid. This effect decreased with a larger number of samples per cycle. The number of points, however, also affected the peak location. The location oscillates about its true value at a frequency of about twice the frequency of the sinusoid, as the length increases.

According to Frost (1976), the resolution of both the Fourier and AR techniques varies with the location of the signal in the frequency band. The variation tends to decrease with increasing model order.

The statistical properties of MEM as an estimator are very difficult to determine analytically. Most of what is known is based on actual use of the method. Gersch and Liu (1978), in their study of MEM's statistical performance, found that the spectral estimates near sharp spectral peaks were likely to have large bias and variance errors. Conventional windowed periodograms, however, also have large variance there. They conclude that AR spectral estimates appear to be asymptotically unbiased and consistent, with variance as low as the best windowed periodogram.

One final problem with AR spectrum estimation is that it has difficulty in measuring the power of sinusoids. Since it is a non-linear procedure, the relative power of two signals cannot be determined by comparison of the relative height of their peaks. Lacoss (1971) states that the amplitude of the response peak is not a good measure of power, but that the area under the peak is. This area, however, would be difficult to measure when two signals are near the resolution limit.

7. Least-squares algorithm

As we have seen, the maximum entropy method of spectrum estimation depends upon the accurate determination of the AR coefficients. In fact, Marple (1980) states that the method used to determine the a's is the key to the performance of the AR technique. Several different methods could be used to determine these coefficients. Although it would be inefficient, the prediction equations could be solved directly, for example. Another, less computationally intensive approach would be to use Burg's recursive algorithm. A third method which has some distinct advantages would be to apply the Least-squares algorithm suggested by Ulrych and Clayton (1976) (independently proposed by A. Nuttall), and extended by Marple (1980).

The Least-squares algorithm is based on an unconstrained least-squares estimation of the AR filter coefficients. Marple (1980) introduced a recursive algorithm for this estimation process. This

algorithm provides an implementation of the least-squares procedure with a computational complexity comparable to that of the Burg algorithm.

The Least-squares algorithm eliminates or reduces several of the problems mentioned in the previous section. Compared to Burg's method, Least-squares has less bias in the frequency estimate, less dependence upon initial phase when dealing with harmonic processes, no line splitting, and somewhat sharper resolution (Marple, 1980). Its algorithm is only slightly more complicated.

Burg's algorithm employs the Levinson recursion, a technique long known in time-series analysis. The algorithm minimizes the prediction error one coefficient at a time. The solution is iterated, starting with an AR model order of one and increasing the order by one each time until the desired order is reached.

In contrast to Burg's approach, Least-squares (LS) minimizes the prediction error by adjusting all of the AR coefficients for a particular filter order at the same time. The prediction error being minimized is taken to be the sum of the forward and backward prediction errors, corresponding to running the AR filter forward and backward over the data, respectively. Because of this, LS has been called forward-backward linear prediction method (FBLP).

As mentioned earlier, the LS estimate for the AR coefficient yields a spectral estimate which is more stable in its peak locations with changes in the initial phase when analyzing sinusoids. This implies

that LS has a big advantage over the Burg algorithm when looking at short realizations of harmonic processes (Ulrych and Clayton, 1976).

According to Marple (1980), the improved resolution of LS over Burg is due to the fact that the locations of the poles are unconstrained. Since they are allowed to approach the unit circle, the peaks can be sharper. On the other hand, nothing prevents them from moving into the unstable region outside of the unit circle. Since the frequency estimate depends only upon the angle the pole location makes with the real axis, and the distance from the unit circle, this effect is not bothersome. In practice, poles appear outside of the unit circle only rarely. It is important, however, if LS were being used to synthesize digital filters.

Another result of the unconstrained nature of pole location is that the amplitude of the peaks has greater variability. Since it is the area under the peak, rather than the peak's amplitude, that is significant, however, this variability is of little concern.

An effect which is bothersome, however, is the tendency of poles not related to a signal to appear close to the unit circle. This yields a spurious peak of significant amplitude. It is, however, the most effective of the linear prediction methods in terms of frequency variability and resolution. Because of this, it has served as the basis for some improvements to the AR methods which will be discussed in the next section.

C. Improvements to MEM: The Work of Tufts and Kumaresan

Most of the problems with the application of maximum entropy spectral analysis techniques are due to the presence of noise in the input sequence. If the performance of these methods is to be improved, then, the key lies in approaching the noiseless case by improving the effective signal-to-noise ratio. Tufts and Kumaresan (1982) have shown that the performance of the linear prediction methods can be improved by using prior knowledge of the rank of the signal correlation matrix. This approach is especially useful in the analysis of a finite number of sinusoids in the presence of noise.

Tufts and Kumaresan proposed a two step procedure for the improvement of the FBLP method at low SNR values.

Step one is to replace the estimated noisy signal correlation matrix of high rank with a smoothed one of the proper rank. In other words, the noisy signal correlation matrix is replaced with a lower rank least-squares approximation to it. This step effectively increases the signal-to-noise ratio.

Step two is to increase the filter order beyond the normal limits to improve the resolution. Since the effect of the noise is reduced, less averaging is needed. This allows the use of a higher filter order with its resultant higher resolution.

Signals like pure sinusoids are of low dimensionality. What this means is that the signal correlation matrix of a single sinusoid has rows and columns that are linearly dependent. It is of rank one. The

matrix for two sinusoids would have two rows and columns which are independent (with the rest dependent on these) and be of rank two, and so on. When noise is present, however, its randomness artificially increases the rank of the signal correlation matrix. This added "information" is what the linear prediction filter tries to account for, degrading the accuracy of the estimation of the actual signal frequencies and magnitudes.

This method can be thought of as an estimation of only the predictable components of the data (the signal), rather than trying to predict all of the data (including the unpredictable noise).

As a result of this modification, the performance of FBLP is brought close to that of true maximum-likelihood estimation (not MLM), even at moderately low signal-to-noise ratios. The unmodified algorithms exhibit a threshold effect at low SNR. Below the threshold, the frequency estimation performance degrades rapidly. This modification lowers that threshold considerably.

Maximum signal-to-noise ratio improvement occurs at a model order of approximately 3N/4. For model orders higher than the optimum, the increased variability brought about by the decrease in the number of terms being averaged overcomes the gain in resolution due to the longer filter.

We are more concerned with the performance of this algorithm in the array signal processing application than with the details of its derivation. Those details will be left to the references (Tufts and

Kumaresan, 1982). A brief discussion of how the filter coefficients are chosen, however, may help explain how these improvements are implemented.

The Tufts and Kumaresan method (TK) is computationally the same as FBLP with one important exception. The vector of coefficients, or prediction vector, is formed from a linear combination of the principal eigenvectors of the estimated signal correlation matrix. Eckart and Young showed in 1936 that the principal eigenvectors, that is, those eigenvectors which correspond to the largest eigenvalues, can be used to compute a matrix of a given rank which is a least-squares approximation to the given matrix. This lower rank matrix is the cleaned up version of the signal correlation matrix discussed earlier. The eigenvalues and eigenvectors of this lower rank matrix fall into two distinct groups or subspaces when the signal is of low dimensionality. One subspace corresponds to the signal, and is characterized by large eigenvalues. The other corresponds to the noise. Noise eigenvalues are usually well separated in magnitude from the signal eigenvalues.

The principal eigenvectors corresponding to the signal eigenvalues are fairly insensitive to the matrix element perturbations (i.e., the noise). Thus, a prediction vector derived from the principal eigenvectors of the noisy signal correlation matrix should be very close to that derived from the noiseless signal. This is the basis of the improvements of this method.

A special case worthy of note called the Kumaresan-Prony (KP) case occurs at the maximum allowable filter order L for a given number of data samples N. In this case, the noise eigenvalues and eigenvectors are automatically excluded, since the actual size of the signal correlation matrix is reduced to the correct rank simply by virtue of the fact that the number of available equations is 2(N-M). Since it requires no eigenanalysis, it is extremely easy to implement, involving only the inversion of a matrix which is usually small. Despite the greater variability brought on by the larger than optimum filter order, it does give improved performance over FBLP with a simple procedure.

TK has several very interesting and very nice characteristics. The poles of the prediction filter which account for the noise tend to be concentrated in clusters at definite, uniform positions removed from the unit circle. The poles for LS, in contrast to this neat arrangement, are scattered all over the unit disk. As mentioned earlier, some get so near the unit circle that they can be mistaken for signals. Thus, TK has a substantial improvement in ambiguity performance over LS. As a result of this pole placement, a spectrum plot from the TK method shows signal peaks which stand out well from a smoothly rippling, relatively uniform noise floor.

Any frequency estimator has a performance limit known as the Cramer-Rao (CR) bound. (For a derivation and discussion of this bound, see Van Trees (1968).) The CR bound places a lower limit on the variance of any unbiased frequency estimator. An estimator which is

able to achieve this bound is known as an efficient estimator. If an efficient estimate exists, it is guaranteed to be the maximum-likelihood (ML) estimate. For this reason, estimation algorithms are often compared to ML performance. Because ML estimation is a highly non-linear problem, any algorithm which has comparable performance is of great interest (Rife and Boorstyn, 1974; Rife and Boorstyn, 1976). The FBLP algorithm as modified by Tufts and Kumaresan effectively achieves the performance limits given by the CR bound for values of SNR below the normal LS threshold. Thus, this method is a viable substitute for the more complex ML estimator.

D. Pisarenko Harmonic Decomposition

Pisarenko Harmonic Decomposition (PHD) was developed by V. F. Pisarenko in 1973. This method models the signal being analyzed as sinusoids with additive white noise. The sinusoids are not necessarily harmonically related as with Fourier analysis. For a high SNR, PHD and MEM give essentially the same results. At lower SNR, PHD has better resolution because its implicit signal model is more appropriate (Frost, 1976).

For an accurate model and with perfect knowledge of a finite number of lags of the autocorrelation function, PHD's two-sinusoid resolution is infinitely fine, regardless of the SNR. For this case, no finer two-sinusoid resolution can be obtained by any other method. PHD has the additional advantage that power measurement accuracy is vastly improved over MEM (Frost, 1976).

The method does have several disadvantages, however, which make it less attractive for array signal processing. First, the autocorrelation function is rarely known exactly, even for a finite number of lags. It must usually be estimated from the data. In addition, the number of sinusoids present in the signal is not usually known a priori. These factors introduce errors and prevent the achievement of perfect resolution (Marple, 1978). Because of this sensitivity to the estimate of the autocorrelation function, PHD is less accurate than MEM for short realizations of a harmonic process (Ulrych and Clayton, 1976). Thus, it would not perform well on arrays with a small number of elements. Finally, like TK, PHD is formulated as an eigenvalue-eigenvector problem, which increases its computational complexity.

In view of these difficulties, PHD will not be considered further in this work.

IV. ARRAY SIMULATION PROGRAMS

This chapter describes the specifications of the antenna array to be simulated, the conditions under which its performance will be evaluated, the algorithms to be studied, and the control programs which implement the simulation. It should make clear how the results and conclusions of the next chapter were obtained.

A. Array Specifications

The computer programs simulate a linear array of isotropic elements 1.5 m long. The element spacing Δz is 0.15 m, which implies a total of eleven elements in the array. With these dimensions, the entire array is one-half wavelength long at 100 MHz. The element spacing is one-half wavelength at 1 GHz. The array's performance was simulated at ten frequencies within these limits, spaced so that the wavelength changes in a fixed ratio at each step.

Frequencies Tested, MHz

1. 100	6. 359
2. 130	7. 464
3. 167	8. 600
4. 215	9. 774
5. 278	10. 1000

Since the resolution of any array varies with angle as well as frequency, the algorithms were evaluated at each frequency for each of

five given look angles θ_0.

Look Angles Tested, degrees

1. 30
2. 37
3. 45
4. 57
5. 90

These angles were chosen because they represent approximately equal changes in the effective aperture of the array as it is steered from one angle to the next.

B. Algorithms Simulated

The information summarized in the preceding chapter provided a reasonably good idea of how well each algorithm would work in the broadband direction finding application. To determine and compare their performance more specifically, several of the algorithms were coded in FORTRAN and run on an HP-1000 minicomputer.

Five algorithms were tested in all. Two of these, the original Burg MEM algorithm as discussed by McDonough (1979), and Prony's method as extended by Hildebrand (1956), were found to give significantly poorer results than the other three in the initial testing. In view of this, they were not considered further.

The other three were more rigorously evaluated. They were the Least-squares (LS) algorithm due to Marple (1980), the Kumaresan-Prony (KP) case, and the more general Tufts and Kumaresan (TK) method, both

due to Tufts and Kumaresan (1982).

This chapter discusses the framework in which the array simulation programs applied each of these algorithms. Two distinct methods were used to compare the results. Each method was useful for the evaluation of different aspects of the direction finding performance. The advantages and disadvantages of each will be discussed in the following sections. The next chapter will discuss the actual results obtained from the simulation.

C. Program ARRAY

1. Overview

Program ARRAY provides a plot of relative power received as a function of angle of incidence with respect to the line of the array. This program gives a good qualitative view of how the selected algorithm performs on a single snapshot of the incident radiation.

2. Program flow

Upon startup, the user is prompted to select which of the algorithms is to be run. All three algorithms share a common input routine. In this routine, up to ten incident plane waves at a single user-selected frequency may be specified. The routine requests the

amplitude, phase, and bearing with respect to the line of the array of each of the incident plane waves. Using this information, the routine computes the resultant complex voltage at each of the eleven elements. A random number generator is then used to add white noise to the real and imaginary parts of the element voltages. These voltages are then stored in a complex vector X for processing. The vector X becomes the input sequence to be analyzed by the selected algorithm.

Control now passes to the algorithm under test. Since the three methods simulated in this study are all of the linear prediction type, the output of each is the vector of prediction coefficients $(a_1, a_2, \cdots, a_m)$. Once the coefficients are computed, the remaining processing is the same for each algorithm.

As we saw earlier, the spectrum of an autoregressive process (Equation 3.1, which is repeated here for convenience) is

$$P(f) = \frac{P_0}{\left| 1 + \sum_{k=1}^{m} a_1 z^{-k} \right|_{z=e^{j2\pi f}}}$$

Since the largest peak in the computed spectrum will be normalized to 0 dB by the output routine, P_0 is set to 1 arbitrarily in the KP and TK algorithms. LS computes a value for P_0, so that value is used, although it too will be normalized later.

The sum in the denominator can be efficiently computed at a finite number of points using an FFT. This approach is used to evaluate the

spectrum at 512 values of wavenumber, the quantity in the spatial domain analogous to frequency. These values are then converted to dB and plotted as a function of bearing θ, according to the relationship

$$\theta = \cos^{-1}\left(\frac{c\,\nu_z}{f}\right)$$

where θ = bearing with respect to the line of the array,

c = velocity of light,

ν_z = component of wavenumber along line of array, and

f = frequency of incoming plane wave.

ARRAY's output, then, is a plot of the relative power received by the array as a function of angle of arrival. This plot is based on a single snapshot, or sample, of the field.

3. Advantages and disadvantages of this format

The plot produced by ARRAY, an example of which is shown in Figure 4.1, is quite easy to interpret. Notice that there are two curves in the figure. The upper, smooth curve is the response of a uniformly weighted phased array scanned across the sources. The lower, more peaked curve is the pattern computed using algorithm LS. The model order M defined in Chapter III (shown on the plots as the variable LA) is six, and the frequency F is 1 GHz. Two plane waves, each of amplitude 100 V/m, with zero relative phase, and at bearings of forty

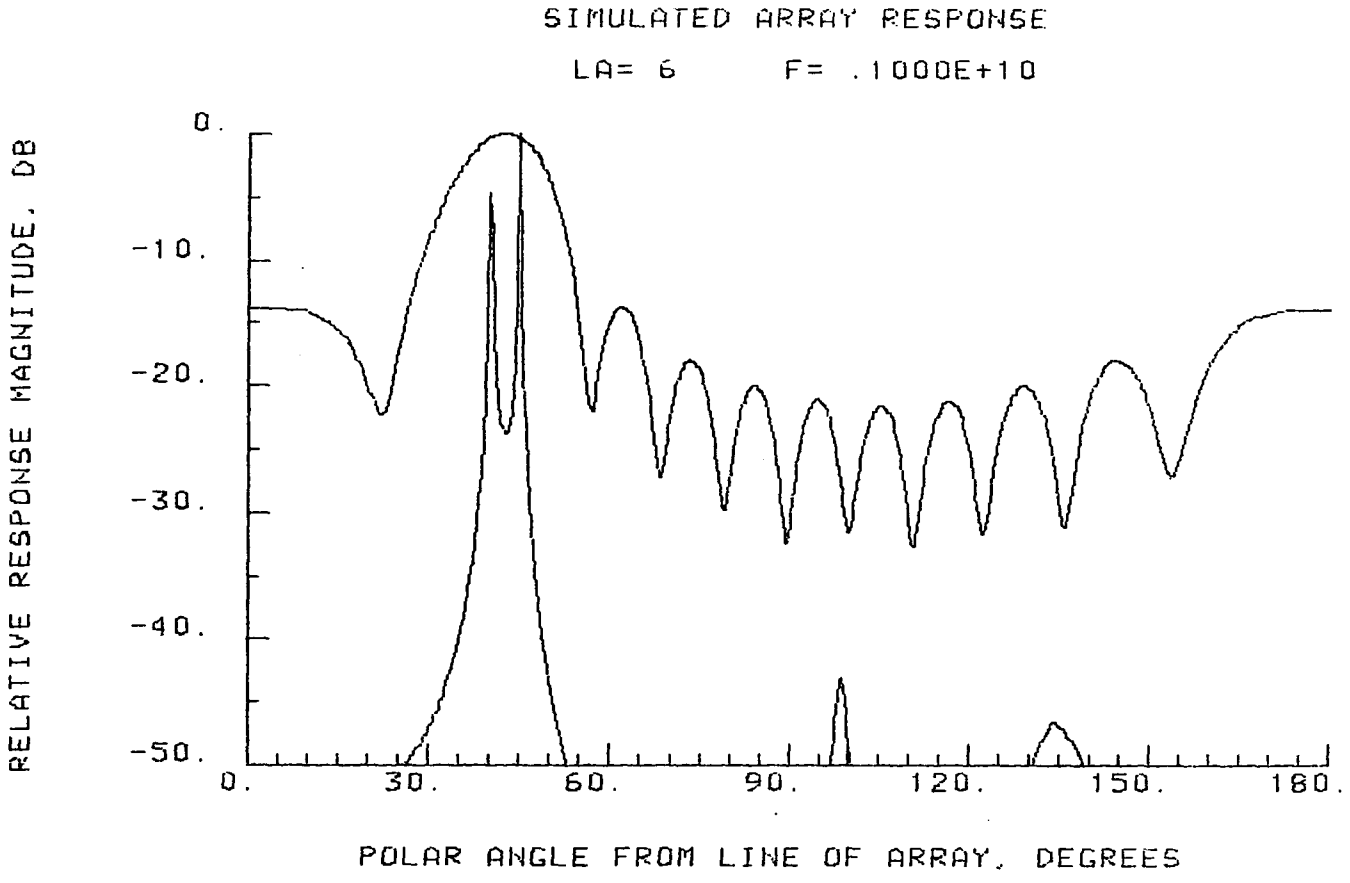


Figure 4.1. ARRAY plot: algorithm LS6, bearings 40°-45°, amplitudes 100 V/m, phase 0°

and forty-five degrees, are striking the array.

The main advantage of this program's format is the ease of interpretation of the output. Relative received power as a function of angle of arrival is a very natural way to look at the problem of resolving sources in angle. It provides the most convenient means of comparing the algorithms' performance to a conventional array pattern. It also allows possible ambiguities to be easily assessed, since they appear in the plots as false peaks.

Some aspects of performance, however, are difficult to evaluate using this format. First, it is difficult to arrive at a suitable criterion for determining the point at which two sources are resolved. There is no first null to determine a specific beamwidth, as there is in a conventional pattern. What is more, since the results are derived from a single snapshot of a field which is corrupted by noise, there can be great variability in the pattern from one snapshot to the next. Two sources which might be perfectly resolved in one snapshot may merge into a single peak in the next. Some other format which takes this variation into account was necessary to arrive at a meaningful evaluation of the resolution of these algorithms.

A second disadvantage of ARRAY stems from the fact that an FFT is used to evaluate the pattern. The number of points evaluated is held constant over the entire frequency band. It turns out that as the frequency of operation is decreased, fewer and fewer of these points correspond to physically real angles. Thus, at low frequencies,

consecutive points on the plot correspond to rather large increments in angle. This makes the plots at low frequencies of operation rather coarse.

ARRAY was used, then, to get a good qualitative idea of the behavior of the various algorithms, and to assess their potential for ambiguity. A second major program, called ZPLOT, was written to get quantitative information concerning two-signal resolution and statistical variability.

D. Program ZPLOT

1. Overview

As discussed earlier, all of the useful spectral information contained in an AR model of a process can be extracted from the locations in the complex plane of the zeros of the prediction error filter transfer function. ZPLOT generates a plot of these zeros for 25 different runs of the algorithm under test. Because of the variability introduced by the noise, the zeros representing a signal do not appear in exactly the same place for each snapshot. Instead, they group together in clusters. The size of the clusters indicates how much variability there is in the angle of arrival and amplitude estimate.

Two signals coming from nearly the same direction would be plotted as two clusters of zeros very close to each other. Resolution, for our

purposes, was defined to be the angle at which the two zero clusters just touch each other. Beyond this point, the separate custers rapidly merge into one. This format allowed a single plot to clearly indicate the variability of the estimate and demonstrate its effect on the two-signal resolution.

2. Program flow

ZPLOT inputs the parameters of the simulation in much the same way as ARRAY. Rather than plotting the spectrum of a single snapshot, however, ZPLOT computes the AR coefficients for twenty-five different realizations of noise added to the element signals. Each set of coefficients represents a slightly different prediction error filter (PEF).

ZPLOT then uses a complex polynomial rooting routine to solve for the zeros of each PEF. These zeros are finally plotted in the complex plane.

