NAVSWC TR 90-236

SPECTRAL ESTIMATION TECHNIQUES WITH EMPHASIS ON SHORT DATA RECORDS-- A PRIMER

BY MANFRED K. PRUESSNER

UNDERWATER SYSTEMS DEPARTMENT

29 JUNE 1990

NAVAL SURFACE WARFARE CENTER

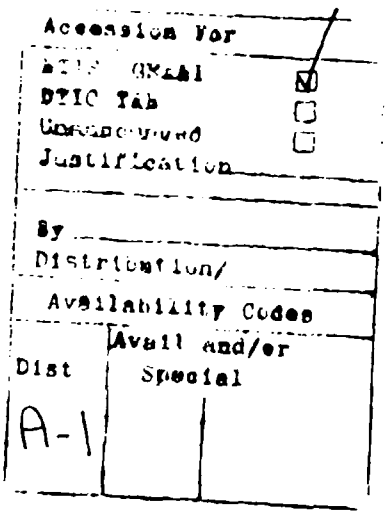
Dahlgren, Virginia 22448-5000 • Silver Spring, Maryland 20903-5000

NAVSWC TR 90-236

SPECTRAL ESTIMATION TECHNIQUES WITH EMPHASIS ON SHORT DATA RECORDS-- A PRIMER

BY MANFRED K. PRUESSNER
UNDERWATER SYSTEMS DEPARTMENT

29 JUNE 1990



NAVAL SURFACE WARFARE CENTER
Dahlgren, Virginia 22448-5000 • Silver Spring, Maryland 20903-5000

FOREWORD

A vast amount of literature exists that concerns itself with spectrum analysis. In particular, the last two decades have seen a great increase of published theoretical work as well as applications of so-called modern spectral estimation techniques based on nontraditional parametric approaches. Many papers dealing with this subject make it difficult even for the initiated to make the connection to his well known and routinely applied methods based on the Fourier Transform when he reads about "modeling a random process to obtain a parametric spectrum" of the observed data.

The report attempts to show the transition from the classical approach of spectral estimation to the basic parametric methods. It also shows the results of these new techniques when applied to short data records and demonstrates their capabilities and limitations.

Approved by:

DR. J. E. GOELLER
Deputy Department Head
Underwater Systems Department

CONTENTS

CONTENTS (Cont.)

ILLUSTRATIONS

ILLUSTRATIONS (Cont.)

ILLUSTRATIONS (Cont.)

ILLUSTRATIONS (Cont.)

ILLUSTRATIONS (Cont.)

CHAPTER 1
INTRODUCTION

Estimation of the power spectral density of a signal is a popular tool in signal analysis for detection purposes (by enhancing the signal-to-noise ratio of an expected signal) and classification (by showing the distribution of signal energy across the frequency band). Although efforts to break a spectrum into its constituting parts go back at least as far as Isaac Newton when he split white light through a prism, Jean Baptiste Fourier (1822) established that any periodic process can be interpreted as an infinite sum of sine and cosine terms whose frequencies are integral multiples of the fundamental frequency of the process (harmonic spectral analysis).

It was Arthur Schuster (1897) who coined the term "periodogram" for the plot of the squared magnitude of the Fourier coefficients to designate the spectrum. For noisy data, however, these spectra tended to be quite random and inconsistent, and many researchers lost interest in the periodogram. In 1930, Norbert Wiener established the fundamental relationship between the autocorrelation function and the spectrum as a Fourier Transform pair which then Robert Blackman and John Tukey used as their approach to the spectrum (1958). They also introduced averaging and windowing to eliminate the randomness and to improve the sidelobe structure of the original spectra. This method became the most popular spectral estimation technique until Jim Cooley and John Tukey developed in 1965 the Fast Fourier Transform (FFT) algorithm which calculates Schuster's periodogram digitally in a very efficient way. This so-called FFT is now the most widely used spectral estimation

technique. It may be of interest here that Karl Gauss apparently found out about the principle of the FFT in 1805 before Fourier even had published his fundamental paper.

Parallel to these developments, an entirely different approach was followed by statisticians, geologists, and economists who tried to use existing test data to generate modeled data with similar spectral characteristics. They used appropriate filters whose parameters (parametric analysis) were determined by the method of least-square errors. The filter is driven by a white noise source. Arthur Schuster tried this approach as far back as 1927 in order to determine the periodicity in the cycle of maximum sunspot numbers. This parametric approach to spectrum analysis has become quite popular since the 1960s because of its promise of greater resolution in transient analysis (John Burg, 1967). A multitude of different algorithms was developed since then, based on various physical interpretations for determining the filter parameters as well as on different mathematical techniques to compute these parameters efficiently (fast programs).

This report attempts to show the transition from the classical approach of spectral estimation to some basic parametric methods and their results when applied to short data records (transients).

Besides numerous papers, two books by Steven Kay[1] and Lawrence Marple[2] have served the author to familiarize himself with the material and to acquire some working knowledge for applying it. Also, the programs developed by those authors have been used to calculate the various signal spectra. These programs have been combined with input (signal generation) and output software (display and file transfers) so that they can be used as a flexible interactive package (see Reference 3).

CHAPTER 2

REVIEW OF CLASSICAL SPECTRUM ANALYSIS

SPECTRA OF ANALOG CONTINUOUS-TIME FUNCTIONS

The Fourier series of a deterministic periodic function $x(t)$ of period $T_1 = 1/f_1$ is:

$$x(t) = \sum_{n=-\infty}^{\infty} F(n) \cdot e^{jn2\pi f_1 t}$$

where the Fourier coefficients $F(n)$ are:

$$F(n) = \frac{1}{T_1} \int_{t=-T_1/2}^{+T_1/2} x(t) \cdot e^{-jn2\pi f_1 t} \cdot dt \quad , \quad n = 0, \pm 1, \pm 2, \ldots .$$

Because of the periodic nature of the signal $x(t)$, the spectrum is a line spectrum which exists only at the discrete frequencies nf_1, $n = 0, \pm 1, \pm 2, \ldots$. In general, the spectrum is a complex amplitude spectrum. As a special case, the spectrum becomes real if the signal is an even function in time.

The Fourier Transform $F(\omega)$ of a deterministic but aperiodic function $x(t)$ is:

$$F(\omega) = \frac{1}{2\pi} \int_{t=-\infty}^{\infty} x(t) \cdot e^{-j\omega t} \cdot dt \quad , \quad \omega = 2\pi f .$$

The inverse transform reestablishes the original time function:

$$x(t) = \int_{\omega=-\infty}^{\infty} F(\omega) \cdot e^{+j\omega t} \cdot d\omega .$$

Here the complex spectrum $F(\omega)$ is an amplitude density spectrum; it is continuous in ω.

$$|F(\omega)|^2$$

is the energy density spectrum and

$$\int_{\omega=-\infty}^{\infty} |F(\omega)|^2 d\omega$$

the total signal energy.

If x(t) is not known exactly at all times, it is a nondeterministic (random) function, and its Fourier Transform (FT) cannot be determined in the time domain (in theory). Therefore, the key to the spectrum is the autocorrelation function ACF. The Wiener-Khinchin theorem states: "The autocorrelation function of a random function and the power density spectrum of the random function form an FT pair" (see Reference 4). The autocorrelation function can be determined from the statistical properties of the random function, especially its first and second probability densities $p(x_1)$, $p(x_1, x_2; t_1, t_2)$, where x_1 and x_2 are the values of x(t) at two positions t_1 and t_2. If the process x(t) is "wide-sense stationary" in the first and second moment (i.e., has constant mean and variance), the probability density function (PDF) is a function of the time difference $\tau = t_2 - t_1$. Then the ACF is $r_{x_1 x_2}(\tau)$ or abbreviated $r(\tau)$:

$$r(\tau) = \int_{x_2=-\infty}^{\infty} \int_{x_1=-\infty}^{\infty} x_1 x_2 \cdot p(x_1, x_2; \tau)\, dx_1\, dx_2 .$$

The parameter τ is also called the "lag" value of the ACF. The ACF defined by the above equation is an ensemble ACF because it uses statistics obtained by observing all possible sample functions of $x(t)$ at time separation τ. This correlation function is a deterministic (generally aperiodic) function which therefore has a Fourier Transform:

$$P(\omega) = \frac{1}{2\pi} \int_{t=-\infty}^{\infty} r(t) \cdot e^{-j\omega t} \cdot dt .$$

$P(\omega)$ is defined as the FT of the aperiodic function $r(\tau)$, and is therefore the "amplitude density spectrum" of $r(\tau)$. From the definition of $P(w)$ follows the inverse transform:

$$r(t) = \int_{\omega=-\infty}^{\infty} P(\omega) \cdot e^{j\omega t} \cdot d\omega .$$

If a random process is "ergodic," the ensemble averages and the corresponding time averages are equal. For this class of signals, the time autocorrelation function and the ensemble autocorrelation function are equal.

The definition of the time ACF of $x(t)$ is given by:

$$r(t) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} x(t) \cdot x(t+\tau) dt$$

when $\tau = 0$, the two above equations give:

$$r(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x^2(t)\,dt = \int_{-\infty}^{\infty} P(\omega)d\omega$$

If x(t) represents a voltage across or current into a 1Ω load, this mean square value is the power taken by the load. Therefore, P(ω) (the FT of the ACF of x(t)) represents the power per hertz of frequency, or the power density spectrum of x(t). Since $r(\tau)$ is aperiodic, its Fourier Transform, i.e., the power density spectrum of x(t), is continuous over frequency.

An equivalent definition of the power spectral density (PSD) is:

$$P(\omega) = E\left\{ \lim_{T\to\infty} \frac{1}{2T} \left| \int_{-T}^{T} x(t)\cdot e^{-j\omega t}\cdot dt \right|^2 \right\}$$

where E means the "ensemble average" or "expectation" or plain, the average over the terms inside the curly brackets. This definition says to take the FT of a given "realization" $x_1(t)$ of the random function x(t) over an interval 2T as if it were an aperiodic function, square it, and divide it by the record length. The result is a continuous function in ω and random. For any similar realization, $x_i(t)$ of x(t), one will get a different FT and its time average squared. Doing this for a infinite number of realizations of x(t) leads to an ensemble of random spectral functions:

$$\frac{1}{2T} \left| \int_{-T}^{T} x_i(t)\cdot e^{-j\omega t}\cdot dt \right|^2 , \quad i = 1,2, \ldots \to \infty .$$

Picking a particular value for ω, the corresponding values of the spectral functions are totally random and will not converge to a stable spectral value for increasing time intervals 2T. Therefore, the ensemble average is taken over this infinite ensemble which will then represent a statistically stable power density spectrum as if it had been derived via the Wiener-Khinchin procedure. The proof for the equivalence between both definitions is not trivial and can be found in References 1 and 2. The spectra derived by both methods can be considered "true spectra" compared to "estimated spectra" which generally result from a practical application, i.e., from one realization $x_n(t)$. In such a case, one disregards or only approximates the averaging over an infinite ensemble and obtains the power spectral estimate

$$\hat{P}_n(\omega) = \frac{1}{2T} \left| \int_{-T}^{T} x_n(t) \cdot e^{-j\omega t} \cdot dt \right|^2$$

which is a "sample spectrum" (also original Schuster periodogram). The hat ($\hat{}$) means "estimate." Although the mean of the sample spectrum will tend to converge to that of the true PSD in the limit ($T \to \infty$), the variance will stay constant and proportional to the mean of the sample spectrum. Figures 1 and 2 depict the superposition of the sample spectra of five 1-Vrms white noise records of 32 and 128 data samples. They show that the size of the statistical fluctuations of the periodogram remains constant independent of the sample size. Only in the 1950s, when statistical smoothing (ensemble averaging) was applied, the variance was reduced resulting in a more stable spectrum leading to the acceptance of the periodogram approach. The periodogram is also called the direct approach to the spectrum (i.e., directly from the data), whereas the Wiener-Khinchin, Blackman-Tukey method via the ACF is called indirect.

SPECTRA OF DISCRETE-TIME FUNCTIONS

Both the indirect and direct methods exist also if the signal x(t) is sampled at intervals Δt to produce a time series of samples x(n) for the integer $-\infty < n < \infty$.

$$P(\omega) = \sum_{k=-\infty}^{\infty} r(k\cdot\Delta t)\cdot e^{-j\omega(k\cdot\Delta t)} \qquad \text{(indirect)}$$

where $k\cdot\Delta t$ corresponds to the lag parameter τ for the case of a continuous-time function x(t), and $r(k\cdot\Delta t)$ is the value of the ACF of $x(n\cdot\Delta t)$ at that lag.

$$P(\omega) = E\left\{ \lim_{M\to\infty} \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x(n\cdot\Delta t)\cdot e^{-j\omega(n\cdot\Delta t)} \right|^2 \right\} \qquad \text{(direct)}$$

These are discrete-time, continuous-frequency spectra that can be shown to be periodic in ω with period $2n/\Delta t$. They are "true spectra" with infinite resolution and an absence of side lobes and require, of course, an infinite data sequence.

If the data sequence x(n) is available only during a limited time window from $n = 0$ to $n = N-1$, the result will be an estimated spectrum:

$$\hat{P}_{\omega} = \sum_{k=-(N-1)}^{(N-1)} \hat{r}(k\cdot\Delta t)\cdot e^{-j\omega(k\cdot\Delta t)} \qquad \text{(indirect)}$$

wher

$$\hat{r}(k\cdot\Delta t) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n)\cdot x(n+k)$$

is called the biased ACF and

$$\hat{R}(k\cdot\Delta t) = \frac{1}{N-k}\sum_{n=0}^{N-1-k} x(n)\cdot x(n+k)$$

is called the unbiased ACF (k = 0,1,2,...,N-1).

The direct sample spectrum (estimate) is:

$$\hat{P}(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)\cdot e^{-j\omega(n\cdot\Delta t)}\right|^2$$

For the indirect method, a practical maximum number of lags k is always chosen smaller than N (suggested by various authors as 10 to 20 percent of N). The reason for this is to avoid the greater variance of the estimated autocorrelation associated with higher lags k. To see this, consider the maximum lag N-1 possible for N data points for which the ACF estimate is

$$\hat{R}(N-1) = \frac{1}{N}\cdot x(0)\cdot x(N-1)$$

which is highly variable due to the lack of averaging (the sum degenerates to one term only), regardless of how large N becomes. On the other hand, the maximum lag should be chosen large enough so that the ACF has decayed close to zero and the ACF outside this lag does not contribute to the PSD.

A practical way to achieve better stability for larger k is to weigh those ACF estimates less by multiplying the correlation function, obtained from the N data points, with a tapered window function $w(k \cdot \Delta t)$, where

$$0 \le w(k \cdot \Delta t) \le w(0) = 1 \quad \text{for} \quad k \le k_{max} = M$$

and

$$w(k \cdot \Delta t) = 0 \quad \text{for} \quad k > M \ , \ M < N\text{-}1 \ .$$

With this condition,

$$\hat{P}_{BT}(\omega) = \sum_{k=-M}^{M} w(k \cdot \Delta t) \cdot r(k \cdot \Delta t) \cdot e^{-j\omega(k \cdot \Delta t)}$$

This is the Blackman-Tukey (BT) spectral estimate (1958). Many window functions exist, but all are characteristic of having their maximum at $k = 0$ and then symmetrically tapering off to a minimum (often zero) at the maximum chosen lag M.

The direct method became popular after the Discrete Fourier Transform (DFT) was calculated by Cooley and Tukey via the FFT (1965). It is based on the discrete-time. discrete-frequency, Fourier-series type approach for a periodic signal. The signal x(n) for $n = 0$ to $n = N\text{-}1$ is stored in memory and implicitly assumed to be repetitive with the fundamental frequency

$$f_1 = \frac{1}{N \cdot \Delta t}$$

and the harmonics $f_k = k \cdot f_1$ so that

$$\omega \rightarrow \omega_k = 2\pi \frac{k}{N \cdot \Delta t}$$

$$\omega \cdot (n \cdot \Delta t) \rightarrow 2\pi \frac{k}{N \cdot \Delta t} \cdot (n \cdot \Delta t) = 2\pi \frac{nk}{N}$$

$$\hat{P}(k) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi \frac{nk}{N}} \right|^2$$

This approach leads to a harmonic analysis of a random signal sequence x(n), and the spectrum is called a periodogram since it is periodic with $k_{max} = N$. This can be seen by calculating the Fourier Transform F(k + N):

$$F(k+N) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi \frac{n(k+N)}{N}}$$

$$= \frac{1}{N} \sum_{n=0}^{N-1} x(n) \cdot e^{-j2\pi \frac{nk}{N}} \cdot e^{-j2\pi n}$$

i.e., $F(k+N) = F(k)$, and $F(N) = F(0)$ since $e^{-j2\pi n} = 1$. This is also equivalent to saying that the spectrum is repetitive with the period $1/\Delta t$.

Smoothing the individual random sample periodograms is done by "pseudo-ensemble averaging." The given data sequence x(0), x(1),....,x(N-1) is divided in P nonoverlapping segments of D samples each so that $P \cdot D = N$. An independent sample periodogram of the Pth segment is then

$$P_D(f_k) = \frac{1}{D} \left| \sum_{m=0}^{D-1} x_p(m) \cdot e^{-j2\pi \frac{mk}{D}} \right|^2$$

The P independent sample spectra are then coherently added to produce a smoothed Bartlett periodogram. Bartlett also applied a triangular window to each data segment. Welch (1967) then went one step further and permitted data segments to overlap, thereby increasing the number of averaged segments. This decreases the PSD estimate variance further, although the segments and sample spectra are not independent any longer. He also applied a variety of better windows to his segments before computing the FFT. This Welch periodogram is the most widely calculated FFT in use today.

Inherent in the concept of the discrete-time FT over a finite data sequence is the multiplication of an infinite data sequence by a uniform (rectangular) window.[5] Figure 3 shows the development of the DFT step by step, the time domain on the left, the frequency domain on the right side. Multiplying the data x(t) by a window w(t) in the time domain is equivalent to a convolution of their Fourier Transforms in the frequency domain. Since the FT of a rectangular window is a (sinπfT)/(πfT) function (Figure 3b), the two lines (positive and negative frequencies) of a narrowband signal (Figure 3a) are convolved into two such (sinπfT)/(πfT) functions resulting in the transform X(f)*W(f) where the symbol * signifies the convolution operation (see Figure 3c). The function (sinπfT)/(πfT) is characterized by a series of side lobes, i.e., energy from a single frequency "leaks" into neighboring frequencies of the spectrum. The width of the main lobe where it crosses the frequency axis is 2/T and, at the 3 dB points down from the peak, the width is about 1/T. This is the resolution of a DFT for a uniform window, i.e., the capability to resolve two frequencies with a separation $\Delta f = 1/T$. The side lobes of a strong signal will interfere with the main lobe of a small

signal if the frequency separation is $\Delta f = 1/T$ or less. In order to suppress the side lobes, one can use tapered windows. This, however, causes the main lobe to widen, resulting in less resolution.

The effect of sampling in the time domain is shown in Figures 3d and 3e, where s(t) is the sampling function with the sample pulses Δt apart and its line spectrum S(f) has a frequency separation of $1/\Delta t$. Figure 3e shows the sampling of Figure 3c and its transform: $X(f)*W(f)*S(f)$, a spectrum function still continuous in frequency. The step to the discrete line spectrum is done in Figure 3f, where the windowed time function in Figure 3c is considered as one period of a repetitive signal. This forces both the time function $x_p(n)$ and its DFT $X(f_k)$ to be discrete. The time sample spacing is Δt, the frequency line spacing is $\Delta f = 1/(N \cdot \Delta t)$, the signal is periodic in the time domain with $T = N \cdot \Delta t$, and the spectrum is periodic with $F = 1/\Delta t$. Both time signal and spectrum contain N terms per basic cycle, i.e., the DFT is simply a one-to-one mapping of N terms of $x_p(n)$ into N terms of $X(f_k)$.

Figures 3d and 3e also demonstrate the effect of the sampling rate. The smaller the sampling rate $1/\Delta t$, the closer the signal lobes move toward each other until neighboring lobes start to overlap. The minimum sampling rate is the Nyquist rate of two samples of the highest frequency component in the signal band (low pass band). Below that, aliasing will occur, i.e., the appearance of out-of-band high frequencies as in-band low frequencies which then do not represent the true signal anymore: the signal recovered by low pass filtering becomes a more and more distorted replica of the original time function.

The main lobe structure of the spectrum calculated for the DFT with a uniform window is shown in Figure 4. The main lobes can be considered to be a bank of

bandpass filters of width 1/T at their 3 dB points. For a sine wave signal with a frequency f_a in Figure 4a, which coincides with one of the computed frequencies f_k, the DFT output would result in a response at the appropriate harmonic and zero at all the other calculated harmonics. A so-called picket fence effect becomes evident when the signal being analyzed lies between two calculated DFT frequencies, e.g., f_b, where signal energy appears to be generated at the two neighboring calculated frequencies with harmonic numbers 6 and 7. If the signal is swept across the band, then the spectral power response will follow the curve of Figure 4b showing a 3 dB ripple. The depth of this ripple can be alleviated through zero-padding which generates an artificial record length $N' > N$ by adding zeroes to the actual data record. The frequency spacing between calculated harmonics becomes $\Delta f' = 1/(N' \cdot \Delta t)$ with the result that the calculated bandpass filters move closer, resulting in a spectral response with less ripple (see Figures 5a and 5b). The width of the main lobe remains the same since the actual data window is that of the original data record. Zero-padding allows one to calculate the power response for any intermediate frequency by adding the proper number of zeroes. It results in an interpolation of the original spectrum, but does not improve the actual resolution. The spectral resolution is equal to the reciprocal of the signal observation time or signal record length which determines the width of the main lobe of the data window transform (width of the bandpass filters). The longer the observation time or actual record length, the narrower this filter becomes with an equivalent improvement in resolution.

CHAPTER 3

APPLICATIONS OF CLASSICAL SPECTRUM METHODS

Some examples of classical spectrum analysis follow. They are based on a Fourier Transform of 256 points in the time and frequency domain. The number of data points N can be chosen up to N = 256. For N < 256, a series of (256-N) zeroes is attached to the data record, so that always 256 FT points cover the total frequency band independent of the number of actual data points. This way a reasonably smooth appearance of the spectrum display is assured. The sampling of the frequency axis is done relative to the sampling frequency $f_r = f/f_s$, so that $-0.5 \le f_r \le 0.5$ where $f_r = 0.5$ represents the maximum signal frequency possible for a chosen f_s without aliasing (Nyquist frequency). The vertical axis represents P(f), with the PSD plotted in dB relative to $1V^2/Hz$ ($PSD = 10 \log P(f)/1V^2/Hz$). This is the display format of MISA (see Reference 3).

The periodogram (regular FT) for a rectangular data window of width 2M over N = 32 data points of a noiseless complex sine wave of amplitude e = 1V is displayed in Figure 6. The relative frequency is chosen as $f_r = 0.25$. The results can be extrapolated to any desired frequency within $\pm 0.5\, f_s$. The data record of N = 32 points represents $N \cdot f_r = 8$ sine wave cycles; the number of data points per cycle is the reciprocal value of f_r. The shape of the spectrum is the result of convolving the FT of the rectangular window $w_R(t)$, $W_R(f)$, with the FT of the nonwindowed signal s(t), i.e., S(f), so that $S_R(f) = W_R(f) * S(f)$, and $P_R(f) = |S_R(f)|^2$. For a sine wave signal, S(f) is a Delta function $\delta(f)$, which effectively samples the window function $W_R(f)$ to generate $P_R(f)$. For the continuous Fourier Transform (CFT), $W_R(f)$ is

$$W_R(f) = CFT\{w_R(t)\} = T\frac{\sin\pi fT}{\pi fT},$$

where T is the window duration, and for the discrete FT it is

$$W_R(n) = DFT\{w_R(n)\} = \Delta t\frac{\sin\pi f_r(2M+1)}{\sin\pi f_r},$$

where Δt is the sample interval $\Delta t = 1/f_s$ and $(2M+1)$ data points lie within the window. The transition from the DFT to the CFT can be made by letting $\Delta t \to 0$, $f_s \to \infty$, $f \ll f_s$ so that $\sin\pi f_r \to \pi f_r$, and $\Delta t/\pi f_r = 1/\pi f_s f_r = 1/\pi f$. Considering that $T = (2M+1)\Delta t$, $\sin\pi f_r(2M+1) \to \sin\pi fT$. CFT becomes zero the first time for $fT = 1$, from which one determines the width of the main lobe as $2/T$. The 3 dB bandwidth is found as $0.89/T$ and the spectrum peak is $10\log T|e|^2 = 15$ dB. The width of the side lobes is $1/T$. For the DFT the corresponding values are obtained by letting $T \to N = 2M+1$. The total number SL of side lobes reveals the length of the data record: $SL+2 = N$. The first side lobe, important for the detection of a weaker source at a slightly different frequency, is down by 13.3 dB from the main lobe.

The effect of the triangular Bartlett window $w_B(t) = 1 - |n/M|$ with $n \le M \le N/2$ can be seen in Figure 7. The spectral shape is given by:

$$CFT\{w_B(t)\} = \frac{T}{2}\cdot\left(\frac{\sin(\pi fT/2)}{\pi fT/2}\right)^2$$

or

$$DFT\{w_B(n)\} = \frac{\Delta t}{M}\cdot\left(\frac{\sin\pi f_r M}{\sin\pi f_r}\right)^2$$

so that the main lobe width 4/T is twice that of the rectangular window and the 3 dB bandwidth is 1.28/T. The first side lobe is down by 26.5 dB. The main lobe peaks at $10 \log aN|e|^2$ with $a = 0.235$.

Spectra for the same sine wave record multiplied by a Hanning and Hamming data window are plotted in Figures 8 and 9, respectively. Both windows are raised cosine functions of the form $(0.5+\alpha) + (0.5-\alpha) \cdot \cos|\pi n/M|$ where $\alpha = 0$ for the Hanning window and $\alpha = 0.04$ for the Hamming window. The 3 dB bandwidth is 1.44/T for the Hanning window and 1.3/T for the Hamming window. The first side lobes are down by 31.5 dB and 43 dB, respectively. The side lobe ripple across the band i st constant for Hamming with a floor of 32 to 36 dB, whereas Hanning provides a steep roll-off of 18 dB/Octave. The main lobe peaks at $10 \log aNe^2$ with $a = 0.235$ for Hanning and $a = 0.275$ for Hamming.

Blackman-Tukey (BT) indirect spectra can be seen in Figure 10, based on the unbiased correlation function (CF), and Figure 11, derived from the biased CF, for a window size $M = 32$ points for $N = 128$ data samples. These two types of BT spectra will be called from now on as the unbiased and biased BT spectra (uBT, bBT). A triangular Bartlett window was chosen over the CF. As seen for the periodogram, the window size M determines the bandwidth also for the BT spectrum. However, whereas the window size for the periodogram is generally chosen as $2M = N$ as to include the entire data record, the window size over the CF for calculating the BT spectrum can vary within $M \leq (N-1)$ since the CF has twice as many samples as the original data record. The largest window chosen then will include all points of the CF.

A comparison of Figure 10 with Figure 6 shows that the uBT spectrum from 64 correlation values (M = 32) is identical with the periodogram having a rectangular window over 32 data points (N = 32, M = 16). This is true for the uBT with any N > M = 32. Therefore, the uBT spectrum has the bandwidth of $0.89/T_M$ for a continuous-time signal or 0.89/M for a discrete-time signal. Choosing the biased CF results in smoothing the BT spectrum as seen in Figure 11. The side lobe fluctuations of the bBT spectrum are calculated as 10 log N/M as can be verified also in Figure 12 for N = 32, M = 6. Three curves are overlaid in Figure 12. One is the uBT spectrum for N = 32, M = 6; the second is the corresponding bBT spectrum showing the reduced negative excursions of the PSD; and the third, in which the minimum possible number of data points N was chosen for M = 6 which drives 10 log N/M to zero and the bBT spectrum acts as an envelope follower for the uBT or the periodogram. The number of side lobes SL across the frequency band is directly related to M, SL = M-2.

In all following plots concerning the periodogram and the BT spectrum, white Gaussian noise of variance $\sigma_n^2 = 1V^2$ was added to the sine wave signal of power $\sigma_s^2 = 1V^2$, with a signal-to-noise ratio (SNR) of 10 log $\sigma_s^2/\sigma_n^2 = 0$ dB. Figure 13 gives a comparison between a periodogram with a rectangular window and one with a Bartlett window. Due to the broader lobe structure, the Bartlett window smoothens the periodogram resulting in lower and wider lobes. This is true also for the signal's main lobe.

The same characteristics can be seen also for the Hanning and Hamming window in Figures 14 and 15 when both are compared to the Bartlett window. In particular, the Hamming and Hanning windows generate almost identical sample

spectra for noisy signals. This is not surprising because of the similar structure of these two cosine windows.

Figure 16 shows overlays of five periodograms for a rectangular window, where each spectrum is based on a different data record of 64 samples to indicate the statistical variation inherent in sample spectra and to gain some insight into the aspects of improving the output SNR and signal detectability with spectral analysis. The mean of the signal-plus-noise at $f_r = 0.25$ is again determined by 10 log a Ne^2 ($a = 1$ for $w_R(t)$). The mean of the spectrum noise m_{PSD} at $f_r \neq 0.25$ for the rectangular window is the mean of the input noise power $\sigma_n^2 = 1V^2$, i.e., 0 dB. The positive noise fluctuations above this mean are best described by the standard deviation σ_{PSD}. The variance is

$$\sigma_{PSD}^2 \simeq \sigma_n^4 \left(1 + \left(\frac{\sin 2\pi N f_r}{N \cdot \sin 2\pi f_r}\right)^2\right).$$

The envelope of the second term within the parentheses approaches 0 for $|f_r| > 0$ rapidly with increasing N so that practically $\sigma_{PSD} \rightarrow \sigma_n^2$. If a threshold is set at a level equal to $m_{PSD} + \sigma_{PSD} = 2\sigma_n^2$ to define the equivalent spectrum noise, then this level would be a constant 3 dB above the mean of the spectral power, i.e., above the mean of the white noise input power. Defining the signal-to-noise (S/N) as the ratio of the mean signal to equivalent noise of the spectrum determines the SNR as

$$SNR = 10 \log(S/N) = 10 \log \frac{Ne^2}{2\sigma_n^2}.$$

Since the SNR of the input is 10 log $(e/\sigma_n)^2$, the SNR improvement or gain through spectral analysis using the periodogram derived for a rectangular data window would be 10 log(N/2), which is approximately 15 dB for N = 64 in Figure 16.

Comparing similar plots of five different data records for the Bartlett, Hanning and Hamming window with N = 64 in Figures 17, 18, and 19, it can be seen that the noise statistics are practically identical which should be expected considering the spectra in Figures 14 and 15. In a series of trials, it was established that for these three windows the mean of the spectrum noise m_{PSD} drops by 6 dB compared to the rectangular window with a standard deviation of 3 dB above this mean. However, since the mean of the signal-plus-noise peak is also about 6 dB down ($a \approx 0.235$, 0.275), the SNR and the gain remain practically the same for all four windows. This result is not surprising either since any window acts identically on the noise as on the signal. The major benefit of windowing with tapered functions then appears to be the side lobe depression for signals of high SNR.

The following plots are related to Blackman-Tukey spectra for noisy sinusoidal signals with both the signal power and the noise power equal to $1V^2$. Figures 20 and 21 indicate that no significant difference exists between the uBT and the bBT spectrum as long as the window size $2M \leq N$. In Figure 20 the data record is N = 128, and in Figure 21 it is N = 64; in both cases M = 32. The plots of Figure 22 compare a uBT and bBT sample spectrum for the limiting data record size N = 33 for M = 32 from which one can conclude that generally the bBT is a smoothed representation of the uBT spectrum for values of M/N such that $0.5 < M/N < 1$ with increased smoothing as $M/N \rightarrow 1$. Figure 23 shows that the uBT spectrum for N = M + 1 is identical to the periodogram for N = M. This compares directly to the signal-only case of Figures 10 and 6.

A high N/M ratio makes the Blackman-Tukey spectrum (biased or unbiased) useful as a substitute for ensemble averaging of sample periodogram spectra. In Figure 24 for N = 128, M = 32, and N/M = 4, this smoothing effect is equivalent to averaging four periodograms of 32 data samples; and in Figure 25 with N/M = 8, the equivalent ensemble average of eight 32-point periodograms is obtained. From an SNR standpoint , of course, it is better to use the data record available and perform the FT operation over the entire data set rather than dividing it in several parts and averaging the sample spectra. The signal gain in the first case achieves more than the noise reduction in the second case. However, in practice, spectrum analyzers (hardware or software) are designed for a maximum number of data samples. If the data record available exceeds this maximum size, then averaging is done profitably.

The results in Figures 26 and 27 give an indication of the variability of uBT sample spectra where five different data records were used. Figure 27 (N = 64), Figure 26 (N = 128), and Figure 25 (N = 256) represent BT spectra (it does not matter if biased or unbiased) for the same window size M = 32 with monotonically decreasing noise empirically established as

$$10 \log\left(\sigma_n^2\left(1 + \frac{2M}{N}\right)\right)$$

whereas the mean of the signal remains constant as 10 log Me^2 so that

$$\text{SNR} \approx 10 \log \frac{e^2}{\sigma_n^2} \frac{M}{1 + 2M/N}$$

and the gain over the input SNR is

$$G = 10 \log \frac{M}{1 + 2M/N}$$

for a gain of 14 dB in Figure 25, 13.3 dB in Figure 26, and 12 dB in Figure 27.

This concludes the application of the classical approach to spectral analysis.

CHAPTER 4

THE PARAMETRIC APPROACH TO SPECTRAL ESTIMATION

Conventional FT spectral analysis is based on a Fourier series model of the data, that is, the signal is assumed to consist of a set of harmonically related sinusoids. The direct approach (FT of the data) leads to higher resolution than the indirect approach (autocorrelation) because the correlation values of only $M < N$ lags are used for estimating the PSD. Data windowing, either purposely in order to decrease the variance of a sample spectrum or unintentionally because the record length is limited, is the fundamental factor that limits the spectral resolution. Windowing of the data or of the estimated correlation function makes the implicit assumption that the unobserved data or ACF values outside the window are either repetitive or zero. A smeared spectral estimate is the consequence.

If it were possible to "predict" or model the signal outside the observation window accurately, then the resolution could be improved. This is what the various parametric methods attempt to do. They try to model future data samples based on past and present data under the constraint that the predicted ACF is the most likely estimate which can be derived from the observed data. Because of the Wiener relationship between the ACF and the spectrum, this is equivalent to modeling the predicted spectrum as closely as possible to the observed spectral estimate.

Any spectrum can be obtained from a white noise source by filtering it with an appropriate filter. The task is then to design this filter so that the output is a good prediction in a least-square error sense of the data to be expected on the basis of past data.

CHAPTER 5

FILTERS WITH RATIONAL TRANSFER FUNCTIONS

Any linear filter with a transfer function H(z) relating the z-transform Y(z) of the output y(n) to the z-transform X(z) of the input x(n) can be constructed from the knowledge of its poles and zeroes[6,7] in the form of a rational function in polynomials of z (see Appendix A)

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^{q} b_i \cdot z^{-i}}{\sum_{i=0}^{p} a_i \cdot z^{-i}} = \frac{\sum_{i=0}^{q} b_i \cdot z^{-i}}{a_0 + \sum_{i=1}^{p} a_i \cdot z^{-i}}$$

Dividing numerator and denominator by a_0 and renaming $b_i/a_0 \rightarrow b_i$ and $a_i/a_0 \rightarrow a_i$ one gets

$$Y(z) + \sum_{i=1}^{p} a_i \cdot z^{-i} \cdot Y(z) = \sum_{i=0}^{q} b_i \cdot z^{-i} \cdot X(z)$$

Each term $z^{-i} \cdot X(z)$ is the z-transform of the input time series x(n) delayed by i sample time intervals, i.e., x(n-i). Equally, $z^{-i} \cdot Y(z)$ and y(n-i) are the z-transform pair of the output. Taking the inverse z-transform on both sides:

$$y(n) + \sum_{i=1}^{p} a_i \cdot y(n-i) = \sum_{i=0}^{q} b_i \cdot x(n-i)$$

$$y(n) = \sum_{i=0}^{q} b_i \cdot x(n-i) - \sum_{i=1}^{p} a_i \cdot y(n-i)$$

This is the general input-output relationship of a digital filter where the present output is a linear combination of the past p outputs and the present and past q inputs. The filter coefficients a_i and b_i are the parameters to be determined. The direct realization of the above expression can be seen in Figure 28 in which the block marked z^{-1} signifies one sample delay.

Two special cases lead to important filter classes. In the first case, all $a_i = 0$ (except $a_0 = 1$) in which case

$$H(z) = \sum_{i=0}^{q} b_i \cdot z^{-i}$$

$$y(n) = \sum_{i=0}^{q} b_i \cdot x(n-i)$$

Its realization is shown in Figure 29. This is the digital equivalent of the classical "Convolution Filter":

$$y(t) = \int_{\tau=0}^{\infty} h(\tau) \cdot x(t-\tau)\, d\tau$$

with the impulse response $h(\tau)$ replaced by the coefficients b_i.

Other names are "Finite Impulse Response" (FIR) filters, "Non-Recursive" filters, "Tapped-Delay Line" filters, "Transversal" filters, and "Moving Average" (MA) filters. If all b_i are equal, a normal average over a $(q+1)$ sample sliding window

results, otherwise a generalized weighted average. Finally, since this class of filters has no poles, it is also known as an "All-Zero" filter. Since no feedback is involved, these filters are always stable.

In the second case, all $b_i = 0$ (except $b_0 = 1$).

$$H(z) = \frac{1}{a_0 + \sum_{i=1}^{p} a_i \cdot z^{-i}}$$

$$y(n) = -\sum_{i=1}^{p} a_i \cdot y(n-i) + x(n)$$

Its realization is shown in Figure 30. This is called an "All-Pole" filter or an "Autoregressive" (AR) filter. It is this type of filter which has found widespread applications in the context of spectral estimation for short data records. It models an output y(n) based on the present input x(n) and the past p outputs. Having defined an AR filter and an MA filter, it follows that, in the general case of Figure 28, the filter is an "Autoregressive-Moving Average" (ARMA) filter. Since for the AR-filters the coefficient a_0 is just a gain factor, it can be set to 1, and

$$H(z) = \frac{1}{1 + \sum_{i=1}^{p} a_i \cdot z^{-i}}.$$

If one observes a random signal s(n) which one wants to substitute or model (in a spectral sense) by an equivalent output y(n) with a modeling filter, then the input should practically be a random process, preferably a white noise process. From such a white noise source, any desired output spectrum $P_y(f)$ can be obtained with a proper

filter that approaches the spectrum $P_s(f)$ of the actually observed signal s(n). From the input-output relationship of a linear filter:

$$Y(z) = H(z) \cdot X(z)$$

one gets the spectral relationship:

$$P_y(f) = |H(f)|^2 \cdot P_x(f) = \frac{P_x(f)}{\left| 1 + \sum_{i=1}^{p} a_i \cdot z^{-i} \right|^2_{z = e^{j2\pi f \Delta t}}}$$

and for a white noise input x = u and power σ_u^2

$$P_y(f) = \frac{\sigma_u^2}{\left| 1 + \sum_{i=1}^{p} a_i e^{-j2\pi f i \Delta t} \right|^2}$$

where Δt is the sampling interval in seconds. If the filter parameters a_i and the input noise power σ_u^2 are known, then the output power density can be determined for any desired frequency as a continuous-frequency spectrum. The filter itself becomes an equivalent representation of the spectrum of the observed signal. The input noise power acts just as a scale factor.

CHAPTER 6
YULE-WALKER EQUATIONS

References 1, 2, 8, 9, and 10 provide the mathematical background for the following material.

In power spectral estimation the autocorrelation function plays a fundamental role (Wiener-Khinchin), regardless if one approaches the problem with one of the classical or parametric methods. It follows that establishing a relation between the filter-parameters (representing the spectrum) and the ACF will provide the solution to determine the AR coefficients a_i. The development of this relation leads to the so-called Yule-Walker normal equations (G. Yule, 1927; G. Walker, 1931) after the two statisticians who used them in their AR filter models to predict trends in economic time series. They were also derived as the Wiener-Hopf equations to solve for the optimum prediction filter.

These equations derived in Appendix B are:

$$r_{yy}(k>0) = -\sum_{i=1}^{p} a_i \cdot r_{yy}(k-i)$$

$$r_{yy}(k=0) = -\sum_{i=1}^{p} a_i \cdot r_{yy}(-i) + \sigma_v^2$$

where $r_{yy}(k)$ is the value of the ACF for a lag $k\Delta t$ of the data modeled from a white noise source and an AR filter with coefficients a_1 through a_p. The constraint is to devise the filter such that the ACF $r_{yy}(k)$ of the modeled data y(n) be equivalent to the ACF $r_{ss}(k)$ of the actually observed data s(n). Therefore, one can replace $r_{yy}(k)$ in the expressions above by $r_{ss}(k)$ subject to the availability of observed data.

The first expression sets up a system of p equations $r_{ss}(k=1)$ to $r_{ss}(k=p)$ with p unknowns a_1 to a_p. The second expression provides one equation with $r_{ss}(k=0)$ for the only unknown σ_u^2. In order to solve for the unknown filter coefficients and the driving noise source, the ACF must be estimated from the available data. Once this has been done, the $(p+1)$ equations can be solved, in principle, by the Gaussian Elimination method; this requires a number of operations proportional to p^3, a rather time consuming operation. Any analysis algorithm to be applied in real-time requires fast numerical algorithms, and so the Durbin-Levinson Recursion method was developed which requires only p^2 operations to solve the Yule-Walker system of equations (J. Durbin, 1960; N. Levinson, 1947).

CHAPTER 7
LEVINSON-DURBIN RECURSION

The proof of this recursive algorithm is lengthy and can be found in Reference 1 (pp. 161-171). Therefore, only the final equations are given here which determine the desired AR coefficients. The number p of the coefficients a_1 through a_p is called the model order. It signifies the number of feedback terms each with its own coefficient. The significance of the recursion of the Levinson method lies in the fact that it evaluates the AR coefficients starting out from the lowest model order p = 2 (after initialization of the procedure for p = 1), then working itself to the next higher model order (p = 3) and continues doing so up to the desired order. At this point, the algorithm has not only designed the filter of the desired order, but also all filters of lower order. This allows one to chose an AR filter whose model order appears to be most effective for the data process on hand.

Because of this recursive feature of the algorithm the coefficients a_i are given two subscripts: p for the model order, and i as the running index from 1 through p. For instance, $a_{p,i} = a_{10,3}$ would be the coefficient a_3 of an AR filter of model order 10.

The Levinson algorithm starts out calculating the highest coefficient (i = p) with one formula and then proceeds to calculate the coefficients for i = 1,2,....,p-1 with a second formula. A third expression determines the power of the driving noise source, also recursively from lower to higher model orders.

The equations are:

$$a_{pp} = - \frac{r_{ss}(p) + \sum_{\ell=1}^{p-1} a_{p-1,\ell} \cdot r_{ss}(p-\ell)}{\sigma^2_{p-1}} \qquad \text{where } \ell \text{ is a dummy parameter}$$

$$a_{pi} = a_{p-1,i} + a_{pp} \cdot a^*_{p-1,p-i} \qquad i = 1,2,3,\ldots,p-1$$

a^* = conjugate complex of a (in case of complex data)

$$\sigma^2_p = (1 - |a_{pp}|^2) \cdot \sigma^2_{p-1}$$

These recursions are initialized with

$$a_{11} = - \frac{r_{ss}(1)}{r_{ss}(0)} \quad \text{and} \quad \sigma^2_1 = (1 - |a_{11}|^2) \cdot r_{ss}(0).$$

The highest-index coefficient a_{pp} plays a significant role in Burg's algorithm, described later. This coefficient is known as "reflection coefficient" k_p.

A sample calculation is given in Appendix C.

CHAPTER 8
LINEAR PREDICTIVE FILTERING

The AR filter approach as described above models the actual signal s(n) by filtering white noise appropriately. It is possible to give this modeling approach a different physical interpretation. This leads to the concept of "Linear Prediction" or "Linear Predictive Coding" (J. Makhoul, 1975, see Reference 11). In linear prediction, one assumes that the input (white noise source u(n) in autoregression) is unknown and that, therefore, information about the actual signal s(n) can be gained only from the past p outputs. A linear combination of these outputs will result in a signal estimate $\hat{s}(n)$ which only approximates the true signal s(n):

$$\hat{s}(n) = -\sum_{i=1}^{p} a_i \cdot s(n-i).$$

Comparing this prediction filter with Figure 30, one can see that it is equivalent to the AR filter with the input removed. The error between the actual value s(n) and the prediction $\hat{s}(n)$ is

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^{p} a_i \cdot s(n-i).$$

The coefficients a_i are determined by minimizing the total squared error e with respect to each of these parameters, i.e., by setting the partial derivative

$$\frac{\partial}{\partial a_i}\left(\sum_n e^2(n)\right) = 0$$

The derivation is given in Appendix D. The result is:

$$\frac{\partial}{\partial a_i}\left(\sum_n e^2(n)\right) = \sum_n s(n)\cdot s(n-k) + \sum_{i=1}^{p} a_i \sum_n s(n-i)\cdot s(n-k) = 0$$

which leads by ensemble averaging to

$$r_{ss}(k) = -\sum_{i=1}^{p} a_i \cdot r_{ss}(k-i) \qquad 1 \le k \le p$$

and the actual minimum error power is

$$e^2_{min} = \sum_{i=1}^{p} a_i \cdot r_{ss}(i) + r_{ss}(0)$$

or

$$r_{ss}(0) = -\sum_{i=1}^{p} a_i \cdot r_{ss}(i) + e^2_{min}$$

These expressions are formally identical with the Yule-Walker equations developed for the autoregressive filter with the only difference that the input noise power σ_u^2 of the AR filter appears as the output error power e_{min}^2 of the predictor. The close relationship between the AR model and the prediction filter is also apparent from the sample error expression

$$e(n) = s(n) + \sum_{i=1}^{p} a_i \cdot s(n-i)$$

which can be written as

$$s(n) = -\sum_{i=1}^{p} a_i \cdot s(n-i) + e(n).$$

This is the same equation used in the AR process, with $e(n) = u$, so that solving for the AR coefficients a_i will determine the identical parameter set of the predictor.

The interpretation of the All-Pole AR filter as an optimum predictive filter is not just a reformulation of the same problem but leads actually to new algorithms and results as seen later on for Burg's method and others.

CHAPTER 9

RESOLUTION ASPECTS FOR SHORT DATA RECORDS

The motivation behind autoregressive spectral estimation or linear prediction filtering is to eliminate the window effect associated with a limited data record that limits the resolution of the analysis and creates distortions through side lobes. AR spectral estimation attempts to extract "enough" information from the limited data record so that the autocorrelation function can be estimated beyond the lags practically possible for the given data record. The extension of the ACF is recursive through the first Yule-Walker equation

$$r_{ss}(k) = -\sum_{a_i=1}^{p} a_i \cdot r_{ss}(k-i)$$

which allows one to calculate the coefficients a_i, given the values $r_{ss}(k)$ for $0 \le k \le p$. Assuming short-term stationarity of the signal beyond the available data record these coefficients can be considered constant, and extended values of $r_{ss}(k)$ can be extrapolated for $k > p$. From an information standpoint, this modeling approach permits a more realistic continuation of the ACF for higher lags rather than setting it to zero or making it repetitive. In fact, Burg (1961) derived the same algorithm through a third interpretation, based on the information concept of "maximum entropy" (see References 10 and 12). He extends the ACF beyond lag p by adding the least amount of new information, therefore "maximizing the entropy" of the process. This extension is the basis for the higher resolution of the AR-PSD estimation compared to the classical FT-based analysis. AR modeling then is a process of

spectrally matching the observed data to modeled data, whereby it should be understood that the modeled time series data are generally random and are therefore in no way matched to the observed data. An infinite number of random noise time series data can produce the same ACF and power spectrum.

CHAPTER 10
ESTIMATION OF THE AR COEFFICIENTS

THE AUTOCORRELATION AND COVARIANCE METHODS

The Yule-Walker equations establish the relationship between the AR filter coefficients and the autocorrelation function of the data. Solving them depends on knowing the exact autocorrelation values $r_{ss}(k)$ for a number of lags k equal to the model order p of the AR filter. Generally, this ACF is not given, and it would require data over an infinite time interval. Usually only a limited amount of data is available from which one can calculate estimated correlation values and then proceed to obtain the coefficients a_i using the Levinson recursion. This is called the Yule-Walker method. Other methods have been developed to derive the filter parameters from the data using the least square error criteria of the prediction filter. As will be shown, the Yule-Walker method can be thought of as a short-time correlation which depends not only on the lag but also on the summation range n.

If N data samples s(0),s(1),....,s(N-1) of a time series are known, then (p + 1) data samples are used to calculate one error sample according to

$$e(n) = \sum_{i=0}^{p} a_i \cdot s(n-i) \qquad \text{with } a_0 = 1 .$$

Appendix E shows the system of linear equations involving all errors e(n) possible for the observed set of data. It can be seen that the available data can be used to form

a maximum of (N + p) error samples: e(0) to e(N + p-1). However, for e(0), only one valid data sample s(0) is available, all others are set to zero, since they are unknown. For e(1) only the first two data samples s(1) and s(0) are non-zero, and so on. The same is true at the end of the error sample series where the last error sample (N + p-1) has been calculated with only the last valid signal sample s(N-1), the one before that, e(N + p-2) with only two: s(N-1), s(N-2), etc. Therefore, the number of valid signal samples in the calculated error sample series can be pictured as in Figure 31.

Now going back to the expression which minimizes the total squared error

$$\sum_{n} |e(n)|^2$$

of a prediction filter over a given range n of data samples s(n), one has:

$$\sum_{i=1}^{p} a_i \sum_{n} s(n-i) \cdot s(n-k) = -\sum_{n} s(n) \cdot s(n-k) \qquad 1 \le k \le p$$

or in matrix notation as found throughout the literature:

$$\begin{bmatrix} r(1,1) & r(1,2) & \cdots & r(1,p) \\ r(2,1) & r(2,2) & \cdots & r(2,p) \\ \vdots & & & \\ r(p,1) & r(p,2) & \cdots & r(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} r(1,0) \\ r(2,0) \\ \vdots \\ r(p,0) \end{bmatrix}$$

where

$$r(i,k) = \sum_{n} s(n-i) \cdot s(n-k)$$

so that

$$\sum_{i=1}^{p} a_i \cdot r(i,k) = -r(k)$$

If n goes to infinity, the sum of the lagged products represents the ACF and the result is the set of Yule-Walker equations. For a finite n, however, with $0 \le n \le (N-1)$, there are four distinct summation ranges over which the total squared error can be minimized, as can be seen from Figure 31. These four ranges are:

$$\sum_{n=0}^{N-1+p} \quad \text{or} \quad \sum_{n=0}^{N-1} \quad \text{or} \quad \sum_{n=p}^{N-1+p} \quad \text{or} \quad \sum_{n=p}^{N-1}$$

Each of these four cases will lead to a different set of p equations and result in a different set of the a_i parameters, i.e., a different AR filter and its associated spectrum. These differences will become less and less significant the larger the data record length becomes compared to p, and all four cases approach the Yule-Walker method of the true ACF.

The first case

$$\left(\sum_{n=0}^{N-1+p} \right)$$

is called the "windowed" method, since all s(n) outside the window $0 \le n \le N-1$ have been set to zero. For the same reason, the second case

$$\left(\sum_{n=0}^{N-1}\right)$$

is called the "prewindowed," the third case

$$\left(\sum_{n=p}^{N-1+p}\right)$$

the "postwindowed," and the fourth

$$\left(\sum_{n=p}^{N-1}\right)$$

the "nonwindowed" method. In this case, none of the data for calculating the error samples has been zeroed out. In much of the literature the first method is also called the "autocorrelation" and the last one the "covariance" method. (This terminology has old roots, and semantics is the reason that makes reading the literature harder than necessary). Of the four methods, these two are found most frequently, where the covariance method leads to better resolution spectra than the autocorrelation method. It is intuitive that the calculation of the AR coefficients based on the largest error series e(0) to e(N + p-1) is not optimum because of the sloped areas of Figure 31 between 0 to p and (N-1) to (N-1 + p) that are based on nonexisting data.

THE FORWARD-BACKWARD PREDICTION METHOD

The prediction filter looked at so far is also known as a forward predictor. It estimates future data from past data, going forward in time. A backward predictor is

then a filter which calculates past data samples s(n-p) from the "future" p data samples s(n-p + 1) to s(n), going back in time. Both predictors use the same data samples, see Figure 32. The forward prediction is

$$\hat{s}(n) = -\sum_{i=1}^{p} a_i^f \cdot s(n-i)$$

with the error

$$e^f(n) = \sum_{i=0}^{p} a_i^f \cdot s(n-i).$$

Similarly for the backward prediction:

$$\hat{s}(n-p) = -\sum_{i=1}^{p} a_i^b \cdot s(n-p+i)$$

with the error

$$e^b(n) = \sum_{i=0}^{p} a_i^b \cdot s(n-p+i).$$

For a stationary process the forward and backward AR coefficients a_i^f and a_i^b are the same (or the conjugate complex of each other, $a_i^b = (a_i^f)^*$, if complex data are handled), because it involves the same statistical information going forward or backward in time. Therefore, one can combine the forward and backward errors in order to get twice the number of error samples compared to the forward-only nonwindowed (covariance) method of range $p \le n \le N-1$. This leads to the so-called

"Modified Covariance" method which is based on the minimization of the average combined forward/backward squared error:

$$\rho = \frac{1}{2}\left[\sum_{n=p}^{N-1} |e^f(n)|^2 + \sum_{n=p}^{N-1} |e^b(n)|^2 \right].$$

Minimizing this expression by differentiating it with respect to the a_i and setting it to zero yields:

$$\left[\sum_{i=1}^{p} a_i \sum_{n=p}^{N-1} s(n-i)\cdot s^*(n-k) + \sum_{n=p}^{N-1} s(n)\cdot s^*(n-k) \right]$$

$$+ \left[\sum_{i=1}^{p} a_i \sum_{n=p}^{N-1} s^*(n-p+i)\cdot s(n-p+k) + \sum_{n=p}^{N-1} s^*(n-p)\cdot s(n-p+k) \right] = 0$$

The expression in the first bracket is the contribution from the forward error which is identical to that of the covariance method, and the expression in the second bracket comes from the backward error. This can be written similarly as for the covariance method:

$$\sum_{i=1}^{p} a_i \cdot r(i,k) = -r(k) \qquad 1 \le k \le p$$

where

$$r(i,k) = \sum_{n=p}^{N-1} \left(s(n-i)\cdot s^*(n-k) + s^*(n-p+i)\cdot s(n-p+k) \right)$$

and

$$r(k) = \sum_{n=p}^{N-1} \left(s(n) \cdot s^*(n-k) + s^*(n-p) \cdot s(n-p+k) \right)$$

THE BURG METHOD—MINIMIZING THE REFLECTION COEFFICIENTS

This method was derived by John Burg (1967) and has become a widely used technique to determine the AR coefficients. It should not be mistaken for Burg's maximum-entropy interpretation of AR spectral estimation, mentioned earlier. It is strongly related to the forward-backward prediction by minimizing the combined forward/backward squared errors. However, while those errors $e_p^f(n)$ and $e_p^b(n)$ were calculated straight forward for a given model order p, in Burg's method they are recursively derived from the errors $e_{p-1}^f(n)$ and $e_{p-1}^b(n)$ of model order (p-1) by requiring that the coefficients a_{pi} follow the Levinson recursion

$$a_{pi} = a_{p-1,i} + k_p \cdot a^*_{p-1,p-i}$$

where $k_p = a_{pp}$ is the "reflection coefficient" as mentioned earlier. In Appendix E, the expressions for the recursive errors are derived:

$$e_p^f(n) = e_{p-1}^f(n) + k_p \cdot e_{p-1}^b(n-1)$$

$$e_p^b(n) = e_{p-1}^b(n-1) + k_p^* \cdot e_{p-1}^f(n)$$

The average combined squared error for the nonwindowed case is then

$$E_p = \frac{1}{2} \sum_{n=p}^{N-1} \left(\left| e^f_{p-1}(n) + k_p \cdot e^b_{p-1}(n-1) \right|^2 + \left| e^b_{p-1}(n-1) + k_p^* \cdot e^f_{p-1}(n) \right|^2 \right)$$

which is only a function of k_p, since the lower order prediction error powers $e_{p-1}{}^f(n)$ and $e_{p-1}{}^b(n)$ are known (having been calculated starting from the first order model). Therefore,

$$\frac{\partial E_p}{\partial (\mathrm{Re}(k_p))} + j \frac{\partial E_p}{\partial (\mathrm{Jm}(k_p))} = 0$$

allows one to solve for k_p as the only unknown. This results in:

$$k_p = \frac{-2 \sum_{n=p}^{N-1} e^f_{p-1}(n) \cdot \left(e^b_{p-1}(n-1) \right)^*}{\sum_{n=p}^{N-1} \left| e^f_{p-1}(n) \right|^2 + \sum_{n=p}^{N-1} \left| e^b_{p-1}(n-1) \right|^2} \qquad \text{(Proof, References 1,2)}$$

The initialization of the Burg method is done by

$$\sigma_o^2 = r_{ss}(0) = \frac{1}{N} \sum_{n=0}^{N-1} |x(n)|^2$$

$$e^f_o(n) = s(n) \qquad n = 1,2,3,\ldots,(N-1)$$

$$e^b_o(n) = s(n) \qquad n = 0,1,2,\ldots,(N-2)$$

Based on these initialization errors the value of k_1 can be calculated. In order to determine k_2, all errors $e_1{}^f(n)$ and $e_1{}^b(n\text{-}1)$ for $n = p, p+1, \ldots, N\text{-}1$ must be calculated from the recursive error expressions

$$e_p^f(n) = e_{p-1}^f(n) + k_p e_{p-1}^b(n-1) \qquad n = p+1, p+2, \ldots, N-1$$

and

$$e_p^b(n) = e_{p-1}^b(n-1) + k_p^* e_{p-1}^f(n) \qquad n = p, p+1, \ldots, N-2$$

With k_2 determined, the errors can be updated again to give the error series for $e_2^f(n)$ and $e_2^b(n-1)$ which are used to compute k_3, and so on, until k_p of the desired model order p has been reached. After each of these cycles, the Levinson recursion is used to find the coefficients a_{pi} from the k's:

$$a_{pi} = a_{p-1,i} + k_p \cdot a_{p-1,p-i}^*$$

and the minimized error power output is

$$\sigma_p^2 = (1 - |k_p|^2) \cdot \sigma_{p-1}^2$$

Finally, the spectral estimate is

$$P_s(f) = \frac{\sigma_p^2}{\left| 1 + \sum_{i=1}^{p} a_i z^{-i} \right|_{z = e^{j2\pi f \Delta t}}^2}$$

$$P_s(f) = \frac{\sigma_p^2}{\left| 1 + a_1 e^{-j2\pi f \Delta t} + a_2 e^{-j4\pi f \Delta t} + \ldots + a_p e^{-j2\pi p f \Delta t} \right|^2}$$

This completes the Burg algorithm. It should be mentioned briefly here that the two error recursion formulae lead to a different interpretation of the prediction error filter, the so-called lattice filter, depicted in Figure 33. The input signal s(n) is the observed signal; the error signals $e_p^f(n)$ and $e_p^b(n)$ are generated as output. The lattice filter is optimally matched to the input signal in the sense that the output error tends to go to zero if the filter is perfectly matched to the input spectrum, i.e., the spectral estimate as calculated above, is equal to the input spectrum. For a signal with slowly changing statistics, the coefficients a_i can be made to change, in effect tracking the input spectrum, such that the output is always minimized. The lattice filter has found its use in adaptive filtering.

CHAPTER 11
APPLICATIONS OF PARAMETRIC SPECTRUM METHODS

RESOLUTION CAPABILITY IN THE (NEARLY) NOISELESS CASE

The following is related to narrowband signals (ideally sinusoidal) in white noise. As stated earlier, the record length T of an observed signal inherently puts a limit on the spectral resolution, i.e., on the capability to resolve two closely spaced frequencies. This limit of resolution is generally taken as

$$(\Delta f)_{res} = \frac{1}{T}$$

For a given signal frequency f and a record length T the number C of cycles in T is

$$C = Tf$$

A higher frequency $(f + \Delta f)$ will allow a larger number of cycles $(C + \Delta C)$ in the same record T

$$C + \Delta C = T(f + \Delta f) \qquad \text{i.e., } \Delta f = \Delta C/T$$

This incremental frequency Δf is equivalent to the limit of resolution for $\Delta C = 1$ cycle. This means two frequencies can be resolved if the higher frequency signal contains at least one more cycle than the lower frequency signal in the record T.

Figure 34 and 35 depict the real part and the magnitude of a complex transient sinusoid containing two components of relative frequency 0.25 and 0.3125 of equal amplitude and with a relative phase shift of 160 degrees for a record length of N = 16 data points. This results in four cycles at the lower and five cycles at the higher frequency and the one-cycle difference necessary for spectrally resolving the two components.

The periodogram is shown in Figure 36, and it indicates the importance of the relative phase between components for short duration signals. For a phase shift of 335 degrees, the two frequencies are clearly resolved, but biased compared to the true frequencies indicated by the two dashed vertical lines. At 270 degrees, both components start to separate, still biased, and at 160 degrees just one spectral peak is located at the average frequency. The periodogram method clearly requires more than a one-cycle difference to reliably resolve two frequencies of difference Δf, or stated differently, the record length must be larger than $1/\Delta f$.

The Burg method has been chosen for calculating the parametric spectrum of Figure 37 for the most difficult phase shift of 160 degrees, showing the superiority of the autoregressive modelling approach.

In the following three illustrations, Figures 38, 39, and 40, the resolution capability of three spectral techniques, the FT based, the Burg, and the Modified Covariance method, is determined more accurately. In each case, the lower relative frequency is 0.25. The upper one is varied in order to accommodate a desired cycle difference ΔC. The two amplitudes are equal, and some noise is added to prevent singularities (SNR = 26 dB).

Figure 38 is a plot of the periodogram for three cases, where ΔC is 1.3, 1.4, and 1.5 cycles, corresponding to a frequency of 0.33125, 0.3375, and 0.34375. The most unfavorable relative phase with respect to resolution was chosen. The PSD is plotted on a linear rather than on the ordinary dB-scale which allows better recognition of the two PSD peaks to develop. It can be seen that a minimum difference of 1.4 to 1.5 cycles is necessary for an FT to resolve two frequencies.

Figure 39 shows the Burg spectrum of order $p = 4$ for $\Delta C = 0.5$ and $\Delta C = 0.7$ cycles, corresponding to $f = 0.28125$ and $f = 0.29375$. An excess of 0.7 cycles will separate the two frequencies, however, with a bias depending on the relative phase $\Delta\Phi$. It can be shown that this bias varies as a function of $\sin \Delta\Phi$.

Figure 40 is a plot of the Modified Covariance spectrum for $\Delta C = 0.3$ and $\Delta C = 0.4$ cycles, corresponding to $f = 0.26875$ and $f = 0.275$. This method then requires an excess of only 0.4 cycles to resolve two frequencies. Also, a significantly smaller frequency bias than for the Burg method exists relative to phase over the entire 2π-range.

RESOLUTION OF SMALL NOISY SIGNALS IN THE PRESENCE OF LARGE SIGNALS

Under noisy conditions, the ideal resolution capability demonstrated above will be degraded. Given are $N = 16$ data points of a sinusoid of relative frequency $f_1 = 0.25$ ($N_1 = 4$ cycles) and 1-volt amplitude. A second sinusoid of the same sample size is chosen such that its frequency f_2 coincides with the frequency of the first

sidelobe maximum of the periodogram at f = 0.34 (N_2 = 5.44 cycles). Its amplitude is 0.1 volt. White noise is added with a standard deviation of 0.1 Vrms, so that SNR(f_1) = 20 dB and SNR(f_2) = 0 dB.

Figure 41 is a representation of the FT based power spectrum for five independent data records indicating the effect of the additive noise. No clue is found in these spectra about the presence of the second sinusoid.

Figure 42 shows the same five simulations using a Hamming window. Again it will be quite difficult to identify and locate the second small signal component. In Figure 43 the Burg algorithm with a model order p = 8 was applied to the same five data records with the result that in four trials the small signal can be identified using simple thresholding and judging the consistency of peaking in the spectrum.

A similar experiment was done selecting the frequency f_2 = 0.405 (N_2 = 6.5 cycles) to coincide with the second sidelobe maximum of the large signal at f_1 = 0.25. The periodogram with a rectangular window was calculated and plotted in Figure 44; the Hamming window was applied in Figure 45. Not much information can be gained from both sets of spectra about the presence of a second signal. The Burg algorithm of model order p = 8, however, provides a spectrum in Figure 46 in which both frequencies can be identified and located.

In a third experiment, the second signal was hidden in the minimum between the main lobe and the first side lobe of the FFT-based spectrum of Figure 47 at f_2 = 0.3125 (N_2 = 5 cycles). Figure 48 shows the spectrum using again the Hamming window. The Burg spectrum (p = 8) is displayed in Figure 49 with a similar result as for the two previous trials.

These experiments allow the conclusion that small noisy signals are difficult to detect across an extended part of the spectrum and particularly so close to the mainlobe of a large signal using the conventional Fourier method. This gap is filled by the parametric method of AR spectral analysis.

COMPARISON OF PARAMETRIC METHODS

The previous series of tests was expanded to include the other parametric algorithms discussed in this report, the autocorrelation, the covariance, and the Modified Covariance method, together with the Burg algorithm. The first relative frequency is again at 0.25, 16 data points form a record of four cycles; for the second signal 5.5 cycles were selected leading to $f_2 = 0.344$, well within the resolution limits of parametric spectra for any phase shift between the two components. The SNR of the first signal is +20 dB, that of the second is 0 dB.

Figure 50 shows what the autocorrelation algorithm can achieve at best. Even at these relatively high model numbers $p = 8$ and 10, this method proves unsatisfactory to represent a realistic spectrum.

The results of the covariance method are plotted in Figure 51 for $p = 4, 5, 6$. At $p = 4$ the resolution is too small, and the spectrum peaks only at the main signal frequency. For the next higher order model, the second component starts to show up, and it is well established for $p = 6$. In either case, the second peak occurs somewhat off the actual frequency. Also note that for $p = 6$ the weak component is calculated to

have a higher spectral value than the main component, indicating that the model order is too large. Going to still higher orders will cause the spectrum to break up into more and more peaks unrelated to the actual signals.

The Modified Covariance method was used for Figure 52. Also for this algorithm, model order 4 generates only the main peak; $p = 5$ indicates the second sinusoid, and $p = 6$ produces a satisfactory spectrum. The Burg algorithm used in Figure 53 produces similar spectra for the same model orders. These two algorithms have been found to be equally effective for various signal and noise conditions; both are relatively robust in the sense that they often (not always) tolerate a higher model order without breaking up. The Modified Covariance method has an edge over Burg with a somewhat better resolution capability, and it is also less affected by phase variations. However, the Burg algorithm is faster. For a data record $N = 64$ it took 5 seconds on a PC to calculate the spectrum for $p = 24$, but 20 seconds for the equivalent Modified Covariance model.

SPECTRAL PERFORMANCE AT LOW SIGNAL-TO-NOISE RATIOS

Generally the benefit of high resolution is derived from AR spectral analysis at high SNR conditions. The following five sets of spectra in Figures 54 through 58 are the result of five simulations of a 16-sample signal at $f = 0.25$ and of white noise with an SNR = 0 dB. In each set, the model order is $p = 6$. The autocorrelation (Figure 55), the Modified Covariance (Figure 57), and the Burg method (Figure 58) show a detection performance comparable to the periodogram (Figure 54), however,

with better resolution. The covariance method of Figure 56 yields highly unstable spectra that are very sensitive to the individual noise record.

Going to still smaller signals with an SNR = -6 dB for the same N = 16 data record, no useful information is obtained any longer from any spectral method, parametric or Fourier based. Figures 59 through 63 show the results. Coherent averaging of the five spectra of each set will improve the performance; however, this would imply longer data records which may or may not exist. This concludes the testing of the AR spectral estimation methods discussed in the report and their comparative performance.

CHAPTER 12
CONCLUSIONS

Whenever new methods are tried and added to the pool of existing signal processing tools, it is desirable to be aware of their capabilities and limitations. What do they allow us to do that could not be accomplished before with the tools available? This is true also when considering the so-called "modern spectral analysis" compared to "conventional Fourier analysis."

FFT-based spectral analysis is the most robust method for widely ranging signal and noise conditions. Particularly for low to moderate SNR's, the periodogram with its numerous window options can be well adapted to the given signal and noise characteristics and will give satisfactory results. Also, it does not make any demands on the bandwidth of the signal and performs well for broadband or narrowband signals. The analysis of a broadband process is done best through the conventional FFT approach. The identification of narrowband components is possible even for transients as long as the cycle excess between components is at least a few cycles, theoretically down to one cycle. All of this can be done in real time because of the fast FFT algorithm.

When all is said, a few well-defined problems remain which cannot be solved conventionally because of properties inherent to the FT when applied to time limited data. The Fourier Transform is strictly defined for unlimited data and is therefore a mathematical rather than an engineering concept. When dealt with in a mathematical context, there are no resolution limits or side lobe effects. However, all

engineering applications involve data that exist only during a given observation time, or that are intentionally truncated, i.e., the data are windowed. A time-limited window function is transformed by the FT into an unlimited frequency space through more or less pronounced side lobes. If somewhere in these side lobes another signal is present smaller than the side lobe level, it cannot be detected.

The modeling approach of the autoregressive analysis is capable of solving such a problem. By substituting the observed data by a model of the process that generated those data, one has access to an unlimited source of equivalent data. This model is derived from the information of the actual signal, essentially its associated correlation function. Since no limits are imposed on the output of the model, the spectral estimation can be done without generating side lobes. A peak in the spectrum related to a sinusoid in the signal will decay asymptotically to zero, so that another smaller sinusoid, even in the vicinity of the main peak, can still be detected.

The model is actually a filter which is designed such that, for each sinusoid contained in the signal, a pole at the corresponding frequency is generated. This gives rise to a resonance effect that is very sensitive to frequency and that generates a strongly pointed peak in the transfer function resulting in better resolution. This high resolution makes it possible to separate two components in frequency even if their cycle difference is less than one cycle, the theoretical limit for conventional analysis.

These are the two essential characteristics of parametric AR spectral analysis that provide a new tool in signal processing and fill a true gap existing in classical analysis. They also indicate the problems associated with this approach. First, it is strictly line oriented due to poles in the transfer function. Second, one has to make

assumptions about the number of lines in the spectrum corresponding to the model order, i.e., one has to have apriori information that is not normally available.

Several criteria exist for choosing the proper model order. No attempt has been made in this report to deal with model order selection since this is a subject by itself. It may suffice here to state that the existing criteria are not hard and fast; they rather provide some guidelines. Generally, parametric spectra do not appear to be very sensitive to model order in large SNR cases (the main working domain for AR analysis) if one is interested primarily in frequency identification; resolution, however, improves with increasing order. Needless to say that computation time balloons for higher model order interfering with many real time applications.

Finally, it should be pointed out that a peak in the AR spectrum resulting from a sine wave component is not proportional to the power of the sinusoid as it is for the periodogram, rather it is proportional to the square of the power. The integrated area under the peak, however, still represents correctly the power as it should be for a power spectral density curve.