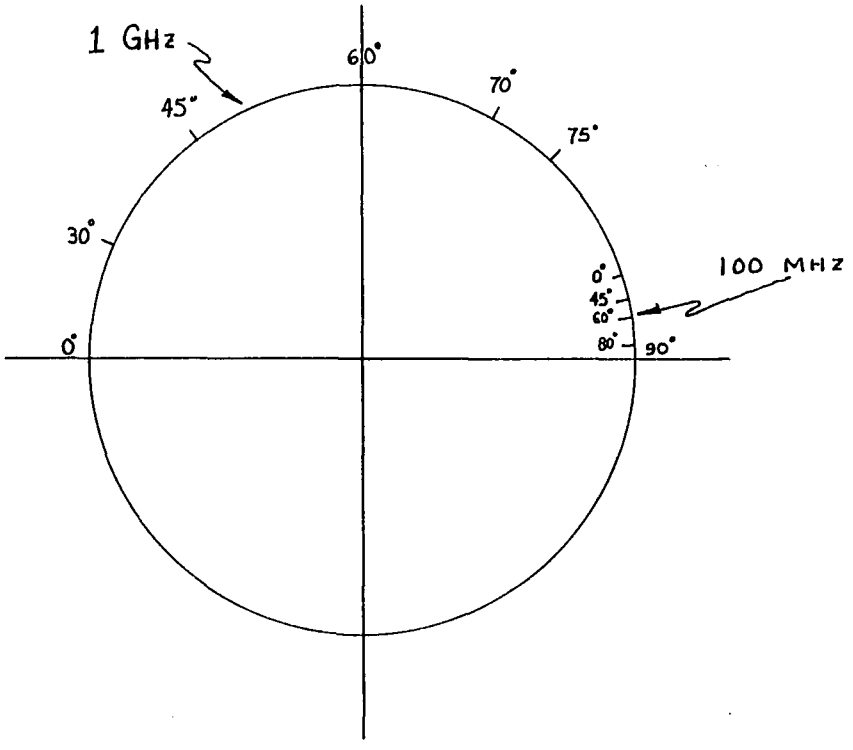
3. Interpretation of the plots

Figure 4.2 shows the zero locations in the complex plane as a function of bearing at 1GHz and at 100 MHz. The expression

$$\Phi = \frac{2\pi}{\lambda} \Delta z \cos \theta, \text{ where}$$

ZERO LOCATION IN Z-PLANE

AS A FUNCTION OF BEARING



$$\Phi = \frac{2\pi}{\lambda} \Delta Z \cos\theta$$

Figure 4.2. Zero location in z-plane as a function of bearing

λ = wavelength of the incoming signals

Δz = interelement spacing

θ = bearing from line of the array,

gives the angle a signal zero makes with respect to the real axis. The closer the zeros lie to the unit circle, the greater the amplitude of the signal peaks they represent.

Figures 4.3 and 4.4 illustrate the output of ZPLOT. Both are runs of the Least-squares algorithm, with model order six. The two incident signals are of amplitude 100 V/m and 0 degrees relative phase. In Figure 4.3, the signal bearings are 70 and 75 degrees. Figure 4.4 shows signals incident from 70 and 72 degrees. Notice the increase in variability, shown by the increase in cluster size, in Figure 4.4. Figure 4.5 shows a case where the signals are no longer resolved. Here, the bearings are 70 and 71 degrees.

The other zeros scattered about the unit disk in Figures 4.3 through 4.5 are placed by the algorithm to most closely account for the noise. The number of zeros of the PEF is equal to the model order. Thus, if there are two signals, and if the model order is six as in these figures, every run will contribute two signal and four noise zeros to the plot. Occasionally, one of these noise zeros will be placed close to the unit circle. This would correspond to a false peak indication in ARRAY. Thus, ZPLOT can also be used to evaluate the ambiguity of the various algorithms.

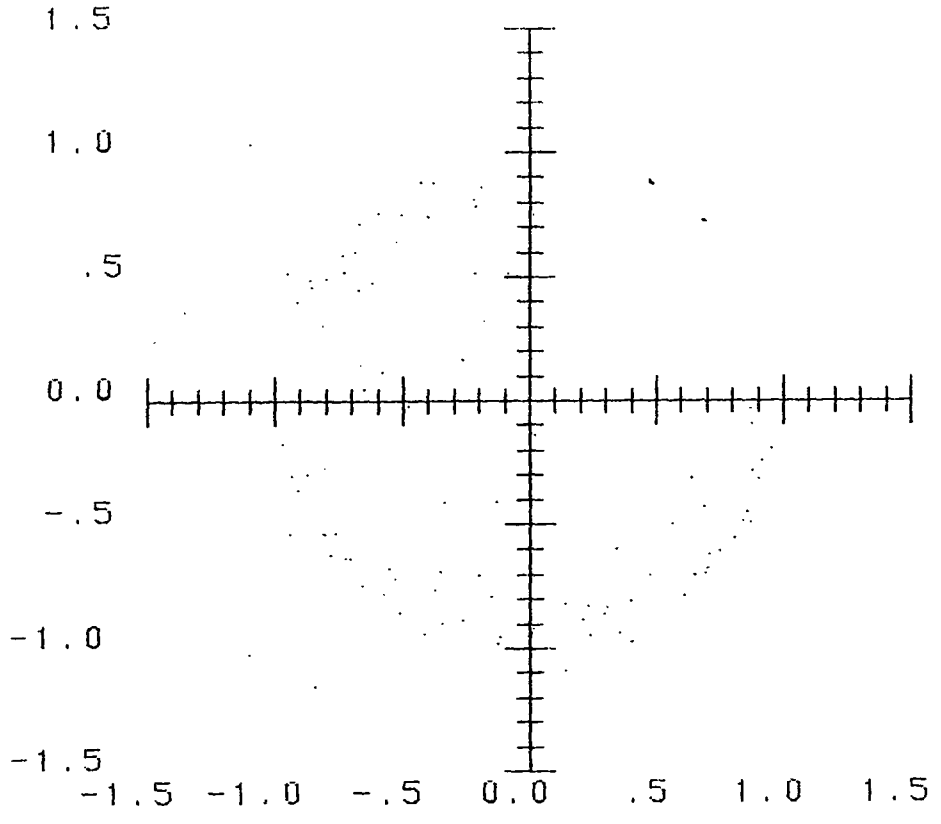


Figure 4.3. ZPLOT plot: algorithm LS6, bearings 70°-75°, amplitudes 100 V/m, phase 0°

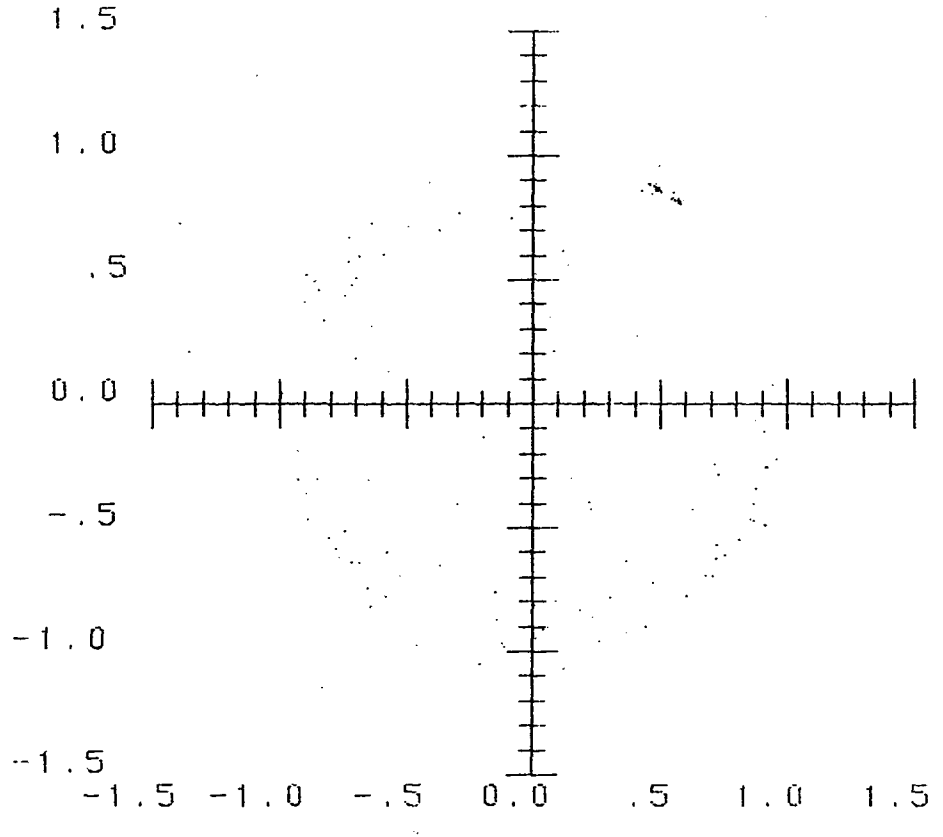


Figure 4.4. ZPLOT plot: algorithm LS6, bearings 70°-72°, amplitudes 100 V/m, phase 0°

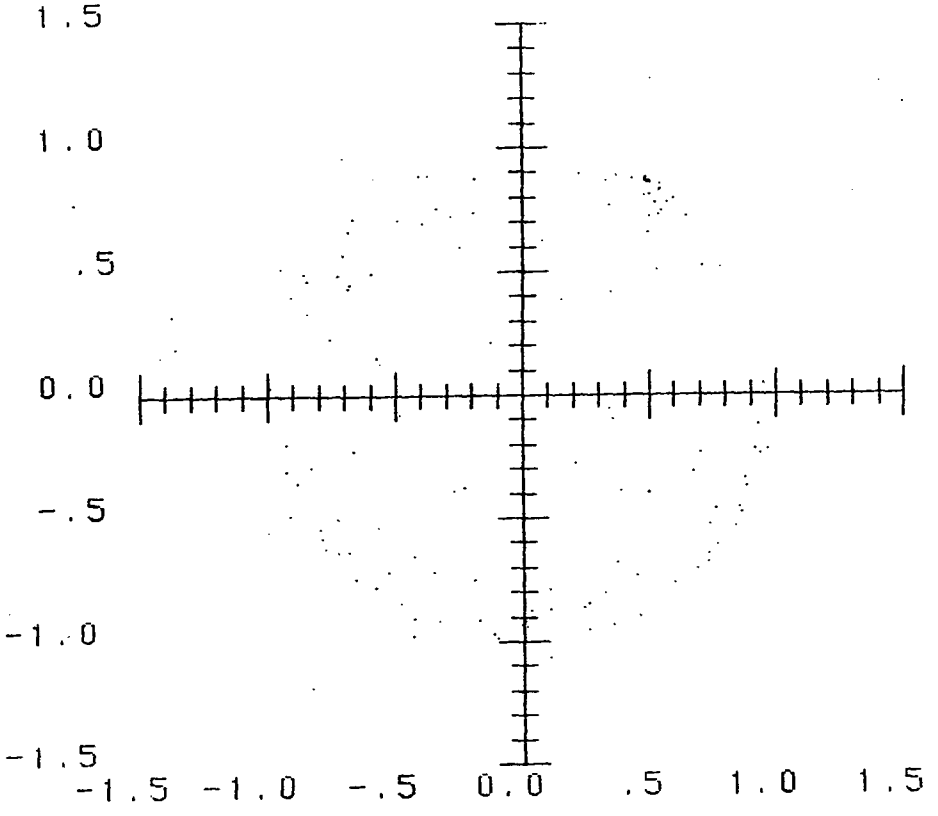


Figure 4.5. ZPLOT plot: algorithm LS6, bearings 70°-71°, amplitudes 100 V/m, phase 0°

E. Summary

ARRAY and ZPLOT were used together to evaluate the broadband direction finding performance of the Least-squares, Kumaresan-Prony, and Tufts and Kumaresan algorithms. Specifically, the smallest angular separation 2Δ was found such that two incoming plane waves were just resolved. The two waves were centered on a given look angle θ_0 so that $\theta_1 = \theta_0 - \Delta$ and $\theta_2 = \theta_0 + \Delta$. Resolution was evaluated at five values of look angle θ_0 ranging from 30° to 90°, at ten frequencies in the range 100 MHz to 1 GHz for each angle. The results of these evaluations along with other observations are discussed in the next chapter.

V. RESULTS AND COMPARISONS

A. Ambiguity Performance

Ambiguity performance is an extremely important criterion in the evaluation of direction finding techniques. No matter how good an algorithm's resolution may be, false or ambiguous indications of signal angle of arrival can render the technique unusable. This is especially true in cases where jamming signals are a part of the signal environment.

As discussed earlier, the techniques studied here have no sidelobes in the usual sense of the word. This fact gives these methods a distinct advantage over conventional beamforming techniques. Some aspects of their behavior, however, can lead to false indications of signal location. In particular, the way in which the algorithm chooses the placement of the extraneous (noise) zeros in the complex plane determines its ambiguity performance. The algorithms evaluated in this study differed greatly in this respect.

The differences can be clearly seen in Figures 5.1 thru 5.3. These figures show the array's response to two 100 V/m plane waves, of 0° relative phase, with bearings of 40 and 45 degrees. Figure 5.1 is KP, and Figure 5.3 is TK8. KP is a special case of TK -- in fact, it is just an efficient method of computing TK10. Thus, the response of the two algorithms is quite similar. But notice the difference between those algorithms and LS6. The resolution of LS appears to be just as

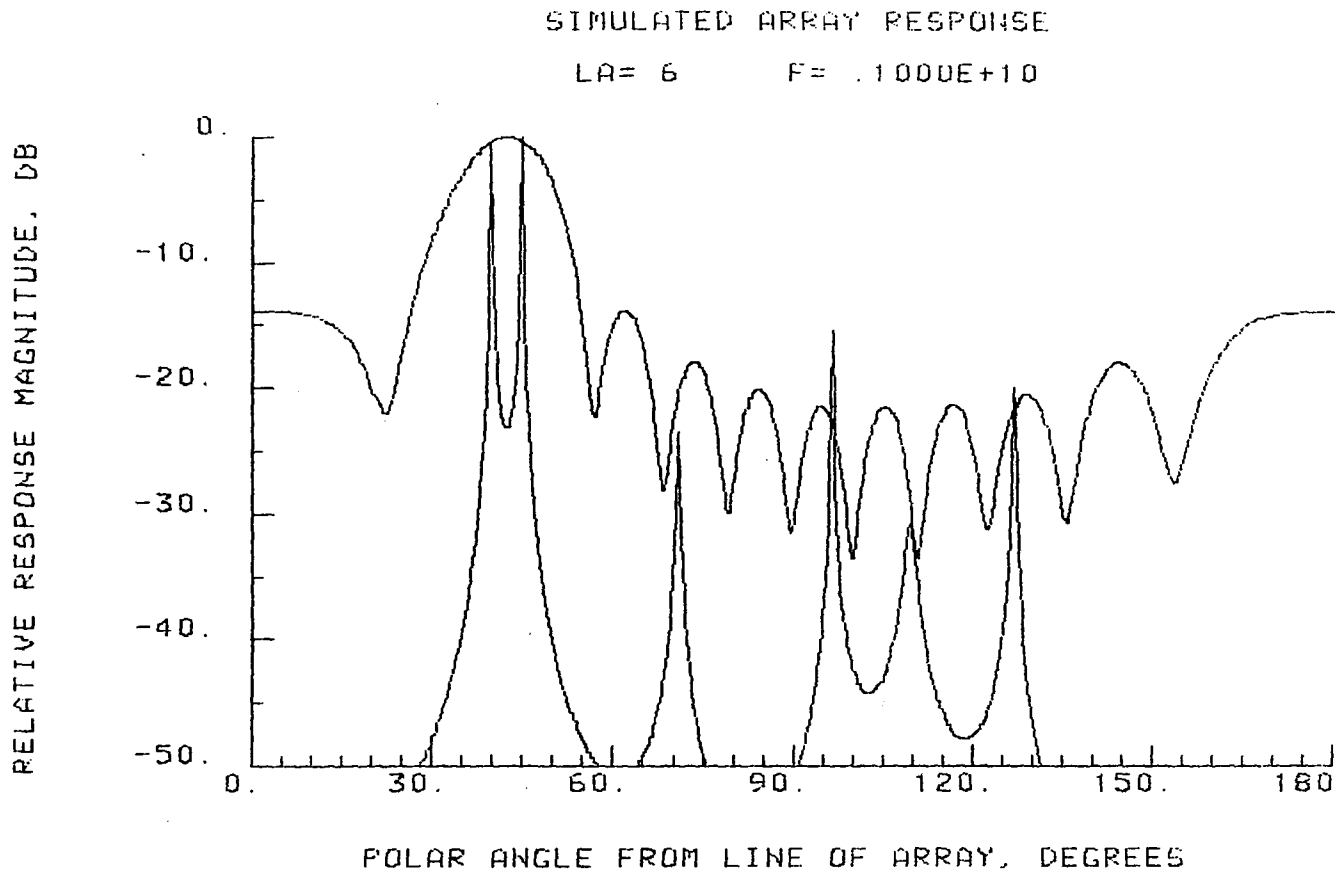


Figure 5.1. ARRAY plot: algorithm LS6, bearings 40°-45°, amplitudes 100 V/m, phase 0°

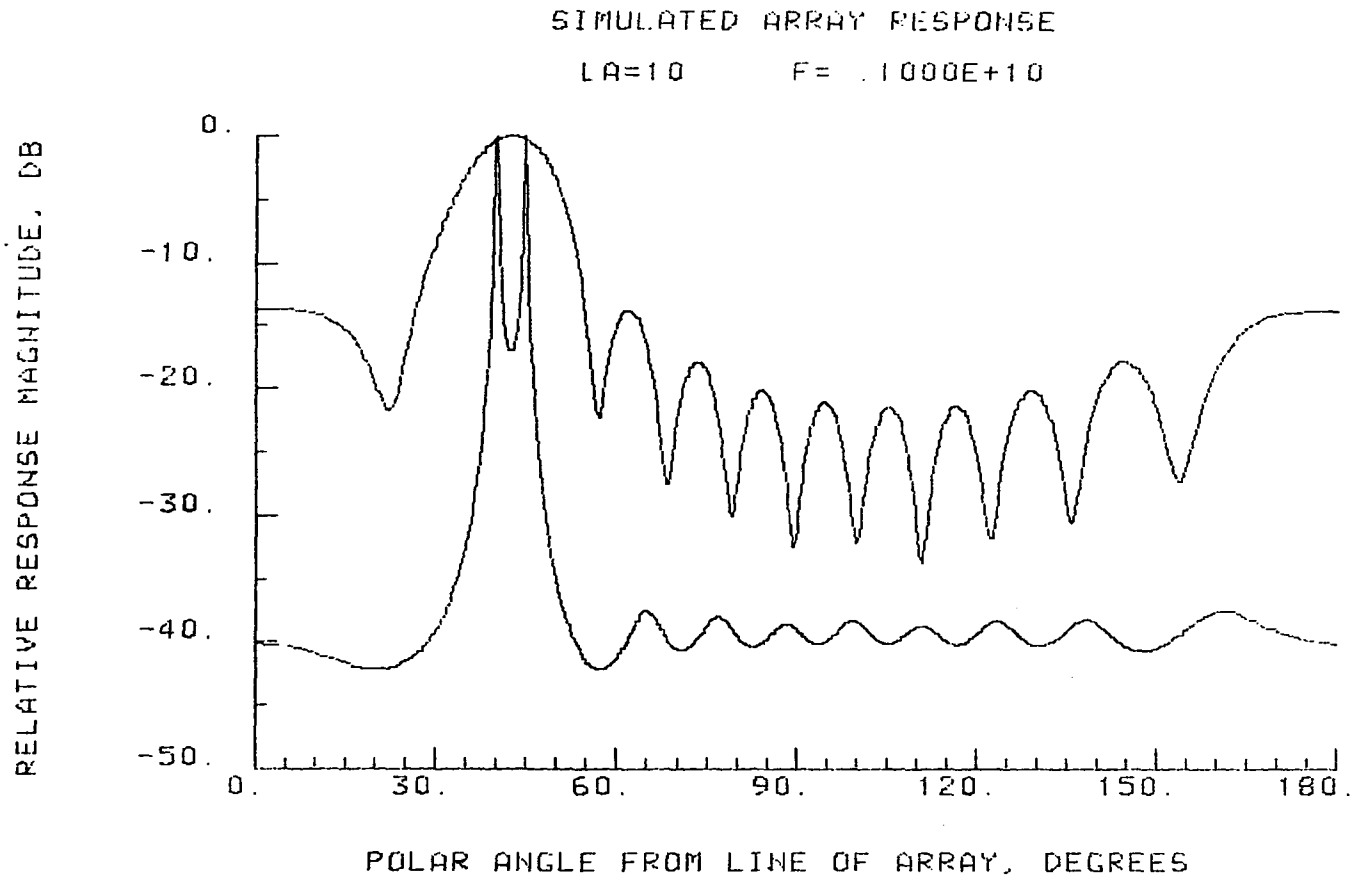


Figure 5.2. ARRAY plot: algorithm KP, bearings 40°-45°, amplitudes 100 V/m, phase 0°

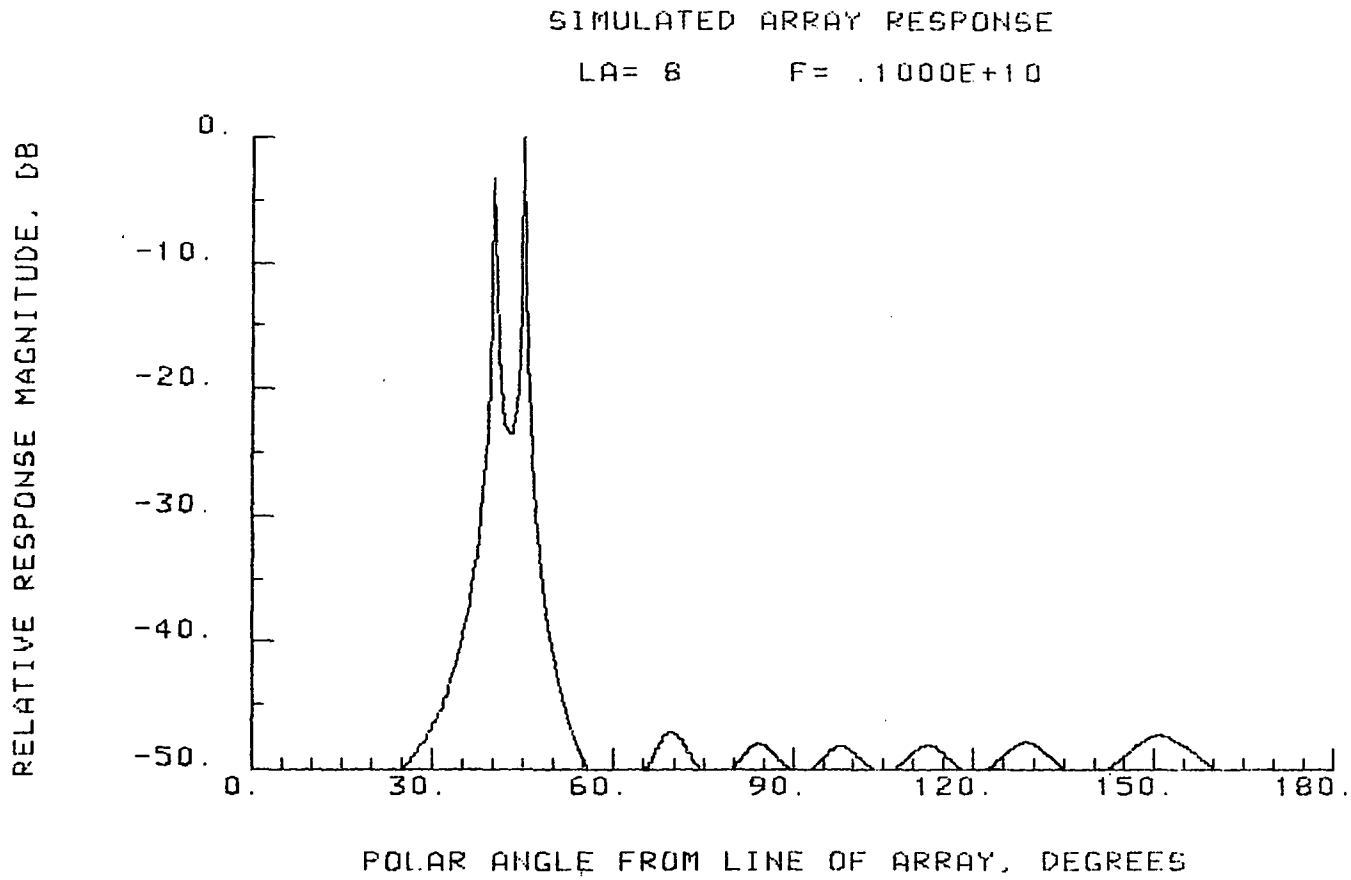


Figure 5.3. ARRAY plot: algorithm TK8, bearings 40°-45°, amplitudes 100 V/m, phase 0°

good as the other two, if not better. In fact, if we were to look at ZPLOT, we would see that the variance of the zero locations is at times actually smaller. The appearance of noise zeros very near the unit circle, however, causes spurious signal indications if their location is within the visible range. The four extra peaks in Figure 5.1 demonstrate this.