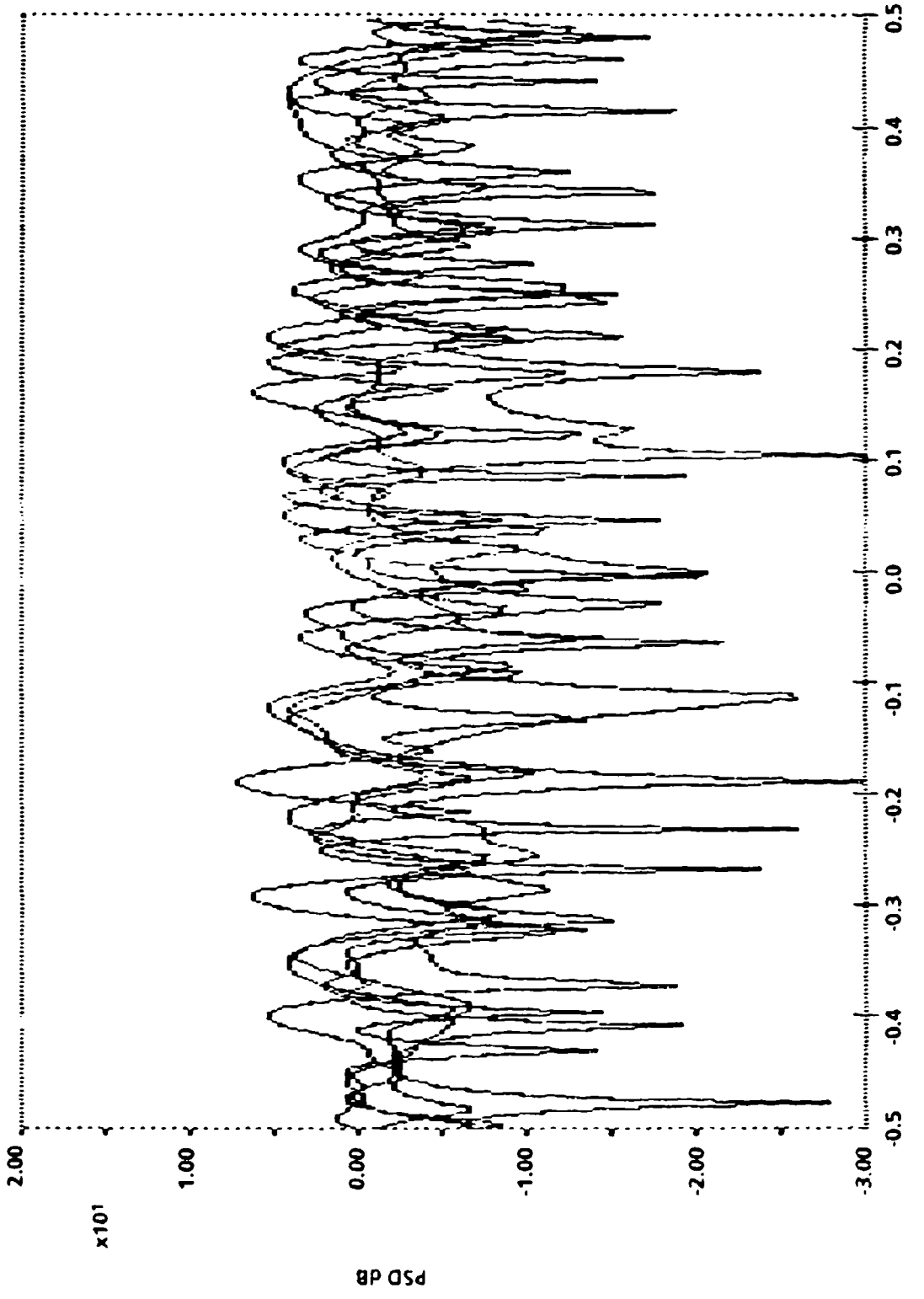


FIGURE 1. FIVE PERIODOGRAMS OF 1 VRMS WHITE NOISE, 32 DATA SAMPLES

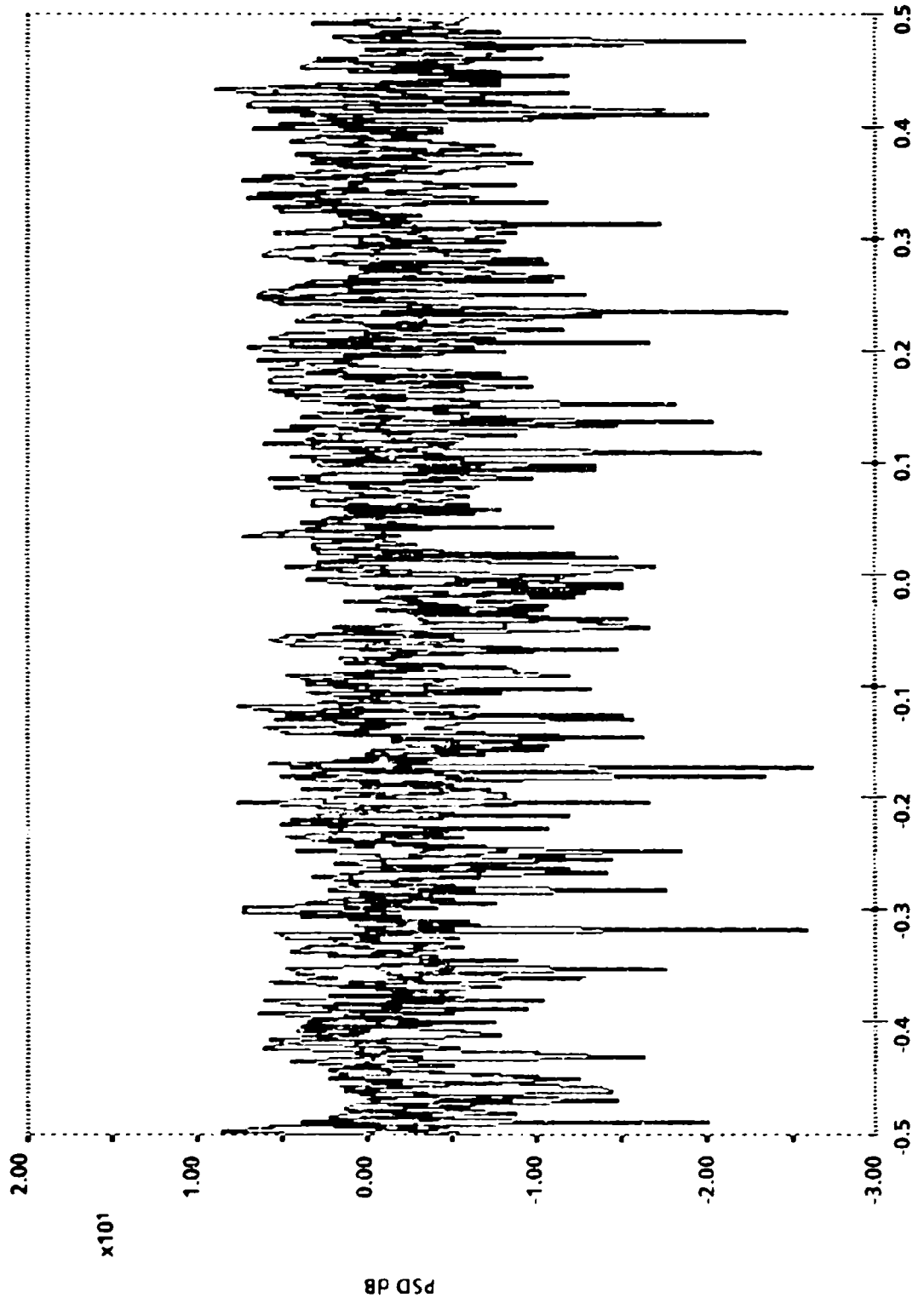


FIGURE 2. FIVE PERIODOGRAMS OF 1 VRMS WHITE NOISE, 128 DATA SAMPLES

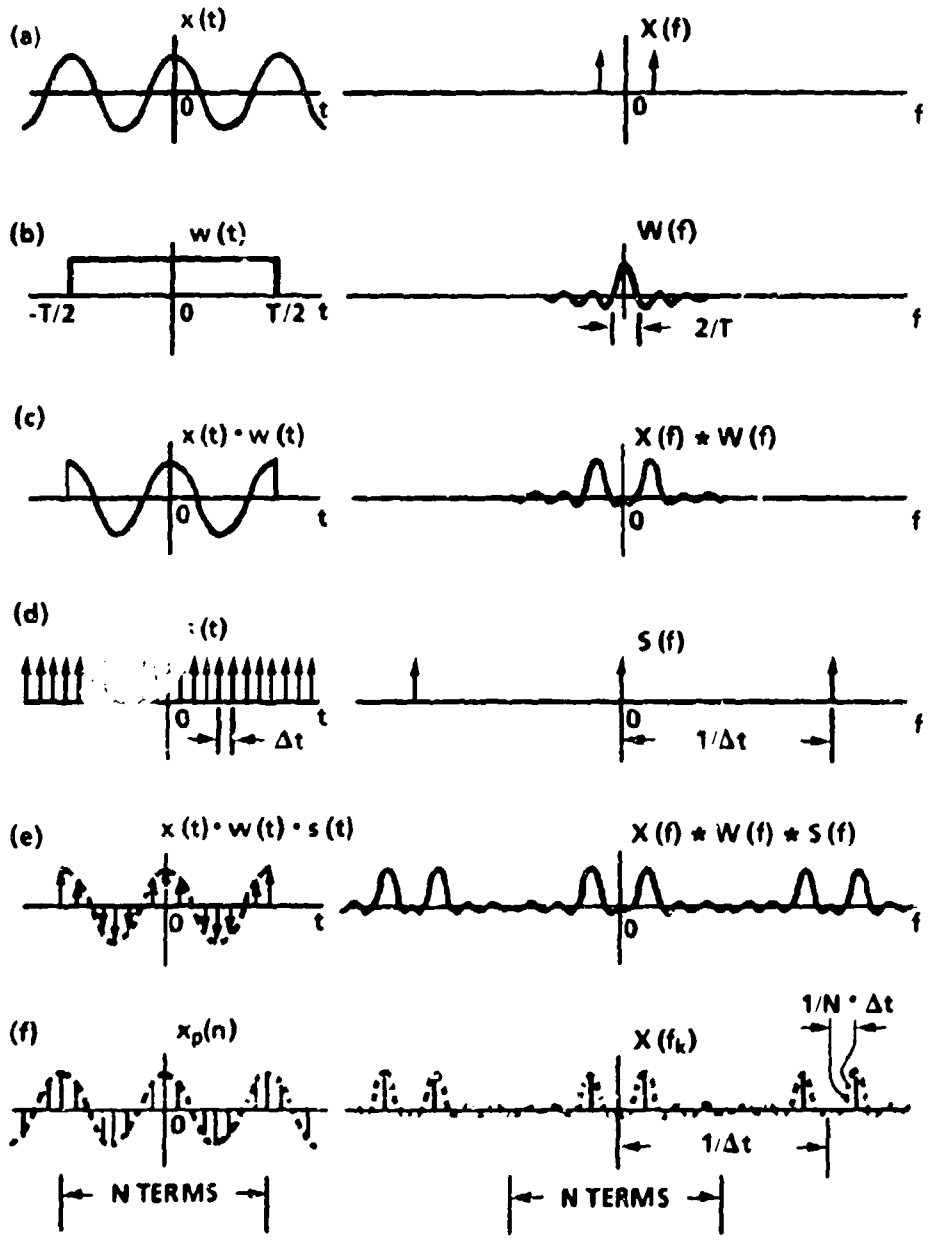
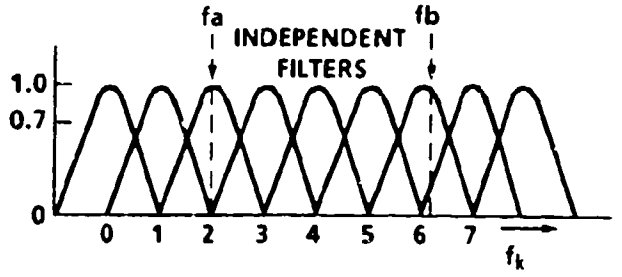
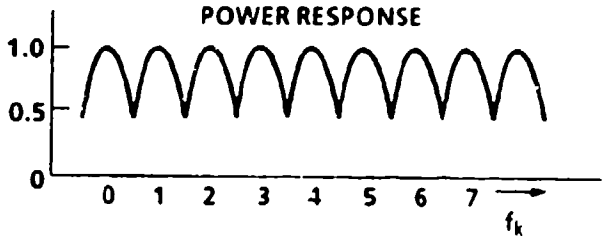


FIGURE 3. DEVELOPMENT OF THE DISCRETE FOURIER TRANSFORM $X(f_k)$ OF THE TIME FUNCTION x(t) USING THE WINDOW FUNCTION w(t)