KP and TK tend to be much more well-behaved as a result of their placement of extraneous zeros. Comparing a run of ZPLOT for LS6 shown in Figure 5.4, with one for KP, in Figure 5.5, illustrates the difference between the two methods. The evenly spaced noise clusters of TK, which are rather far removed from the unit circle, produce the low level ripple seen in Figure 5.2. The randomly scattered noise zeros of LS, on the other hand, are placed so close to the unit circle that they produce responses comparable in magnitude to the actual signals. Thus, the ambiguity performance of LS is greatly inferior to KP and TK. The problem gets worse as the model order is increased, as discussed by Tufts and Kumaresan (1982).

B. Two-Signal Resolution

1. Effect of the signal-to-noise ratio

One of the greatest weaknesses of high resolution spectrum estimation algorithms is their sensitivity to the signal-to-noise

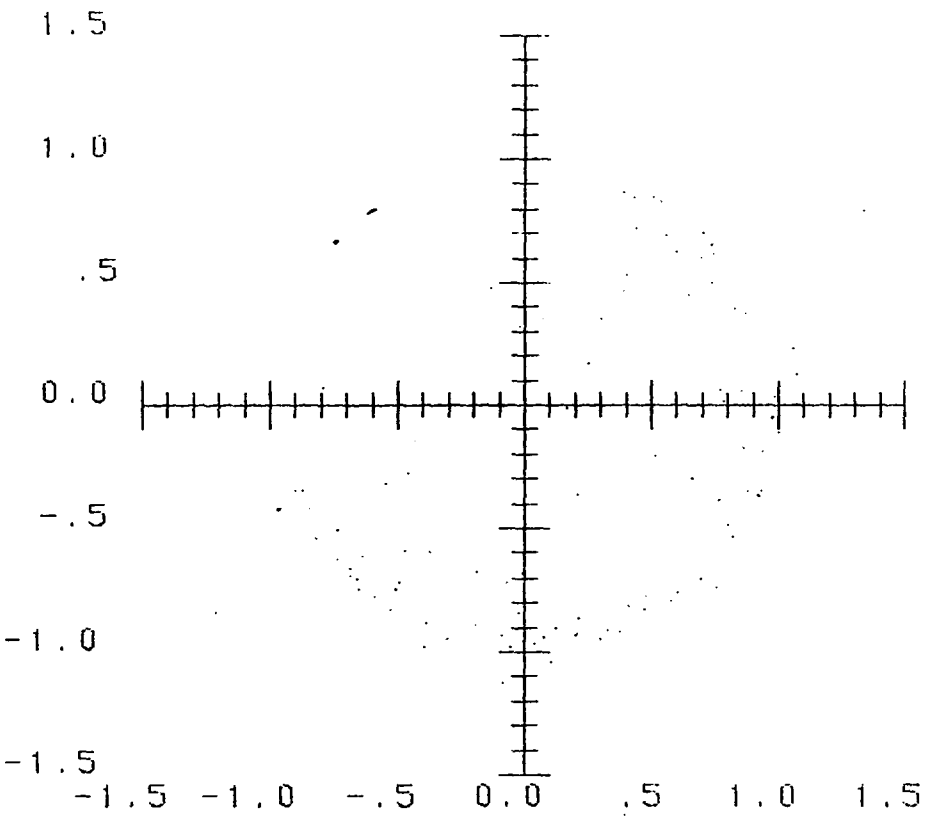


Figure 5.4. ZPLOT plot: algorithm LS6, bearings 40°-45°, amplitudes 100 V/m, phase 0°

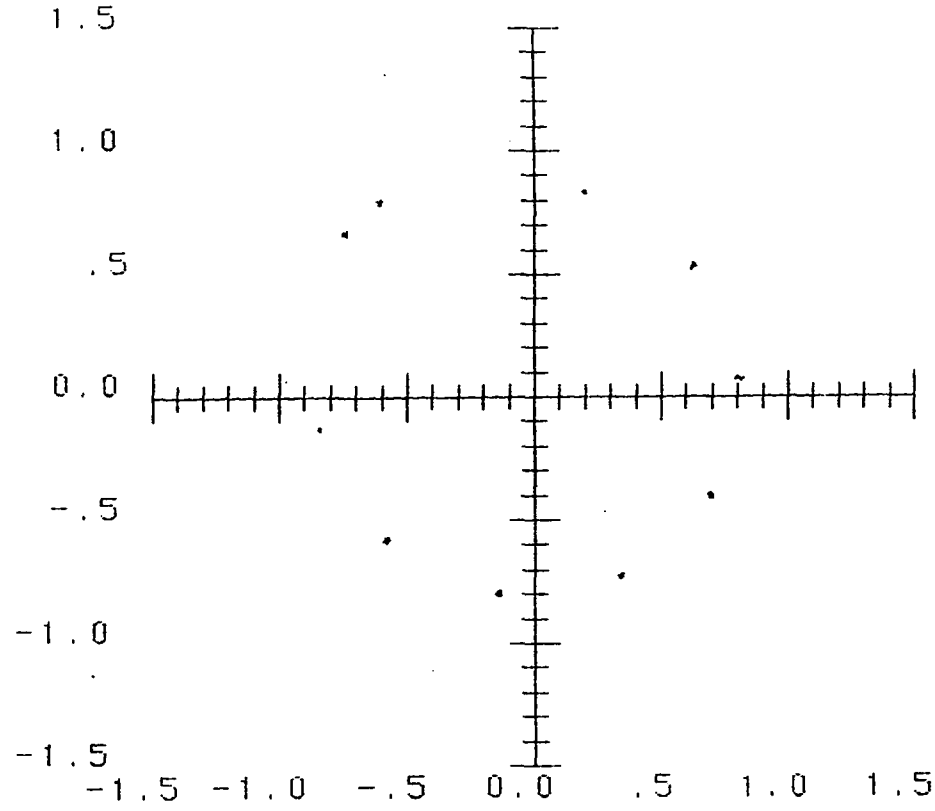


Figure 5.5. ZPLOT plot: algorithm KP, bearings 40°-45°, amplitudes 100 V/m, phase 0°

ratio. As we saw in Chapter III, at low SNR the implicit signal model of most methods no longer applies, and the results become poor. In this study, we are more concerned with obtaining sufficient operating bandwidth than we are with performance at low SNR. For completeness, however, the behavior of the algorithms was evaluated at different values of SNR.

A control program was written to calculate the variance of the spatial frequency estimate $\frac{2\pi}{\lambda} \cos \theta$ for comparison with the Cramer-Rao (CR) bound on the variance of this estimate. Since the angle ϕ that the signal zeros make with the real axis in ZPLOT is $\frac{2\pi}{\lambda} \Delta z \cos \theta$, the variance of $\phi/\Delta z$ was computed. Program flow is much like ZPLOT. An outer loop gradually increases the SNR from 0 to 60 dB. For each execution of the outer loop, the algorithm under test is run twenty-five times at that particular SNR. The variance of $\phi/\Delta z$ is then computed, as well as the CR bound. After the variance is computed for all values of SNR, the results are plotted.

Because the variability of the noise zero location is so large for LS, the simple method used to separate signal zeros from noise zeros did not work. The approach did work well for KP and TK, however, since their zeros are more well-behaved.

Two examples of SNR performance are shown in Figures 5.6 and 5.7. Both illustrate the threshold effect discussed by Tufts and Kumaresan. Both figures correspond to two incident plane waves 1.8 degrees apart, centered on $\theta = 90°$ and at a frequency of 1 GHz. The dashed line in

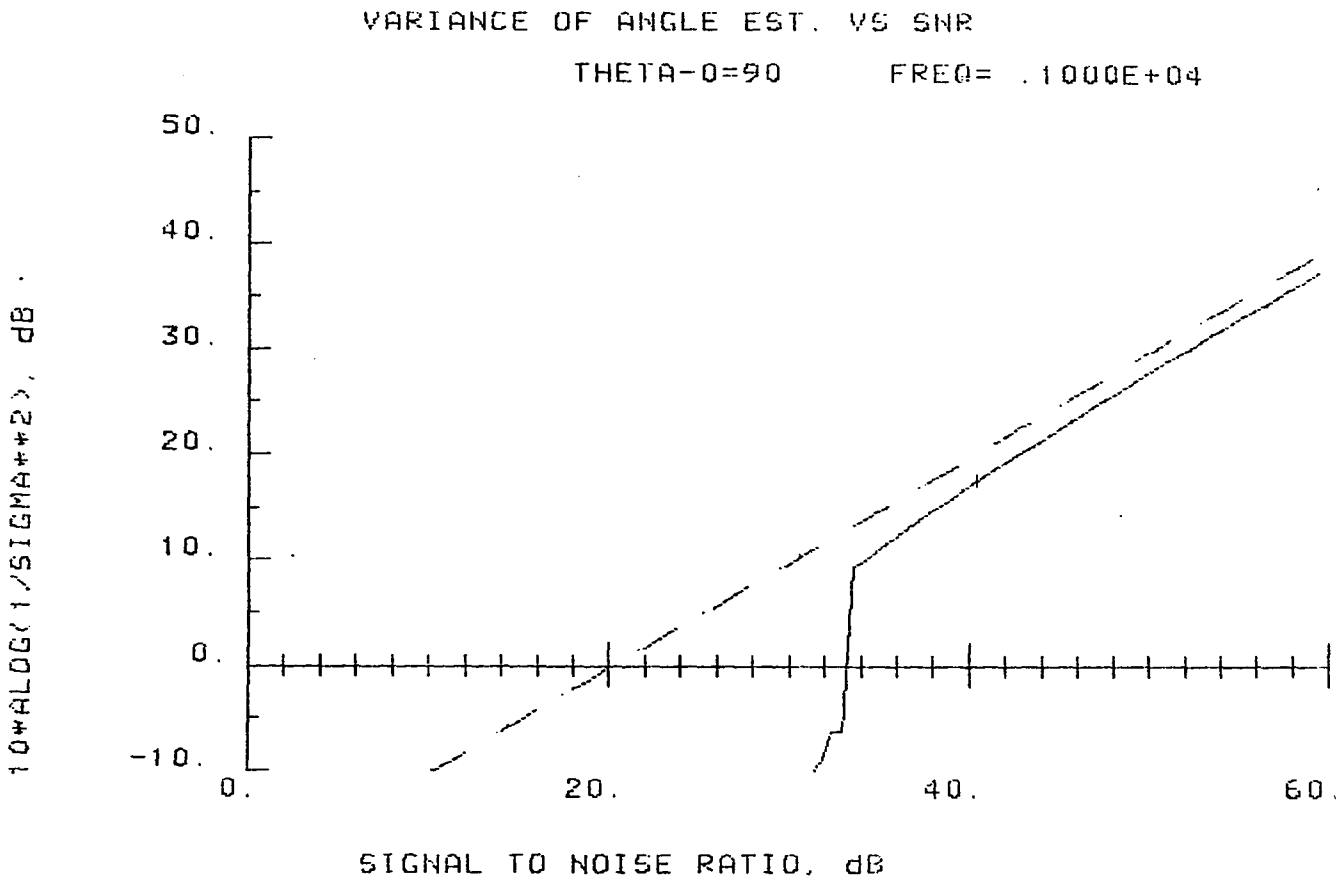


Figure 5.6. Variance of spatial frequency estimate vs. SNR, algorithm KP, $\theta_0 = 90°$, $\Delta = 0.9°$

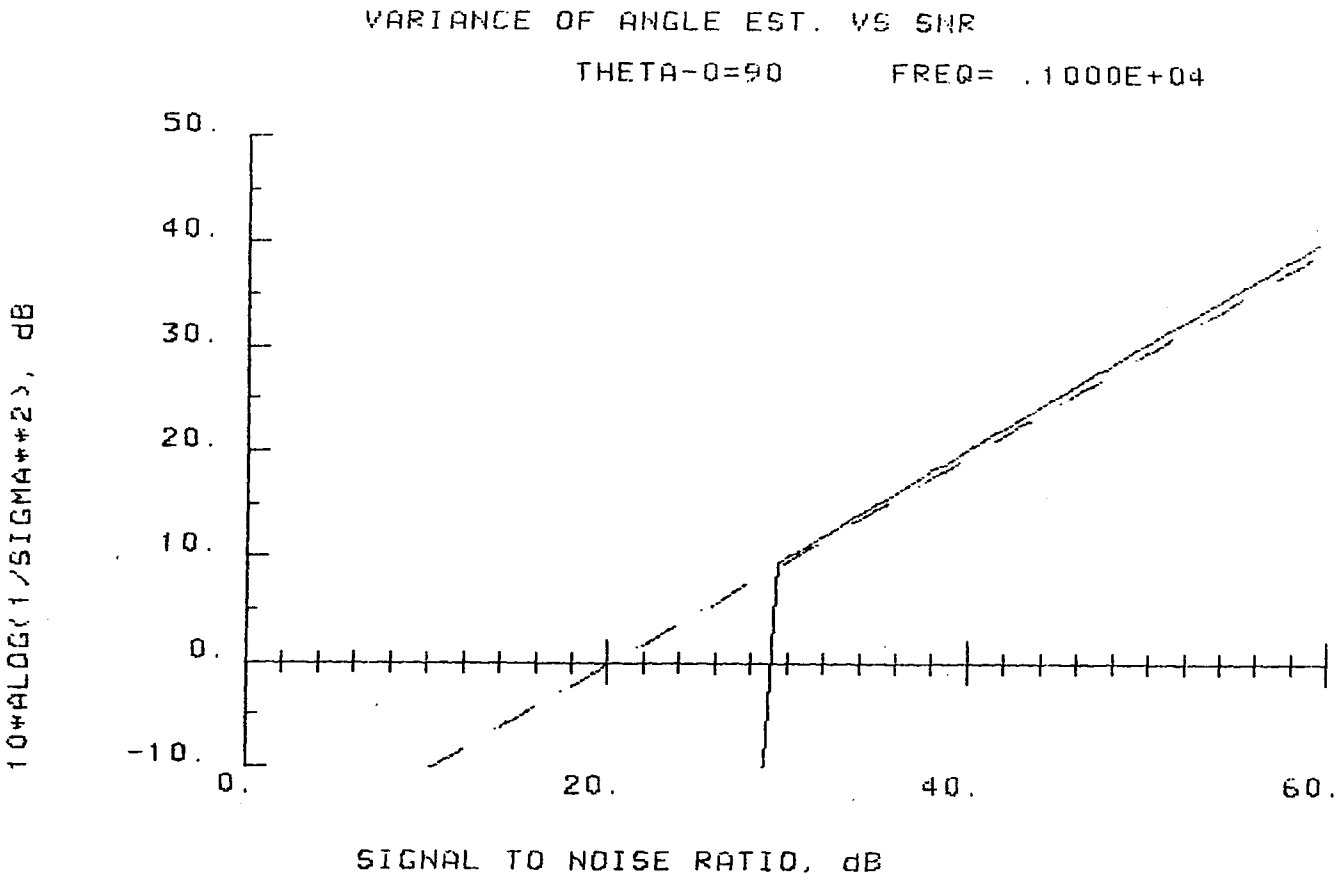


Figure 5.7. Variance of spatial frequency estimate vs. SNR, algorithm TK8, $\theta_0 = 90°$, $\Delta = 0.9°$

each plot is the CR bound. The solid line is the actual performance of the algorithm. Notice that higher values along the ordinate correspond to smaller values of variance. That is, the larger the value of this function, the more consistent the spatial frequency estimate.

Figure 5.6 is a plot of the performance of the KP algorithm. Note that, at high SNR, KP runs roughly parallel to the bound. Its performance is roughly two dB less than the theoretical maximum. At 33 dB SNR, however, a threshold is reached. Below this threshold, the performance deteriorates very rapidly. For perspective, the point where the signal zeros just touch in ZPLOT corresponds on this plot to an ordinate value of 17 dB. This value is achieved for an SNR of 40 dB. Using ZPLOT at signal levels of 40 dB, Δ was determined to be 0.9°. Thus, the results of ZPLOT and the variance vs SNR computation are consistent.

Figure 5.7 shows the results when TK8 is used. This algorithm, for practical purposes, achieves the CR bound above threshold. Its threshold is also lower than KP by approximately 3 dB. The fact that it effectively achieves the CR bound is important. More complicated spectrum estimation algorithms exist, but none of them can do better than the CR bound. This result tells us that TK is "good enough", at least in terms of its variance.

To summarize, these results are consistent with those of Tufts and Kumaresan. TK was found to have the lowest variance over the largest range of SNR values. The optimum model order did turn out to be approximately 3N/4 (in this case, the optimum is eight).

2. Effect of model order

Since the variance of the spatial frequency estimate and the two-signal resolution are intimately related, the model order which gives the lowest variance also gives the best resolution. That statement must be qualified somewhat, in that the definition of resolution we are using in this study differs from the usual.

The usual qualitative definition of resolution is that when there are two signals present, two peaks can be seen in the received power vs. angle plot. This definition does not take into account the variation of the angle estimate from one run to the next. Our definition is more restrictive. Physically, it says that allowable values of the angle estimate lie within a range centered on the actual signal direction. Furthermore, the range of one signal may touch but not intersect the range of the other signal. This is what occurs when the signal zero clusters in ZPLOT just touch.

A high model order will allow two closely spaced peaks to be separated better than a low order. Since a high model order also increases the variance of the estimate and increases the probability of seeing spurious peaks, the optimum model order must be a compromise.