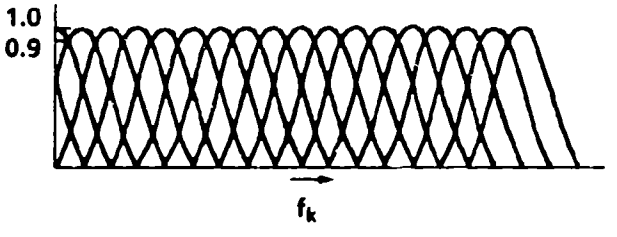


(A) DISCRETE FOURIER TRANSFORM AS A SET OF BANDPASS FILTERS

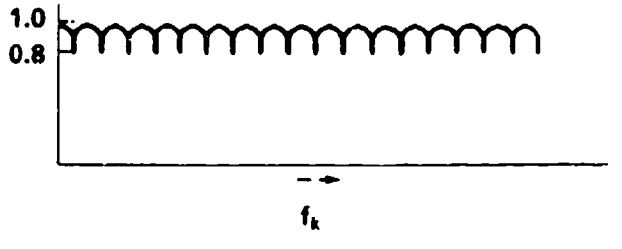


(B) RESULTING PICKET FENCE EFFECT

FIGURE 4. DISCRETE FOURIER TRANSFORM RESPONSE



(A) CLOSER SPACING OF FILTERS THROUGH ZERO PADDING



(B) REDUCTION OF PICKET FENCE EFFECT

FIGURE 5. DISCRETE FOURIER TRANSFORM RESPONSE WITH ZERO PADDING

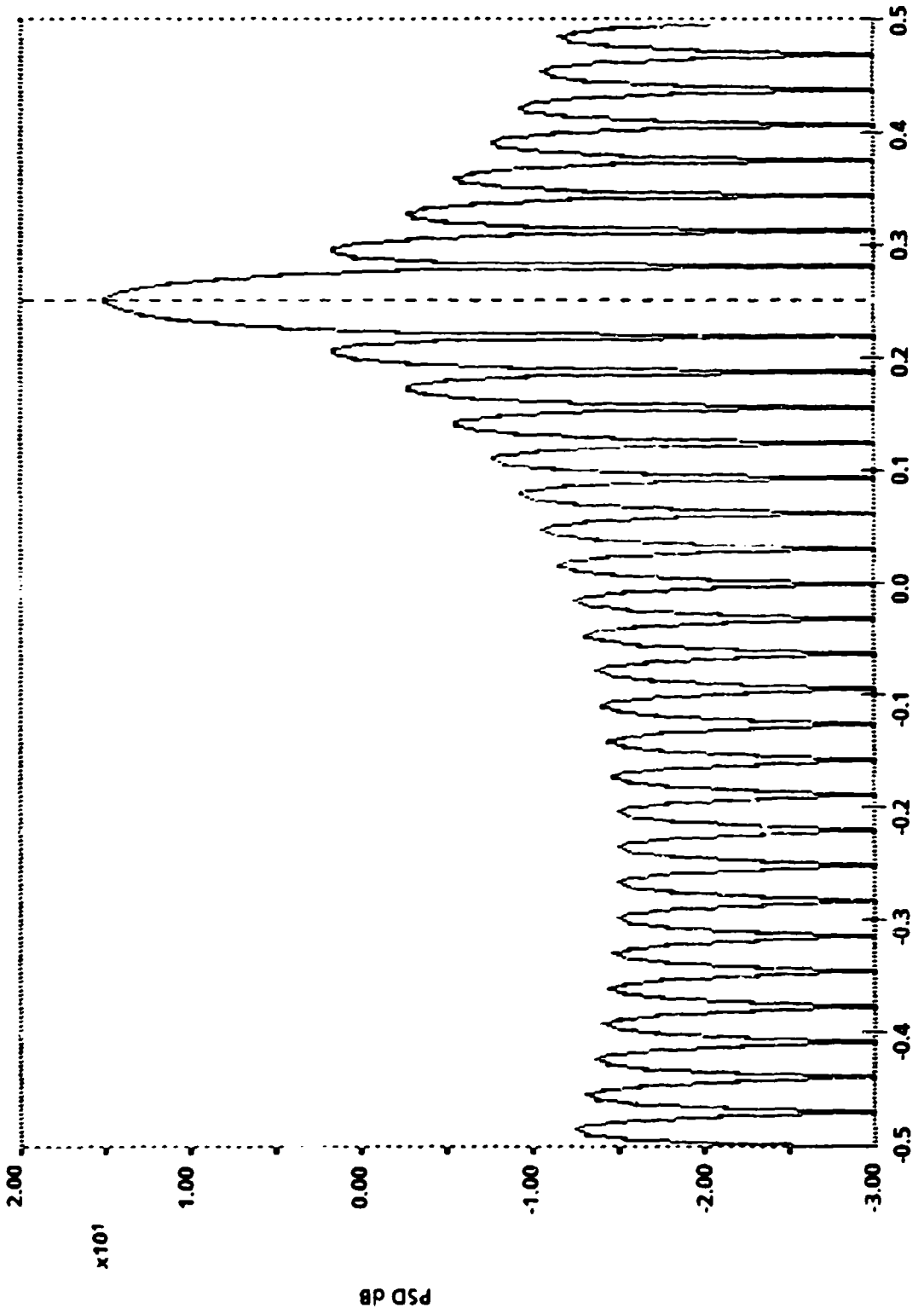


FIGURE 6. PERIODOGRAM OF A COMPLEX SINUSOID, 32 SAMPLES, UNIFORM WINDOW

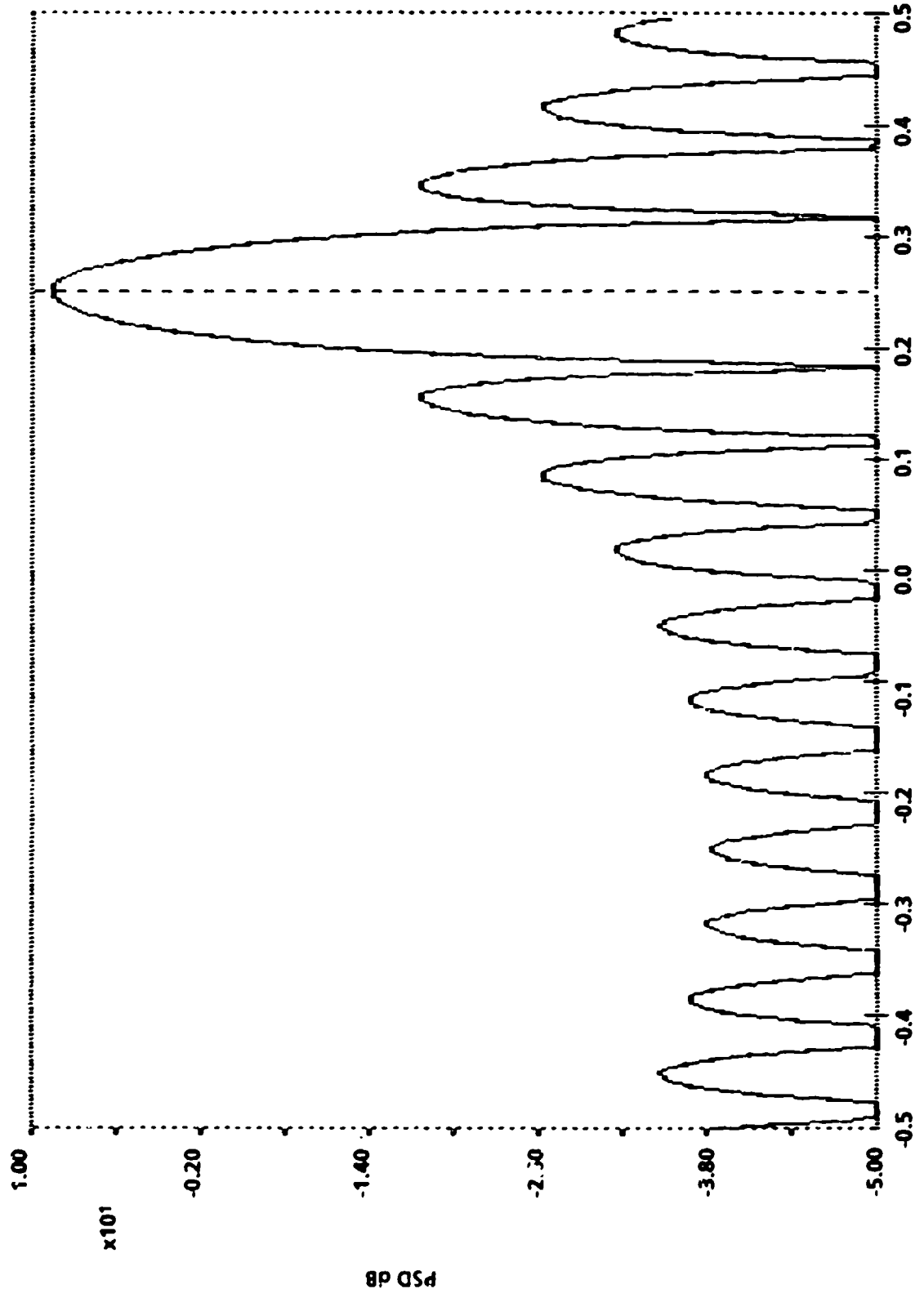


FIGURE 7. PERIODOGRAM OF A COMPLEX SINUSOID, 32 SAMPLES, BARTLETT WINDOW

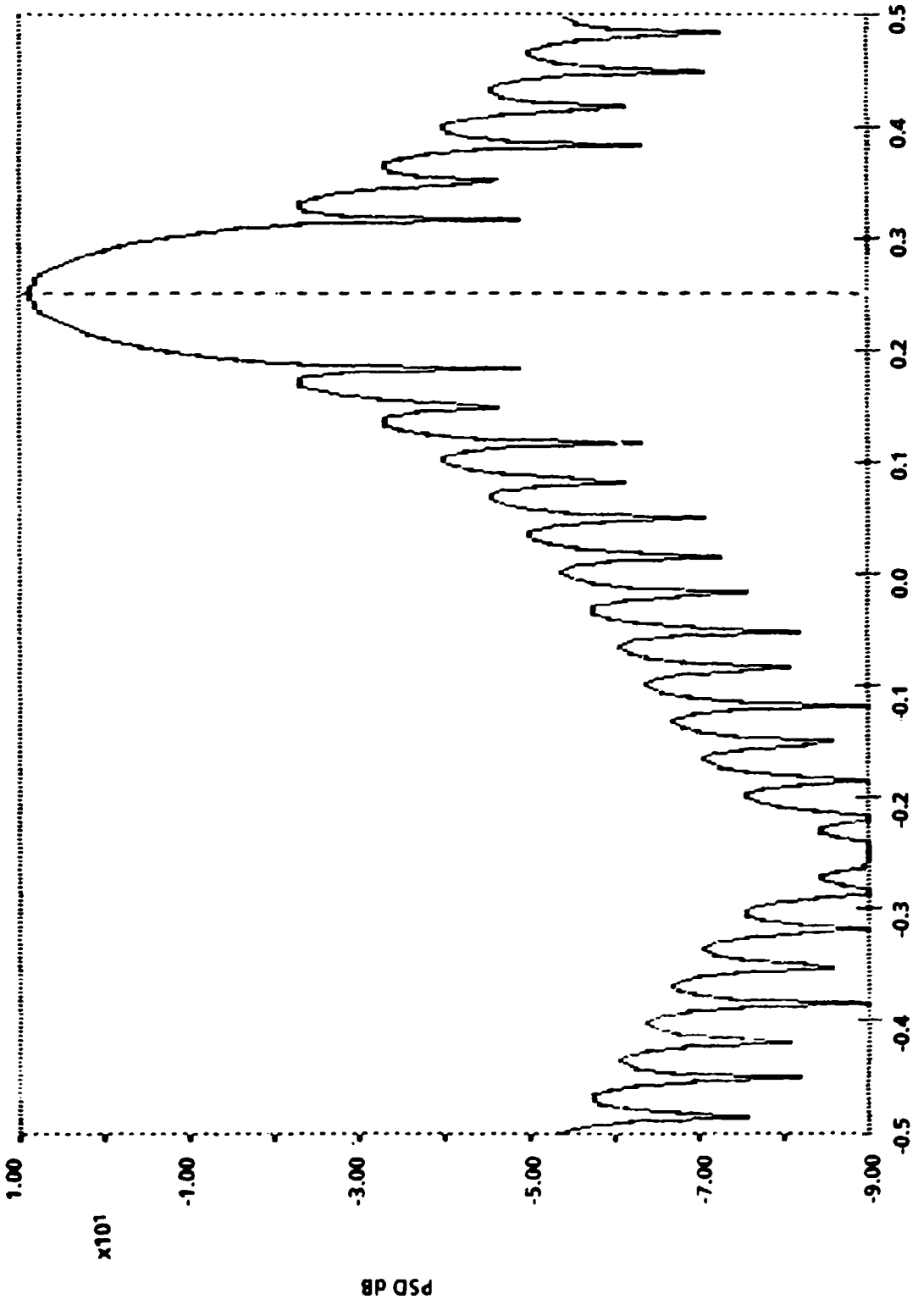


FIGURE 8. PERIODOGRAM OF A COMPLEX SINUSOID, 32 SAMPLES, HANNING WINDOW

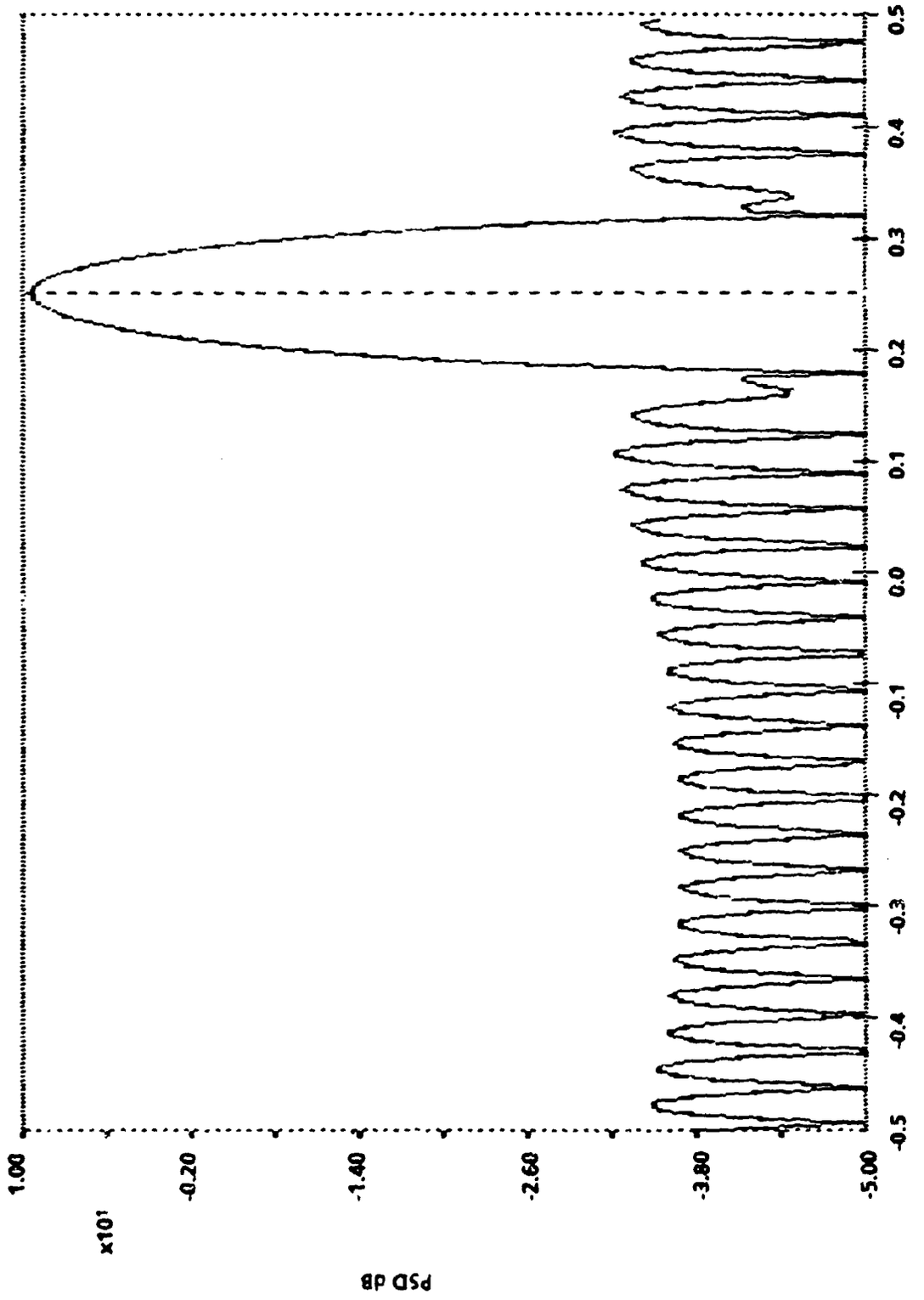


FIGURE 9. PERIODOGRAM OF A COMPLEX SINUSOID, 32 SAMPLES, HAMMING WINDOW

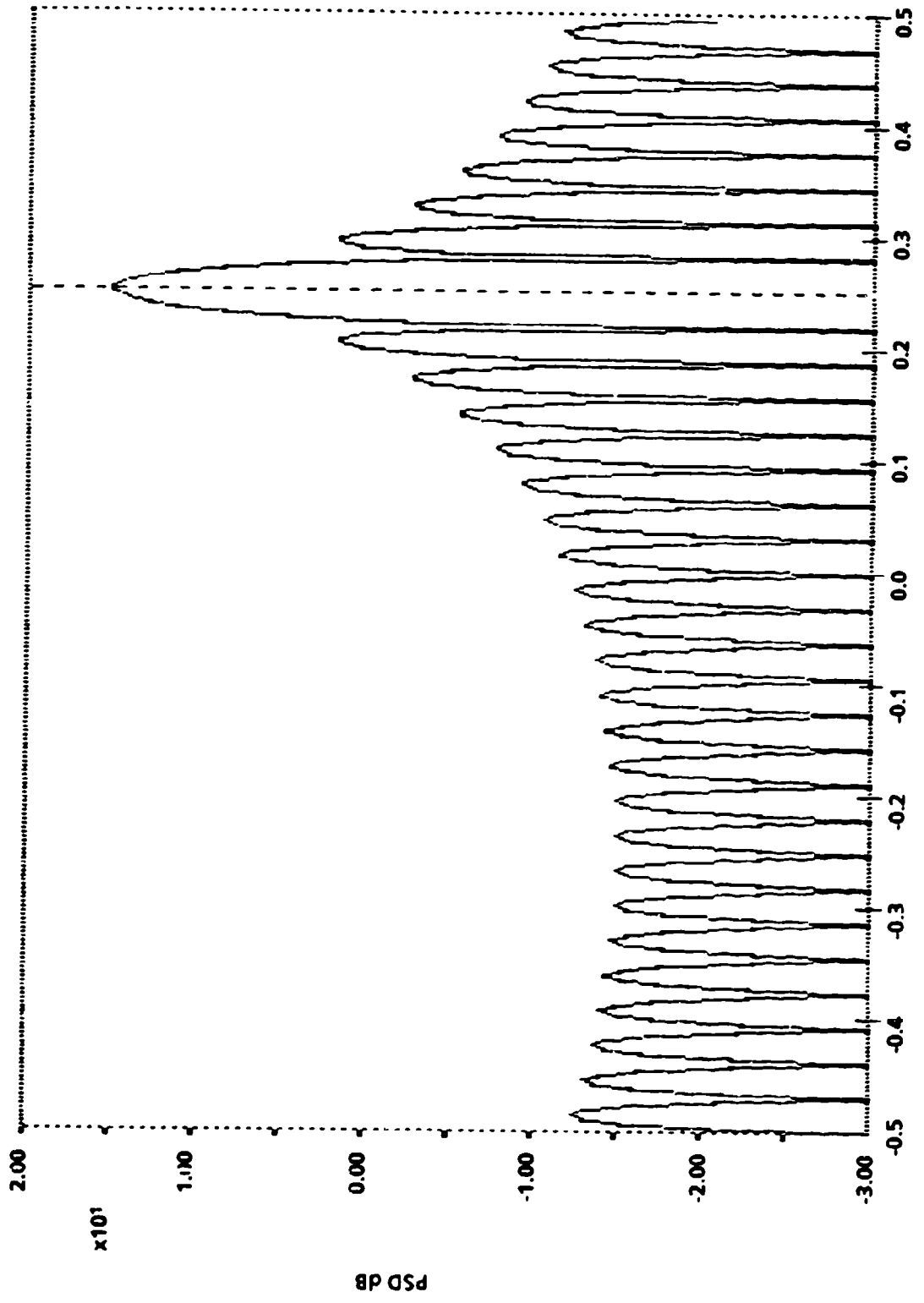


FIGURE 10. BLACKMAN-TUKEY SPECTRUM, UNBIASED, 128 DATA SAMPLES, 64 SAMPLE BARTLETT WINDOW

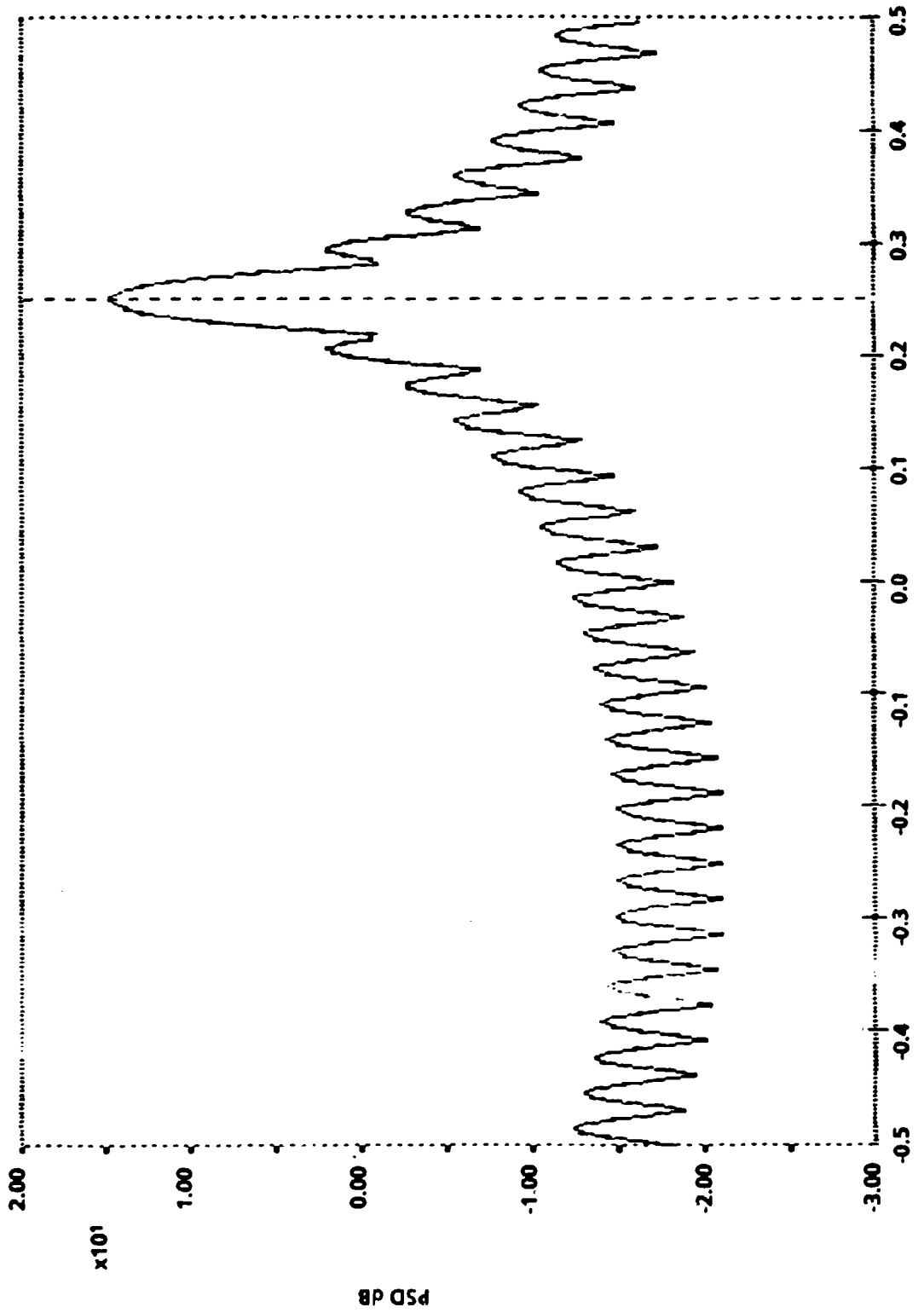


FIGURE 11. BLACKMAN-TUKEY SPECTRUM, BIASED, 128 DATA SAMPLES, 64 SAMPLE BARTLETT WINDOW

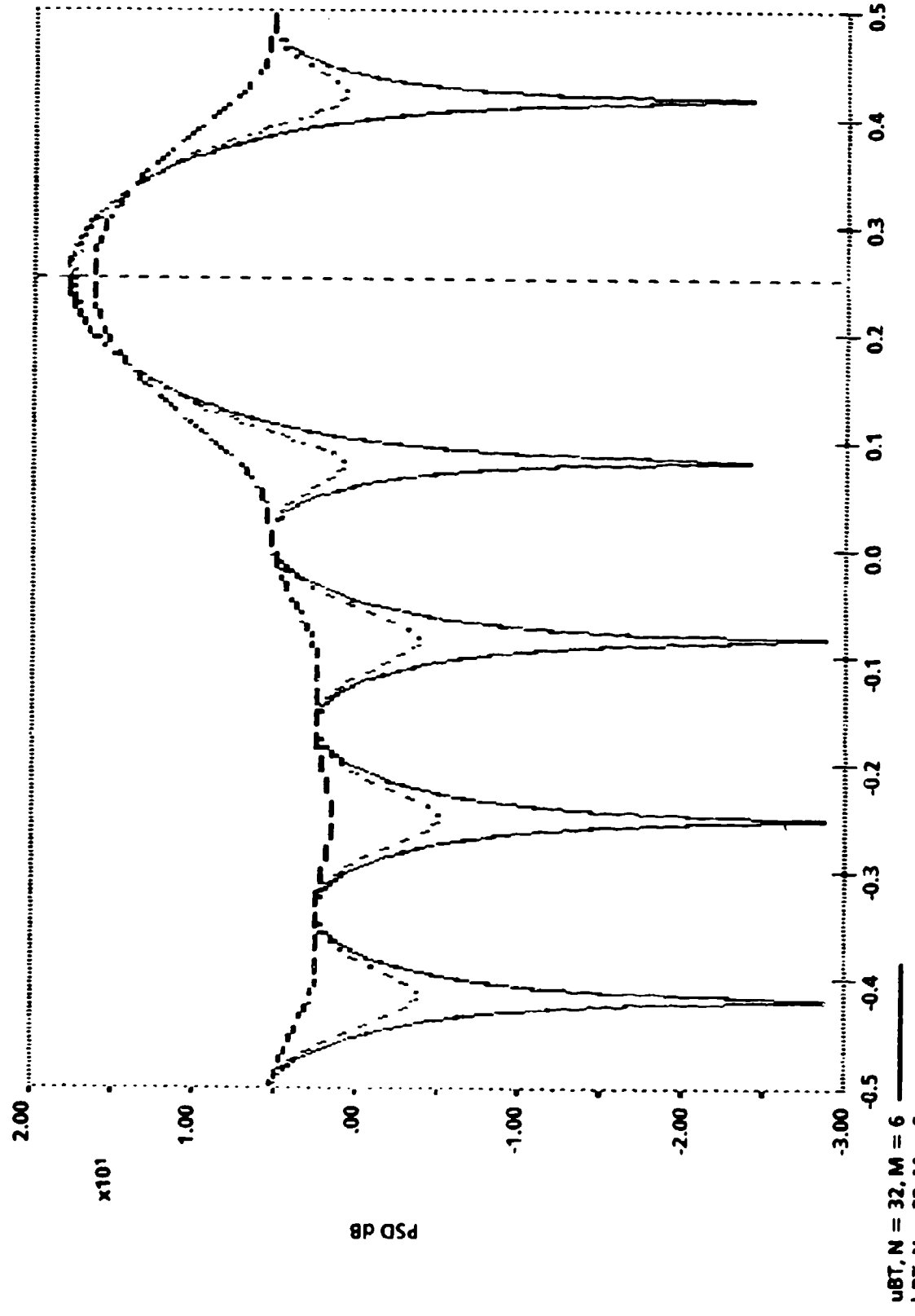


FIGURE 12. UNBIASED AND BIASED BLACKMAN-TUKEY SPECTRA, 12 SAMPLE BARTLETT WINDOW

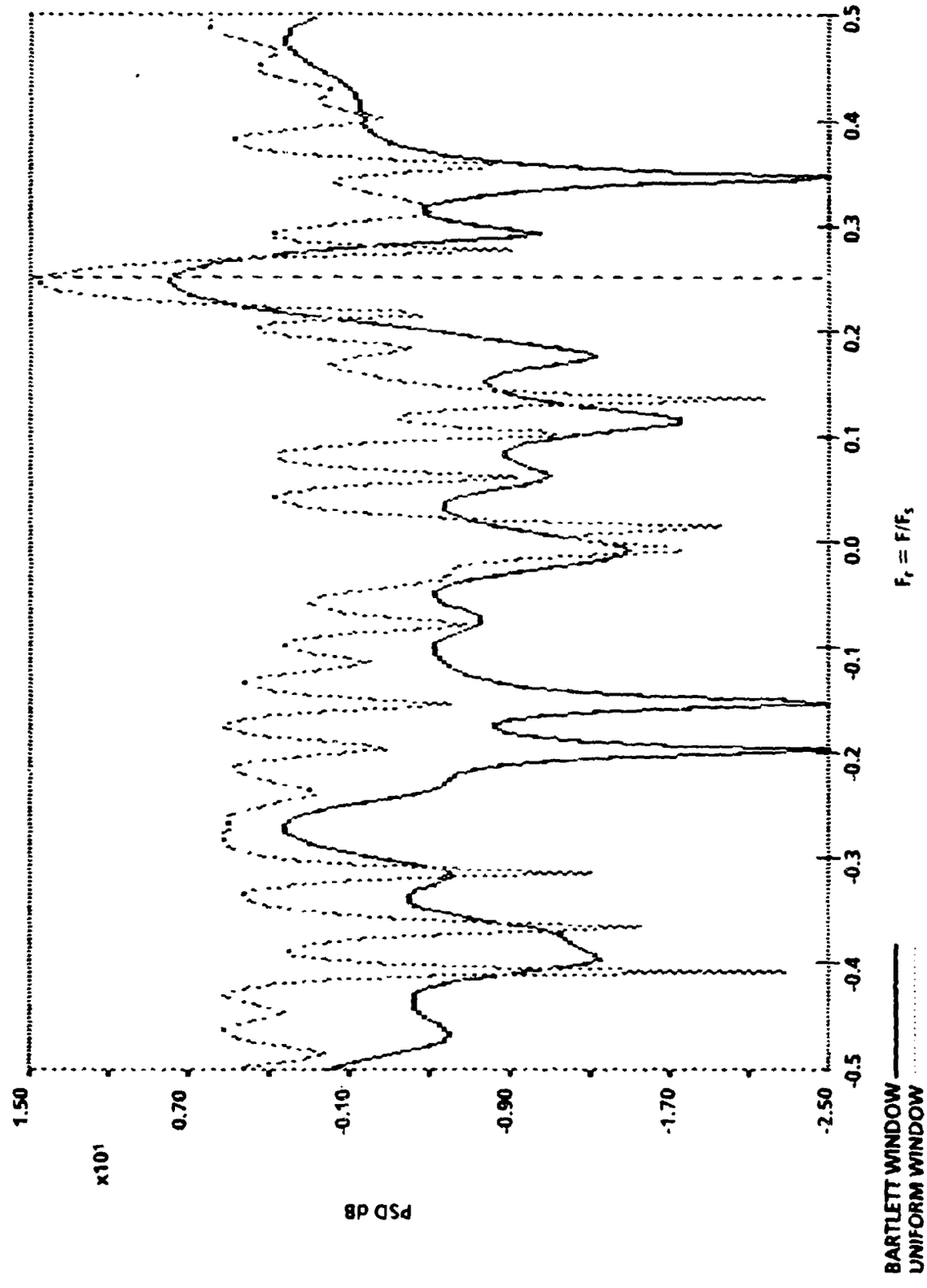


FIGURE 13. PERIODOGRAM OF A SINUSOID AND WHITE NOISE, SNR = 0 dB, 32 DATA SAMPLES, BARTLETT VS UNIFORM WINDOW

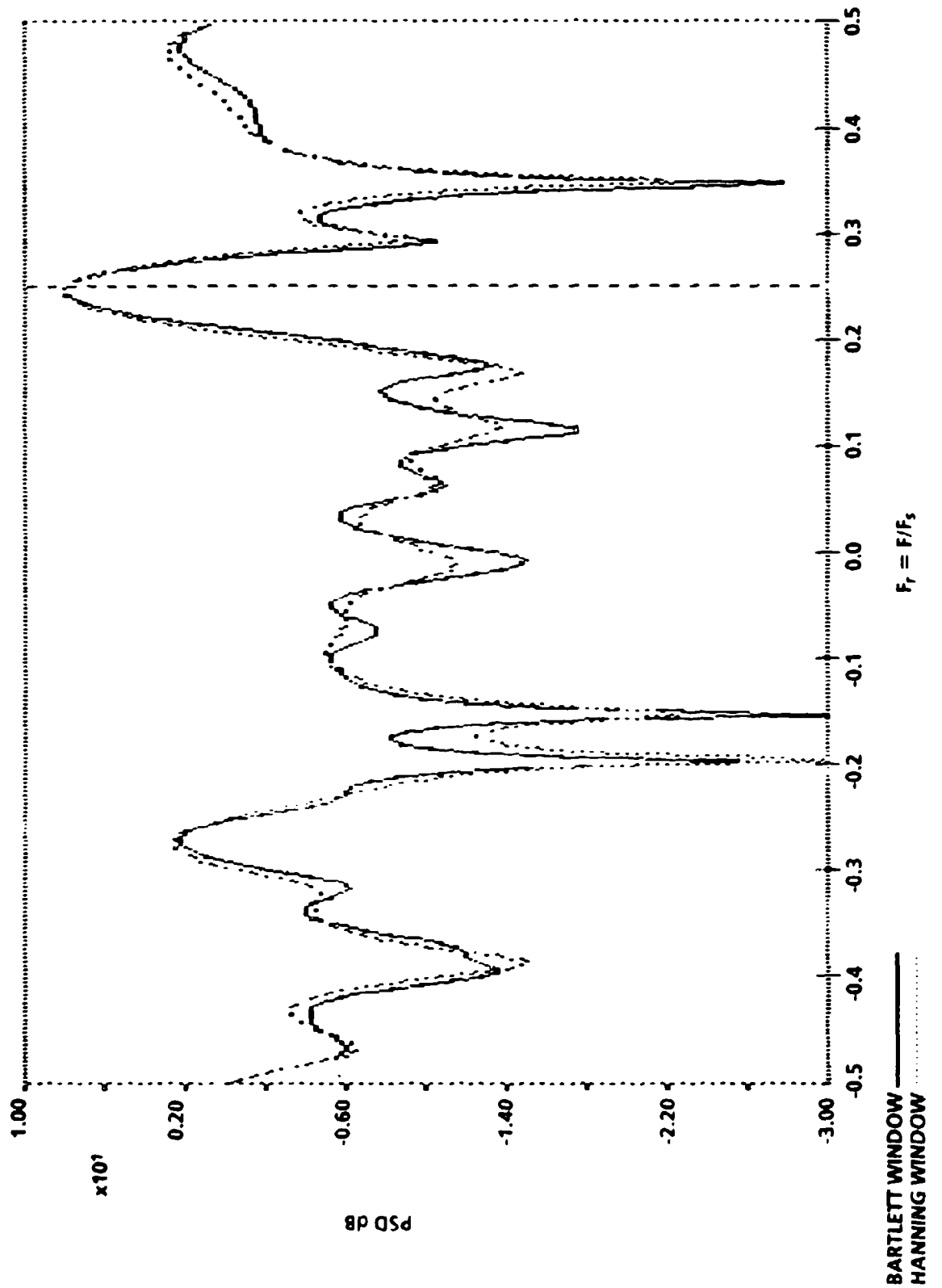


FIGURE 14. PERIODOGRAM OF A SINUSOID AND WHITE NOISE, SNR = 0 dB, 32 DATA SAMPLES, BARTLETT VS HANNING WINDOW

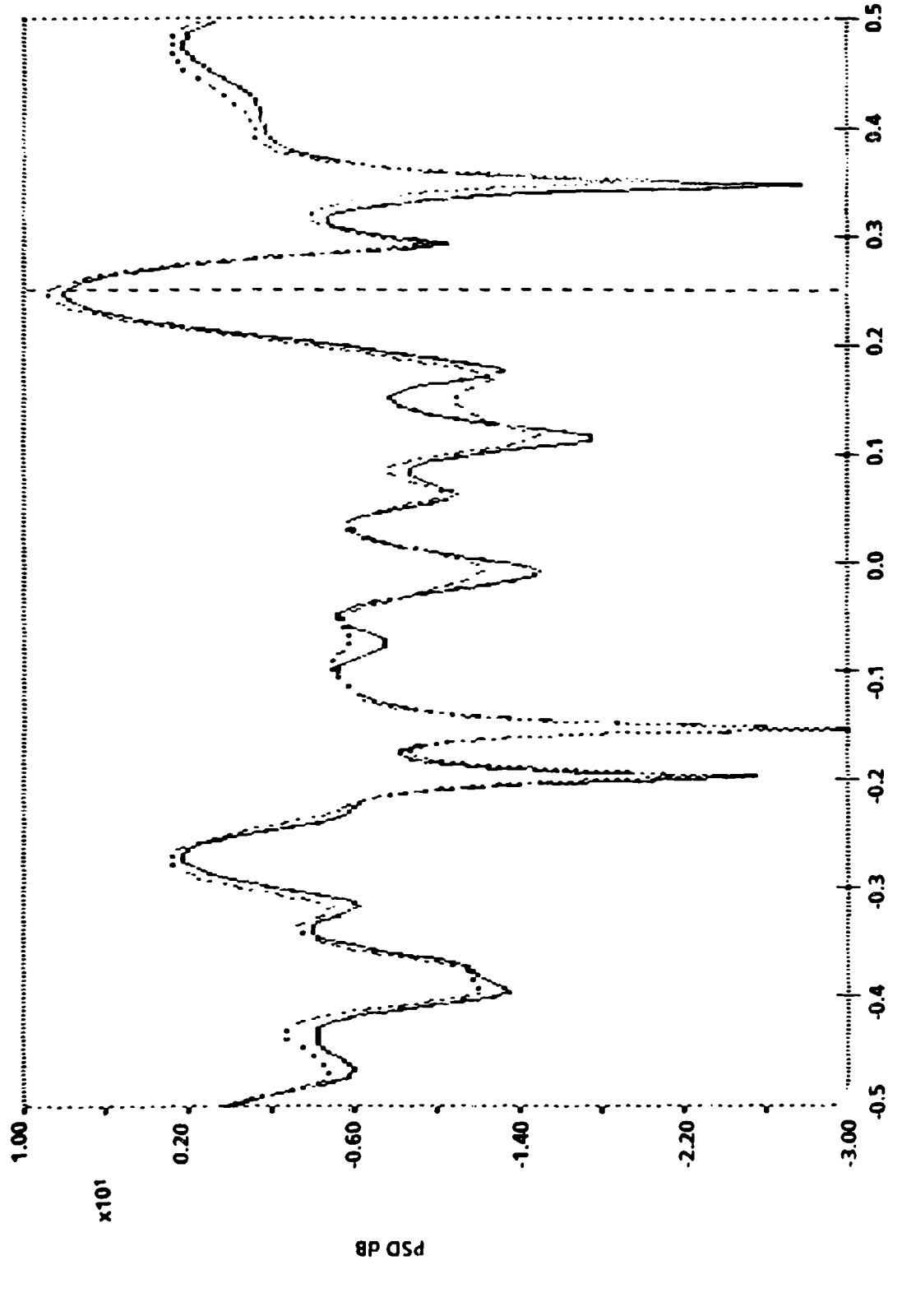


FIGURE 15. PERIODOGRAM OF A SINUSOID AND WHITE NOISE, SNR = 0 dB, 32 DATA SAMPLES, BARTLETT VS HAMMING WINDOW

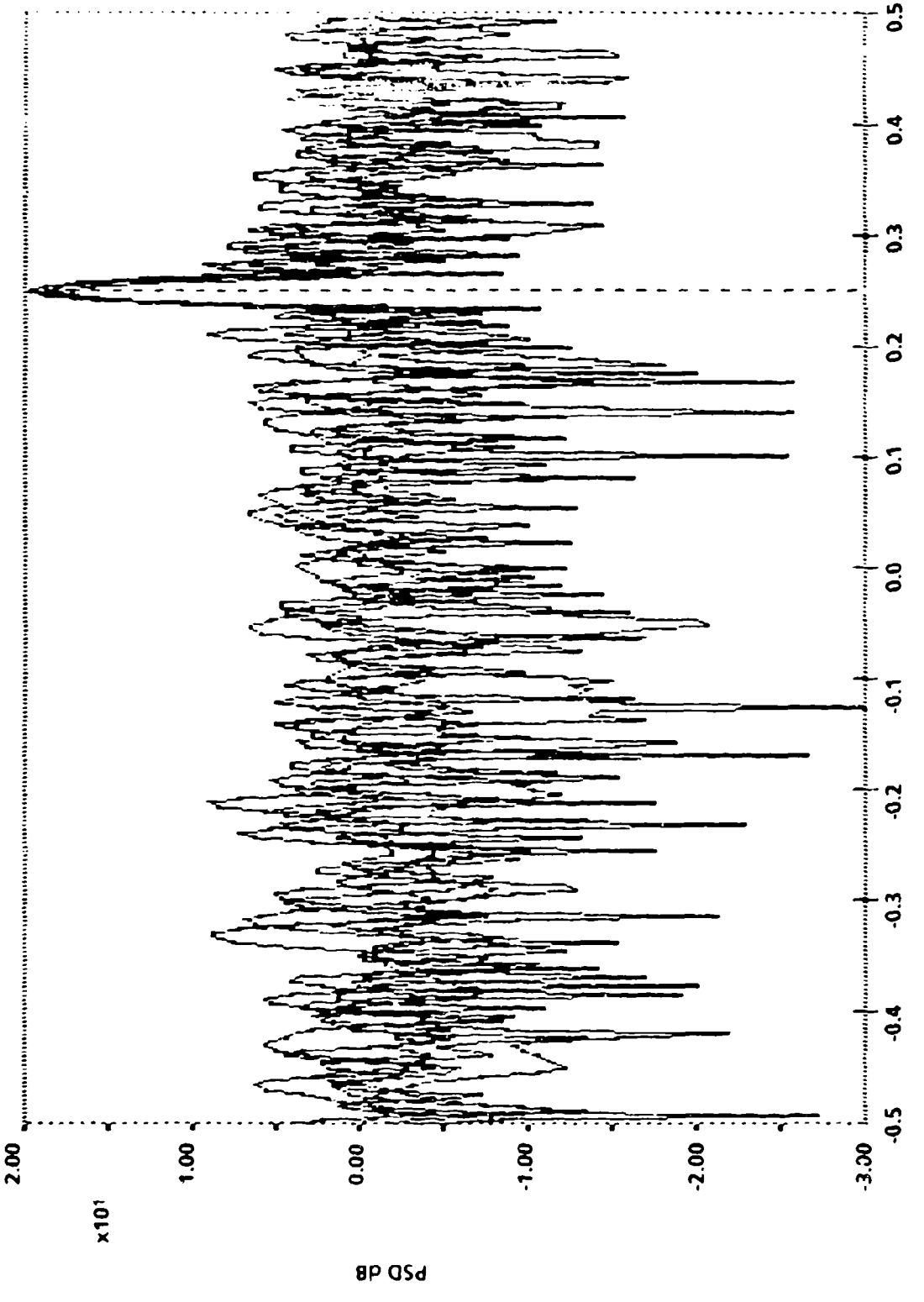


FIGURE 16. PERIODOGRAM OF SINUSOID AND WHITE NOISE, SNR = 0 dB, 5 RECORDS, 64 SAMPLES EACH, UNIFORM WINDOW

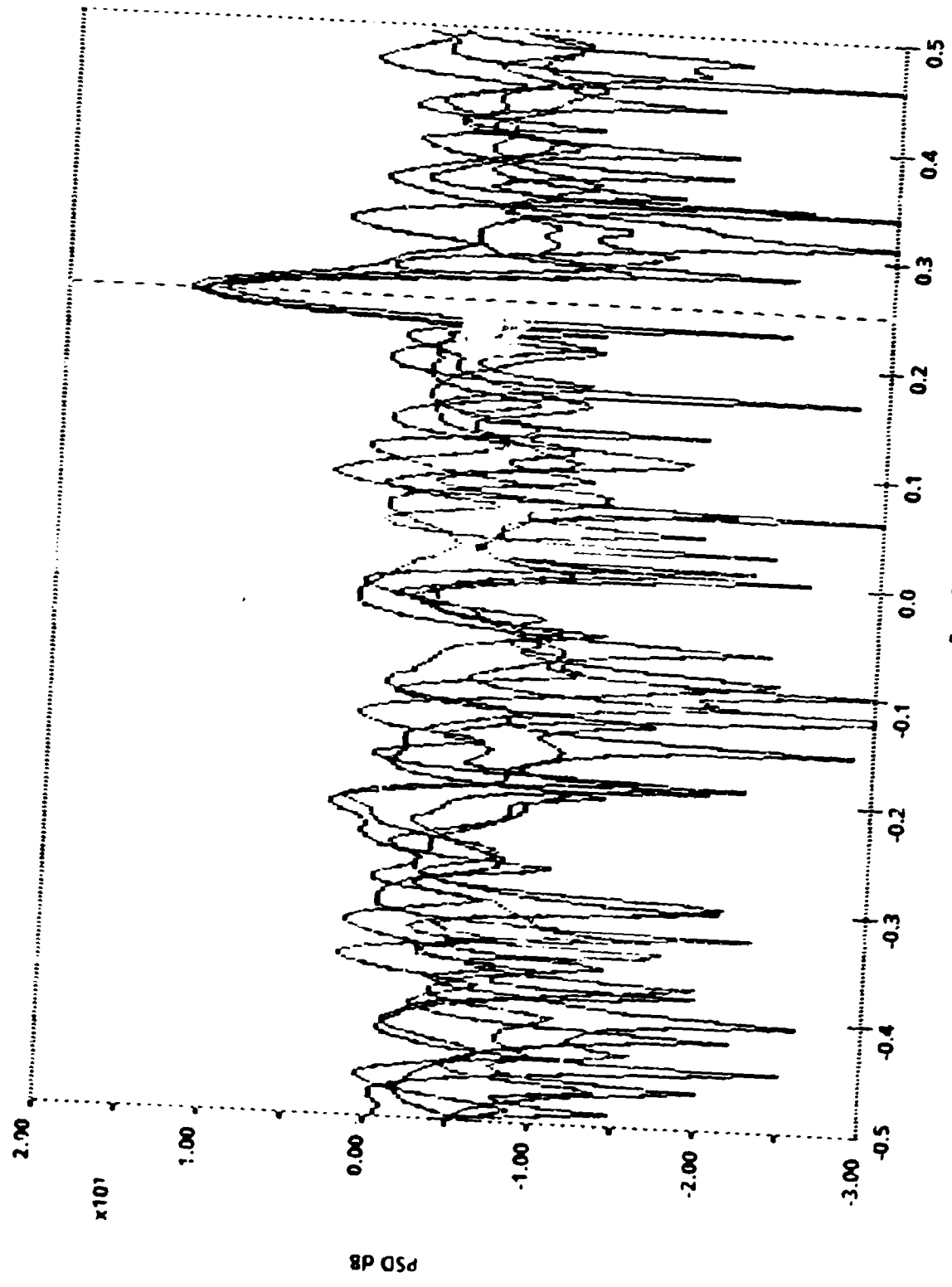


FIGURE 17. PERIODOGRAM OF SINUSOID AND WHITE NOISE, SNR = 0 dB, 5 RECORDS, 64 SAMPLES EACH, BARTLETT WINDOW

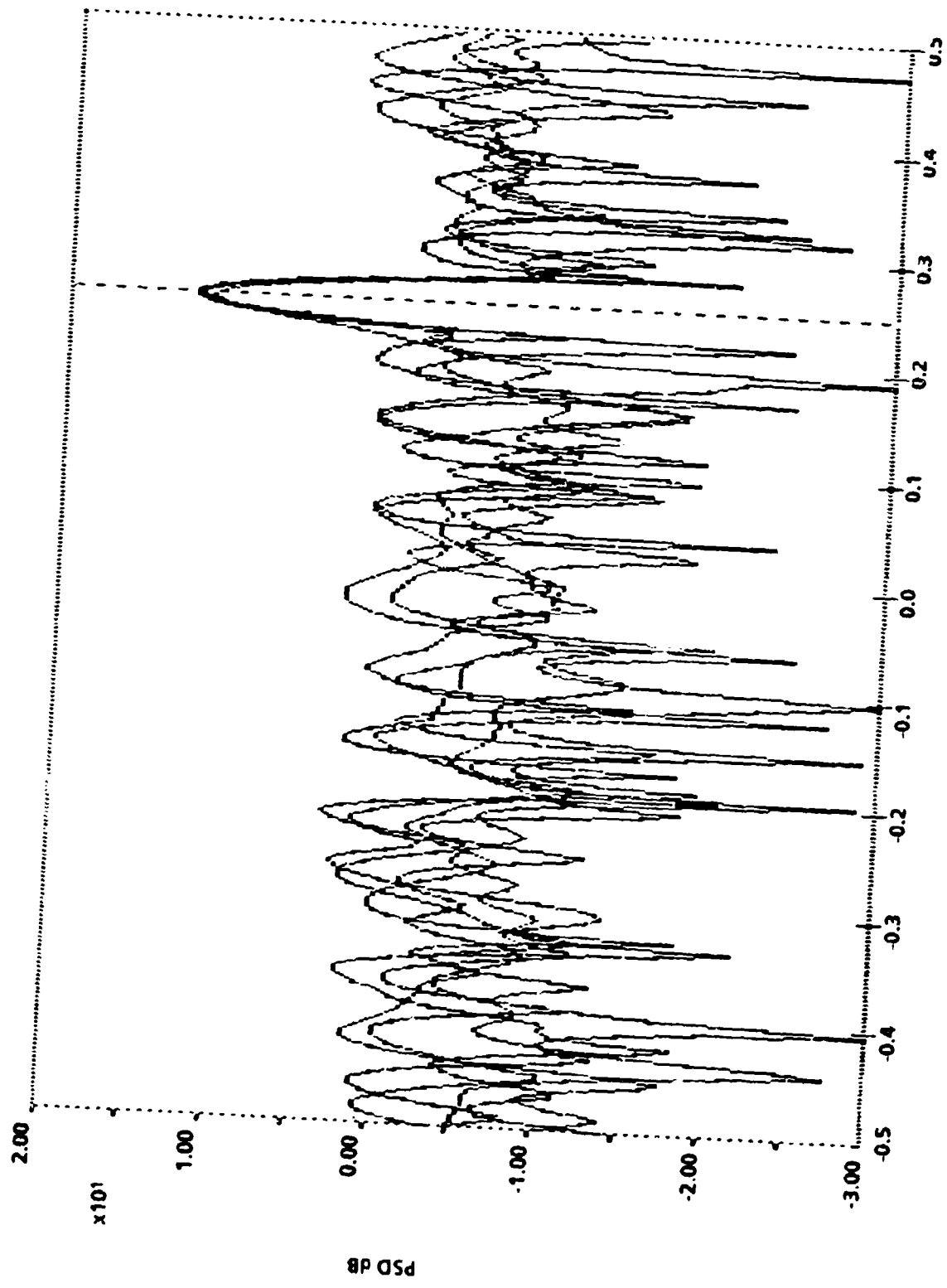


FIGURE 18. PERIODOGRAM OF SINUSOID AND WHITE NOISE, SNR = 0 dB, 5 RECORDS, 64 SAMPLES EACH, HANNING WINDOW

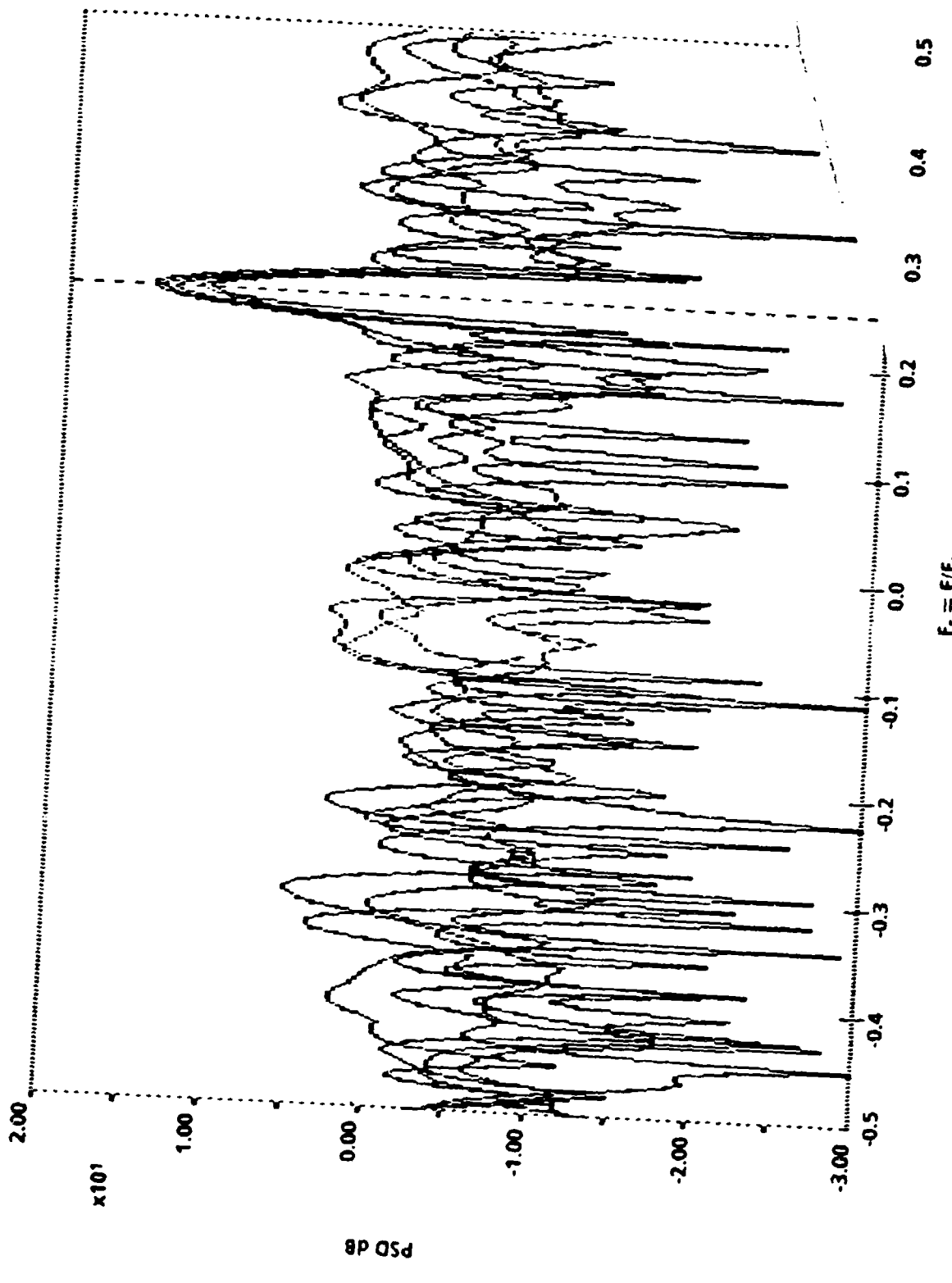