This effect can be seen in the sequence of Figures 5.8 through 5.11. These figures show the results of ZPLOT for 100 V/m plane waves incident from 70° to 75°. The algorithm is TK, and model orders of 4, 6, 8, and 10 are shown. Notice how the variance, that is, the size of

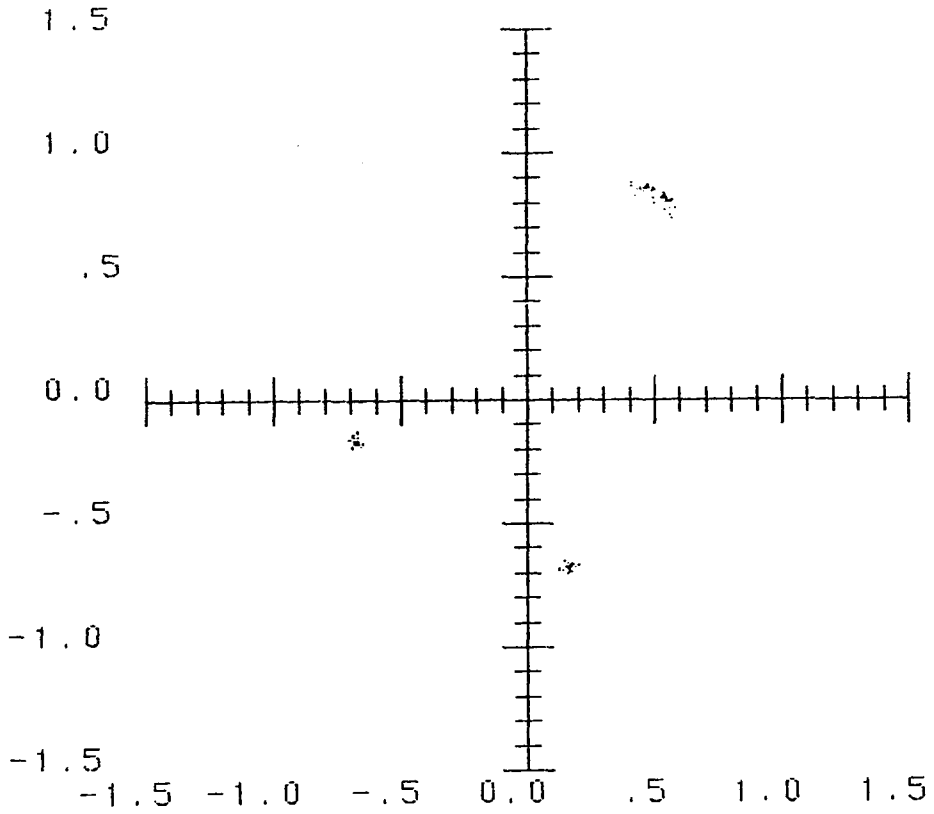


Figure 5.8. ZPLOT plot: algorithm TK4, bearings 70°-72°, amplitudes 100 V/m, phase 0°

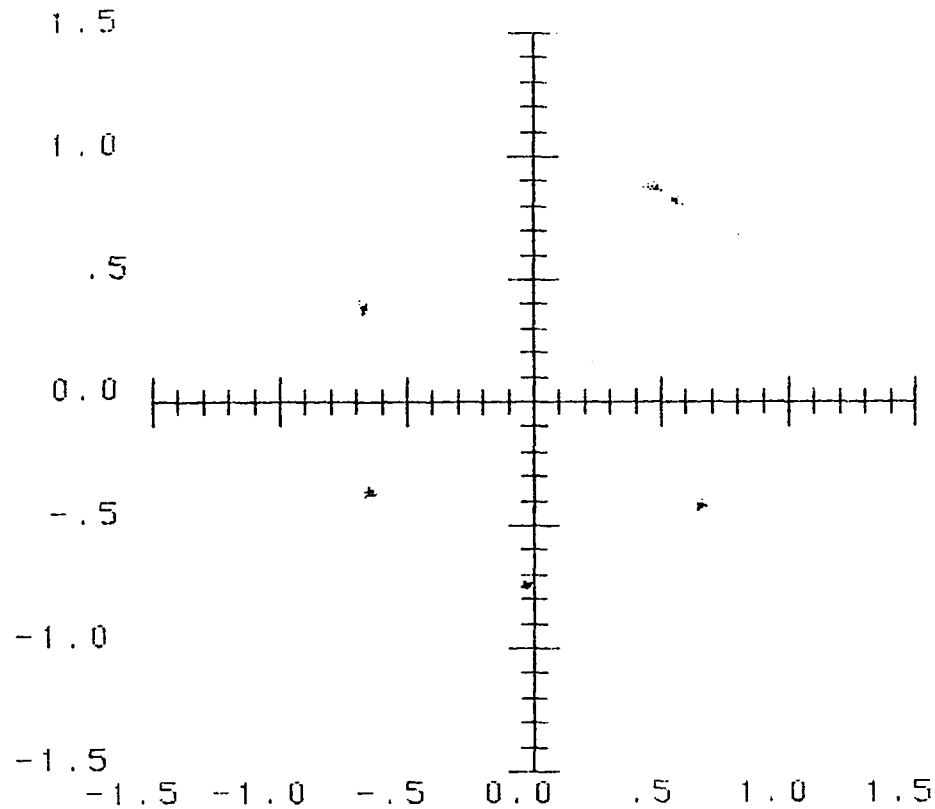


Figure 5.9. ZPLOT plot: algorithm TK6, bearings 70°-72°, amplitudes 100 V/m, phase 0°

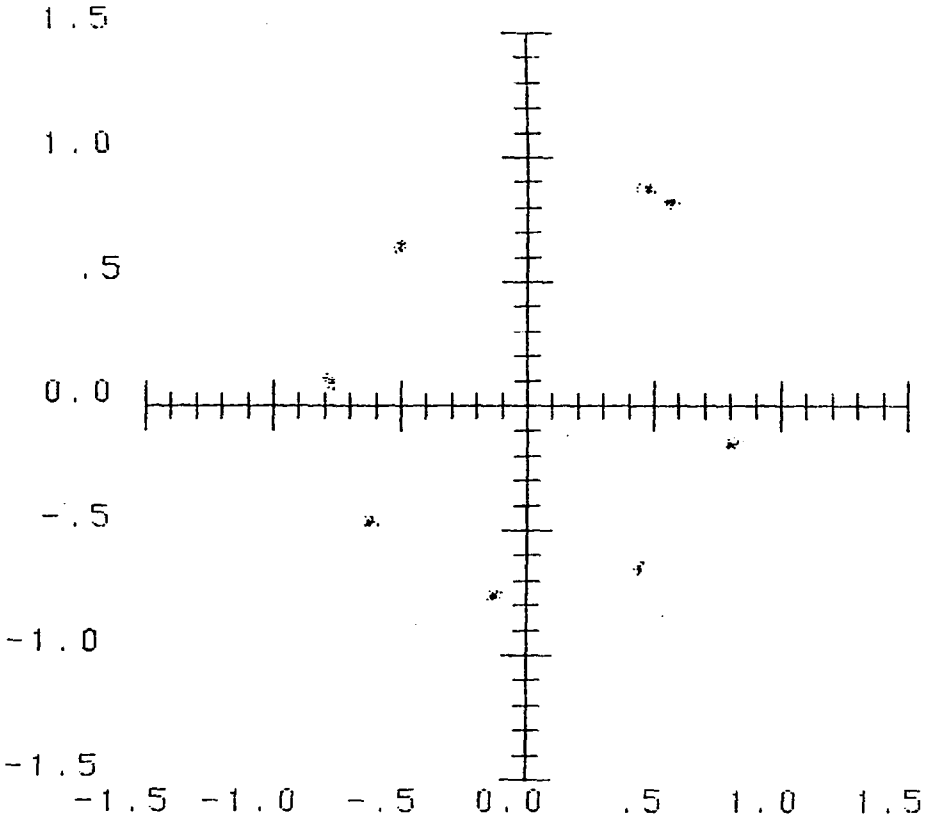


Figure 5.10. ZPLOT plot: algorithm TK8, bearings 70°-72°, amplitudes 100 V/m, phase 0°

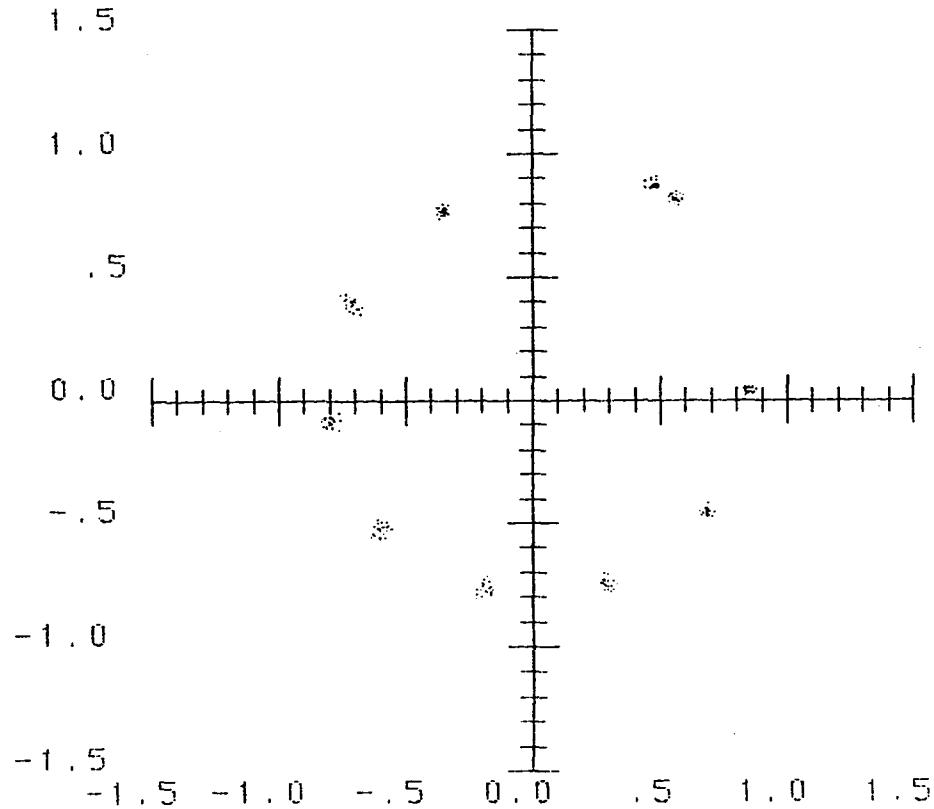


Figure 5.11. ZPLOT plot: algorithm TK10, bearings 70°-72°, amplitudes 100 V/m, phase 0°

the signal zero clusters, decreases through model order eight. Past this optimum value, it begins to increase.

This result holds true for LS as well. The optimum occurs at lower model orders, however, due to the noise effects we discussed in Chapter III.

3. Effect of signal frequency and angle for a fixed length array

The primary thrust of this study has been the evaluation of how well these array direction finding algorithms maintain their resolution over the full range of operating frequencies. Since the effective array length (and consequently the resolution) depend upon the angle of arrival of the signal, the look angle θ_0 must also be considered.

ZPLOT was used to evaluate the resolution of each of the three methods we have compared so far. In looking over the measured resolution at the different frequencies and look angles, a regularity appeared in the ratio of a particular combination of array parameters. This regularity led to the characterization of algorithm resolution for all values of frequency and look angle by a single parameter. This section discusses the reasoning behind this characterization, how it can be used to predict array performance with a given algorithm, and how well it worked in the computer simulation. This characterization is probably the most significant result of this work.

a. Relationship between two-signal resolution and modulation envelope length Rife and Boorstyn, in their paper on digital estimation of the frequencies of multiple tones (1976), found a very interesting fact from their computer calculations. They were computing the CR bounds on the variance of the frequency estimate for two signals. As the two signals were brought closer together in frequency, they found that there was a critical frequency separation below which the CR bounds increase very rapidly. (Remember, the CR bound is a lower bound on the variance of the estimator, so an increasing bound means poorer performance.) This minimum frequency separation was $4\pi/NT$ radians/second, modulo $2\pi/T$. Here, N is the number of samples, and T is the intersample spacing.

It seemed likely that this particular value of difference frequency, $4\pi/NT$, must have some physical significance. It turns out that it does.

Consider the sum of two complex tones at frequencies ω_1 and ω_2, respectively:

$$V(\omega) = \exp(j\omega_1 t) + \exp(j\omega_2 t)$$

Taking a hint from the identity

$$\cos \omega_1 t + \cos \omega_2 t = 2 \cos\left(\frac{\omega_1 + \omega_2}{2}\right) \cos\frac{\omega_1 - \omega_2}{2},$$

write

$$\exp(j\omega_1 t) = \exp\left(j\frac{(\omega_1 + \omega_2)\,t}{2}\right) \exp\left(j\frac{(\omega_1 - \omega_2)t}{2}\right)$$

$$\exp(j\omega_2 t) = \exp\left(j\frac{(\omega_1 + \omega_2)t}{2}\right) \exp\left(j\frac{(\omega_2 - \omega_1)t}{2}\right)$$

Then

$$V(\omega) = \exp\left(j\frac{(\omega_1 + \omega_2)t}{2}\right)\left[\exp\left(j\frac{\omega_d\, t}{2}\right) + \exp\left(-j\frac{\omega_d t}{2}\right)\right]$$

where $\omega_d = \omega_1 - \omega_2$.

$$V(\omega) = 2\cos\left(\frac{\omega_d t}{2}\right) \exp\left(j\frac{(\omega_1 + \omega_2)t}{2}\right)$$

The result is a complex tone at a frequency which is the average of the two tones, amplitude modulated by the cosine of one-half of the difference of the frequencies of the two tones.

What is the period T_d of the modulation envelope?

$$T_d = \frac{2\pi}{\frac{\omega_d}{2}} = \frac{4\pi}{\omega_d}.$$

At the critical frequency separation, $\omega_d = \frac{4\pi}{NT}$. Thus, $T_d = NT$, approximately the total length of the sampling interval. That is, the critical frequency separation occurs when the modulation envelope length

is approximately equal to the length of time spent sampling. When the frequency separation is less, the period of the modulation envelope is longer than the total sampling time. When this occurs, the performance begins to degrade sharply.

These results from the time domain can be directly transferred to the spatial domain and the array direction finding problem. Consider a linear array along the z-axis with two plane waves striking it at angles from the line of the array of θ_1 and θ_2. Let the number of elements be N, the interelement spacing be Δz, the length of the array $(N-1)\Delta z$ be L, and the wavelength of the plane waves be λ.

Then

$$V(z) = \exp\left(j\frac{2\pi}{\lambda} z \cos\theta_1\right) + \exp\left(j\frac{2\pi}{\lambda} z \cos\theta_2\right)$$

Following the same procedure as before,

$$V(z) = 2\cos\left(\frac{\pi}{\lambda} z\,(\cos\theta_1 - \cos\theta_2)\right) \exp\left(j\frac{\pi}{\lambda} z(\cos\omega_1 + \cos\theta_2)\right)$$

Continuing the analogy from the time domain, the critical spatial frequency separation should occur at $\omega_{sd} = \frac{4\pi}{N\Delta Z} \simeq \frac{4\pi}{L}$, where

$$\omega_{sd} = \frac{2\pi}{\lambda}(\cos\theta_1 - \cos\theta_2)$$

The length of the modulation envelope is

$$Z_d = \frac{2\pi}{\frac{\omega_{sd}}{2}} = \frac{4\pi}{\omega_{sd}}.$$

At the critical frequency, the modulation envelope length $Z_d \simeq L$, the array length.

These arguments suggest that the ratio of the physical length of the array to the length of the spatial modulation envelope caused by the interference of the two incident plane waves could be an important quantity.

Let this ratio be represented by the function

$$K(\theta_1, \theta_2, \lambda) = \frac{L}{Z_d} = \frac{L(\cos\theta_1 - \cos\theta_2)}{2\lambda}$$

When ZPLOT was used to determine the resolution at a given look angel θ_0, the program was run with $\theta_1 = \theta_0 - \Delta$ and $\theta_2 = \theta_0 + \Delta$ for various values of the angle Δ. The smallest Δ was found such that signals at θ_1 and θ_2 could be resolved. Substituting these relations for θ_1 and θ_2 into $K(\theta_1,\theta_2,\lambda)$ gives

$$K(\theta_0,\Delta,\lambda) = \frac{L}{\lambda} \sin\theta_0 \sin\Delta. \qquad (5.1)$$

Equation 5.1 was evaluated at the point of resolution for each look angle and each frequency of each algorithm. For each method, $K(\theta_0,\Delta,\lambda)$ turned out to be very nearly constant. Consider the results for the KP algorithm with $\theta_0 = 37°$:

Freq (MHz)	MOD.ENV.(Wavelengths)	ARRAY LENGTH(Wavelengths)	RATIO K(%)
100	7.39	0.500	6.77
130	9.57	0.650	6.79
167	10.6	0.835	7.86
215	15.9	1.08	6.76
278	19.1	1.39	7.29
359	23.8	1.80	7.54
464	31.7	2.32	7.31
600	47.6	3.00	6.30
774	54.1	3.87	7.11
1000	79.3	5.00	6.30

Average K = 7.00%

Standard deviation = 0.511%

Since Δ was determined only to the nearest degree for all but the two highest frequencies, the relative consistency of the values of K was considered to be significant. Knowledge of the value of K for a particular algorithm allows us to predict Δ for a given look angle and frequency. This, in turn, can be fed back into the simulation to determine whether the conclusion that K is a constant is valid. This, as it turns out, will help us to explain why the measured value of K varies as much as it does.

b. Prediction of resolution Using Equation 5.1, Δ can quite easily be predicted from the look angle, frequency, array length, and K. Solving for Δ in 5.1:

$$\Delta = \sin^{-1}\left[\frac{K\lambda}{L \sin \theta_0}\right] \qquad (5.2)$$

Given an appropriate value of K, the resolution 2Δ can be computed for any frequency and look angle. This K will depend upon the algorithm used, the array (possibly the number of elements and element spacing as well as just the length L), and the criterion used to determine when two signals are resolved.

Figure 5.12 shows a plot of Δ versus frequency for algorithm KP. Here, θ_0 is 37°. The plus signs indicate the measured values, and the solid curve is a graph of Equation 5.2. K is taken to be 0.07. This plot is typical; the measured values lie both above and below the predicted curve by a small amount, but they follow the functional form quite well. Those points which differ by a significant amount, such as the point at F = 167 MHz in the figure, are not due simply to measurement error. The simulation was repeated for these points, and the same results were obtained. A possible explanation for these anomalous points will be explored in Section 3.d.

c. An array theory approach to delta An expression for Δ of the same form as Equation 5.2 can be derived by a completely different approach using array theory. Recall from Chapter I that the output of a linear array can be written in the form

$$AF(\theta) = \sum_{n=0}^{N-1} a_n \exp(jn\beta d\cos\theta + \alpha)$$

where a_n = the element weighting coefficients

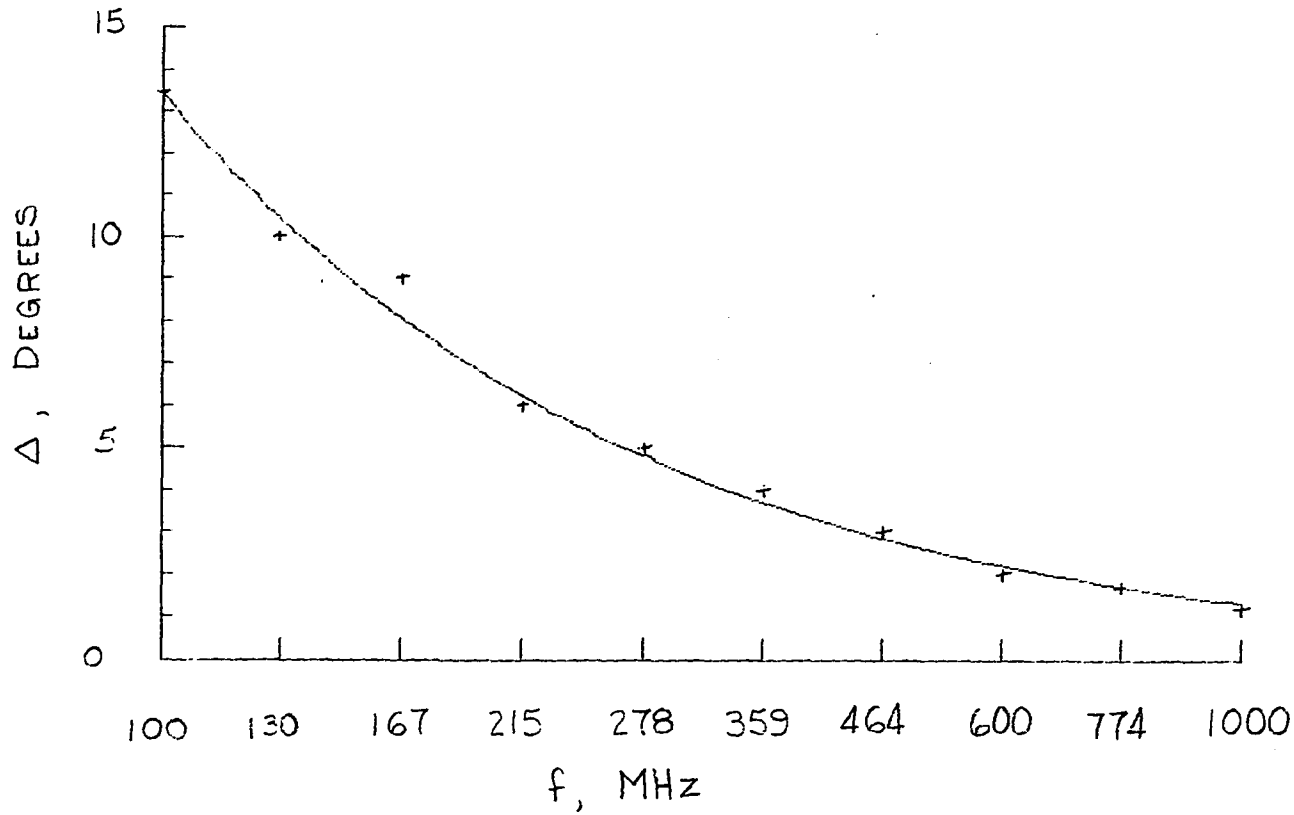


Figure 5.12. Measured and predicted Δ vs. frequency, algorithm KP, θ_0 = 37°, K = 0.07

β = $2\pi/\lambda$

d = interelement spacing (Δz, in our case)

θ = angle with respect to line of array

α = element to element phase shift

If θ_m is the position of the desired maximum, α should be set equal to $-\beta \cos \theta_m$. Setting the array weights all equal,

$$AF(\theta) = A_0 \frac{\sin \frac{1}{2} N (\beta d(\cos \theta - \cos\theta_m)}{\sin \frac{1}{2} (\beta d(\cos \theta - \cos \theta_m)}$$

Suppose we use the Rayleigh criterion for resolution, i.e., the sources are resolved when they are separated by 1/2 BWFN. Let the array be pointed in direction θ_2. Then the array factor at θ_1 should be zero:

$$A_0 \frac{\sin \frac{1}{2} N (\beta d(\cos\theta_1 - \cos \theta_2))}{\sin \frac{1}{2} (\beta d(\cos \theta_1 - \cos \theta_2))} = 0$$

This implies

$\sin \frac{1}{2} N (\beta d (\cos \theta_1 - \cos \theta_2)) = 0$, which requires

$$\frac{N}{2} \beta d(\cos \theta_1 - \cos \theta_2) = \pm m\pi \quad m = 0,1,2,3,\ldots$$

In this case, m is one. Since

$$\cos \theta_1 - \cos \theta_2 = 2 \sin \theta_0 \sin \Delta,$$

$$N\beta d \sin \theta_0 \sin \Delta = \pi$$

$$\sin \Delta = \frac{\pi}{N\beta d \sin \theta_0}$$

$$\Delta = \sin^{-1} \left[\frac{\pi}{(Nd)\left(\frac{2\pi}{\lambda}\right) \sin \theta_0}\right]$$

$$\simeq \sin^{-1} \left[\frac{\left(\frac{1}{2}\right) \lambda}{L \sin \theta_0}\right] \qquad (5.3)$$

Equation 5.3 is of exactly the same form as Equation 5.2, even though the approaches used to derive each were completely different.

d. Variance as a function of frequency at resolution The resolvability criterion we have established effectively states that two signals are resolved when the variance of their signal zeros is below a given threshold. The threshold itself corresponds to a fixed value of variance, regardless of which algorithm or set of input conditions is being used. Thus, the expression for delta given in Equation 5.2 may be verified by sweeping over the operating frequency range of the array, computing delta at each frequency, and simulating the array's performance at this value of delta. If the variance is a constant as the frequency is changed, we have a good indication that Equation 5.2 gives the correct functional dependence of the resolution on the input parameters.

This check on Equation 5.2 was implemented, first for the CR bounds at the predicted value of delta, and then for the actual algorithms. The CR bound result is shown in Figure 5.13. Signal amplitudes equal 100 V/m, with 0° relative phase. The look angle for this run is 90°, and K is 0.07. Notice that the bound on the variance is indeed a constant, at a level of about 18 dB. This level would vary with different amplitude signals or a different value of K, but the variance is still a constant with frequency. This result holds true for the other look angles as well. Equation 5.2 predicts the proper value of delta to just meet the criterion for resolvability.