FIGURE 19. PERIODOGRAM OF SINUSOID AND WHITE NOISE, SNR = 0 dB, 5 RECORDS, 64 SAMPLES EACH, HAMMING WINDOW

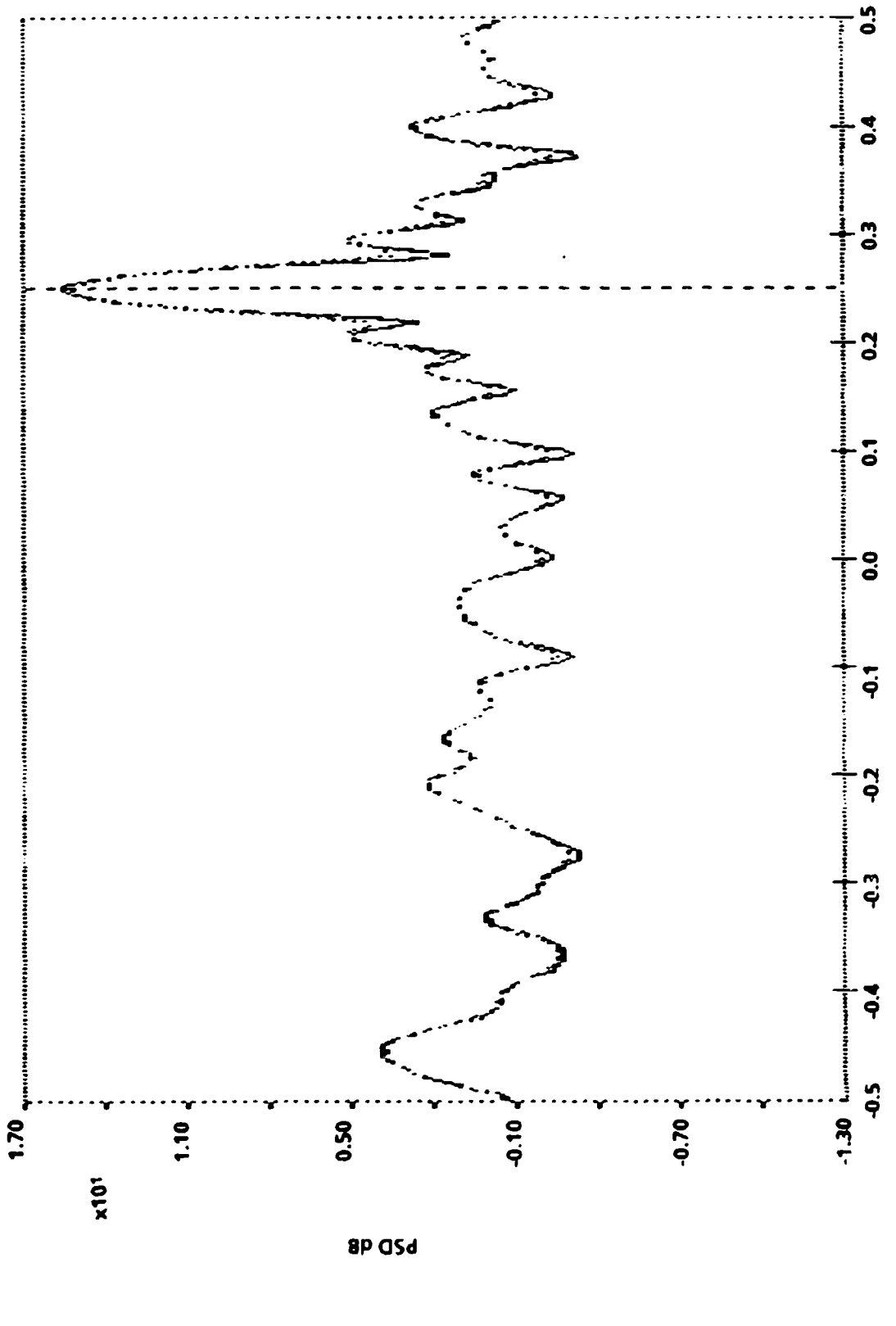


FIGURE 20. BLACKMAN-TUKEY SPECTRA OF A SINUSOID AND WHITE NOISE, SNR = 0 dB, 128 DATA SAMPLES, 64 SAMPLE BARTLETT WINDOW

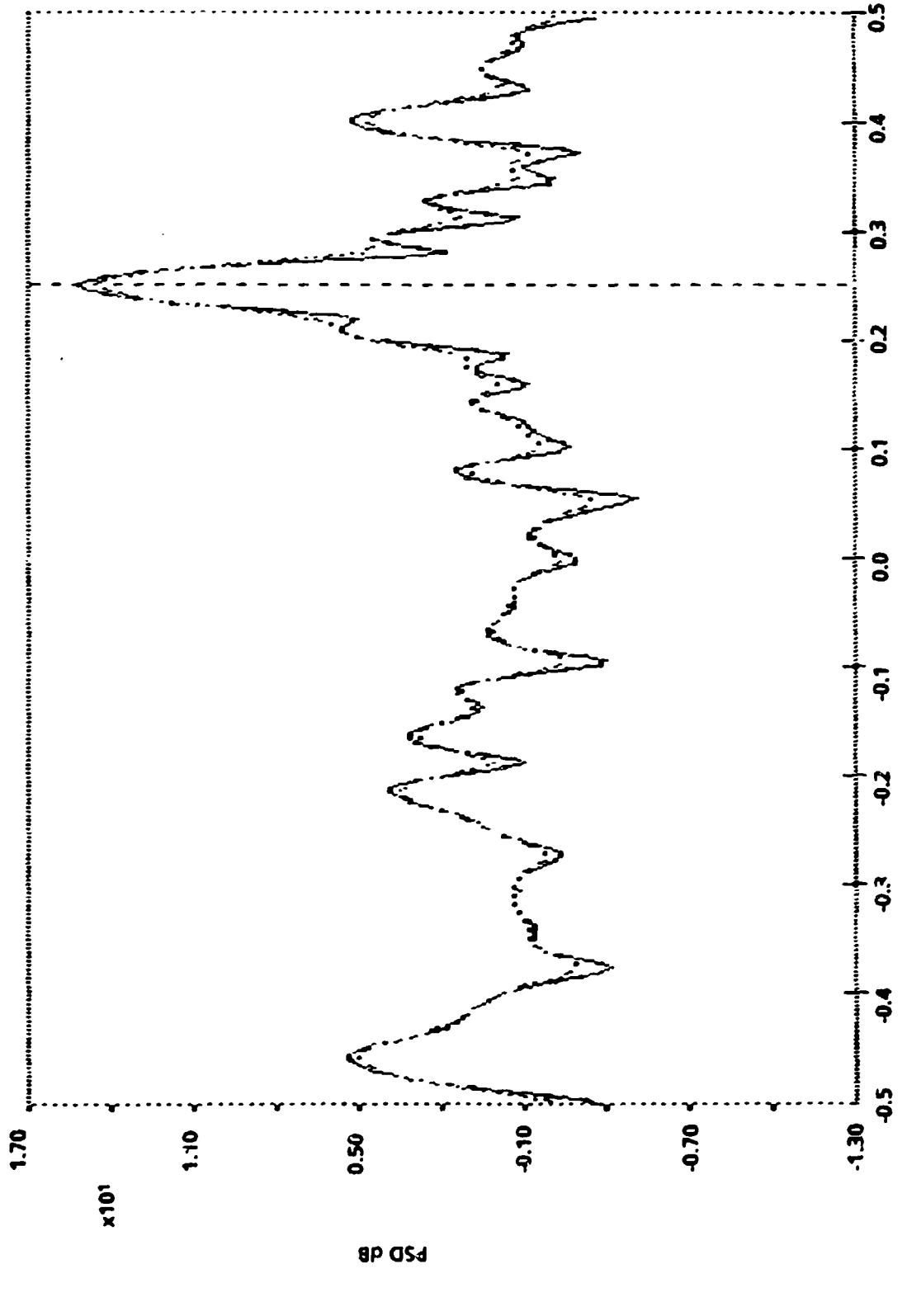


FIGURE 21. BLACKMAN-TUKEY SPECTRA OF A SINUSOID AND WHITE NOISE, SNR = 0 dB, 64 DATA SAMPLES, 64 SAMPLE BARTLETT WINDOW

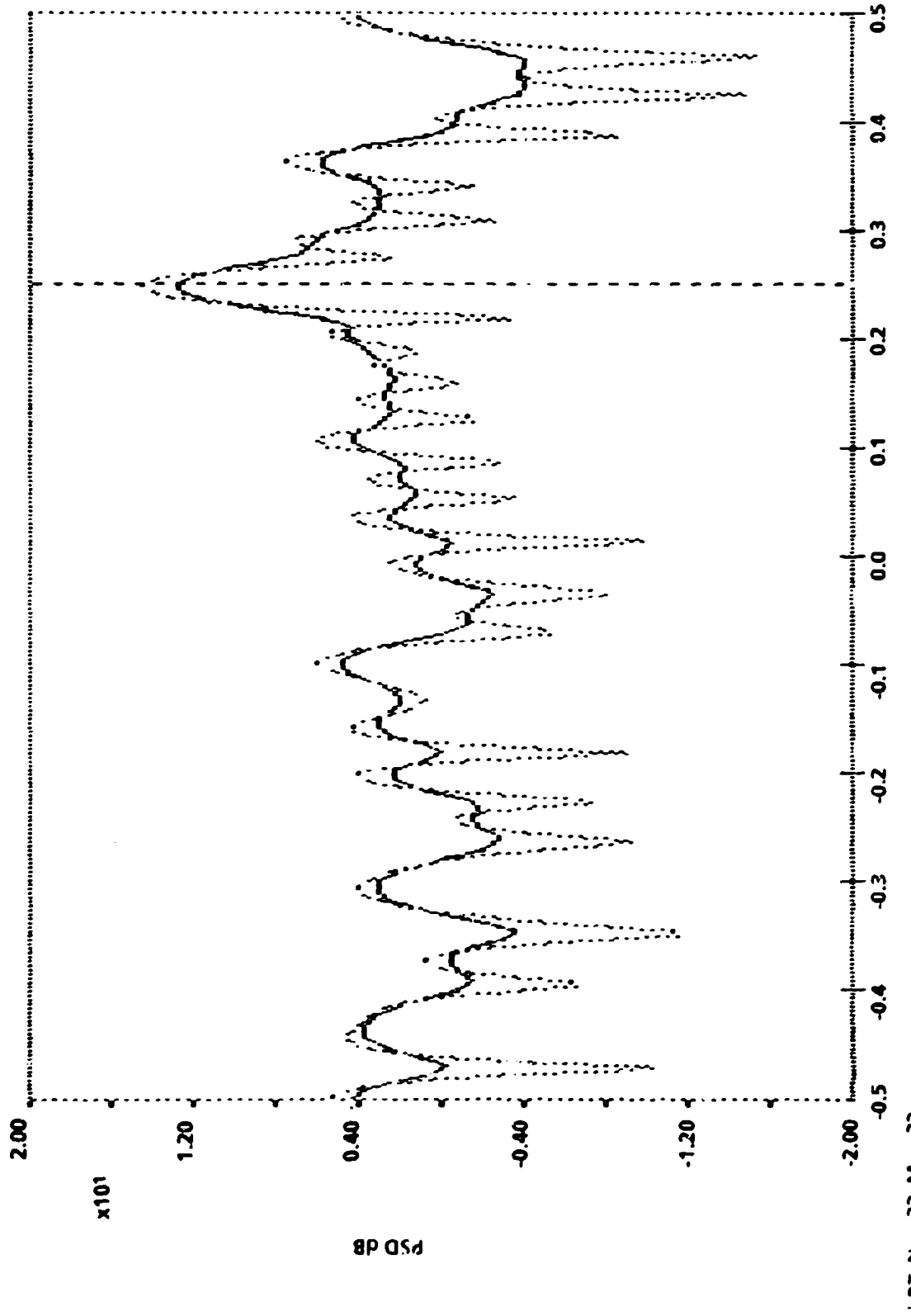


FIGURE 22. BLACKMAN-TUKEY SPECTRA OF A SINUSOID AND WHITE NOISE, SNR = 0 dB, 33 DATA SAMPLES, 64 SAMPLE BARTLETT WINDOW

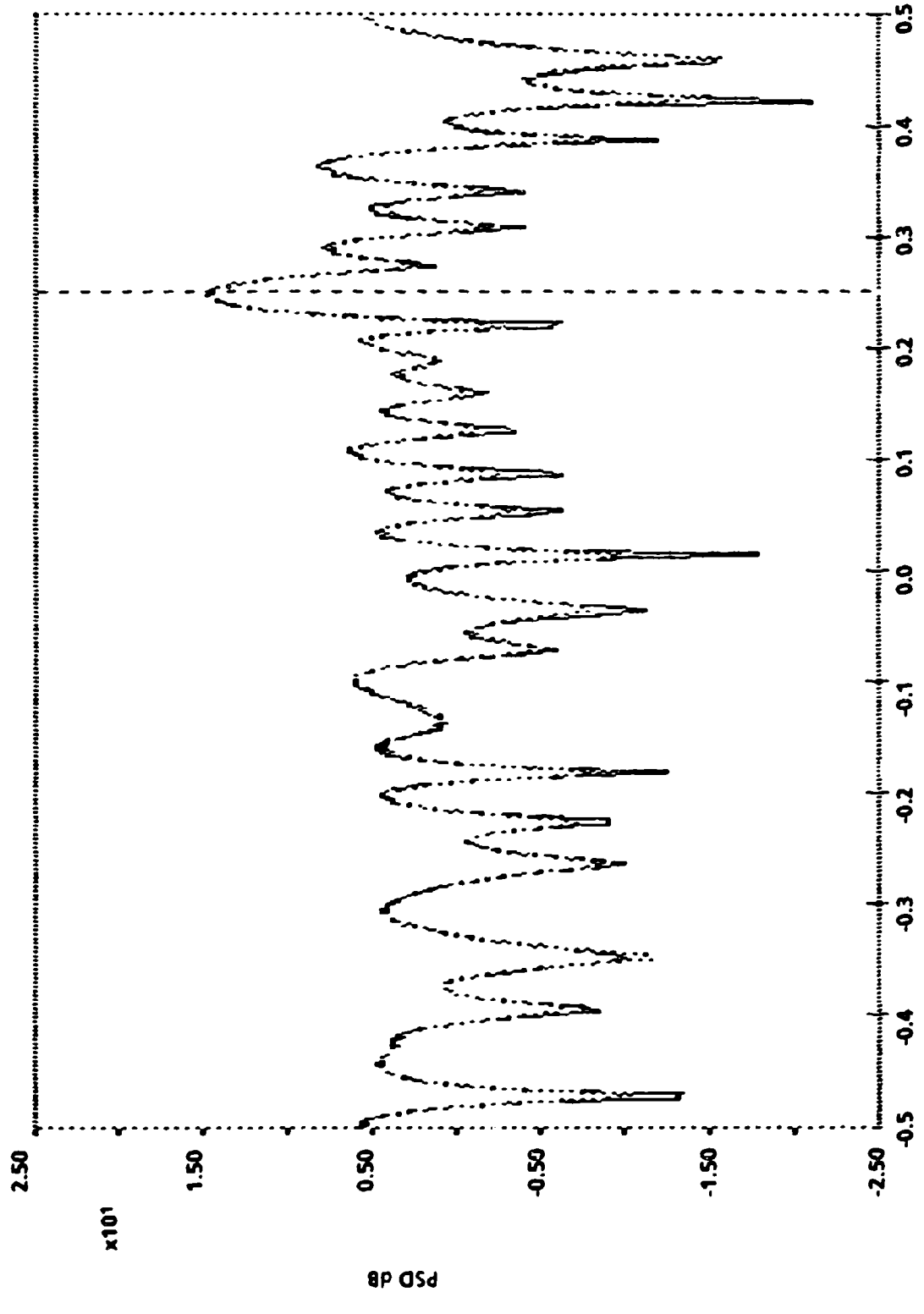


FIGURE 23. PERIODOGRAM (32 DATA SAMPLES) WITH UNIFORM WINDOW AND BLACKMAN-TUKEY SPECTRUM (33 DATA SAMPLES) WITH 64 SAMPLE BARTLETT WINDOW, SNR = 0 dB

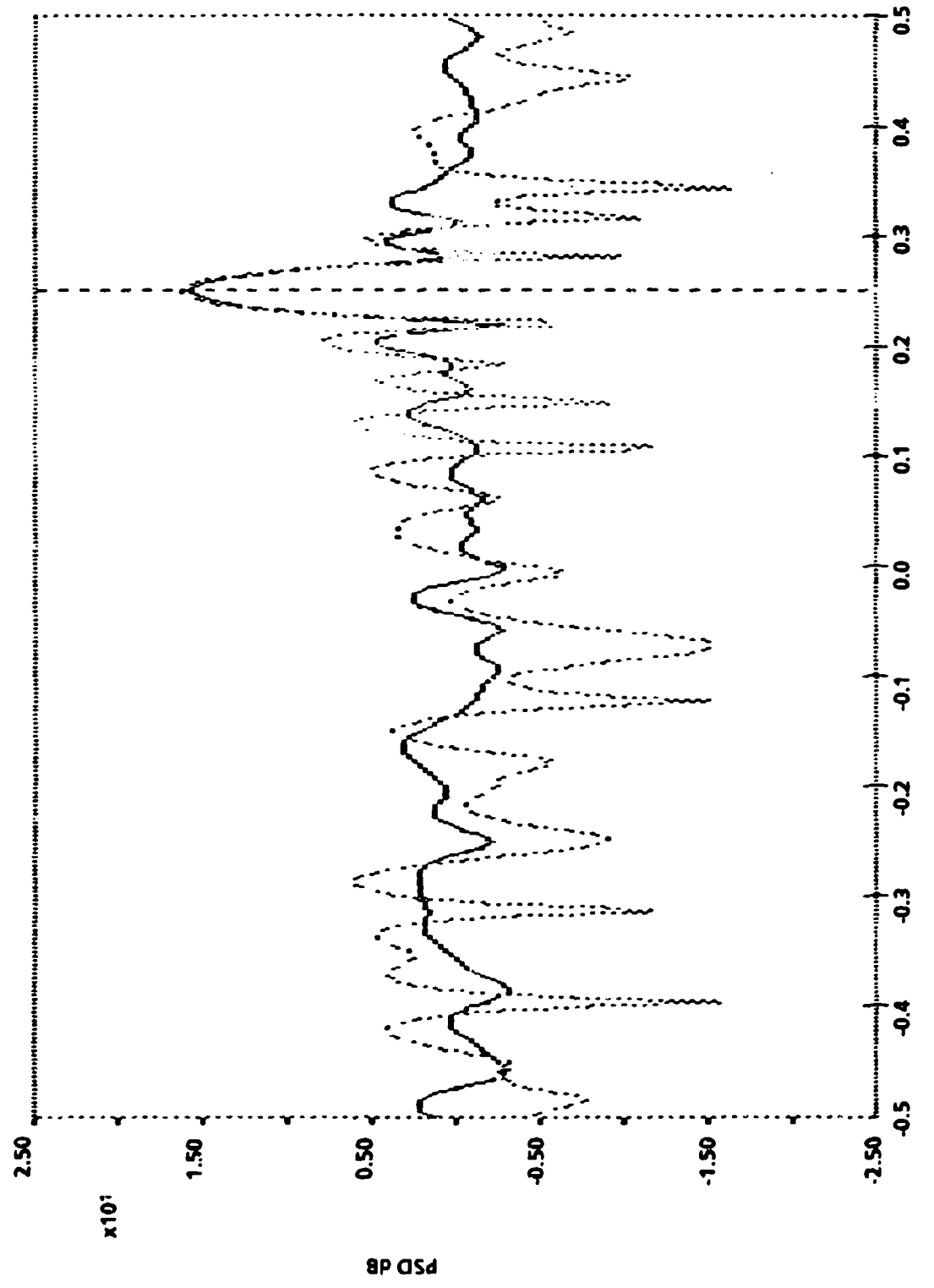


FIGURE 24. PERIODOGRAM (32 DATA SAMPLES) WITH UNIFORM WINDOW AND BLACKMAN-TUKEY SPECTRUM (128 DATA SAMPLES) WITH 64 SAMPLE BARTLETT WINDOW, SNR = 0 dB

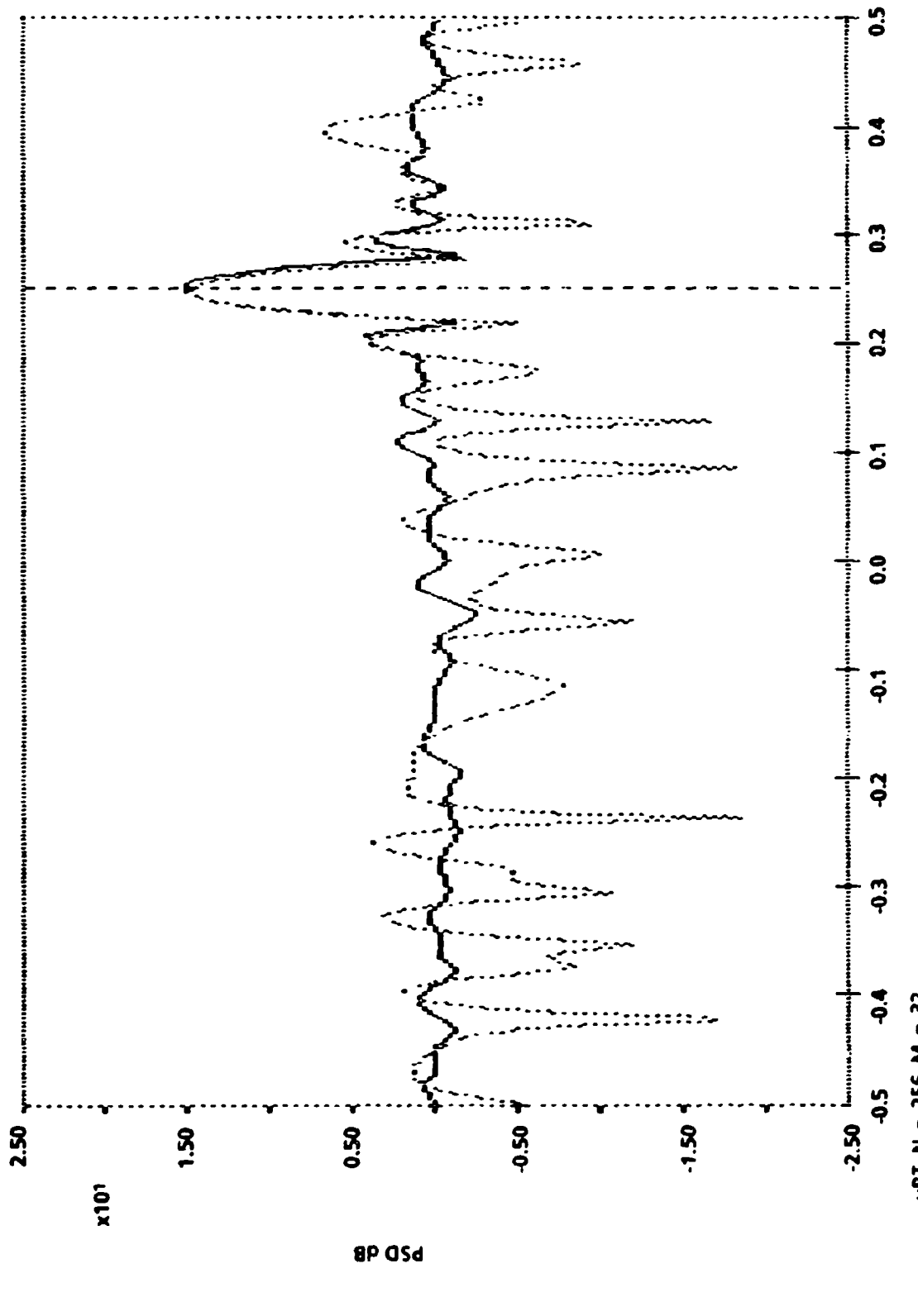


FIGURE 25. PERIODOGRAM (32 DATA SAMPLES) WITH UNIFORM WINDOW AND BLACKMAN-TUKEY SPECTRUM (256 DATA SAMPLES) WITH 64 SAMPLE BARTLETT WINDOW, SNR = 0 dB

FIGURE 26. BLACKMAN-TUKEY SPECTRA, 5 RECORDS, 128 DATA SAMPLES EACH, WITH 64 SAMPLE BARTLETT WINDOW, SNR = 0 dB

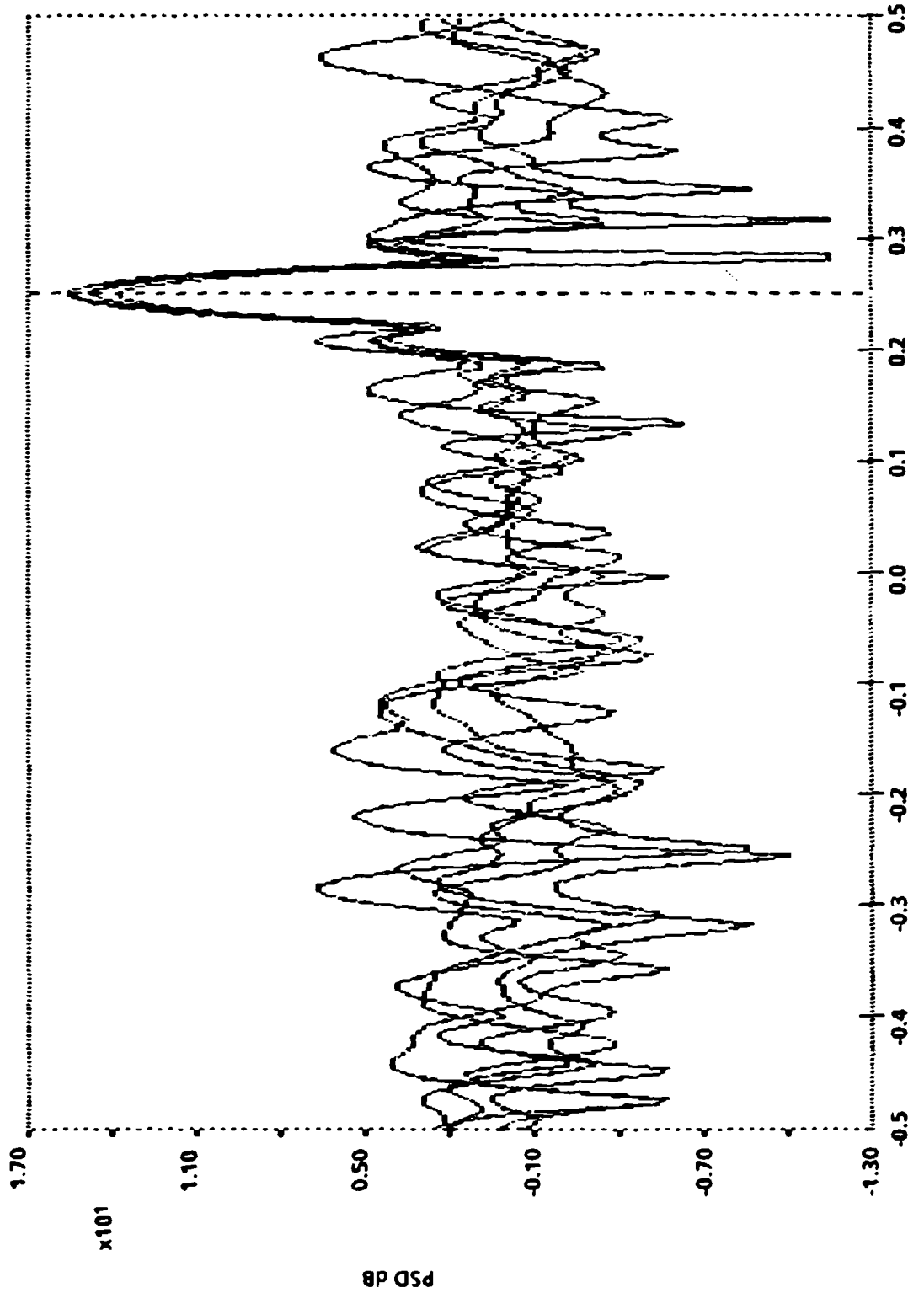


FIGURE 27. BLACKMAN-TUKEY SPECTRA, 5 RECORDS, 64 DATA SAMPLES EACH, WITH 64 SAMPLE BARTLETT WINDOW, SNR = 0 dB

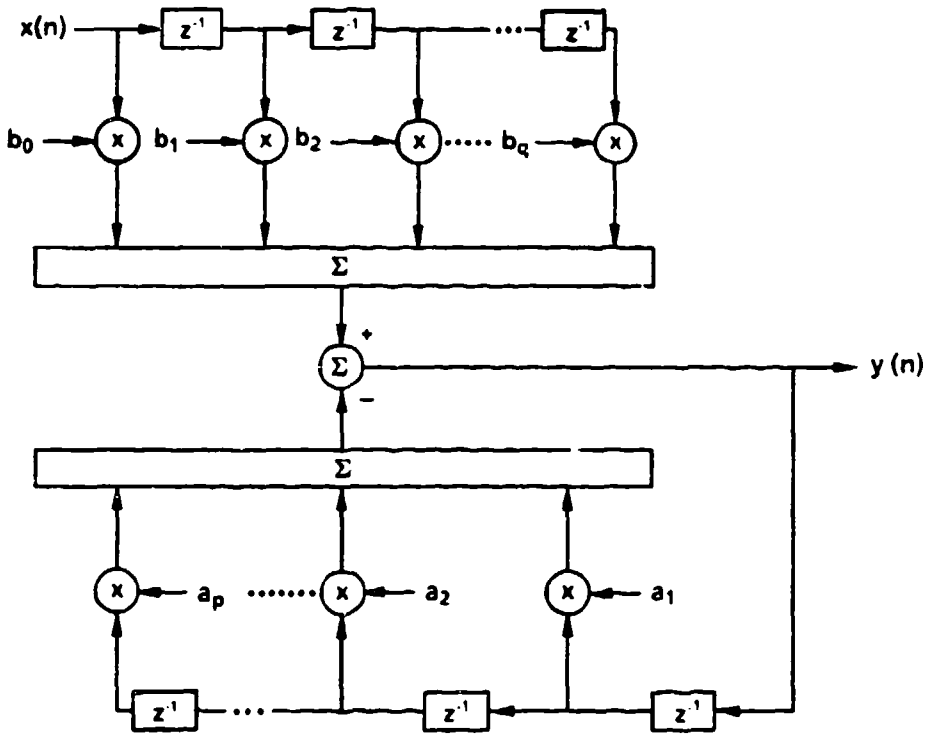


FIGURE 28. GENERALIZED LINEAR FILTER WITH RATIONAL TRANSFER FUNCTION (p POLES, q ZEROES)

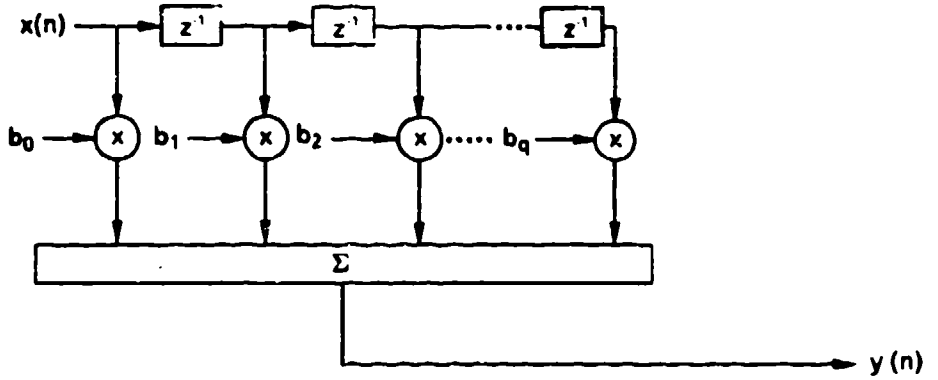


FIGURE 29. MOVING AVERAGE FILTER

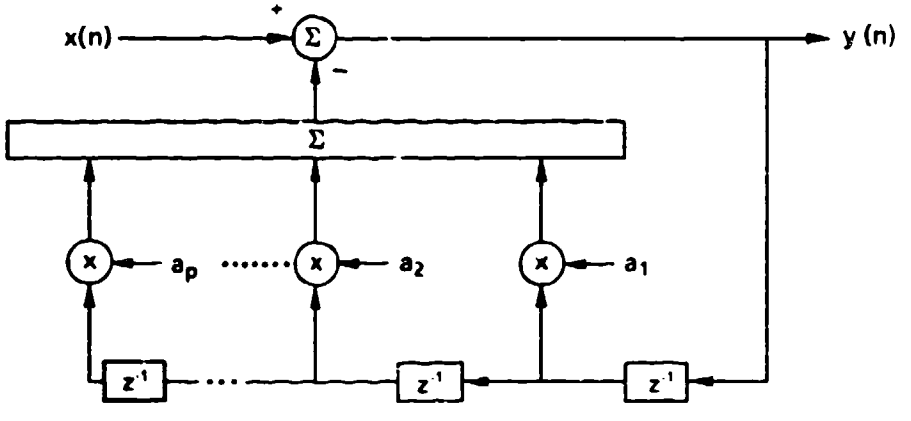


FIGURE 30. AUTOREGRESSIVE FILTER

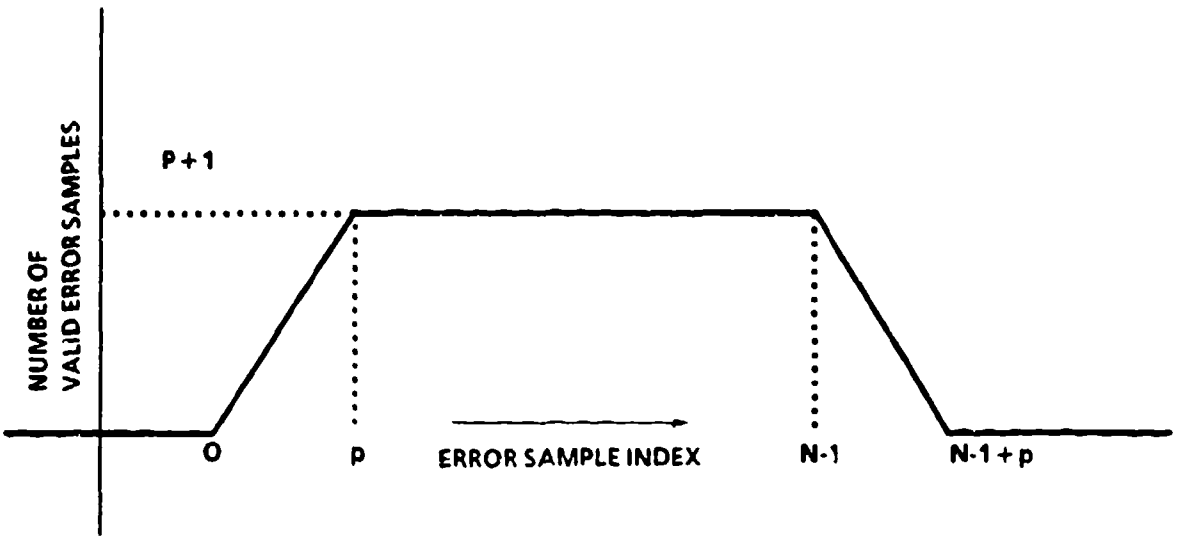


FIGURE 31. VALID ERROR SAMPLES IN ERROR SERIES

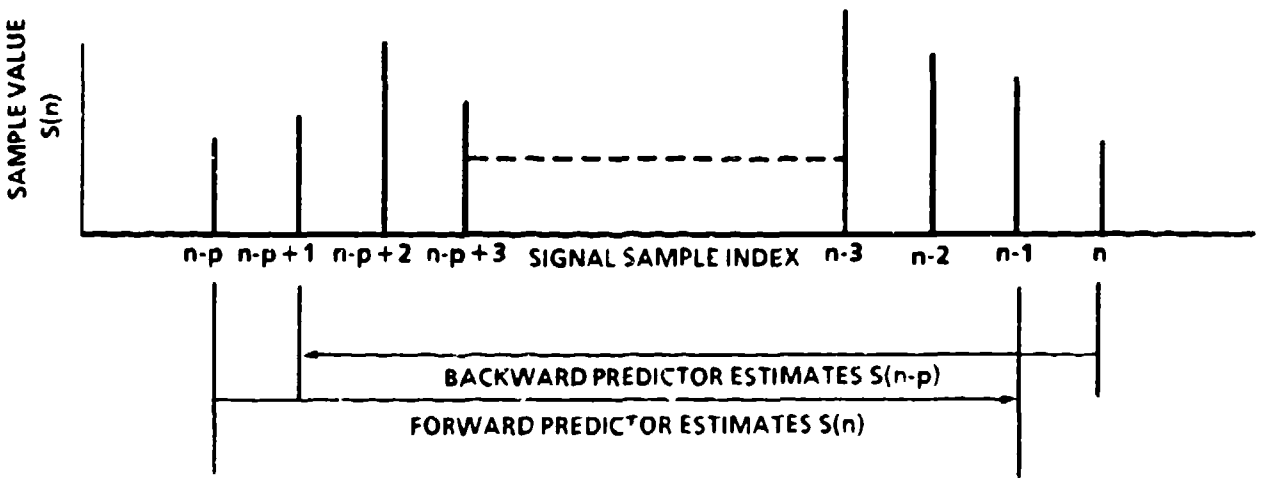


FIGURE 32. FORWARD/BACKWARD ESTIMATION

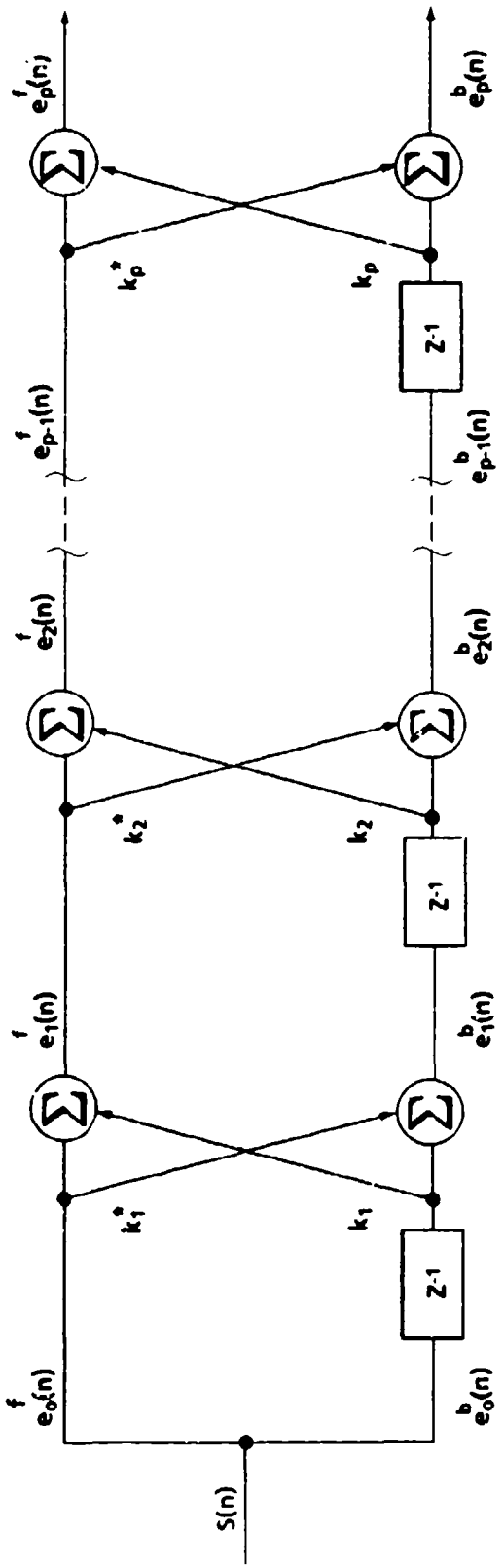


FIGURE 33. LATTICE FILTER WITH FORWARD AND BACKWARD PREDICTION ERROR

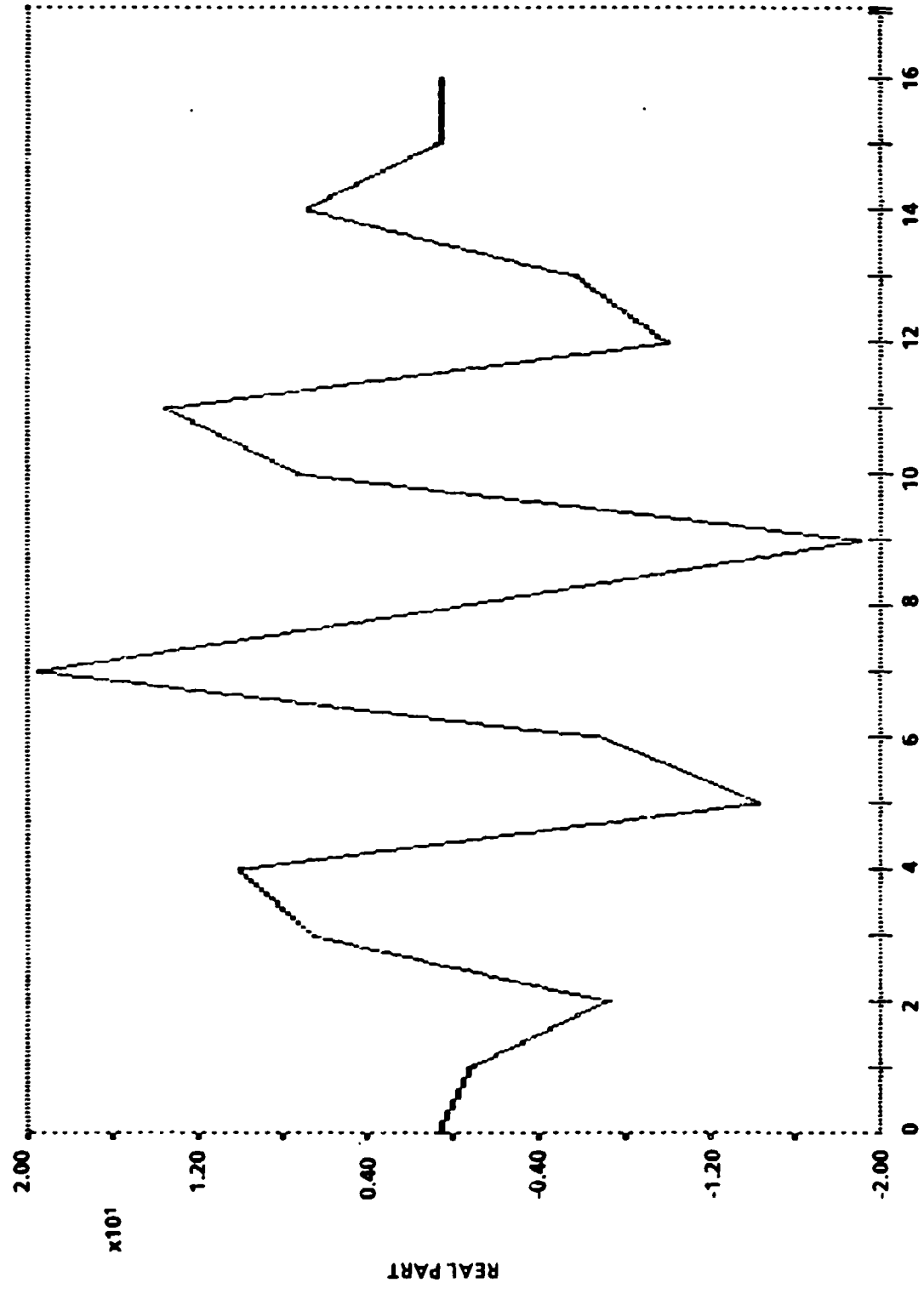


FIGURE 34. REAL PART OF A COMPLEX TRANSIENT (16 DATA SAMPLES) COMPOSED OF THE FREQUENCIES 0.25 AND 0.3125 OF EQUAL AMPLITUDE AND 160-DEGREE PHASE SHIFT