Figure 5.14 shows the actual variance for algorithm TK8. In this case, all of the input parameters are the same as for Figure 5.13. With this look angle (θ_0 = 90°), the results are just as predicted. The actual variance is a constant for the predicted value of delta as the frequency sweeps over the operating range. It appears that Equation 5.2 can predict the resolution attainable for a given algorithm from the input conditions.

Things are not quite so simple, however, at values of look angle other than 90°. Figure 5.15 shows TK8's performance at a look angle of 57°. The variance, although relatively constant and with no particular trends, does wander over about a 4 dB range with no immediately apparent pattern. The results are repeatable, however. This means that the variations are due to characteristics of the algorithm itself, not just random fluctuations.

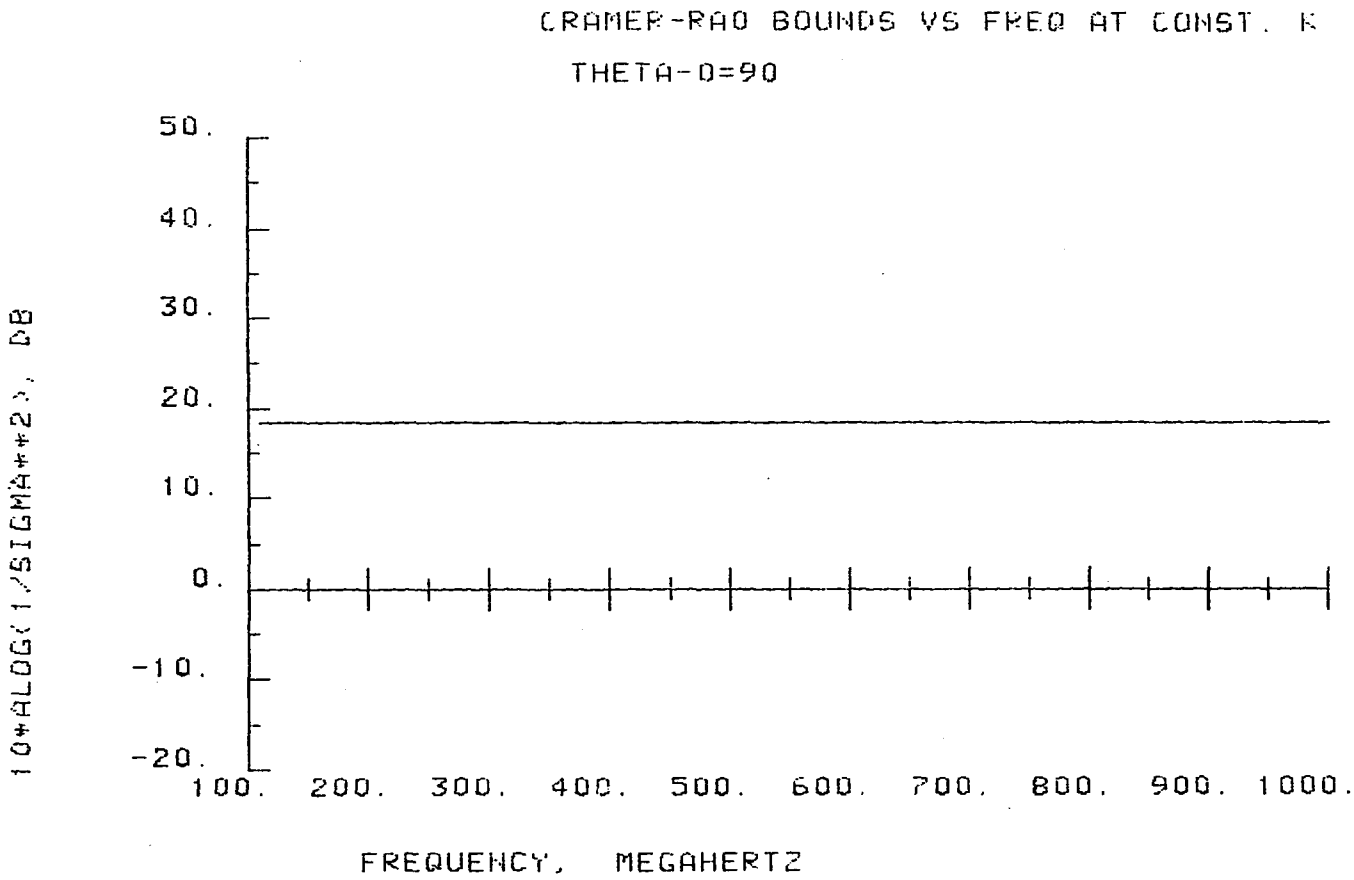


Figure 5.13. Cramer-Rao bounds vs. frequency, $\theta_0 = 90°$, K = 0.07

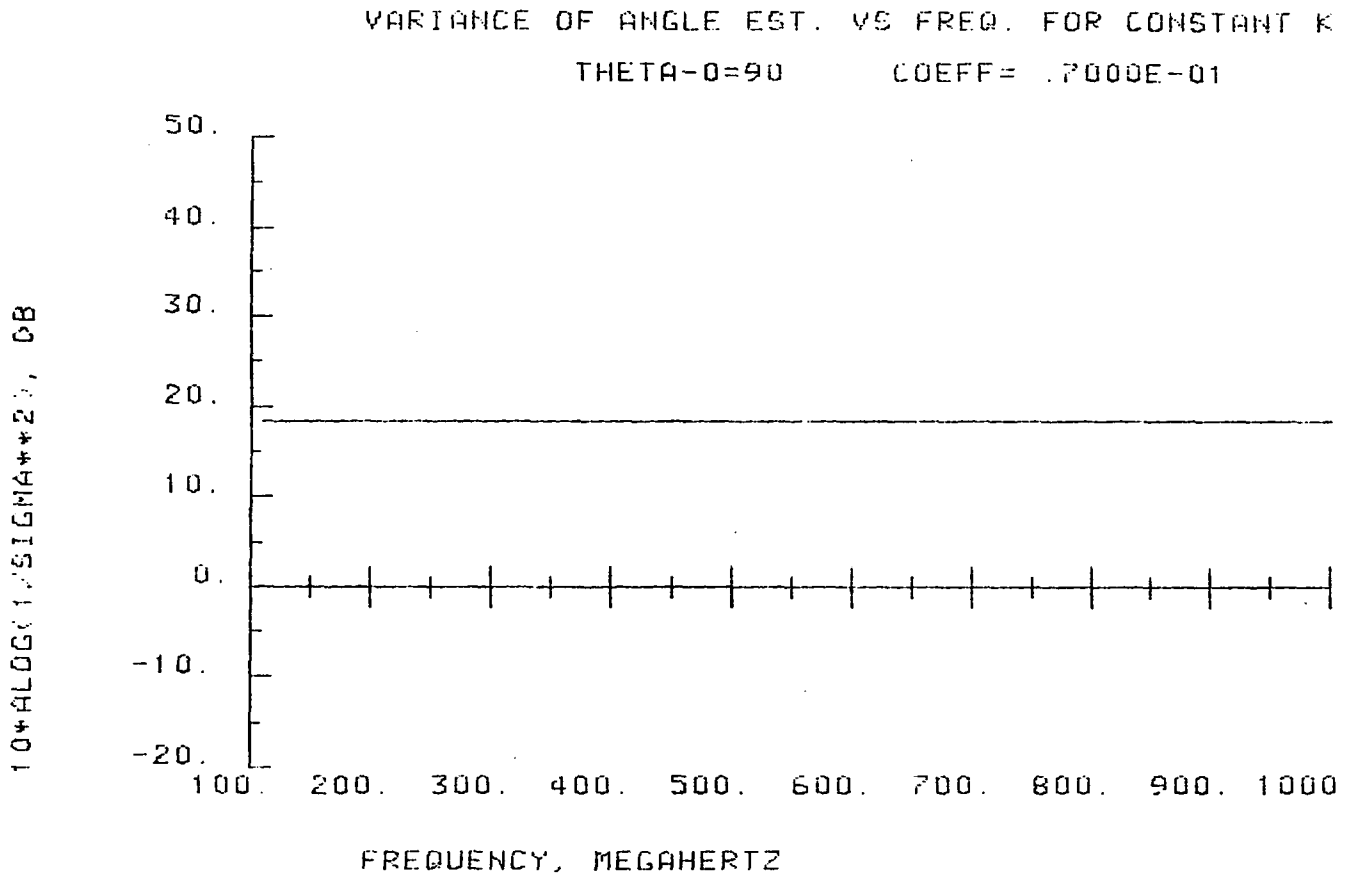


Figure 5.14. Variance vs. frequency at predicted value of Δ, algorithm TK8, $\theta_0 = 90°$, K = 0.07

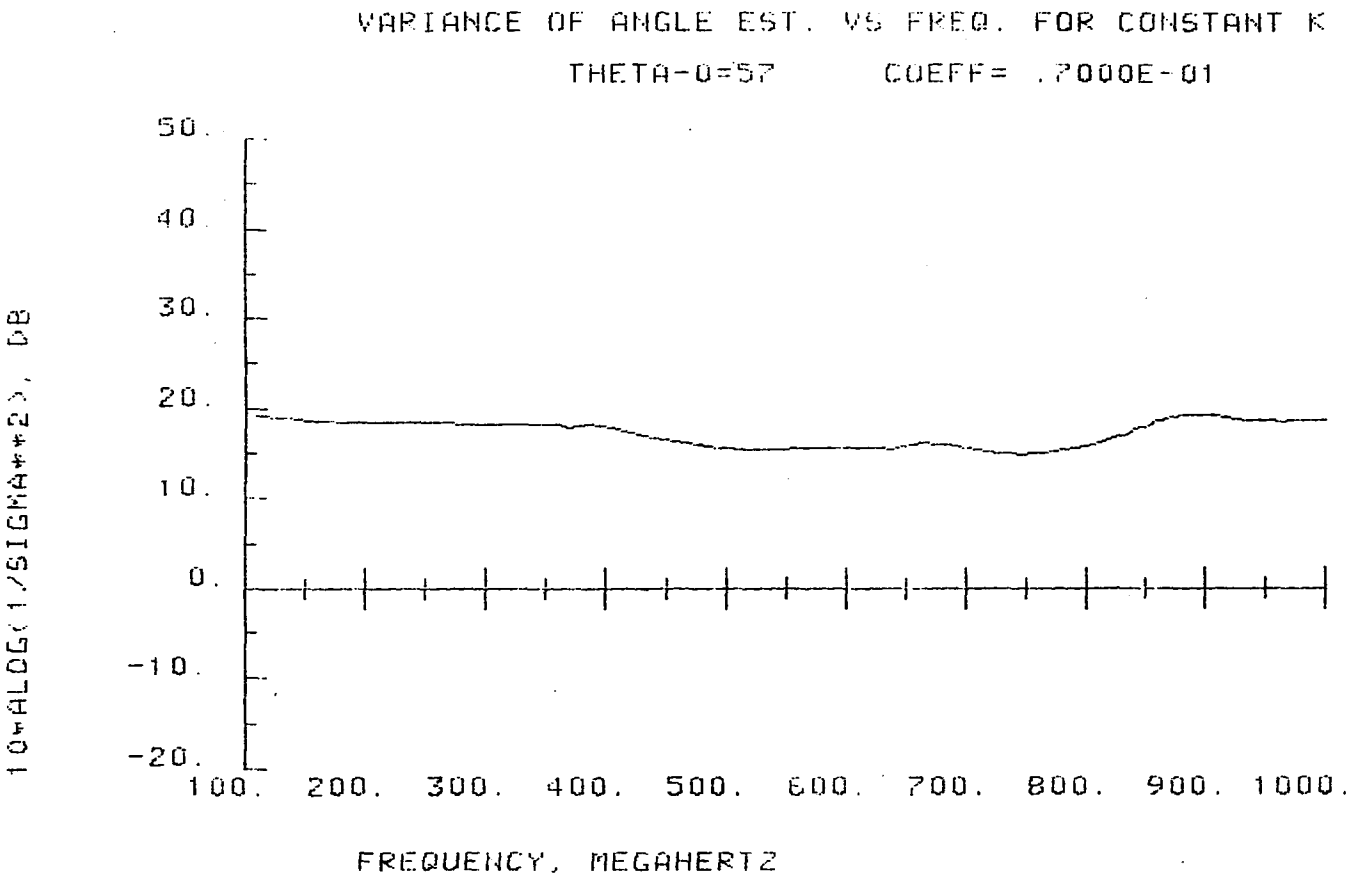


Figure 5.15. Variance vs. frequency at predicted value of Δ, algorithm TK8, θ_0 = 57°, K = 0.07

A possible explanation for this behavior lies in the sensitivity of the algorithms to initial phase of the sinusoids. At a look angle of 90°, the two signals are arriving from angles 90° - Δ and 90° + Δ. The relative phase at the center of the array, and thus, the array's location within the modulation envelope, remain the same as Δ changes. This is not true for other look angles. Initial phase effects can enter in and degrade the performance. This may be the reason for the anomalous points in Figure 5.12.

This explanation is supported by Figures 5.16 through 5.18, which show the actual variance of the TK8 algorithm for look angles of 85, 80, and 75 degrees, respectively. If the variation in performance were due to phase effects, the problem would increase with decreasing angle of incidence. As shown in the figures, this is what actually occurs.

e. Variance as a function of angular separation of sources The performance of the algorithms as the angular separation of the sources changes is another example of the influence of the position along the modulation envelope. To test this aspect of the algorithms' performance, a control program was written to vary delta from zero to some maximum value. Both the CR bound and the actual variance were computed for each delta. The frequency of operation and the look angle are fixed.

This program produced some interesting results. For example, consider Figure 5.19, which shows TK10 at 1 GHz. In this and the following figures, there are two sources present, each with amplitude

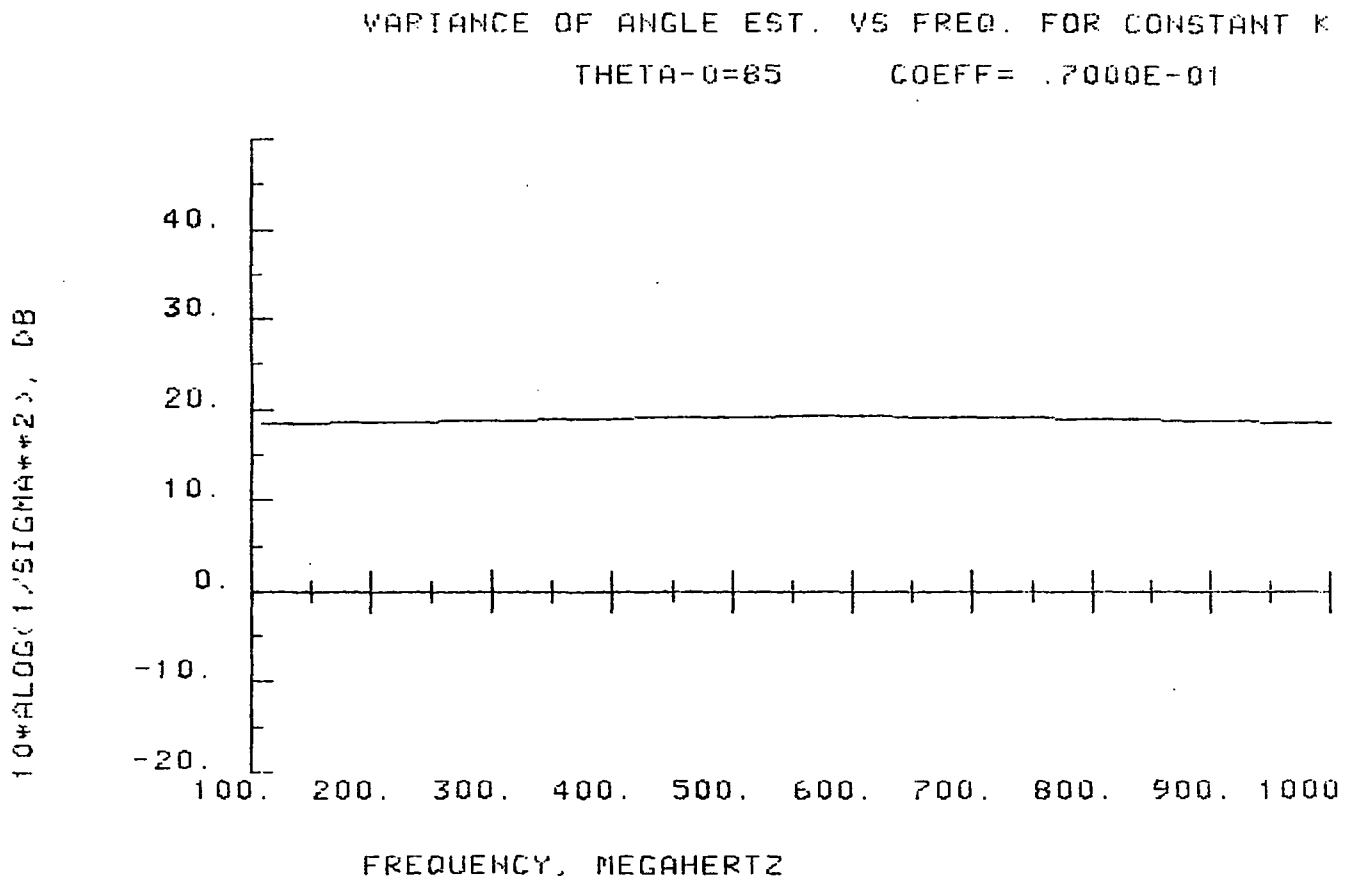


Figure 5.16. Variance vs. frequency at predicted value of Δ, algorithm TK8, θ_0 = 85°, K = 0.07

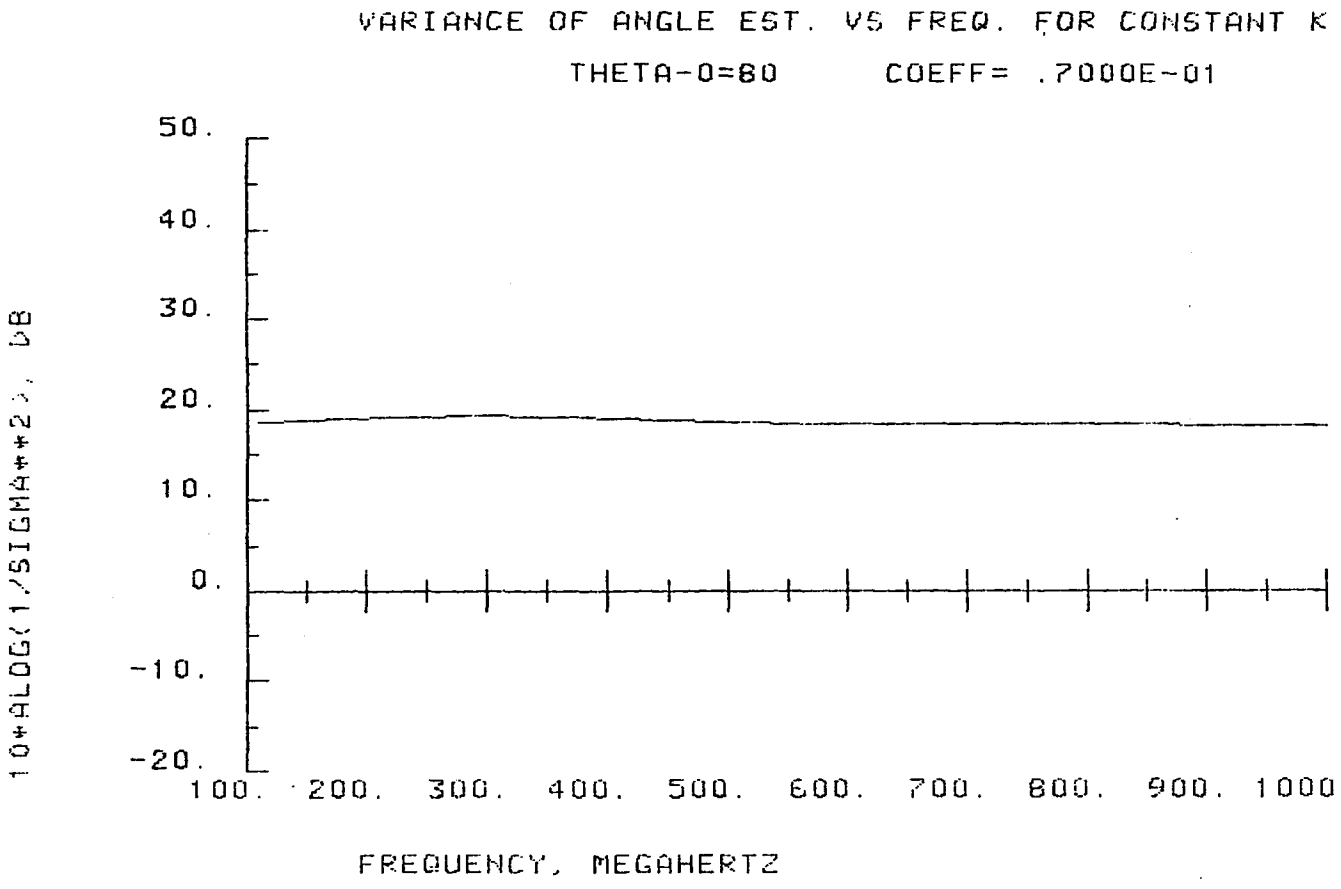


Figure 5.17. Variance vs. frequency at predicted value of Δ, algorithm TK8, θ_0 = 80°, K = 0.07