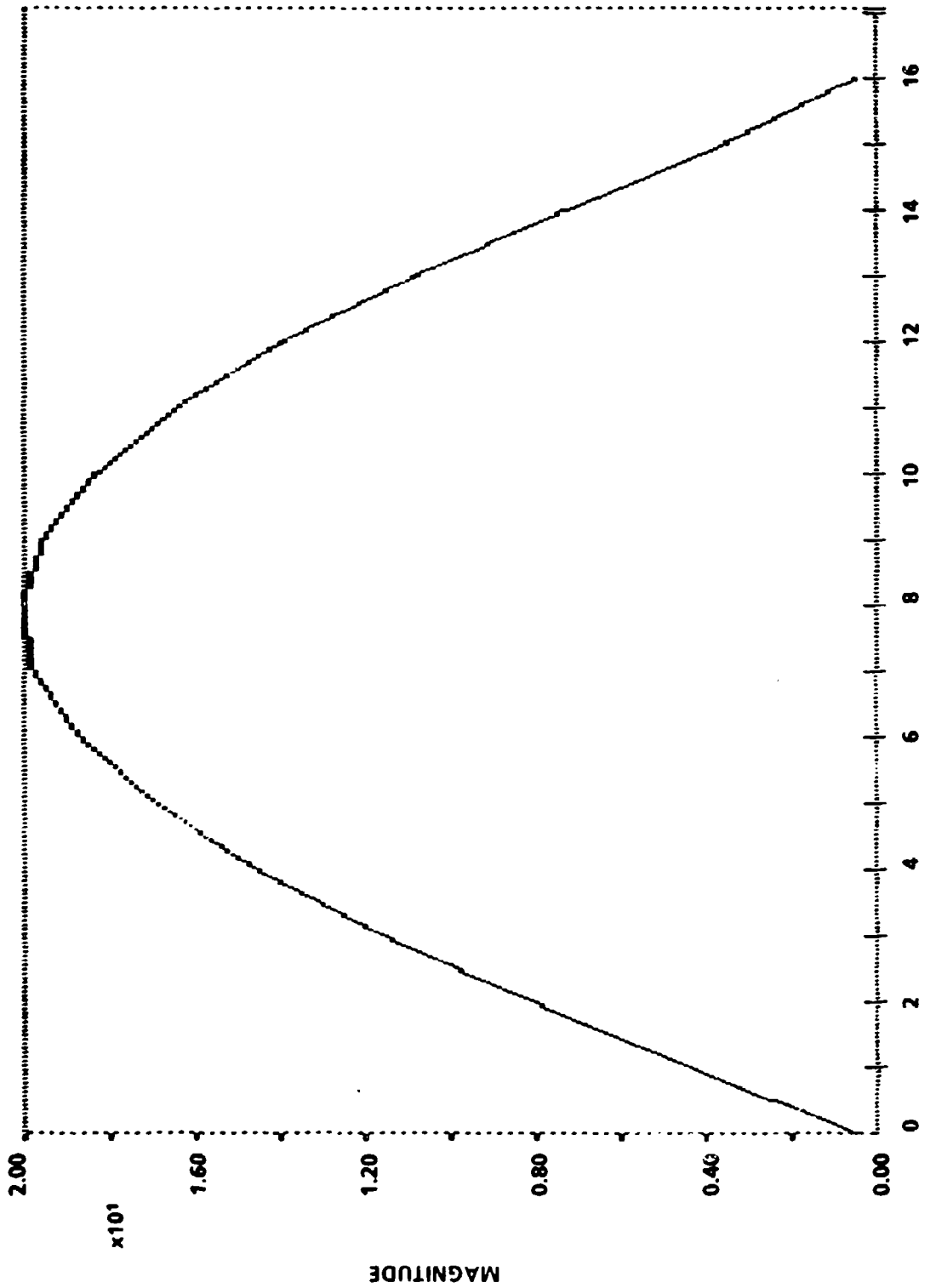


FIGURE 35. MAGNITUDE OF A COMPLEX TRANSIENT (16 DATA SAMPLES) COMPOSED OF THE FREQUENCIES 0.25 AND 0.3125 OF EQUAL AMPLITUDE AND 160-DEGREE PHASE SHIFT

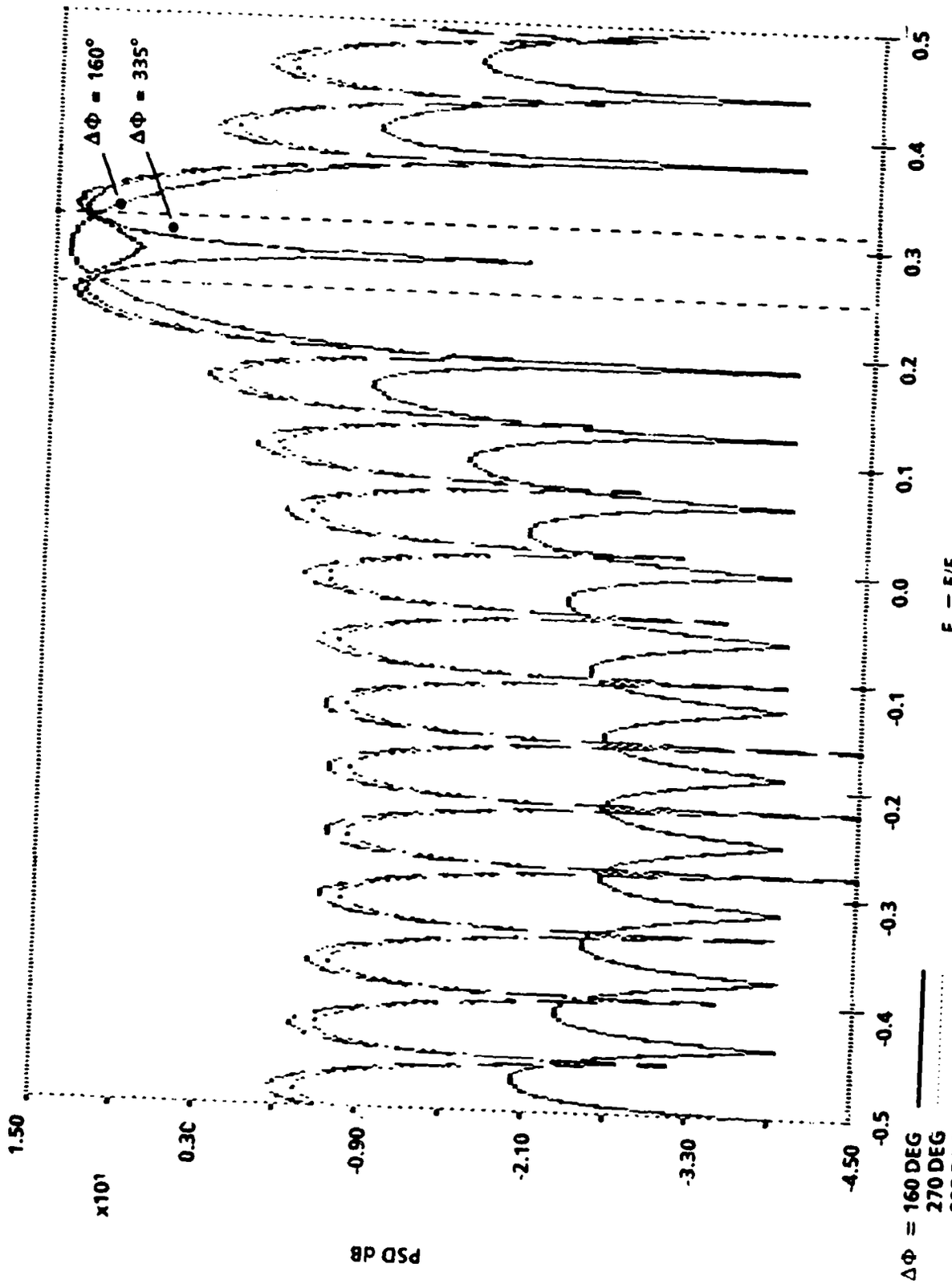


FIGURE 36. PERIODOGRAMS OF TRANSIENT OF FIGURES 34, 35 WITH THREE DIFFERENT PHASE SHIFTS ΔΦ

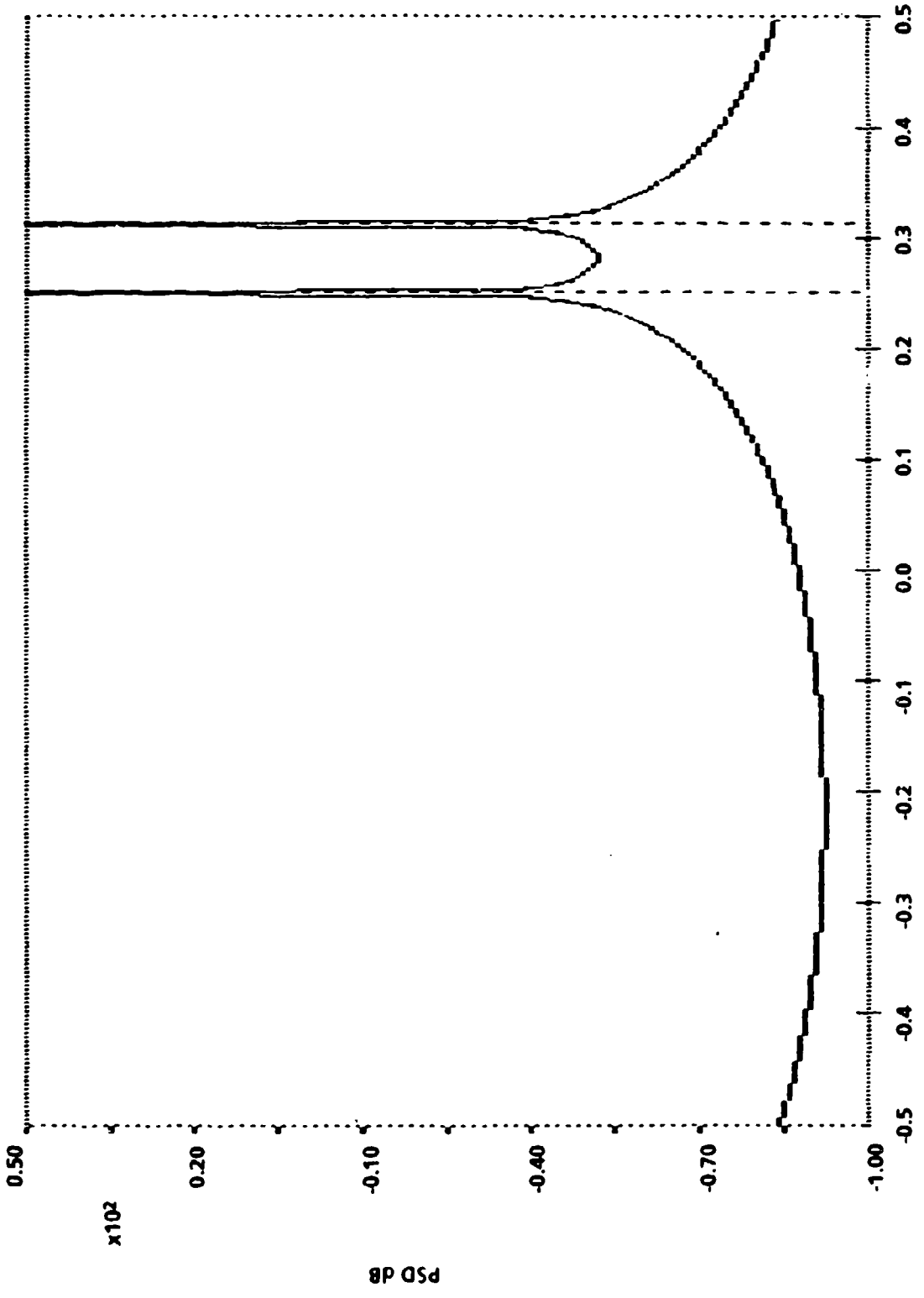


FIGURE 37. BURG SPECTRUM (P = 2) OF TRANSIENT OF FIGURES 34, 35

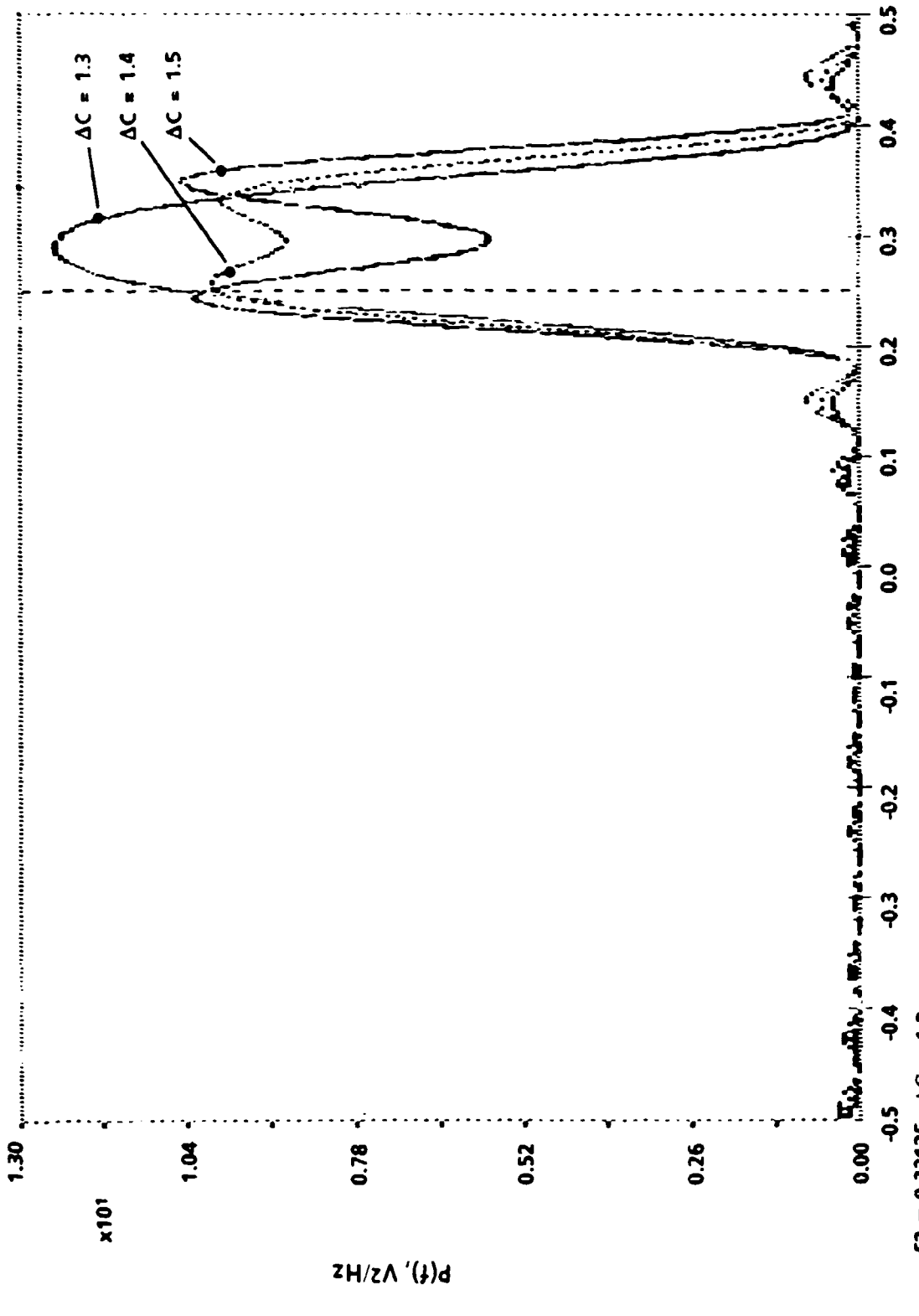


FIGURE 38. PERIODOGRAM OF TRANSIENT OTH OF EQUAL AMPLITUDE16 DATA SAMPLES, COMPOSED OF TWO FREQUENCIES F1 = 0.25 AND F2, BOTH OF EQUAL AMPLITUDE

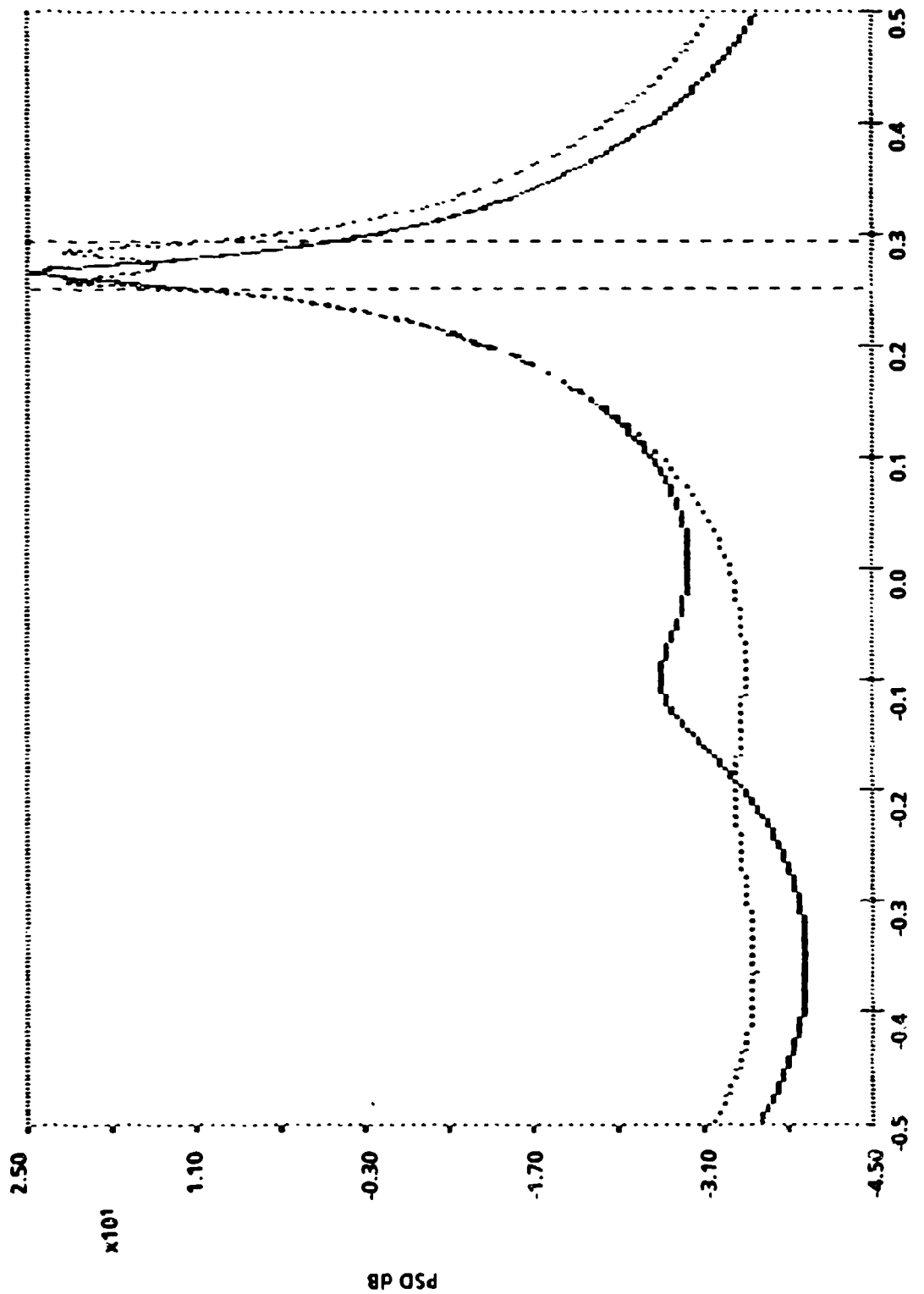


FIGURE 39. BURG SPECTRA (P = 4) OF A 16 SAMPLE TRANSIENT COMPOSED OF TWO FREQUENCIES F1 = 0.25 AND F2, BOTH OF EQUAL AMPLITUDE

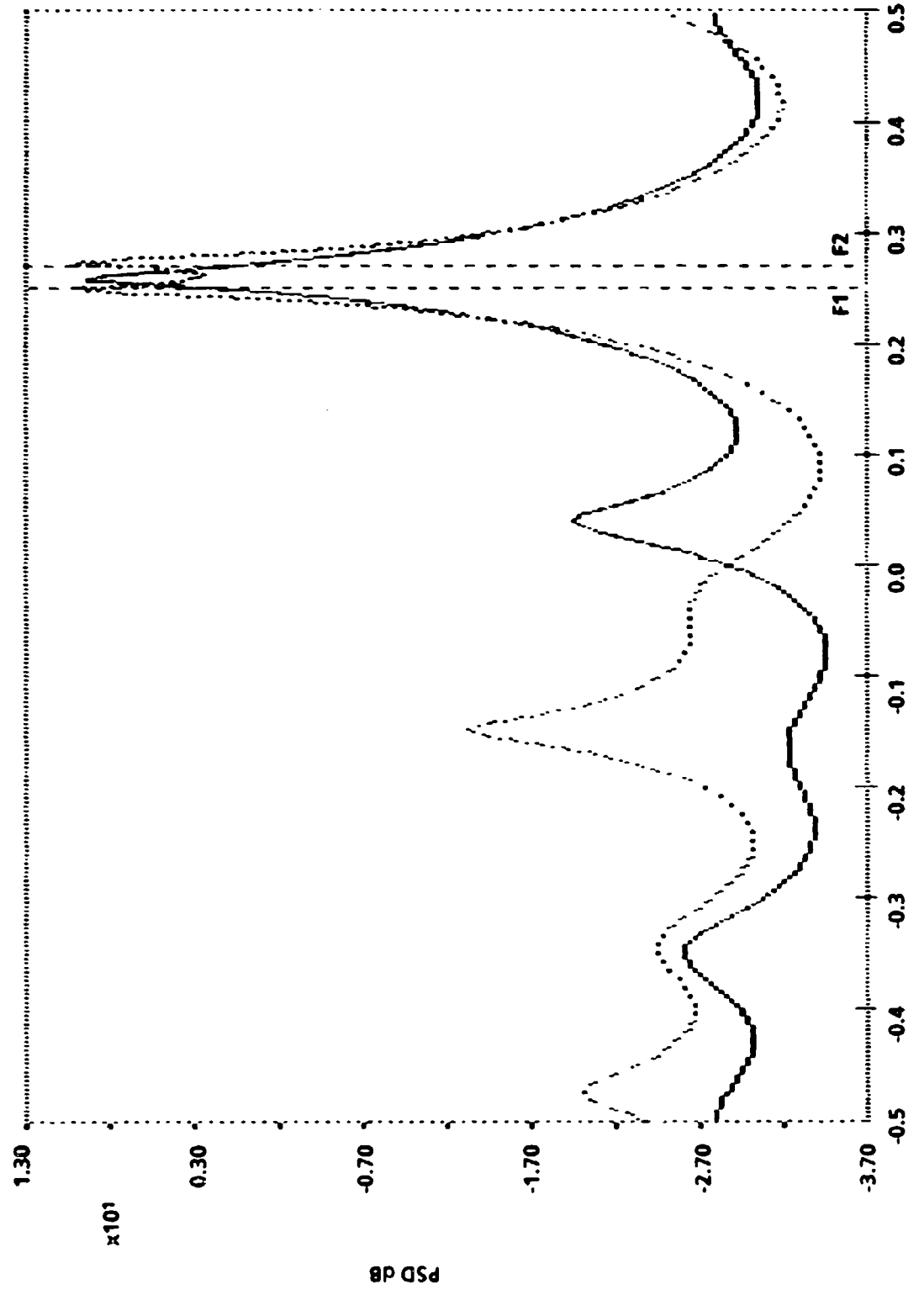


FIGURE 40. MODIFIED COVARIANCE SPECTRA (P = 4) OF A 16 SAMPLE TRANSIENT COMPOSED OF TWO FREQUENCIES F1 = 0.25 AND F2, BOTH OF EQUAL AMPLITUDE

FIGURE 41. PERIODOGRAM OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, UNIFORM WINDOW, SNR (F1) = 20 dB, SNR (F2) = 0 dB

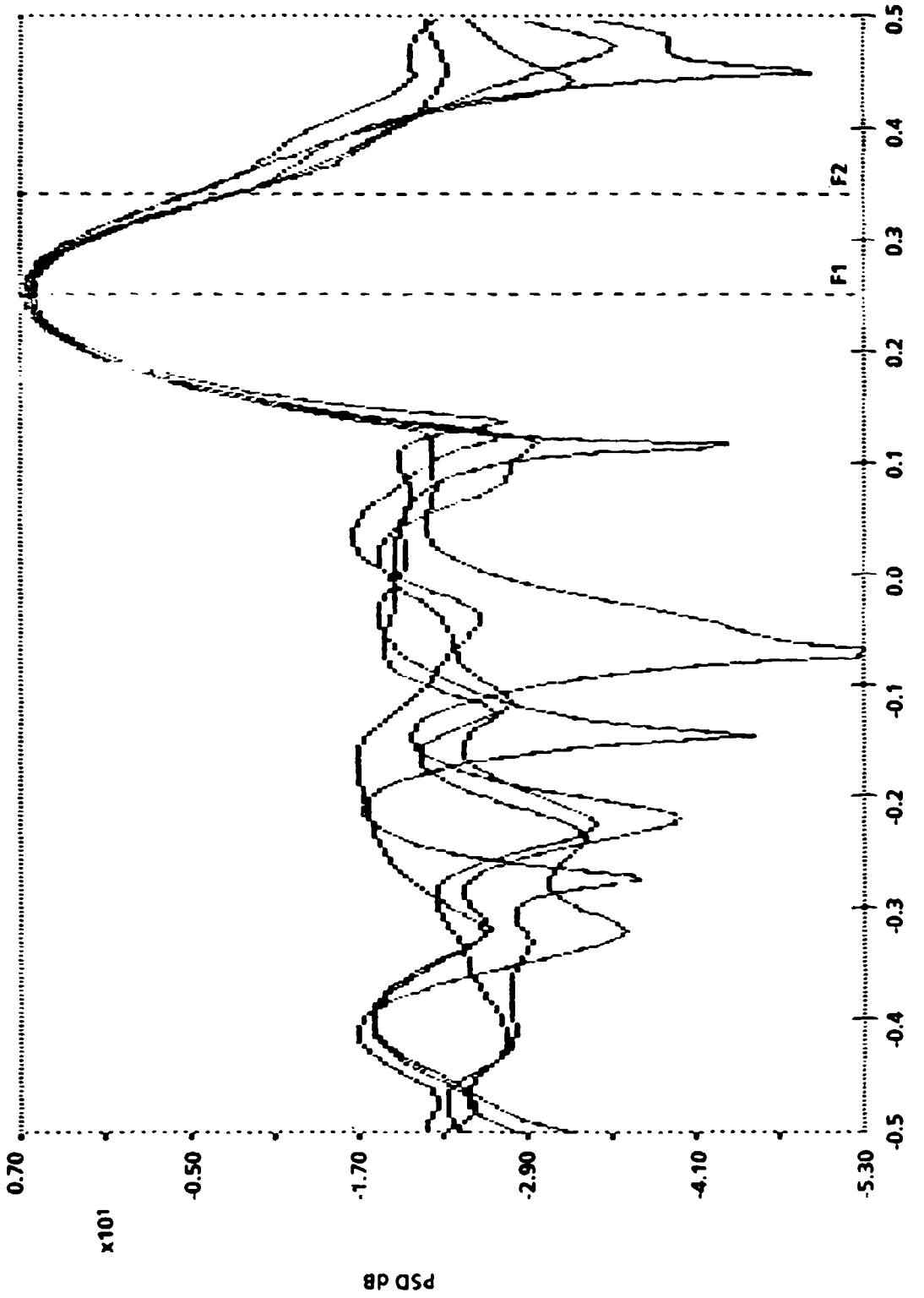


FIGURE 42. PERIODOGRAM OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, HAMMING WINDOW, SNR (F1) = 20 dB, SNR (F2) = 0 dB

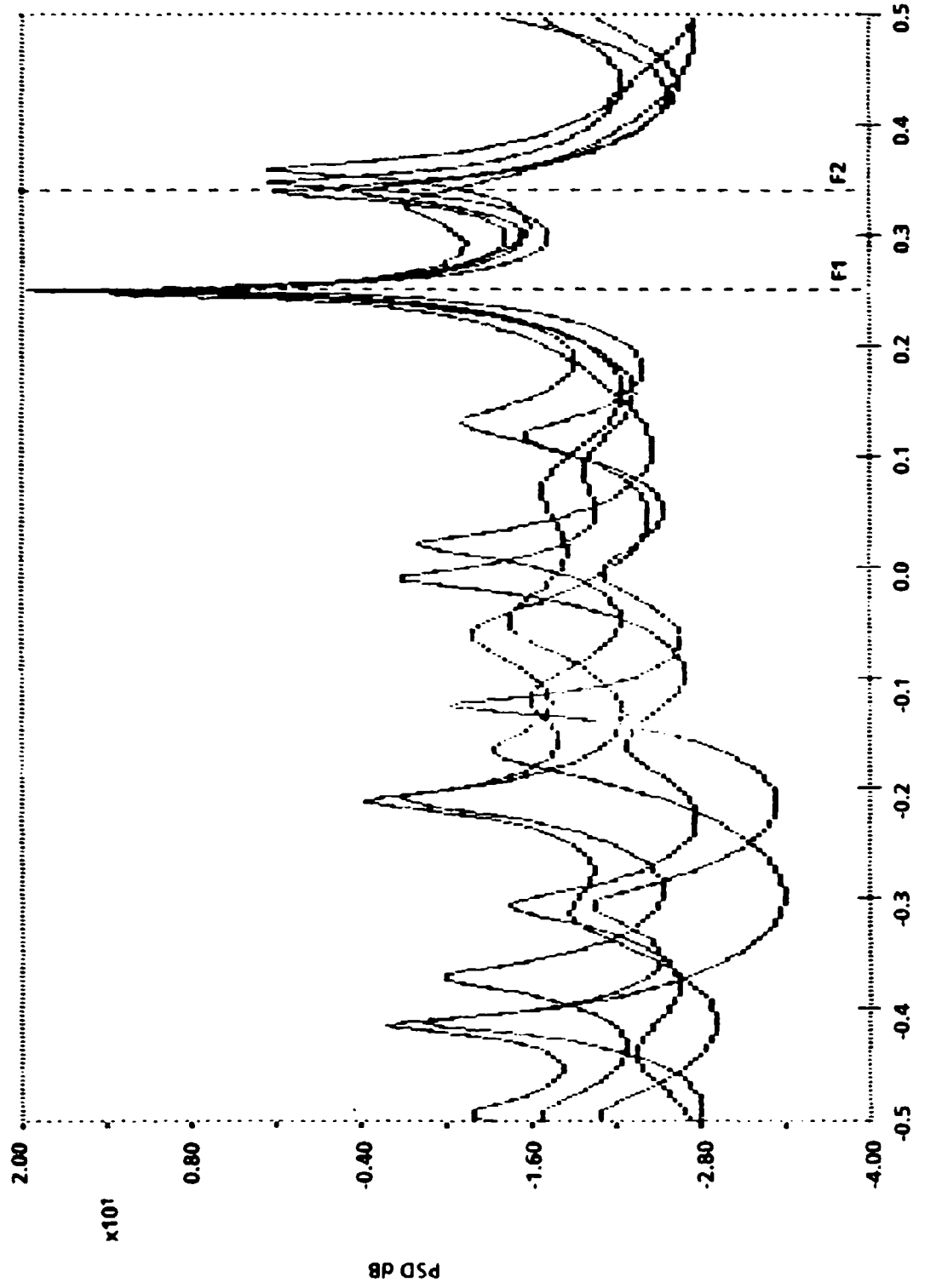


FIGURE 43. BURG SPECTRA (P = 8) OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR (F1) = 20 dB, SNR (F2) = 0 dB

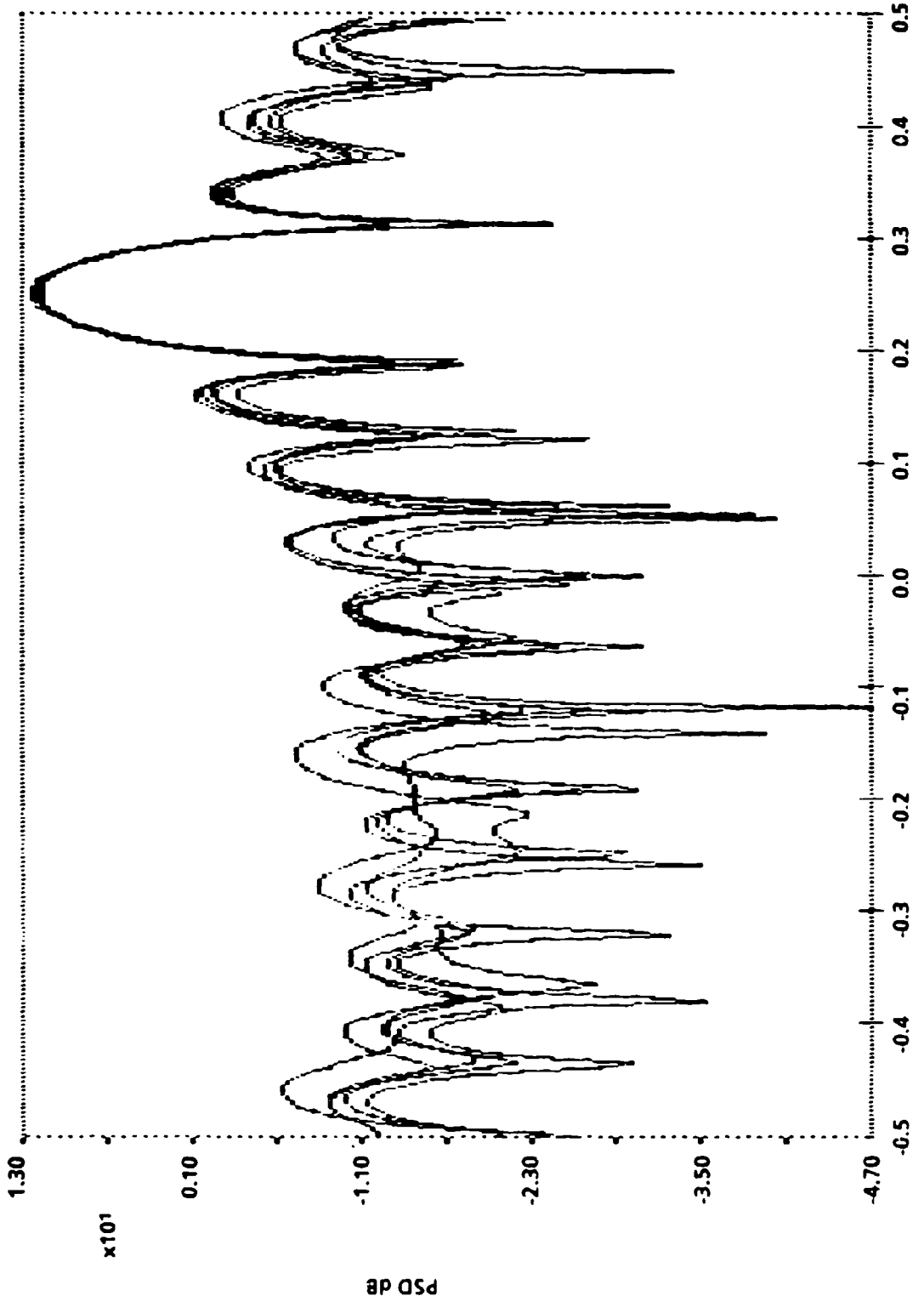


FIGURE 44. PERIODOGRAM OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.405) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, UNIFORM WINDOW, SNR (F1) = 20 dB, SNR (F2) = 0 dB

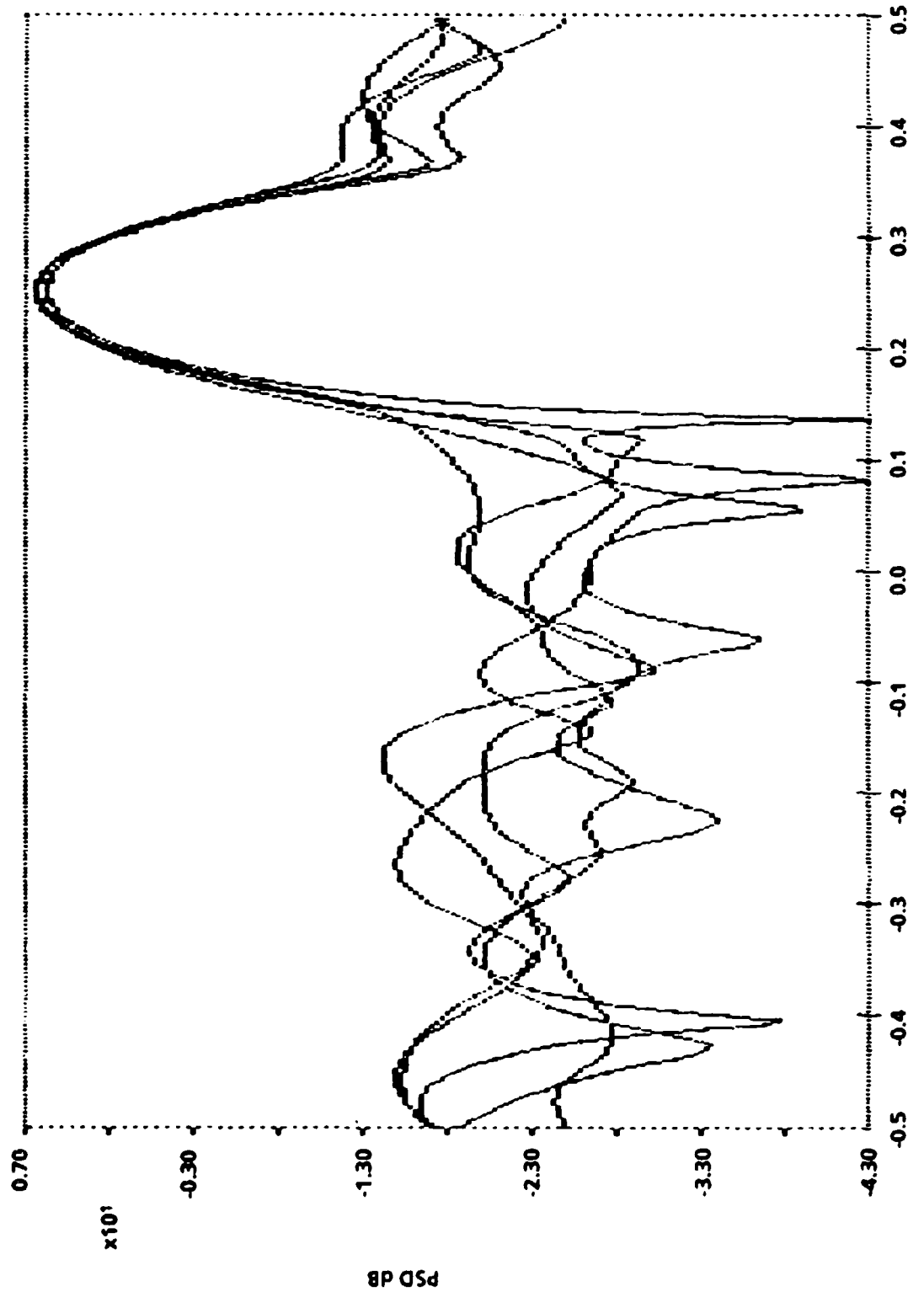


FIGURE 45. PERIODOGRAM OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.405) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, HAMMING WINDOW, SNR (F1) = 20 dB, SNR (F2) = 0 dB

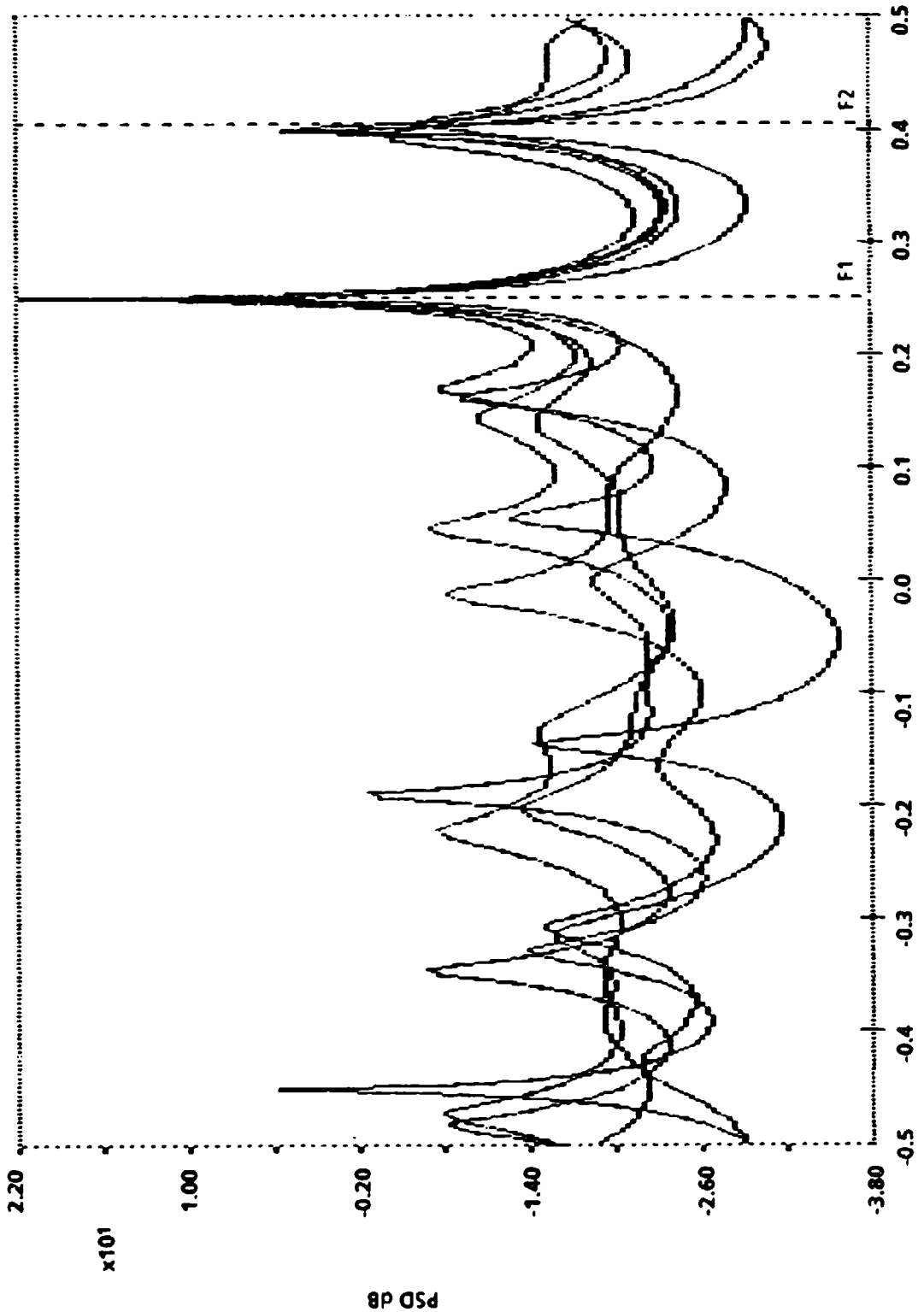


FIGURE 46. BURG SPECTRA (P = 8) OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.405) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR (F1) = 20 dB, SNR (F2) = 0 dB

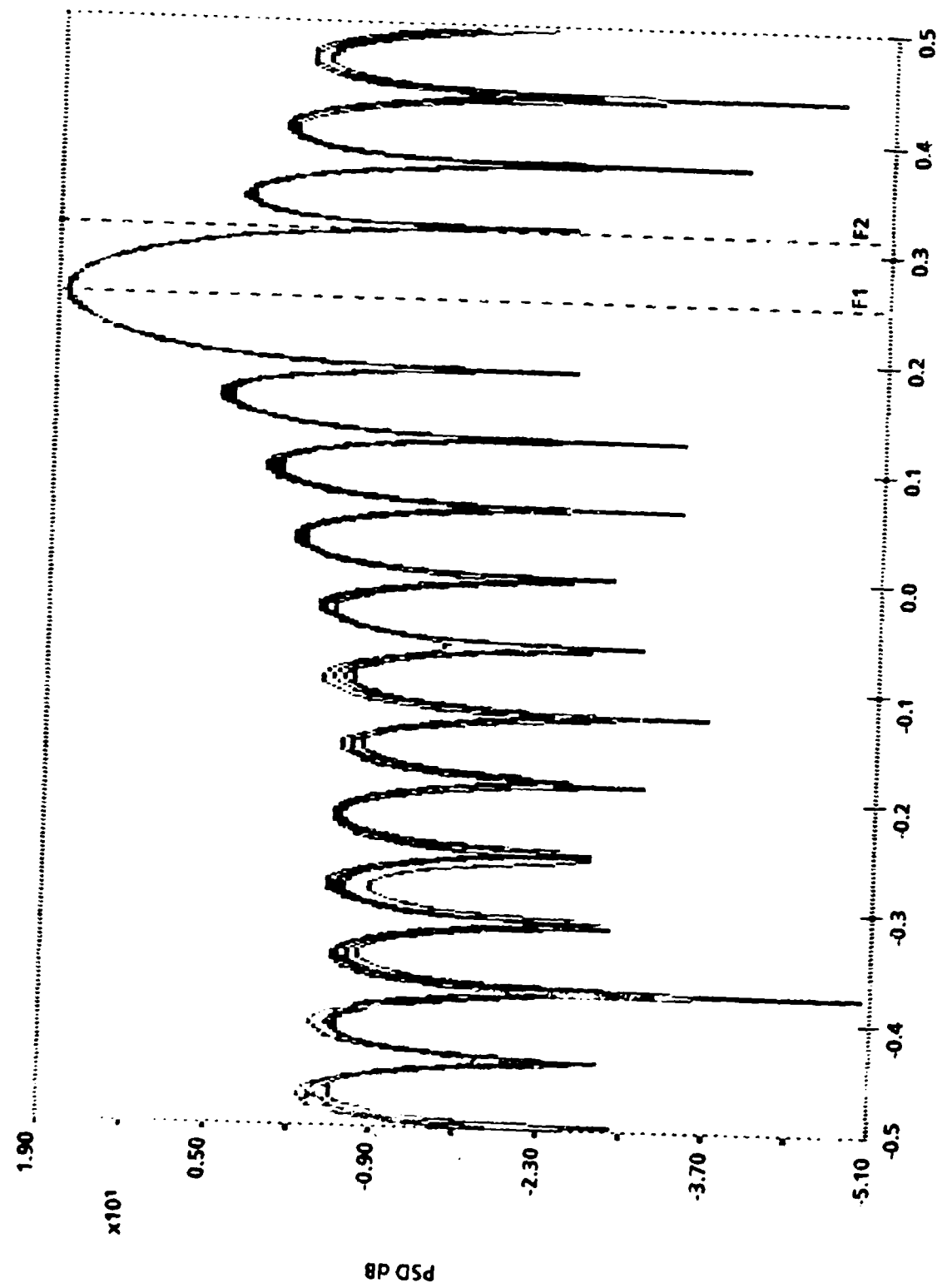


FIGURE 47. PERIODOGRAM OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.312) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, UNIFORM WINDOW, SNR (F1) = 20 dB, SNR (F2) = 0 dB

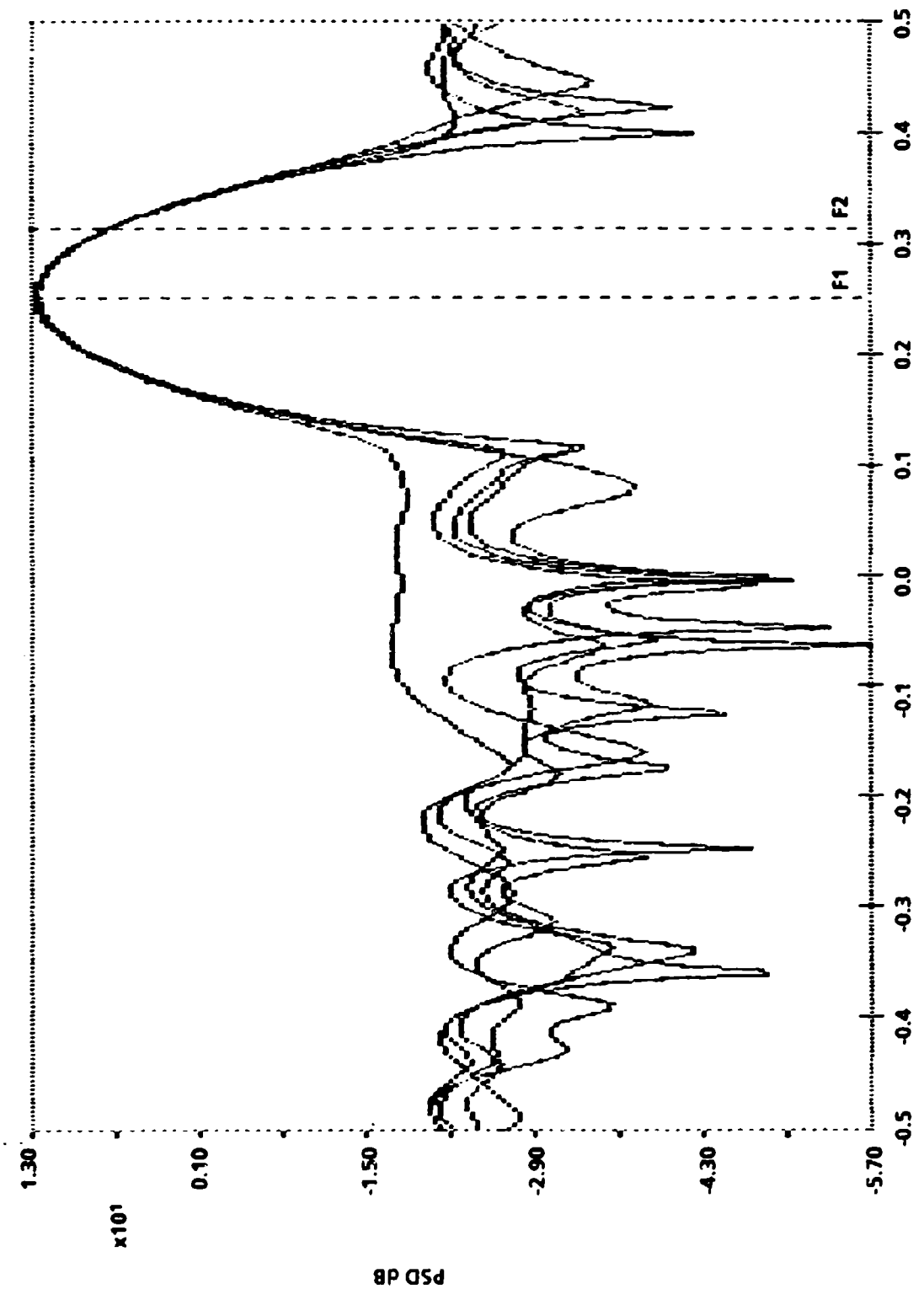


FIGURE 48. PERIODOGRAM OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.312) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, HAMMING WINDOW, SNR (F1) = 20 dB, SNR (F2) = 0 dB

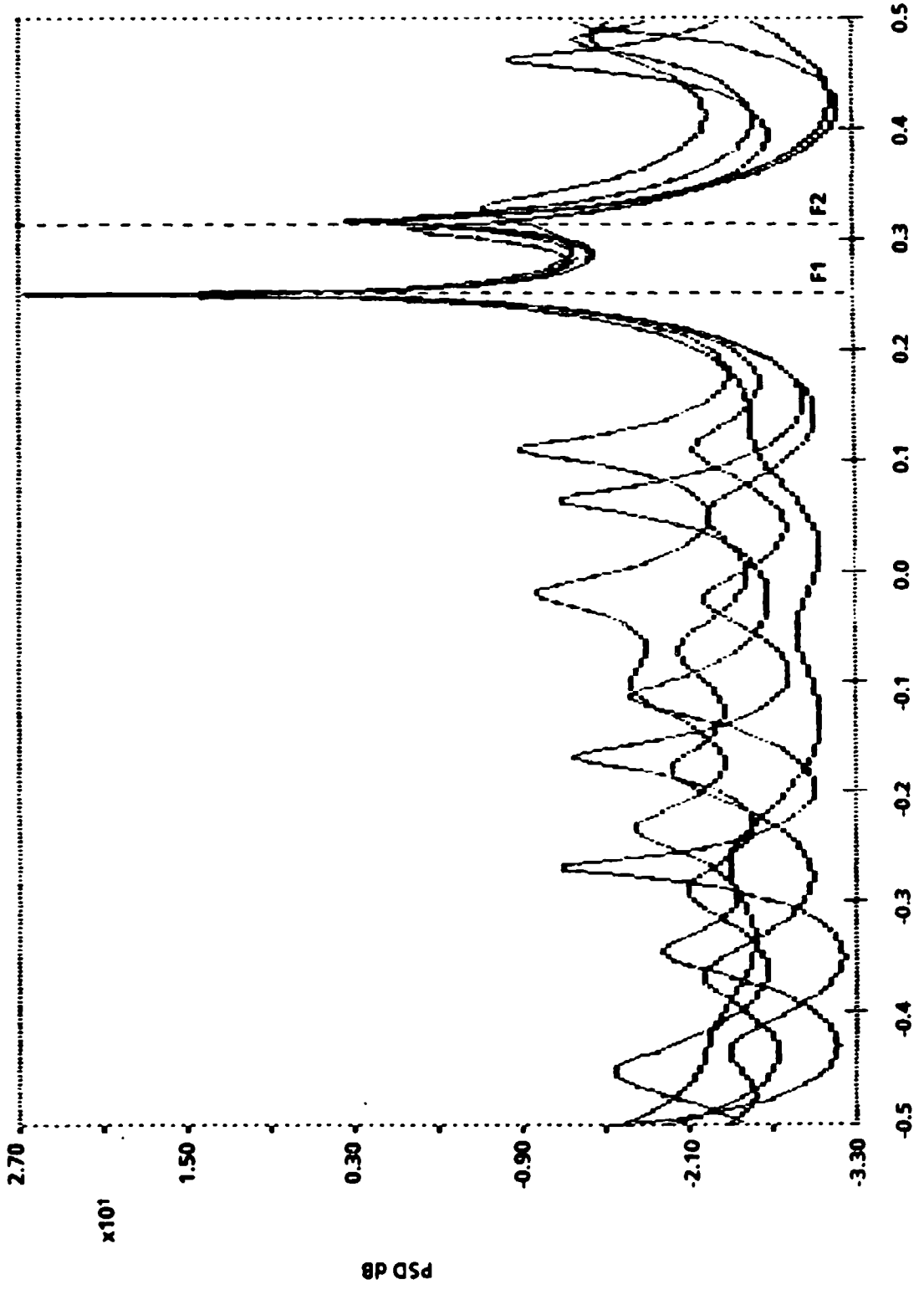


FIGURE 49. BURG SPECTRA (P = 8) OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.312) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR (F1) = 20 dB, SNR (F2) = 0 dB

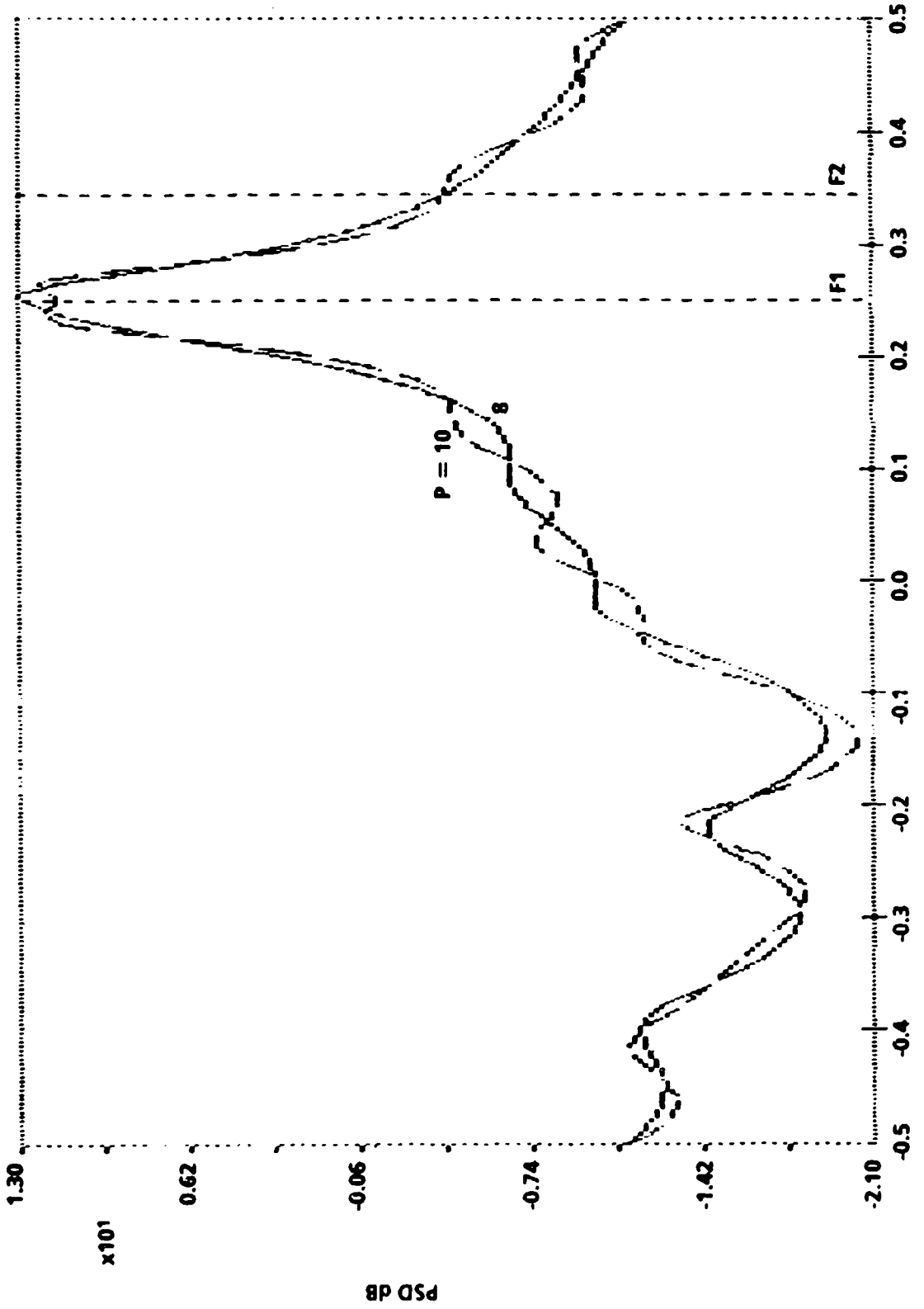


FIGURE 50. AUTOCORRELATION SPECTRA OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 16 SAMPLES, SNR (F1) = 20 dB, SNR (F2) = 0 dB

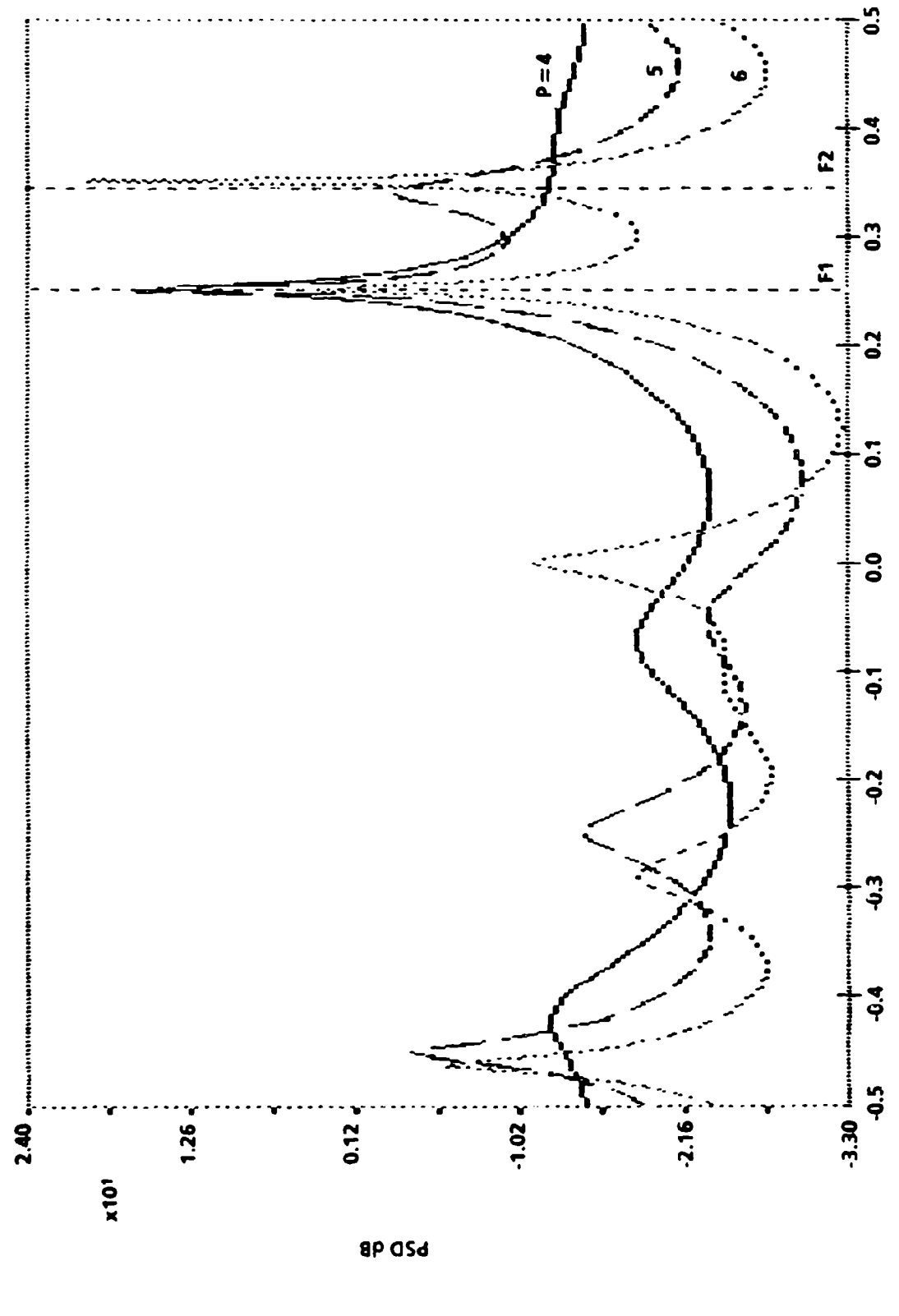


FIGURE 51. COVARIANCE SPECTRA OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 16 SAMPLES, SNR (F1) = 20 dB, SNR (F2) = 0 dB

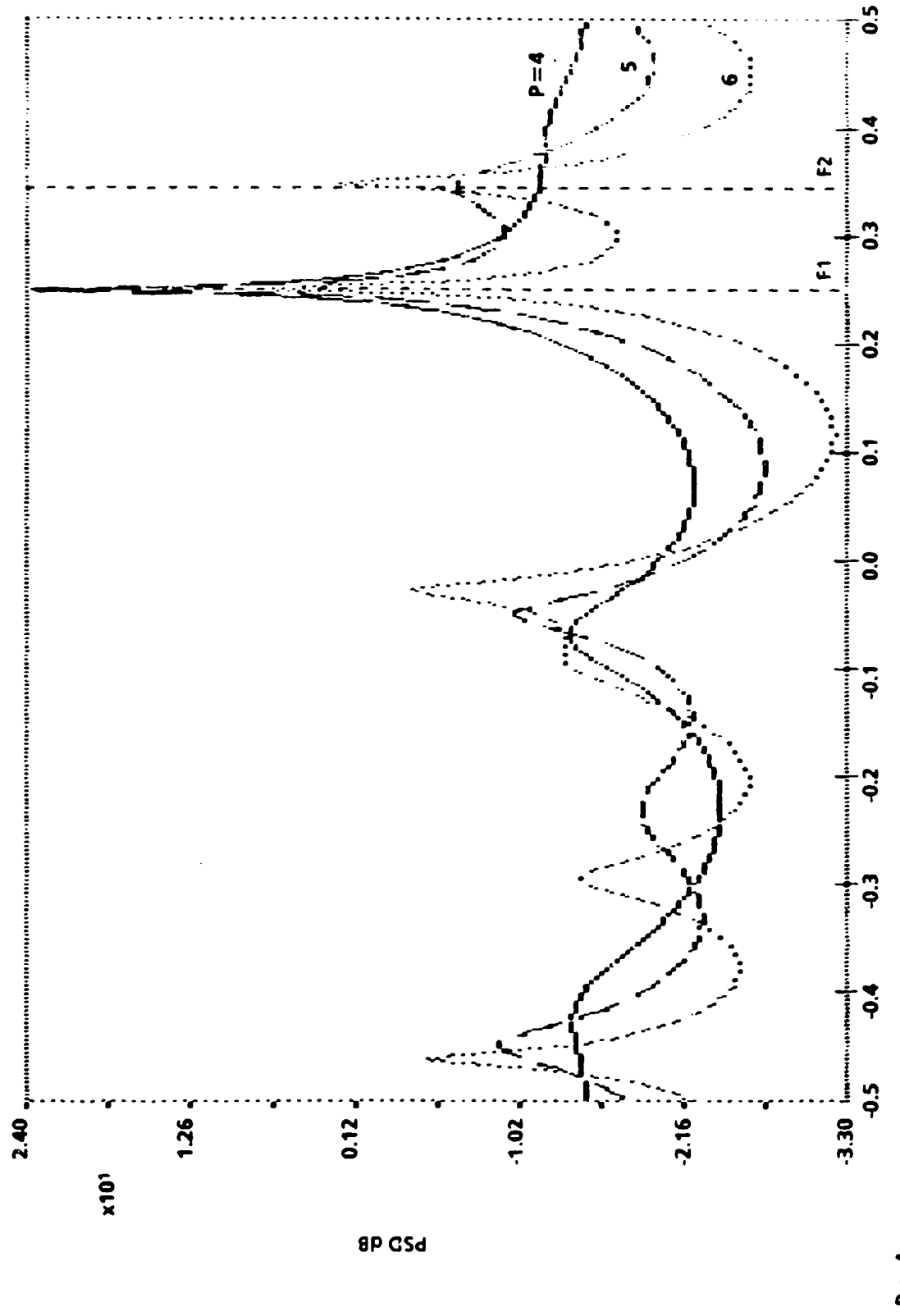


FIGURE 52. MODIFIED COVARIANCE SPECTRA OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 16 SAMPLES, SNR (F1) = 20 dB, SNR (F2) = 0 dB

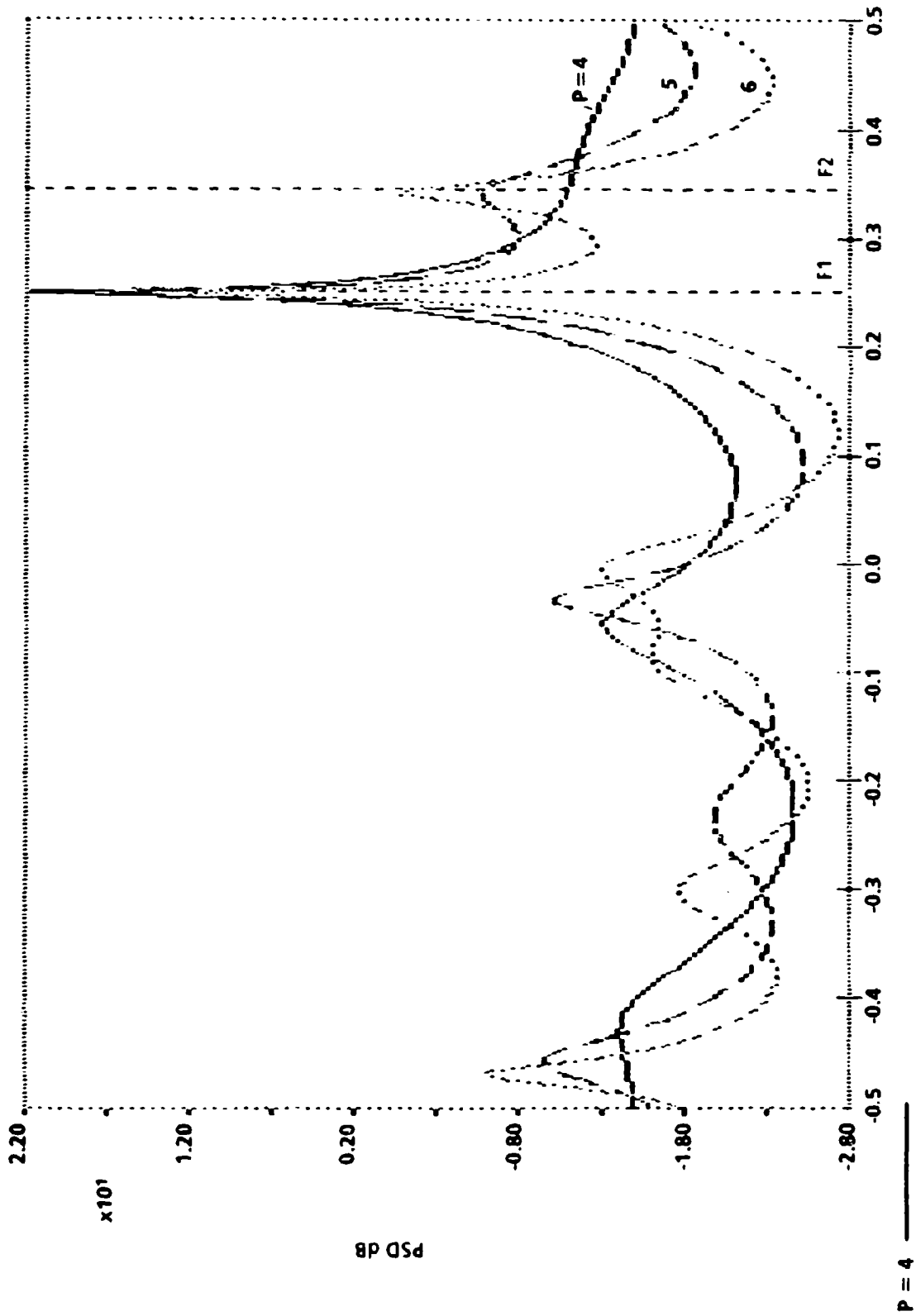


FIGURE 53. BURG SPECTRA OF TWO SINUSOIDS (F1 = 0.25, F2 = 0.34) AND WHITE NOISE, 16 SAMPLES, SNR (F1) = 20 dB, SNR (F2) = 0 dB