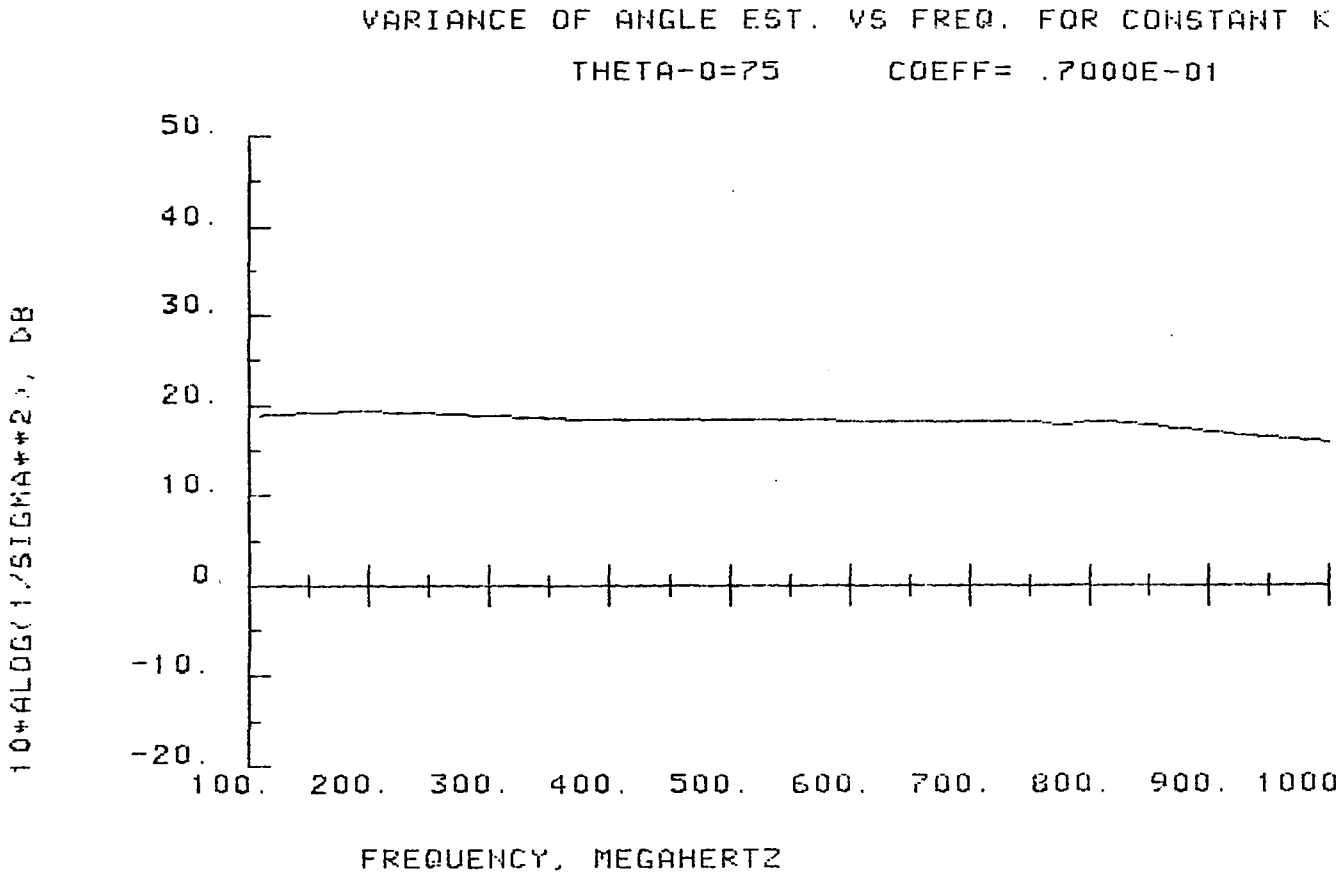


Figure 5.18. Variance vs. frequency at predicted value of Δ, algorithm TK8, θ_0 = 75°, K = 0.07

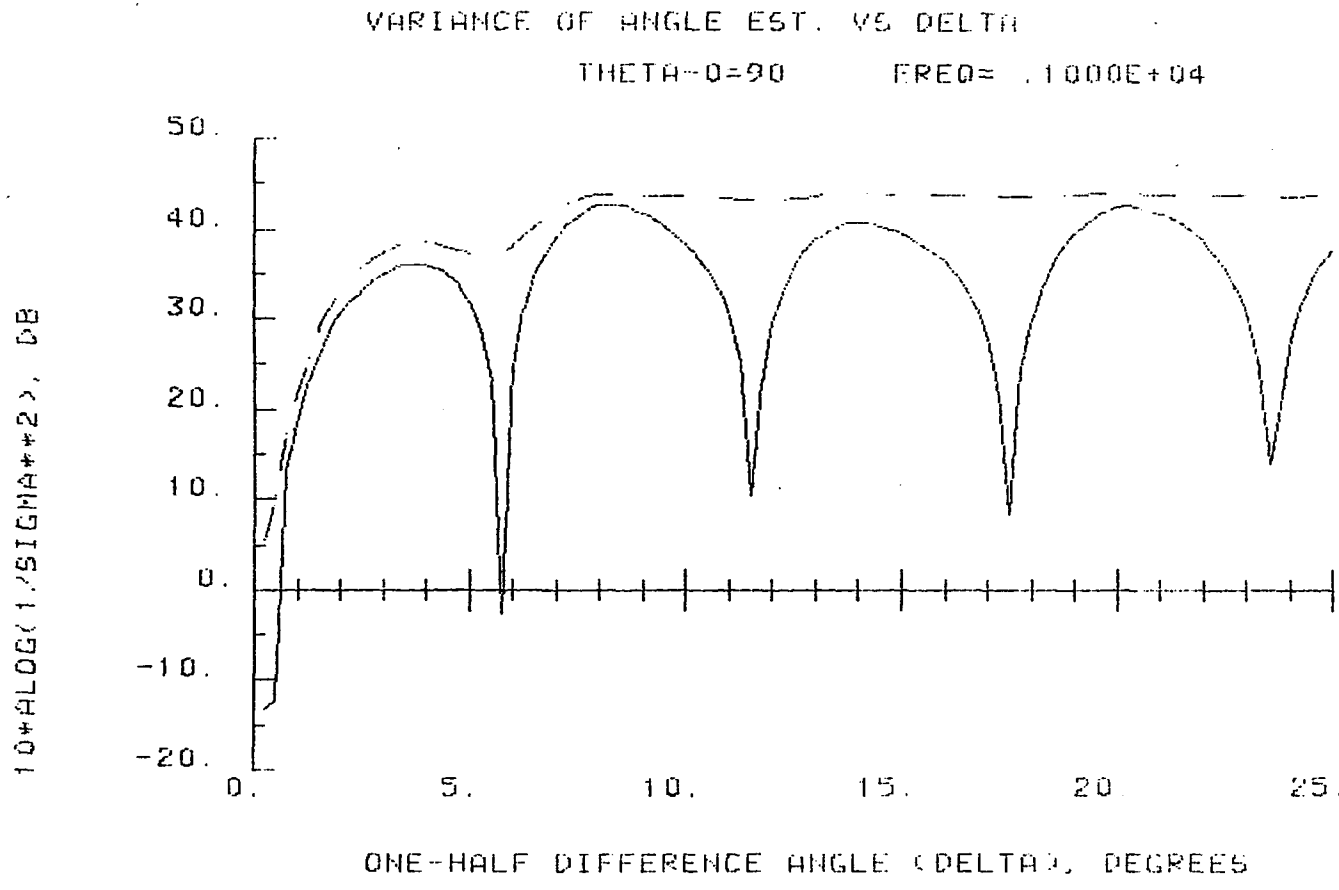


Figure 5.19. Variance vs. delta, algorithm TK10, frequency 1 GHz, amplitudes 100 V/m, phase 0°

100 V/m. The dashed curve is the CR bound, and the solid curve is the actual performance. The relative phase between the two sources is zero degrees.

As would be expected, the variance for small values of delta is quite large, leading to a negative value of $10\log(1/\sigma^2)$. As the source separation increases, the variance decreases. Once the curves shown exceed approximately 17 dB, our criterion for resolution is met. The variance continues to decrease (shown as higher values of $10\log(1/\sigma^2)$ in this format) as the sources are spread further apart. With algorithm TK10 expecting two signals (or equivalently, the KP algorithm), however, a curious thing happens. A point is reached as the separation increases where the variance stops improving and starts getting worse. As shown in the figure, nulls appear at certain critical angles. The algorithm's performance becomes very poor at these values of delta. They are effectively blind angles. Figures 5.20 and 5.21 illustrate what happens to the zeros of the PEF at the blind angles. Figure 5.20 shows TK10 expecting 2 signals at $\Delta = 3.75°$, $\theta_0 = 90°$. This is near the first peak in Figure 5.19. Figure 5.21 shows what happens when delta is increased to 5.75°. Instead of improving the variance, the performance is worse. The signals would be visible as two separate peaks by virtue of their large spacing, but the angle of arrival estimate would vary greatly from snapshot to snapshot.

The position of the first null (blind angle) depends upon the relative phase of the two signals. For in-phase signals, the first null

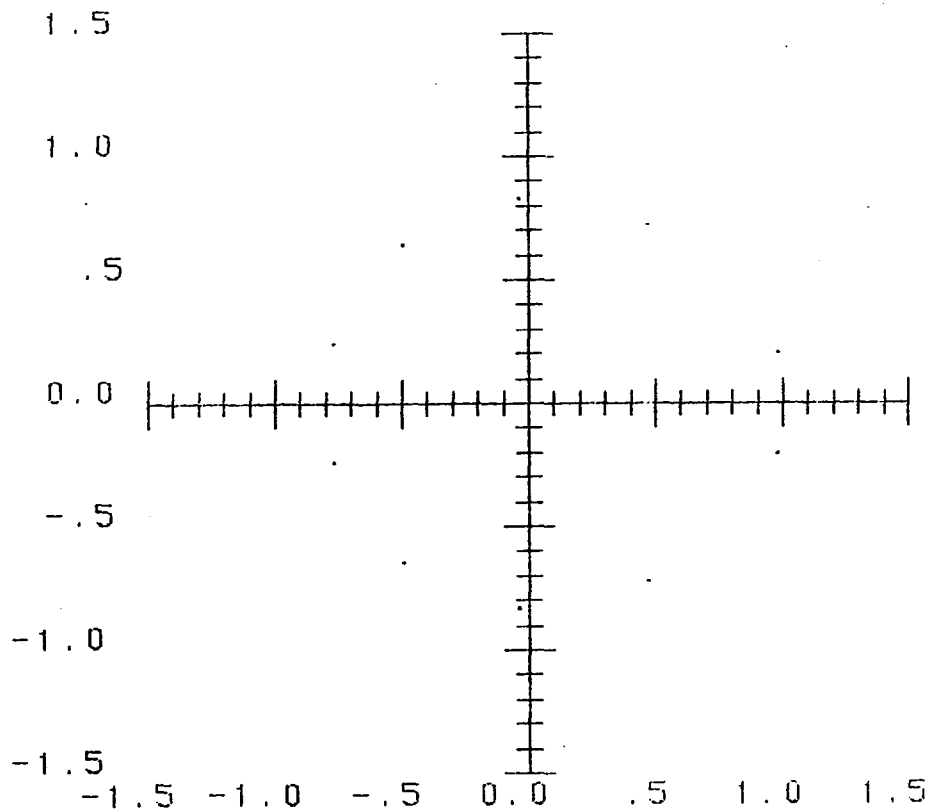


Figure 5.20. ZPLOT plot: algorithm TK10, $\theta_0 = 90°$, $\Delta = 3.75°$, amplitudes 100 V/m, phase 0°

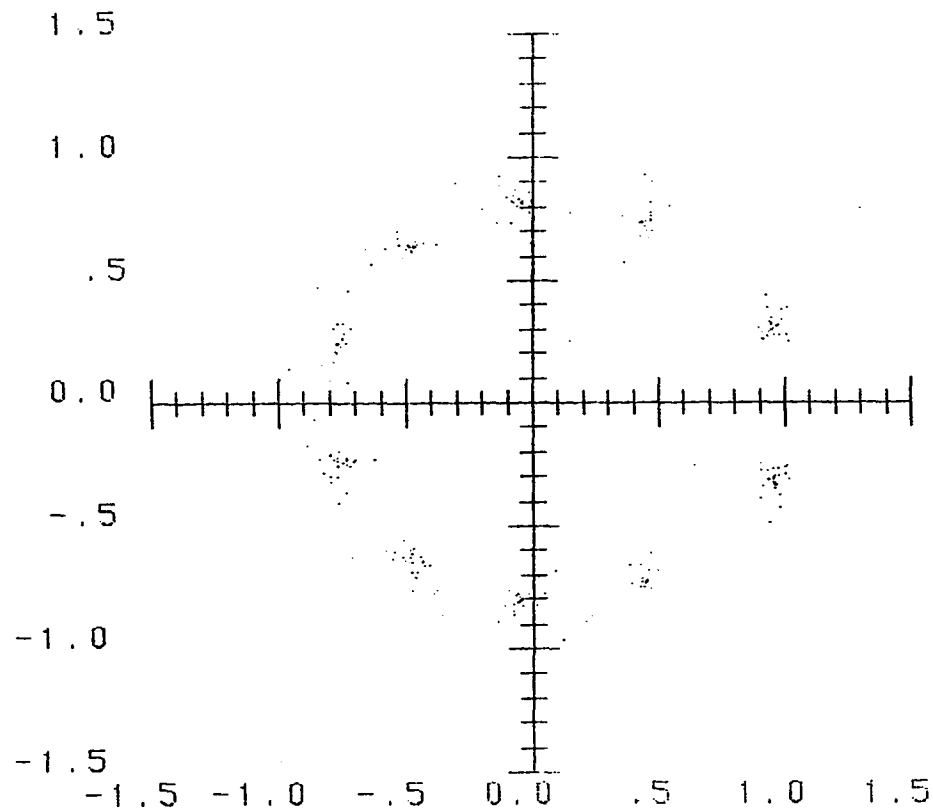


Figure 5.21. ZPLOT plot: algorithm TK10, $\theta_0 = 90°$, $\Delta = 5.75°$, amplitudes 100 V/m, phase 0°

corresponds to $\Delta=0$. As the phase difference increases, the null moves out toward higher values of delta, as shown in Figure 5.22. There, the phase difference is 90°. The situation at 180° phase difference is the same as that at 0°.

Once again, the ratio of the actual length of the array to the modulation envelope length appears to be an important parameter. For example, when the phase difference is 45°, or one-eighth of a cycle, the first null appears at a value of delta such that $L/Z_d = 0.125$ (one-eighth). When it is 90°, or a quarter cycle, the value of delta at the first null is such that the ratio is one-fourth, and so on. The nulls following the first are separated from it by multiples of one-half of the modulation envelope length.

The problem of blind angles appears to affect only the KP case (TK10). TK9 has some moderate ripple, as seen in Figure 5.23, but nothing as severe as KP. Lower model orders exhibit less ripple.

f. Dependence of TK and KP on number of signals expected A qualitative look at how these algorithms are affected by an incorrect assumption of the number of signals present yields some interesting results. In the light of the discussion of these methods in Chapter III, the results are entirely reasonable.

TK and KP do not appear to be overly sensitive to the assumed number of signals, as long as it is greater than or equal to the actual number of signals present. The maximum number of signals indicated is the assumed value. Thus, if the assumption is low, some of the signals

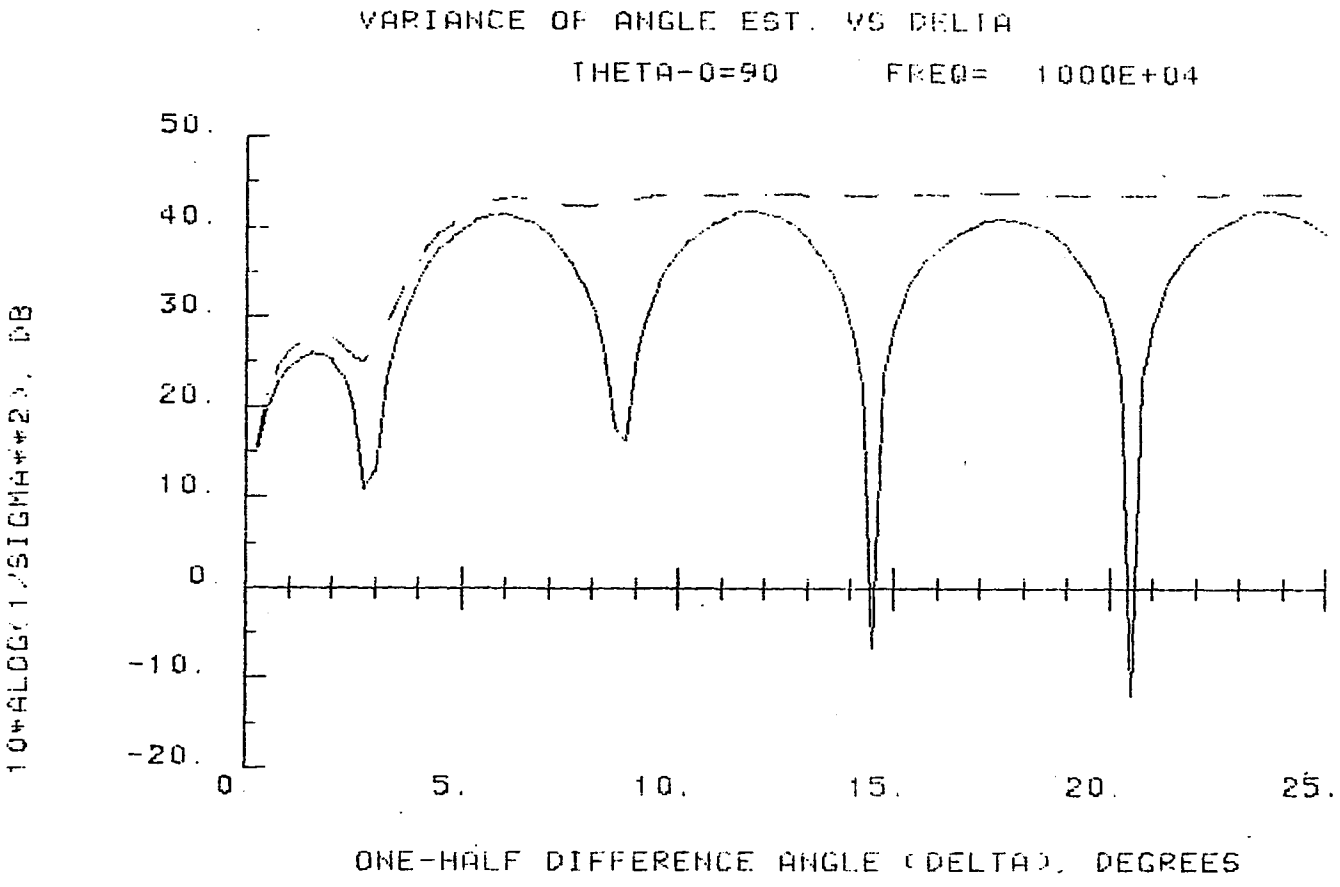


Figure 5.22. Variance vs. delta, algorithm TK10, frequency 1 GHz, amplitudes 100 V/m, phase 90°

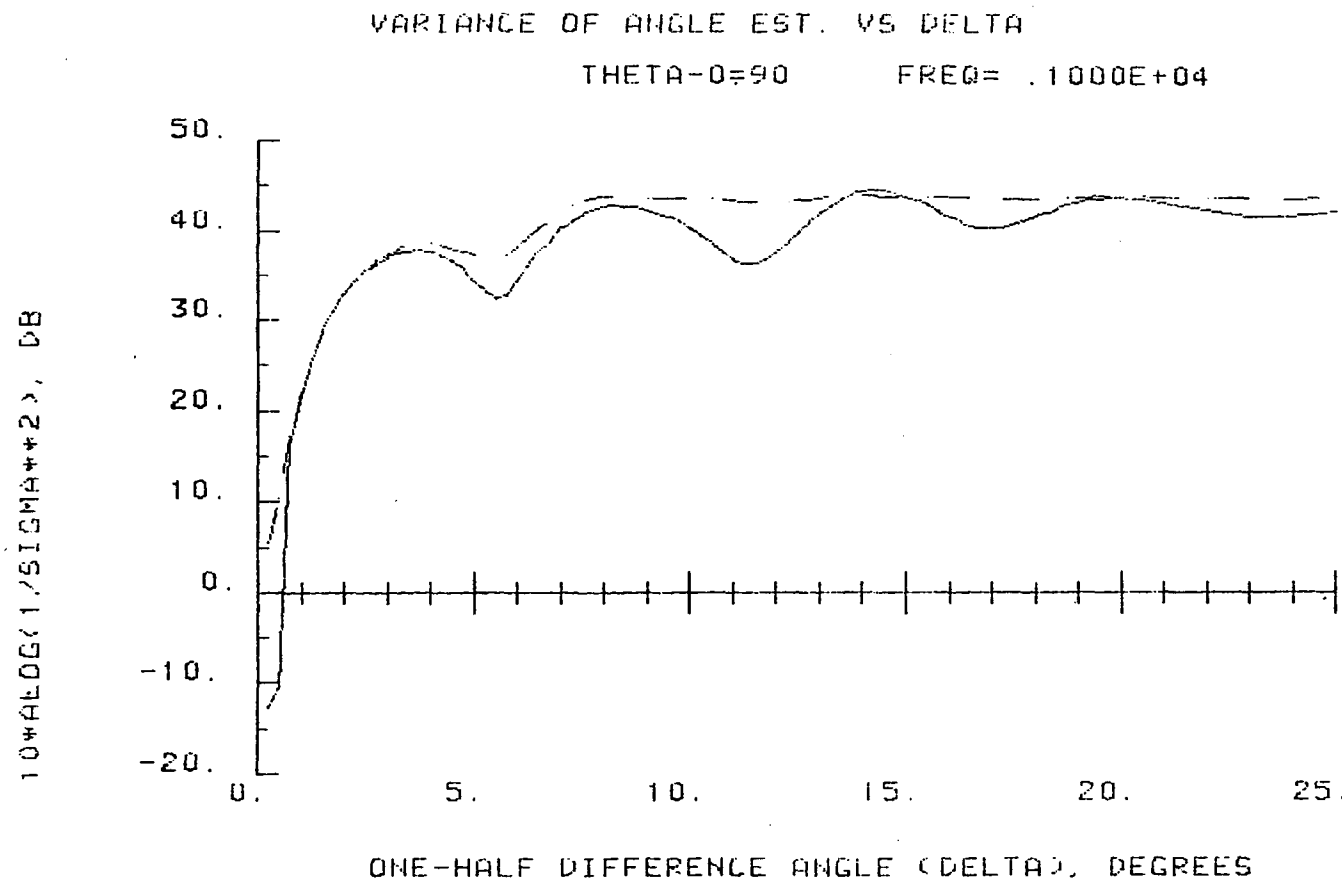


Figure 5.23. Variance vs. delta, algorithm TK9, frequency 1 GHz, amplitudes 100 V/m, phase 0°

will be missed. If it is too high, the resultant spectrum is noisier than it needs to be, and ambiguity performance is degraded. But, the correct number of signals appear in their proper positions.

Recall that these algorithms fit a smoothed matrix of the assumed rank to the actual signal correlation matrix. If the assumed rank is too small, that smaller number of signals will be fit to the data in the best way possible. Their positions will be more or less an average of the true locations. If the rank is too large, not all of the noise effects will be cancelled.

These conclusions are illustrated by Figures 5.24 through 5.33. In all of these runs, two 1 GHz, 100 V/m signals of 0 degrees relative phase are actually present. Their bearings are 75 and 80 degrees. The figures show first ARRAY and then ZPLOT for the assumption of one, two, three, four, and five signals present, respectively. The algorithm is TK8.

Notice especially the change between Figures 5.27 and 5.29. It is interesting that the variance of the noise zeros in Fig. 5.29 is increased so much more than that of the signal zeros. Figure 5.30 is beginning to look very much like an LS plot, complete with a spurious peak. And, indeed, that is what it is. With two signals present but four expected, much of the noise eliminating property of the TK algorithm is gone. The result degrades to the FBLP algorithm upon which TK is based. There is still some advantage in using TK, however, even when five signals are expected with only two present, as in Figures 5.32

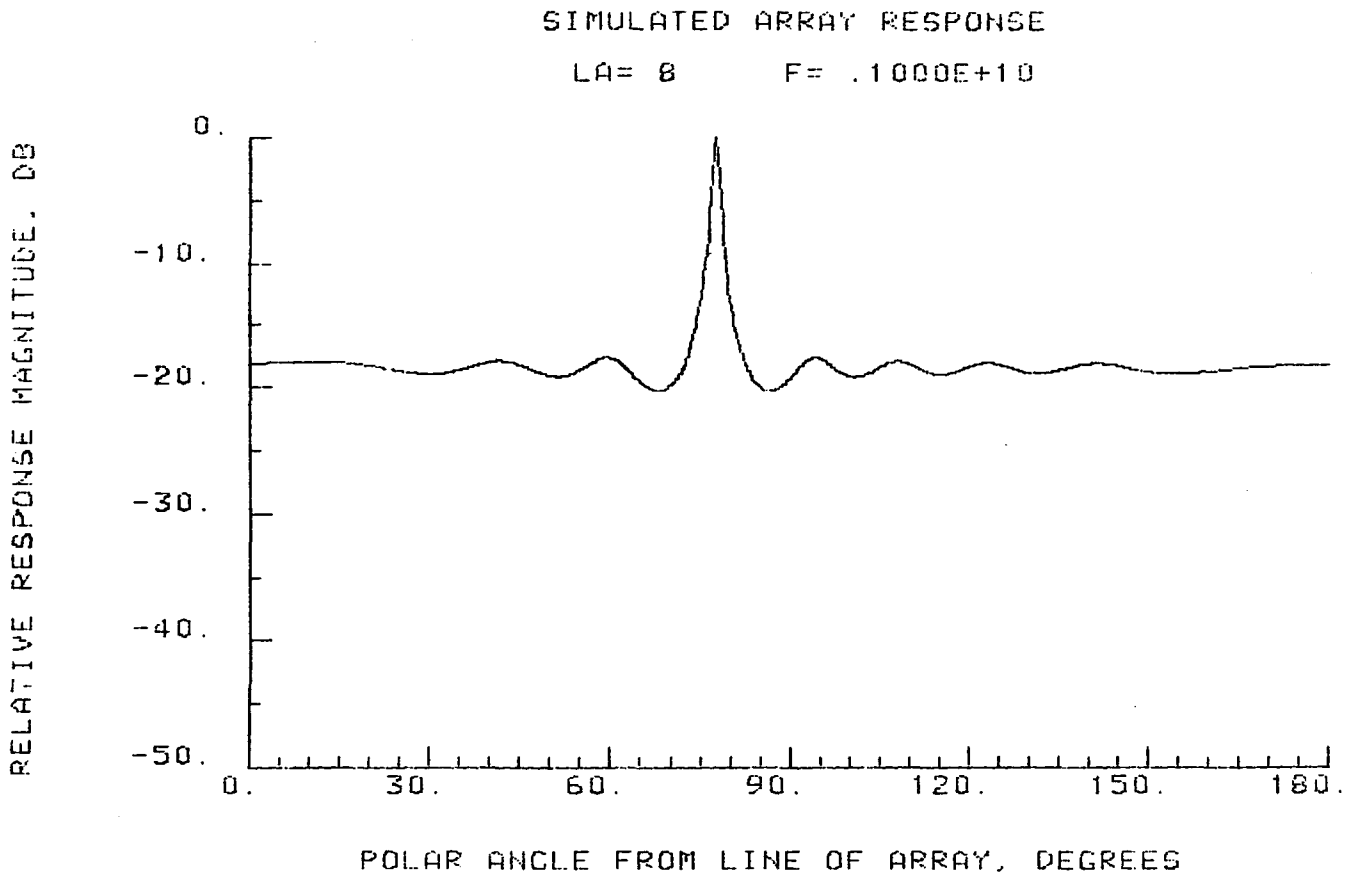


Figure 5.24. ARRAY plot: two signals present, one expected

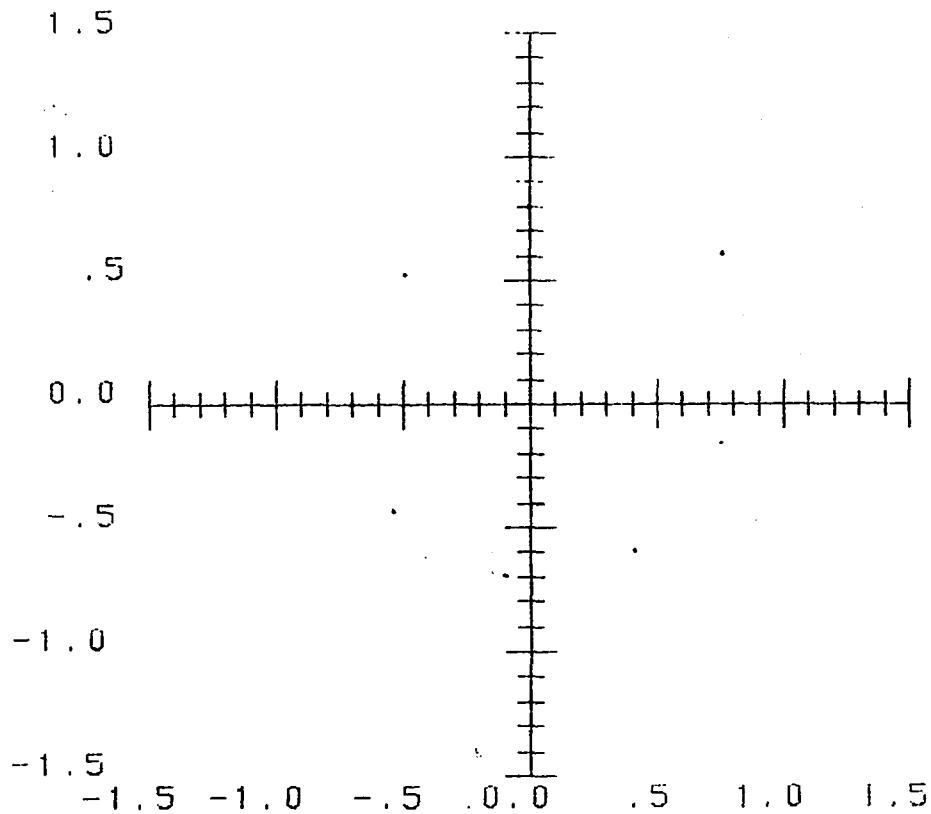


Figure 5.25. ZPLOT plot: two signals present, one expected

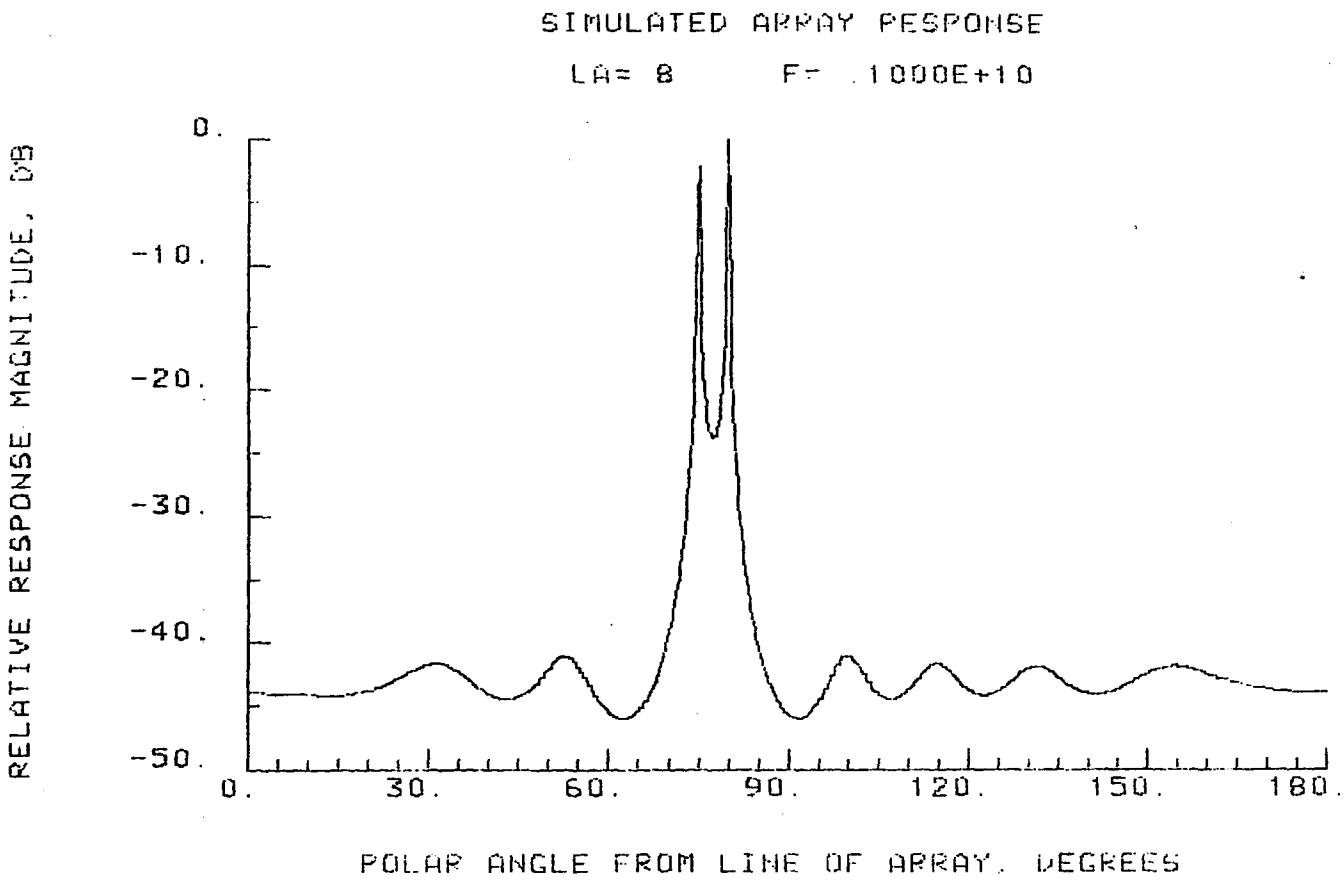


Figure 5.26. ARRAY plot: two signals present, two expected

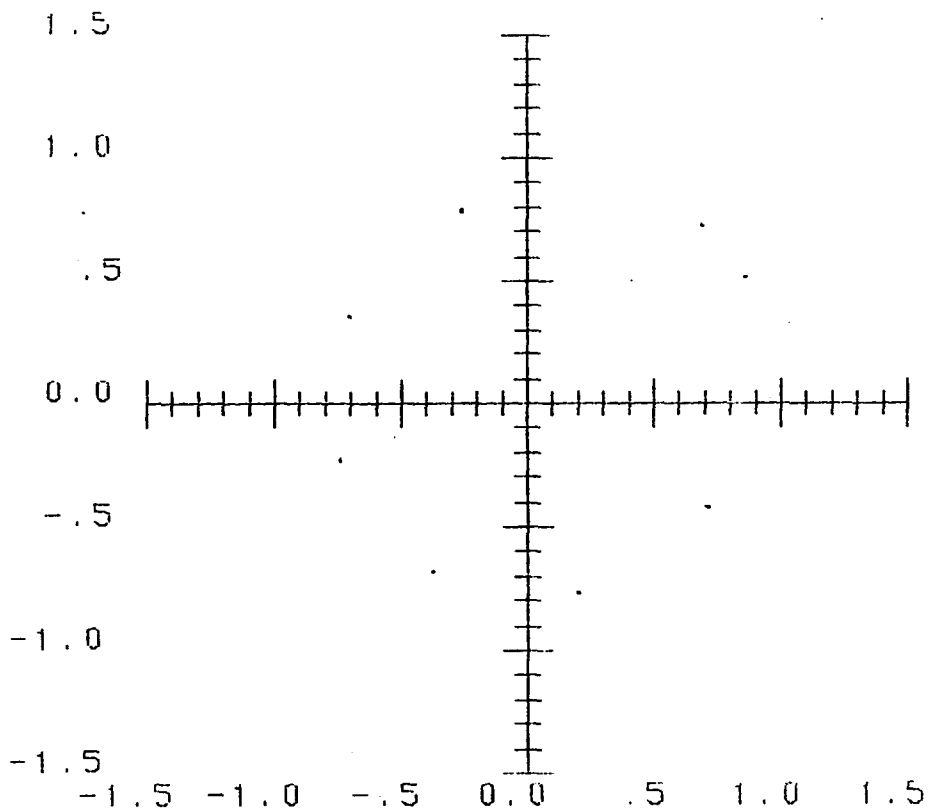


Figure 5.27. ZPLOT plot: two signals present, two expected

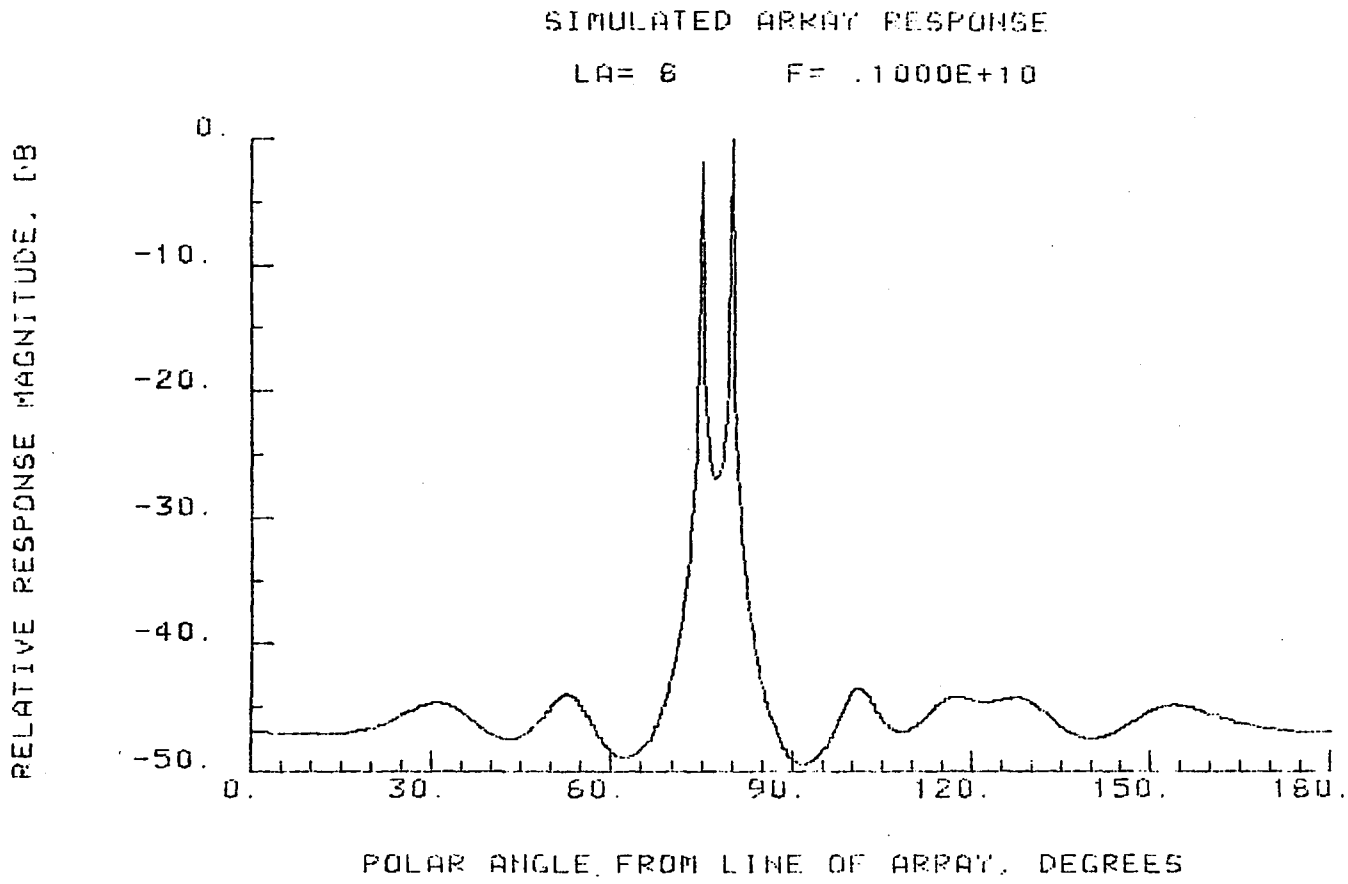


Figure 5.28. ARRAY plot: two signals present, three expected

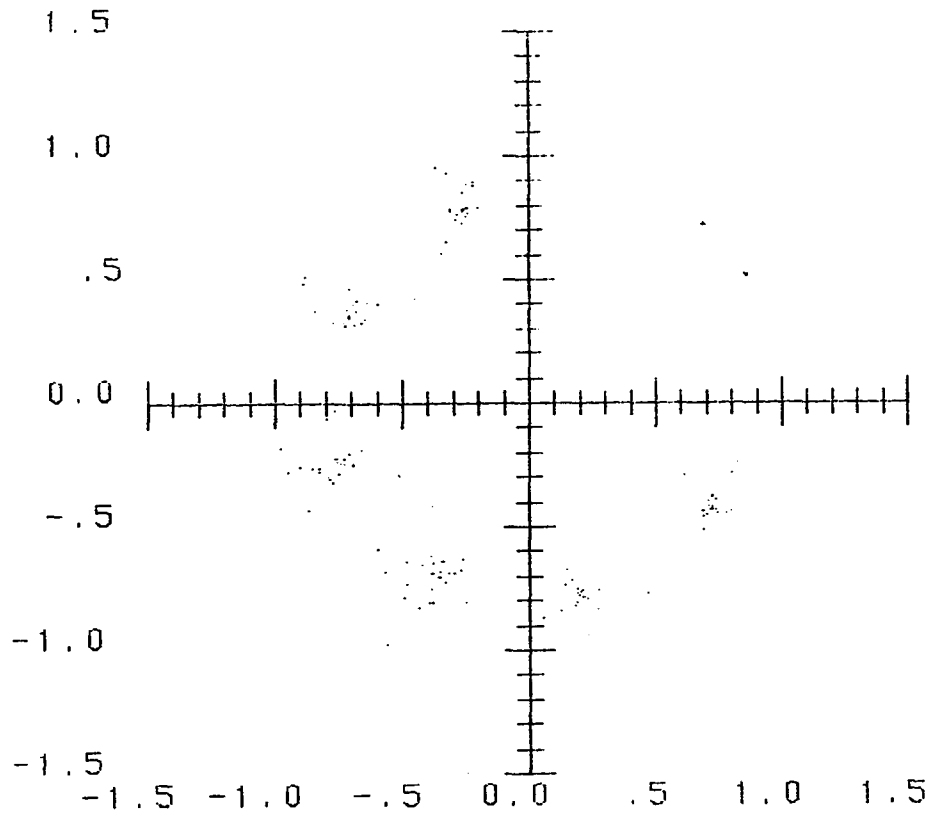


Figure 5.29. ZPLOT plot: two signals present, three expected

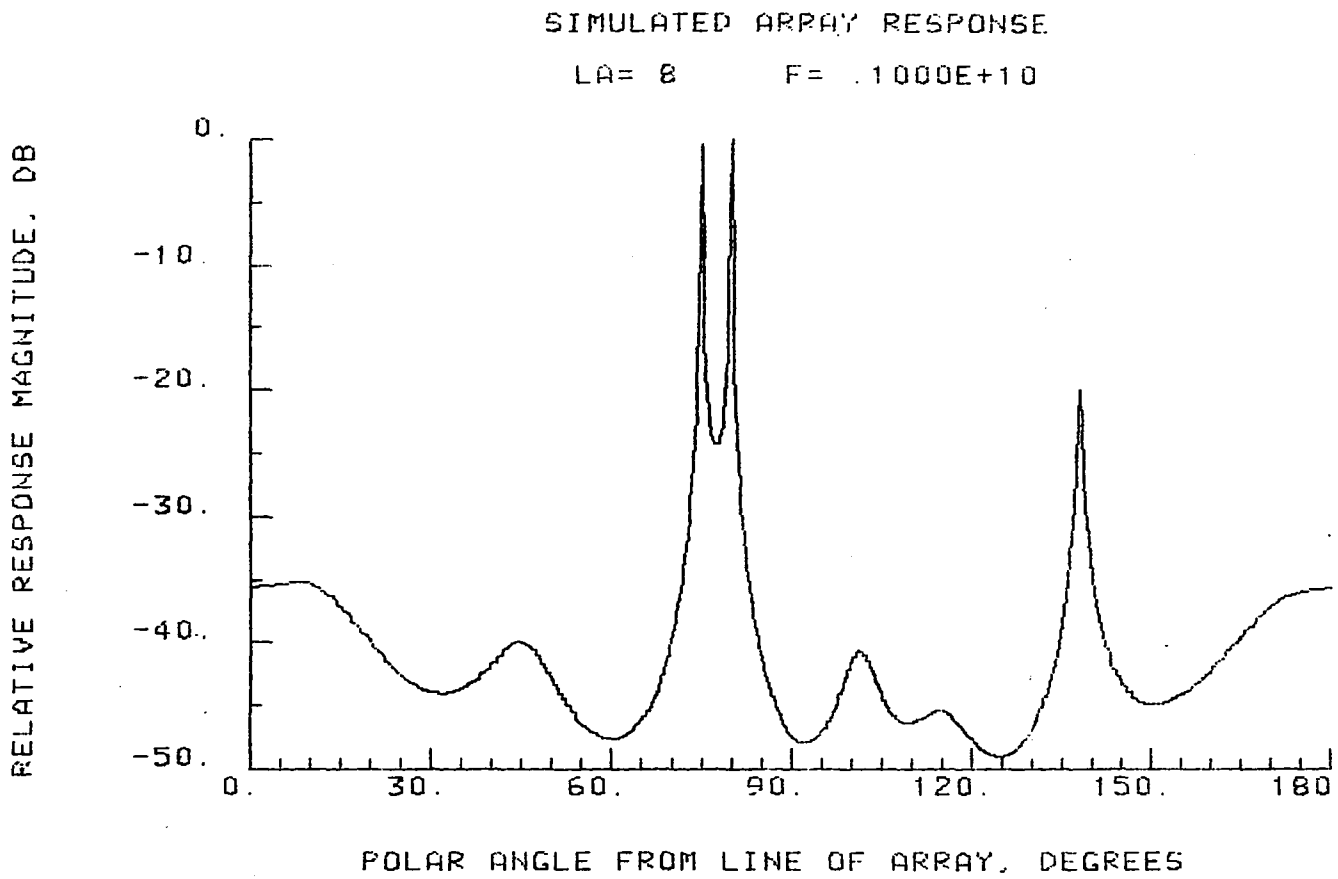


Figure 5.30. ARRAY plot: two signals present, four expected

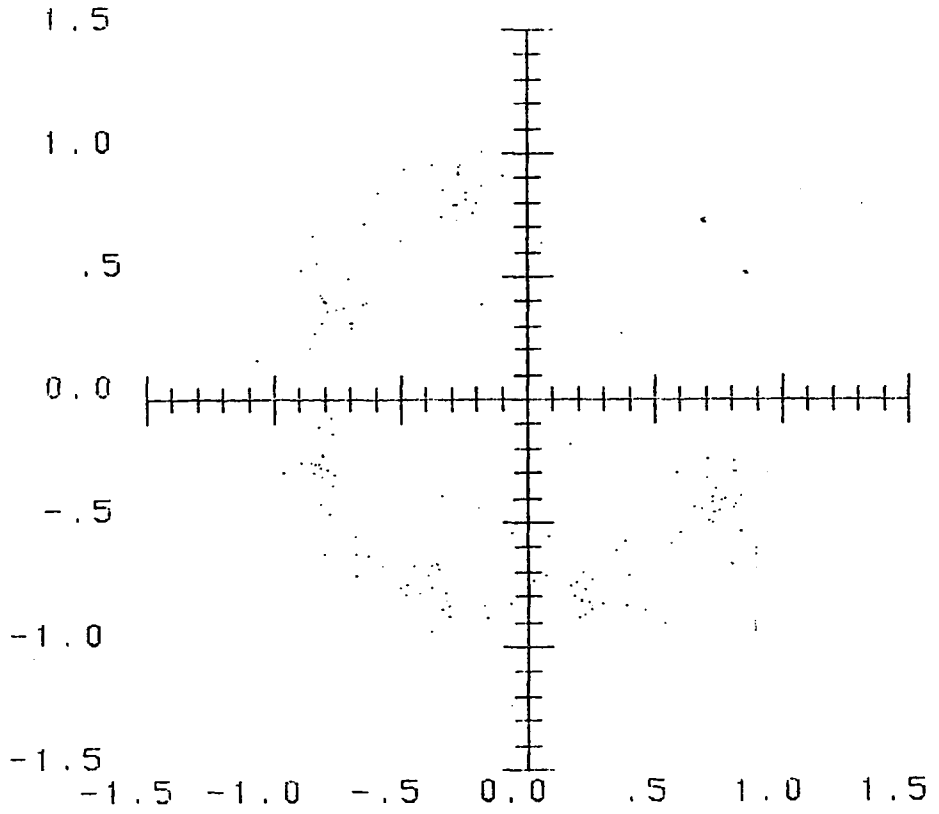


Figure 5.31. ZPLOT plot: two signals present, four expected

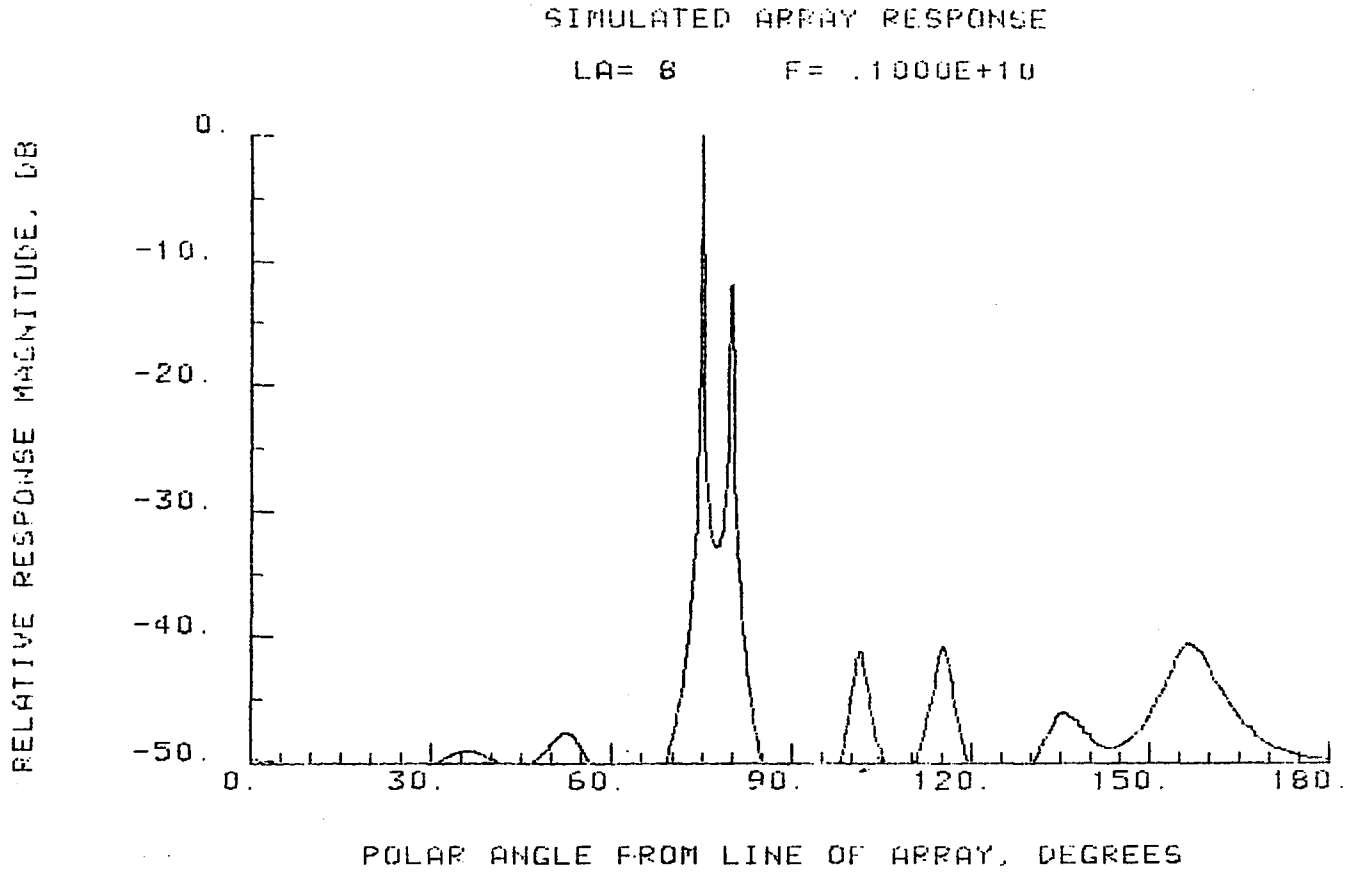


Figure 5.32. ARRAY plot: two signals present, five expected

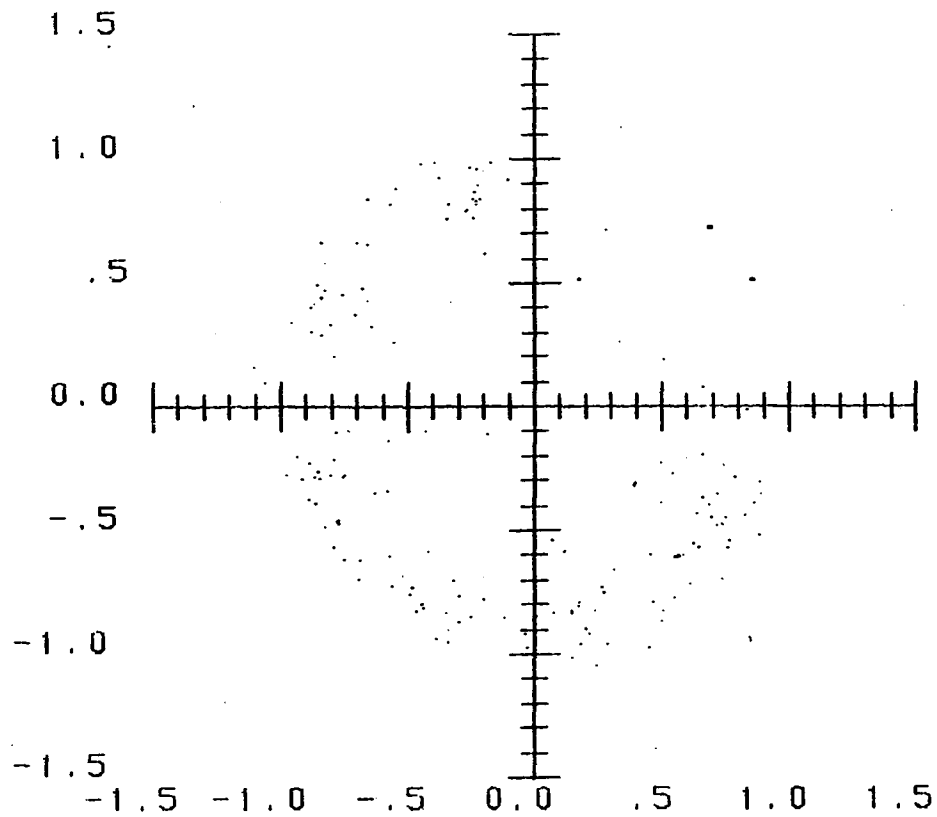


Figure 5.33. ZPLOT plot: two signals present, five expected

and 5.33. First, although they look very much like LS plots, the model order here is higher than LS would allow. Secondly, the variance of the signal zeros remains quite small -- significantly smaller than LS under these conditions. Thus, the direction estimate would be more consistent from snapshot to snapshot.

Similar runs with three signals present led to the same conclusions.

VI. CONCLUSIONS AND RECOMMENDATIONS

This work is a study of the broadband direction finding capabilities of array signal processing algorithms. It has sought to determine whether these algorithms are able to achieve satisfactory broadband direction finding performance with a given fixed array size. The performance of several of the new high resolution spectral analysis algorithms was evaluated and compared by computer simulation. The performance characteristics of concern in this application are high resolution, low ambiguity, and broad operating bandwidth.

The results of this investigation support the following conclusions. First, under the proper conditions, these algorithms can achieve greatly superior broadband direction finding performance compared to conventional linear beamforming techniques.

The qualification "proper conditions" is necessary because of a second conclusion: the algorithms are sensitive to both the element signal-to-noise ratio and the initial phase of the waves striking the array. The algorithms simulated exhibited the threshold effect with decreasing signal-to-noise ratio discussed by Tufts and Kumaresan (1982). Performance depended upon both the model order and the noise handling properties of the algorithm. The improvements suggested by Tufts and Kumaresan did lead to better noise performance in this application.

Phase effects appeared in algorithm KP as blind angles -- particular separation angles where performance is poor even though the separation angle is quite large. It appears that phase effects also entered into the variation in performance with frequency of both KP and TK.

A third, very interesting conclusion is that the ratio of the physical length of the array to the modulation envelope length is a parameter of fundamental importance in this application. (The modulation envelope is that of the interference pattern set up by the two incident plane waves.) The fact that this ratio is a constant at the point of resolution allows both the comparison of different algorithms using a single parameter and the prediction of array performance as a function of frequency and angle given the array specifications.

The simulation results lead to the following specific conclusions:

1. Algorithm TK8 gave the best performance in terms of its resolution, behavior at low SNR, and sensitivity to initial phase. It effectively achieved the theoretical bound on the variance of the angle estimate for the two signal case.

2. Optimum model order for algorithm TK was approximately 3N/4, as suggested by Tufts and Kumaresan.

3. Although it gives the best performance, algorithm TK was the most complex computationally. It requires the computation of the eigenvalues and eigenvectors of the signal correlation matrix. It, of course, takes the longest time to run.

4. Algorithm KP, a special case of TK, is simple to compute and retains some of the noise canceling properties of TK. Since its model order is not optimum, however, its resolution is slightly poorer than TK8. In addition, it exhibits blind angles where

performance is poor even though angular separation is quite wide.

5. The statistical properties of LS could not be effectively evaluated. The variance of the noise zeros was so great that the program could not automatically separate them from the signal zeros. This fact alone, however, indicates that LS is inferior to TK and KP in terms of the variance of its angle estimate. In addition, its placement of noise zeros close to the unit circle leads to poor ambiguity performance.

As is often true, this study has raised more questions than it has answered--some theoretical, some requiring additional simulation, and some important in a practical implementation of a high resolution signal processing array.

Perhaps the most fundamental theoretical question raised by these results is, why does the ratio of the physical length of the array to the modulation envelope length play such an important role? Its influence can be seen in both the CR bounds, which are algorithm independent, and in the actual performance of the algorithms. Although it is perhaps reasonable that this ratio should be important, the specific physical reasons for why it is are not clear.

Further simulation work is needed to answer at least two questions:

1. What is the effect of the number of elements and their spacing on algorithm performance? This simulation left those parameters fixed and considered only the total length of the array.

2. How does the assumption of the number of signals present affect the quantitative performance of TK and KP? The qualitative remarks in the preceding chapter should be verified and quantified.

A question which should be looked at both theoretically and by

simulation is the effect that multiple snapshots could have on algorithm performance. How should multiple snapshots of the field be best combined? Could they be used to enhance the effective SNR? Would the variation in initial phase from snapshot to snapshot eliminate the problem of blind angles?

Finally, before these high resolution algorithms could be used to advantage in a real direction finding system, several practical aspects of implementation must first be considered. The first aspect would be the specification of the system itself. What are the size and weight constraints? How fast would the processing element (or elements) have to be? How much precision is needed? How would variations in element spacing or gain affect the performance? This investigation has dealt only with single snapshots of unmodulated signals. What if the signals were modulated? What kinds of countermeasures might be encountered, and what could be done to minimize their effects? These are just a few of the questions and problems which a practical implementation would have to address.

This study has been significant in that it has shown that high-resolution spectral analysis algorithms have the capability to improve the broadband direction finding performance of an array. The potential improvement is great enough that practical implementation of these techniques should be actively pursued.

This practical implementation should be accomplished in two steps. First, the questions raised above should be answered by further

simulation. The additional simulation could also be used to test algorithm optimization and display formats. Secondly, a prototype signal processing array should be built and tested using the algorithms simulated here. This step would make clear which of the practical problems mentioned above are most serious. Differences in simulated and actual performance could be evaluated and problems identified. This test array would lay the groundwork for future practical high resolution broadband direction finding systems.

VII. LITERATURE CITED

Ables, J. G. 1974. Maximum entropy spectral analysis. Astronomy and Astrophysics Supplement Series 15:283-393.

Barnard, T. E. 1982. Two maximum entropy beamforming algorithms for equally spaced line arrays. IEEE Trans. Acoust., Speech, Signal Process., ASSP-30(2):175-189.

Booker, H. G., and P. C. Clemmow. 1950. The concept of an angular spectrum of plane waves and its relation to that of polar diagram and spectrum distribution. Proc. IEE A III 97:11-17.

Burg, J. P. 1972. The relationship between maximum entropy spectra and maximum likelihood spectra. Geophysics 37:375-376.

Burg, J. P. 1978. Maximum entropy spectral analysis. Pages 34-41 in D. G. Childers, ed. Modern spectrum analysis. IEEE Press Selected Reprint Series, New York, New York. 334 pp.

Capon, J. 1969. High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE 57(8):1408-1481.

Capon, J. 1979. Maximum-likelihood spectral estimation. Pages 155-179 in S. Haykin, ed. Nonlinear methods of spectral analysis. Springer-Verlag, Berlin, Germany.

Chen, W. Y., and G. R. Stegen. 1974. Experiments with maximum entropy power spectra of sinusoids. Journal of Geophysical Research 79:3019-3022.

Childers, D. G., ed. 1978. Modern spectrum analysis. IEEE Press Selected Reprint Series, New York, New York. 334 pp.

Eckart, C., and G. Young. 1936. The approximation of one matrix by another of lower rank. Psychometrica 1(3):211-218.

Evans, J. E. 1979. Aperture sampling technique for precision direction finding. IEEE Trans. Aerospace Electronic Systems AES-15(6):891-895.

Frost, O. L. 1976. Power-spectrum estimation. Pages 125-162 in G. Tacconi, ed. Aspects of Signal Processing, Part I. NATO Advanced Studies Institutes Series, Series C. Vol. 33. D. Reidel, Dordrecht-Holland.

Gabriel, W. F. 1980. Spectral analysis and adaptive array superresolution techniques. Proc. IEEE 68(6):654-666.

Gerhardt, L. 1978. Approaches to spectral analysis--New and used. Pages 1-18 in Proceedings of the RADC Spectrum Estimation Workshop, Rome Air Development Center, Griffiss Air Force Base, Rome, N.Y., May 24-26, 1978. Defense Technical Information Center Technical Report AD A054 650.

Gersh, W. and R. Z. Liu. 1978. Autoregressive model spectral estimation, some simulation study statistical performance results. Pages 37-49 in Proceedings of the RADC Spectrum Estimation Workshop, Rome Air Development Center, Griffiss Air Force Base, Rome, N.Y., May 24-26, 1978. Defense Technical Information Center Technical Report AD A054 650.

Gersh, W. and D. R. Sharpe. 1973. Estimation of power spectra with finite order autoregressive models. IEEE Trans. Automatic Control AC-18:367-369.

Haykin, S., ed. 1979. Nonlinear methods of spectral analysis. Springer-Verlag, Berlin, Germany. 247 pp.

Haykin, S., and S. Kesler. 1976. The complex form of the maximum entropy method for spectral estimation. Proc. IEEE 64(5):822-823.

Haykin, S., and S. Kesler. 1979. Prediction-error filtering and maximum-entropy spectral estimation. Pages 9-72 in S. Haykin, ed. Nonlinear methods of spectral analysis. Springer-Verlag, Berlin, Germany. 247 pp.

Hildebrand, F. B. 1956. Introduction to numerical analysis. McGraw-Hill Book Co., New York, NY. 511 pp.

Kay, S. M., and S. L. Marple. 1981. Spectrum analysis--a modern perspective. Proc. IEEE 69(11):1380-1419.

Kesler, S. B., and S. Haykin. 1978. The maximum entropy method applied to the spectral analysis of radar clutter. IEEE Trans. Information Theory IT-24(2):269-272.

Kraus, J. D. 1966. Radio astronomy. McGraw-Hill Book Co., New York, NY. 486 pp.

Lacoss, R. T. 1971. Data adaptive spectral analysis methods. Geophysics 36:661-675.

Lacoss, R. T. 1976. Autoregressive and maximum likelihood spectral analysis methods. Pages 591-615 in G. Tacconi, ed. Aspects of signal processing, Part II. NATO Advanced Studies Institute Series, Series C. Vol. 33. D. Reidel, Dordrecht-Holland

Marple, L. 1978. Frequency resolution of high-resolution spectrum spectrum analysis techniques. Pages 19-35 in Proceedings of the RADC Spectrum Estimation Workshop, Rome Air Development Center, Griffiss Air Force Base, Rome, N.Y., May 24-26, 1978. Defense Technical Information Center Technical Report AD A054 650.

Marple, L. 1980. A new autoregressive spectrum analysis algorithm. IEEE Trans. Acoust., Speech, Signal Process. ASSP-28(4):441-454.

McDonough, R. N. 1974. Maximum-entropy spatial processing of array data. Geophysics 39(6):843-51.

McDonough, R. N. 1979. Application of the maximum-likelihood method and the maximum-entropy method to array processing. Pages 181-244 in S. Haykin, ed. Nonlinear methods of spectral analysis. Springer-Verlag, Berlin, Germany.

Pisarenko, V. F. 1973. The retrieval of harmonics from a covariance function. Geophysical Journal of the Royal Astronomical Society 33(3):347-366.

Rife, D. C., and R. R. Boorstyn. 1974. Single tone parameter estimation from discrete-time observations. IEEE Trans. Information Theory IT-20(5):591-598.

Rife, D. C., and R. R. Boorstyn. 1976. Multiple tone parameter estimation from discrete time observations. Bell System Technical Journal 55(9):1389-1410.

Robinson, E. A. 1979. Iterative least squares procedure for ARMA spectral estimation. Pages 127-153 in S.Haykin, ed. Nonlinear methods of spectral analysis. Springer-Verlag, Berlin, Germany.

Seligson, C. D. 1970. Comments on "High-resolution frequency-wavenumber spectrum analysis." Proc. IEEE 58(6):947-949.

Skolnik, M. I. 1980. Introduction to radar systems. 2nd ed. McGraw-Hill Book Co., New York, New York. 581 pp.

Tufts, D. W., and R. Kumaresan. 1982. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. Proc. IEEE 70(9):975-989.

Ulrych, T. J., and R. W. Clayton. 1976. Time series modeling and maximum entropy. Physics of the Earth and Planetary Interiors 12(2-3):188-200.

Ulrych, T. J., and T. N. Bishop. 1975. Maximum entropy spectral analysis and autoregressive decomposition. Reviews of Geophysics and Space Physics 12(1):183-200.

Van den Bos, A. 1971. Alternative interpretation of maximum entropy spectral analysis. IEEE Trans. Information Theory IT-17:493-494.

Van Trees, H. L. 1968. Detection, estimation, and modulation theory, Part I. John Wiley and Sons, Inc., New York, N.Y. 697 pp.

Weeks, W. L. 1968. Antenna engineering. McGraw-Hill Book Co., New York, New York. 370 pp.

White, W. D. 1979. Angular spectra in radar applications. IEEE Trans. Aerospace Electronic Systems AES-15(6):895-899.

VIII. ACKNOWLEDGEMENTS

I am indebted to Dr. R. E. Post, my major professor, for his guidance during the preparation of this manuscript. Special thanks are due to Mr. James M. Roe and Mr. E. Ward Hall for their very valuable suggestions during the course of this study. Finally, I would especially like to thank my wife Kae and both of our parents for their consistent support and encouragement.