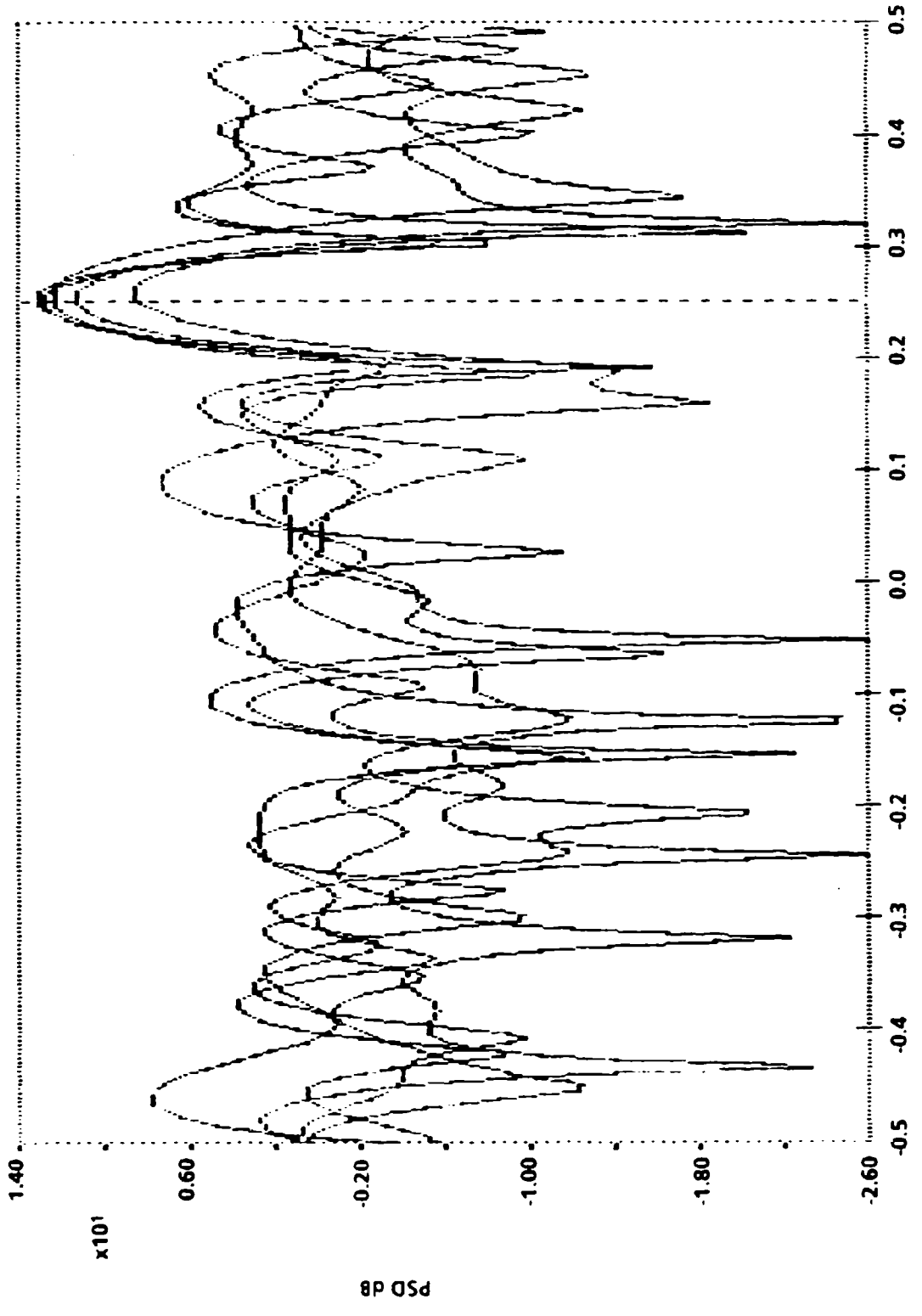


FIGURE 54. PERIODOGRAM OF SINUSOID ($F_r = 0.25$) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, UNIFORM WINDOW, SNR = 0 dB

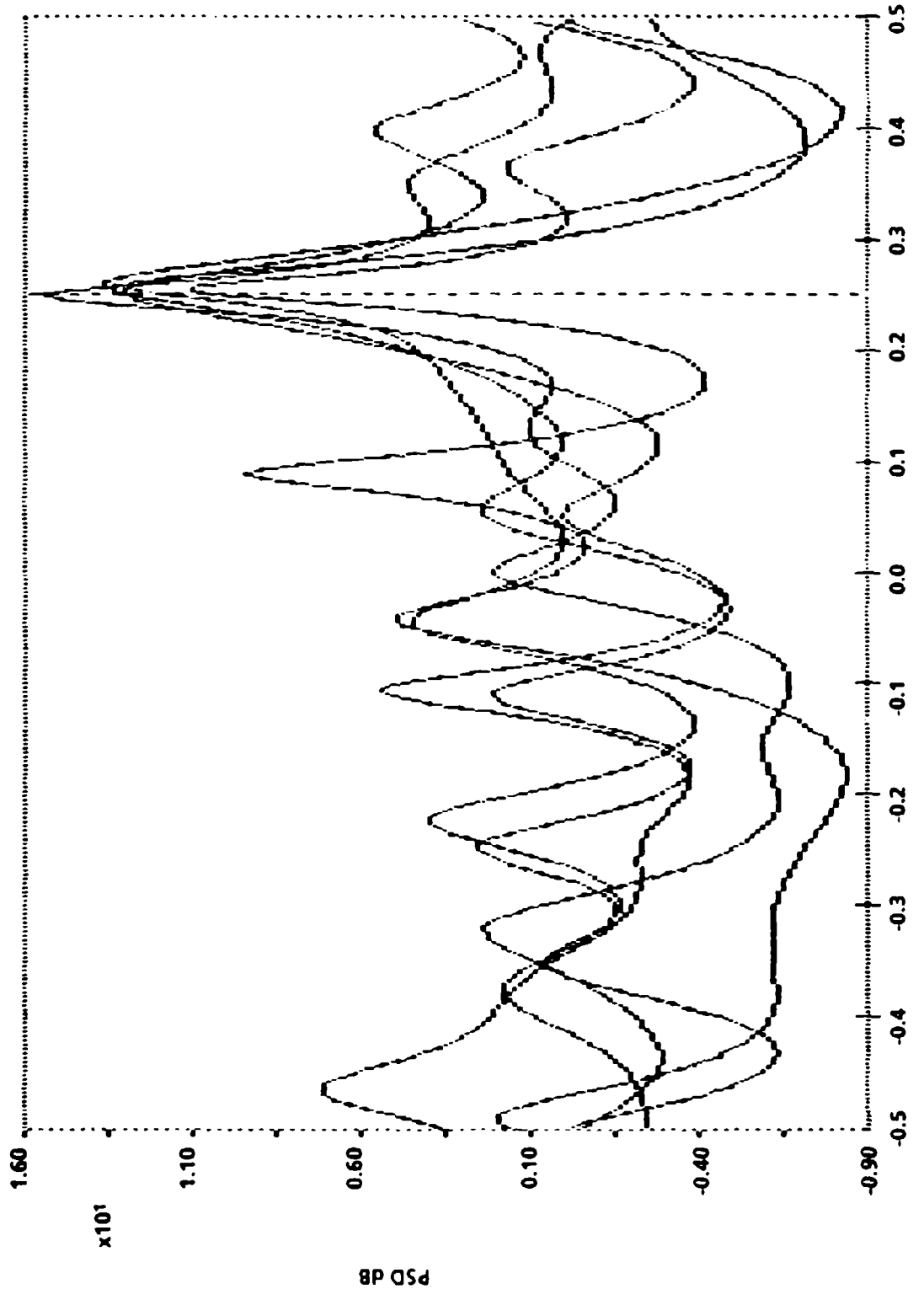


FIGURE 55. AUTOCORRELATION SPECTRA (P = 6) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = 0 dB

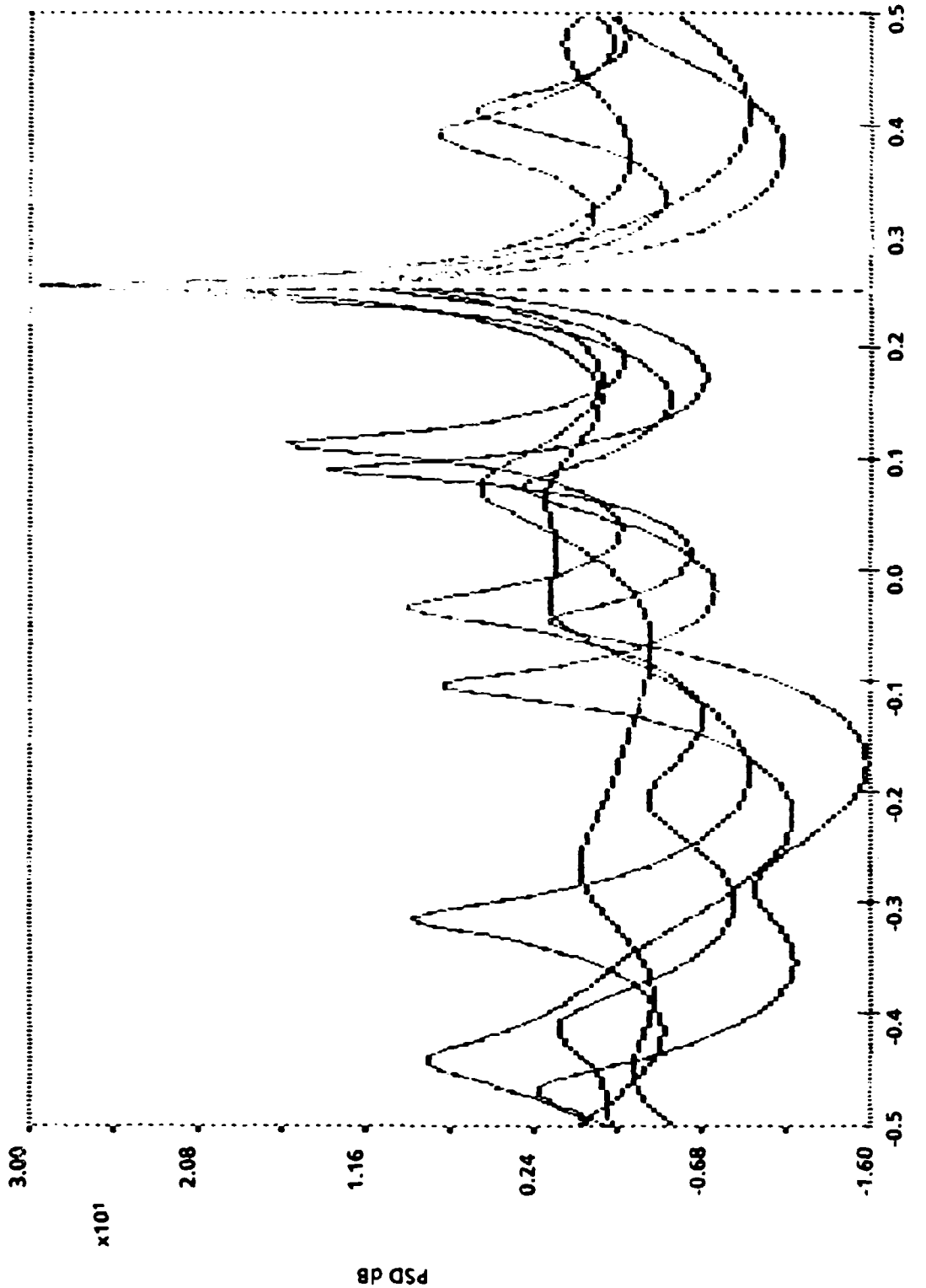


FIGURE 56. COVARIANCE SPECTRA (P = 8) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = 0 dB

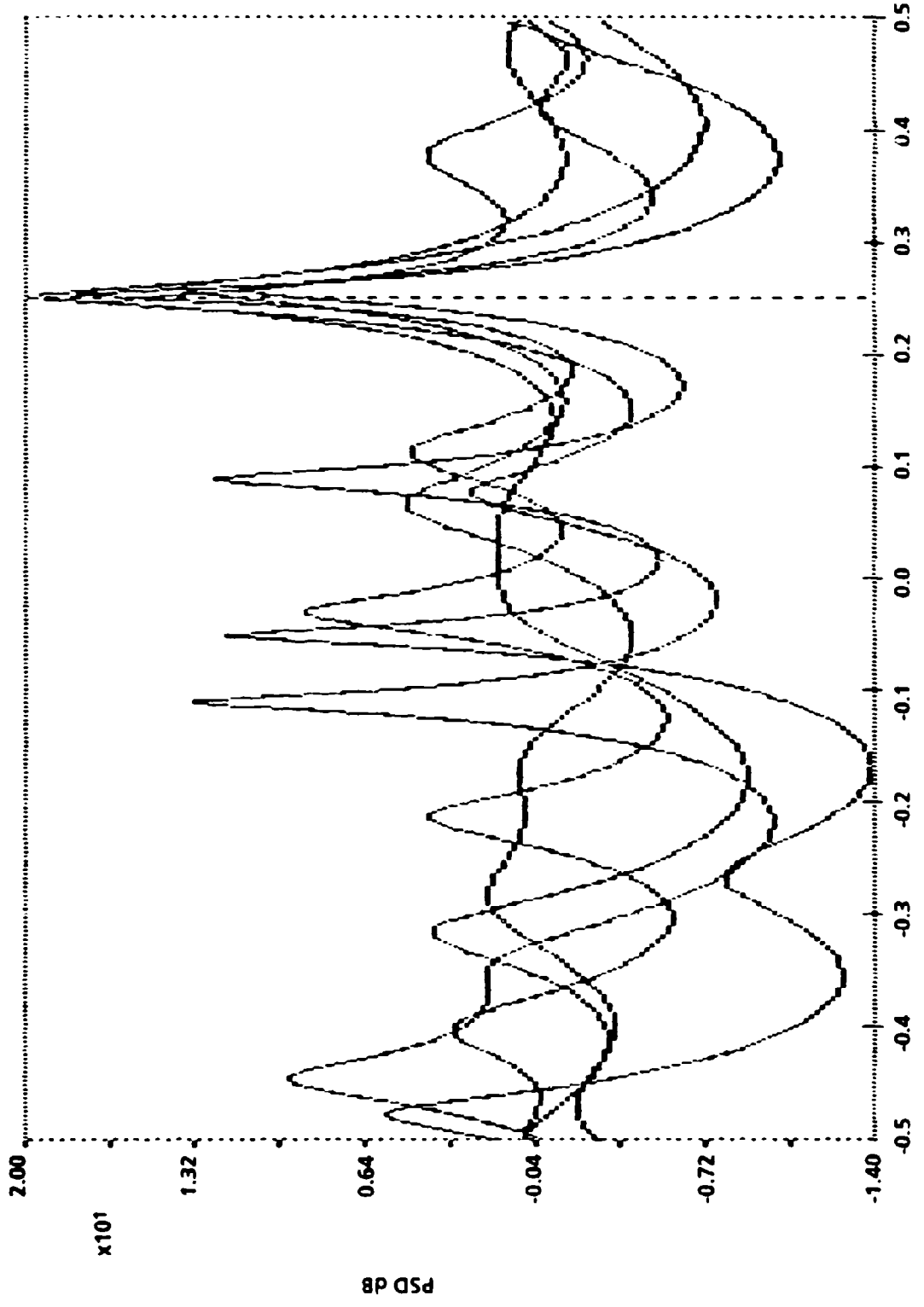


FIGURE 57. MODIFIED COVARIANCE SPECTRA (P = 6) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = 0 dB

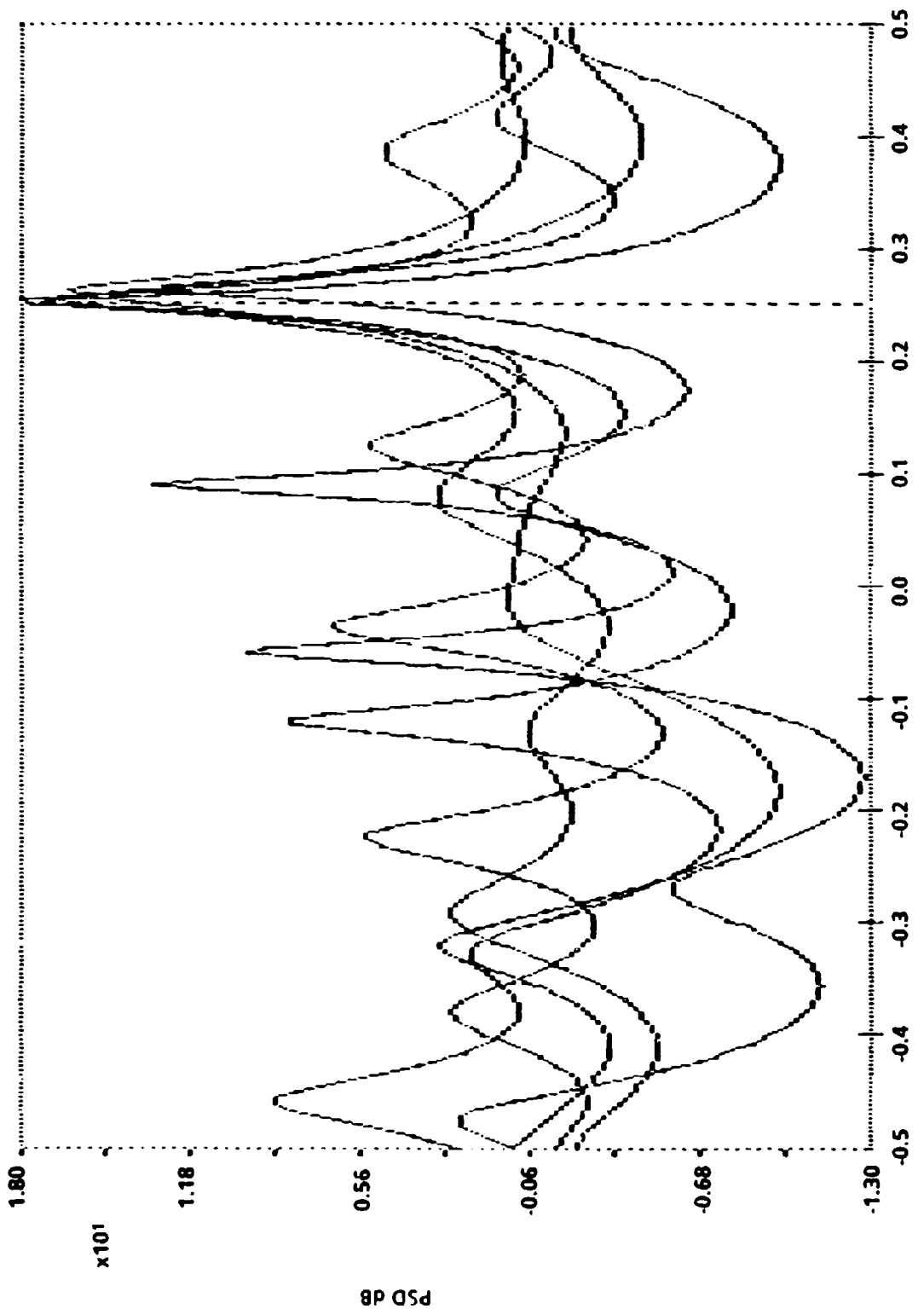


FIGURE 58. BURG SPECTRA (P = 6) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = 0 dB

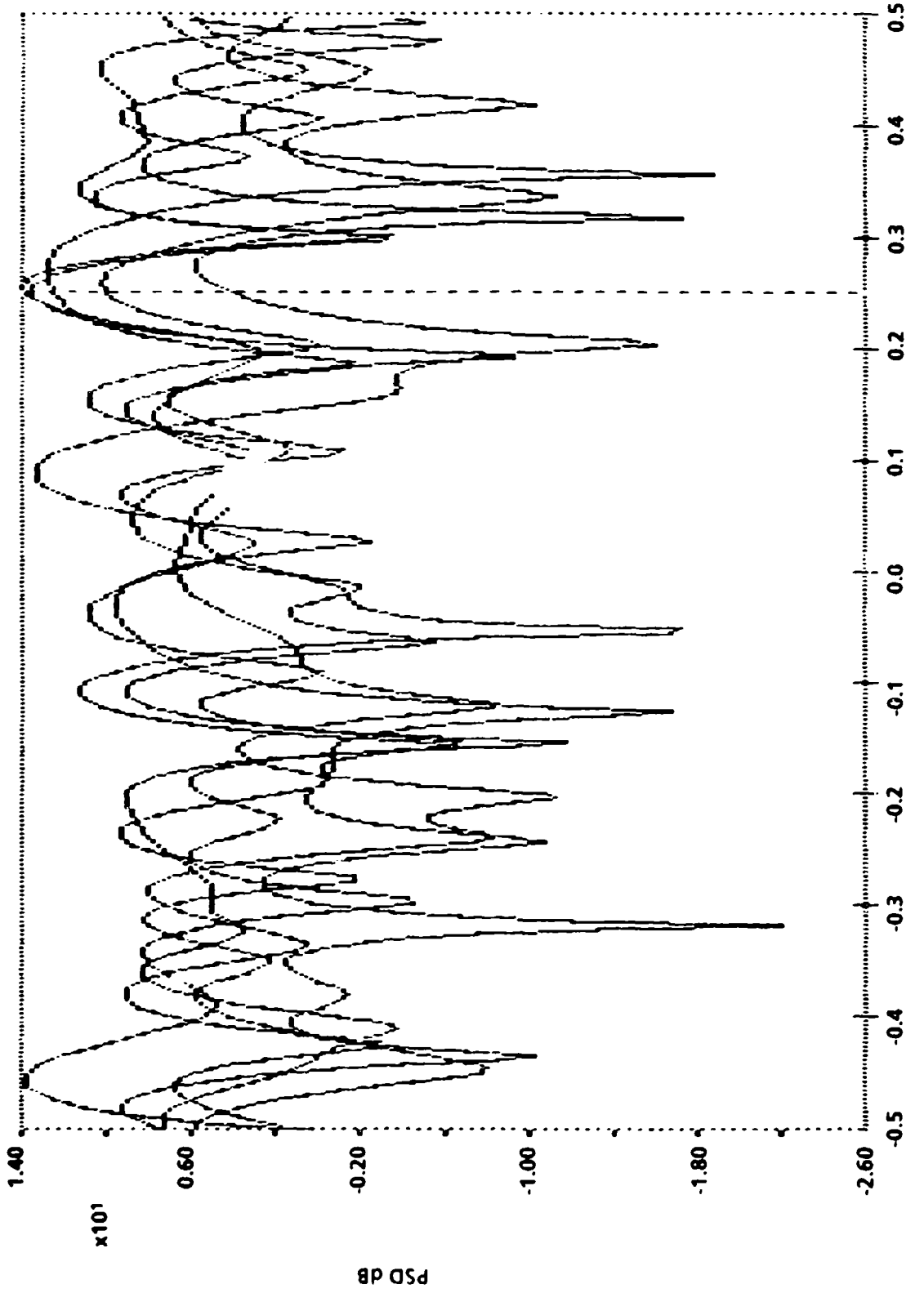


FIGURE 59. PERIODOGRAM OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, UNIFORM WINDOW, SNR = -6 dB

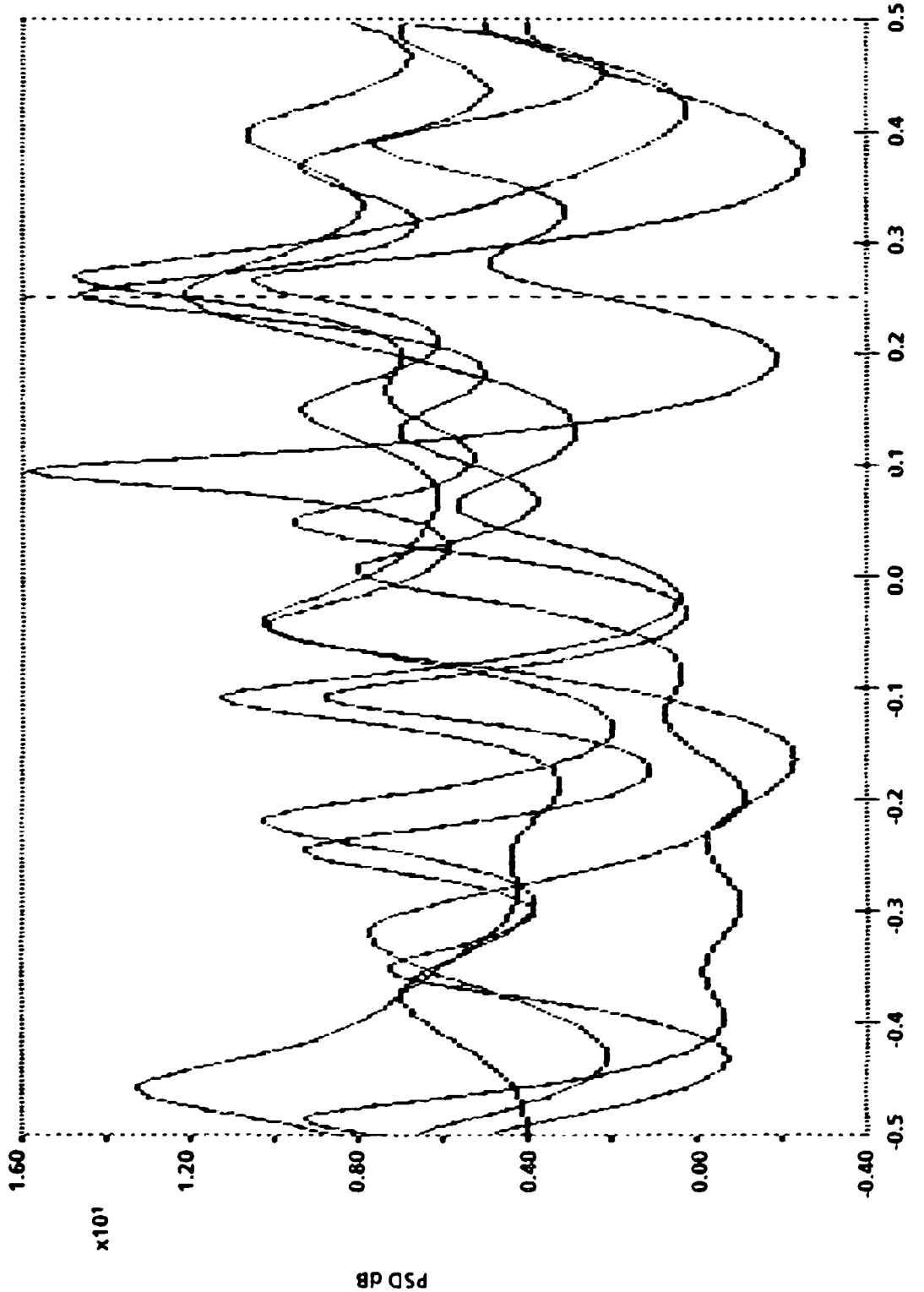


FIGURE 60. AUTOCORRELATION SPECTRA (P = 8) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = -6 dB

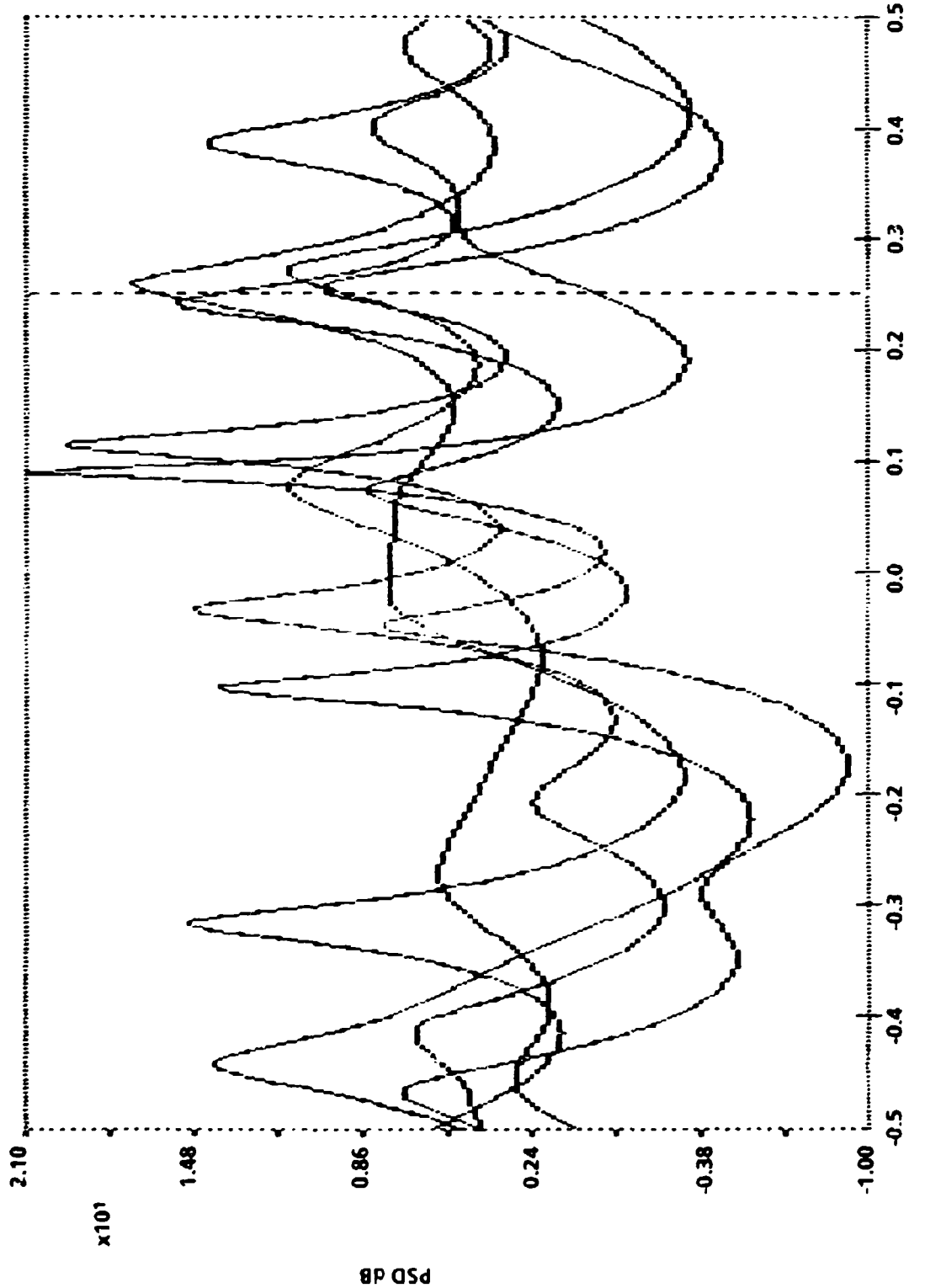


FIGURE 61. COVARIANCE SPECTRA (P = 6) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = -6 dB

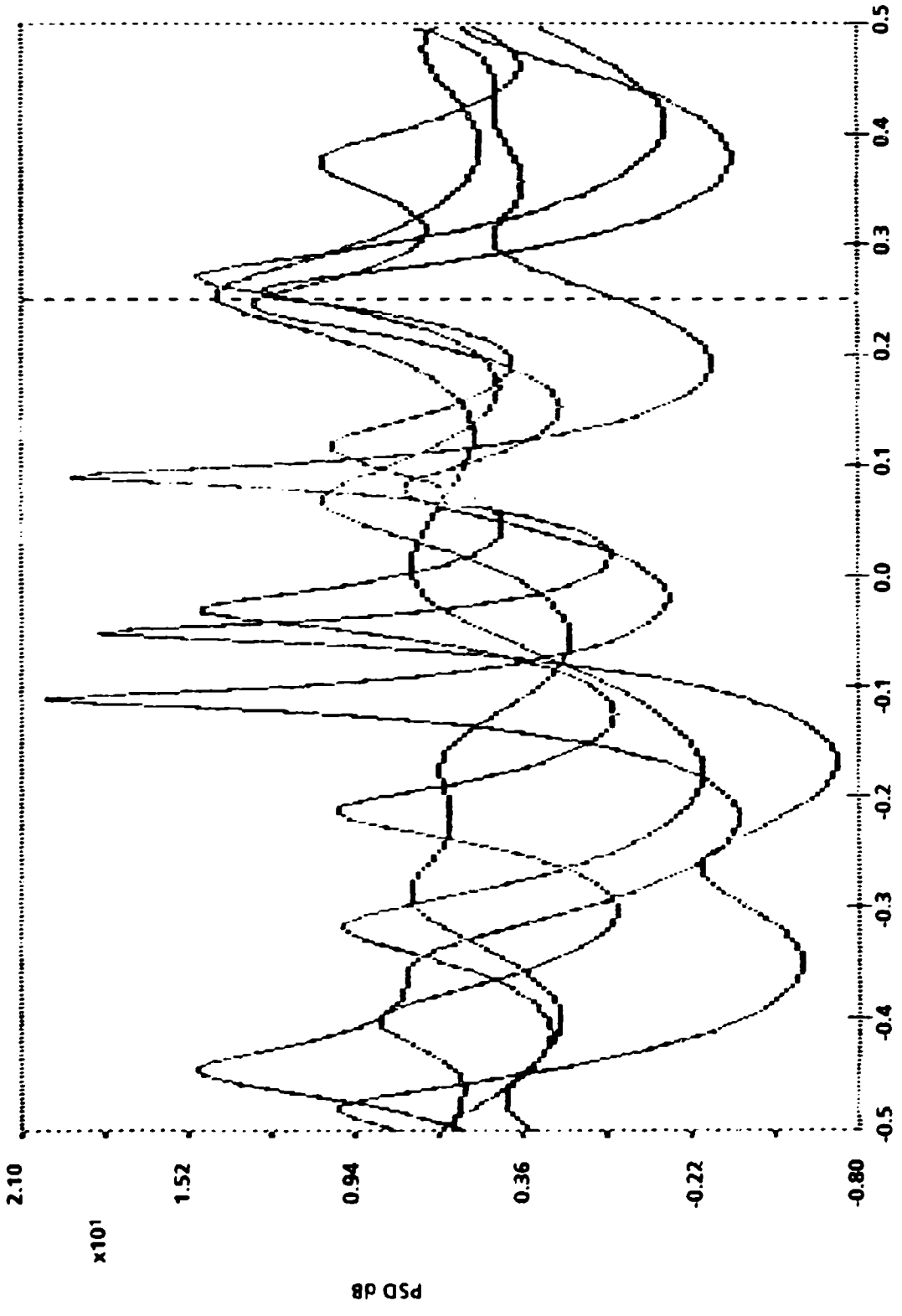


FIGURE 62. MODIFIED COVARIANCE SPECTRA (P = 6) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = -6 dB

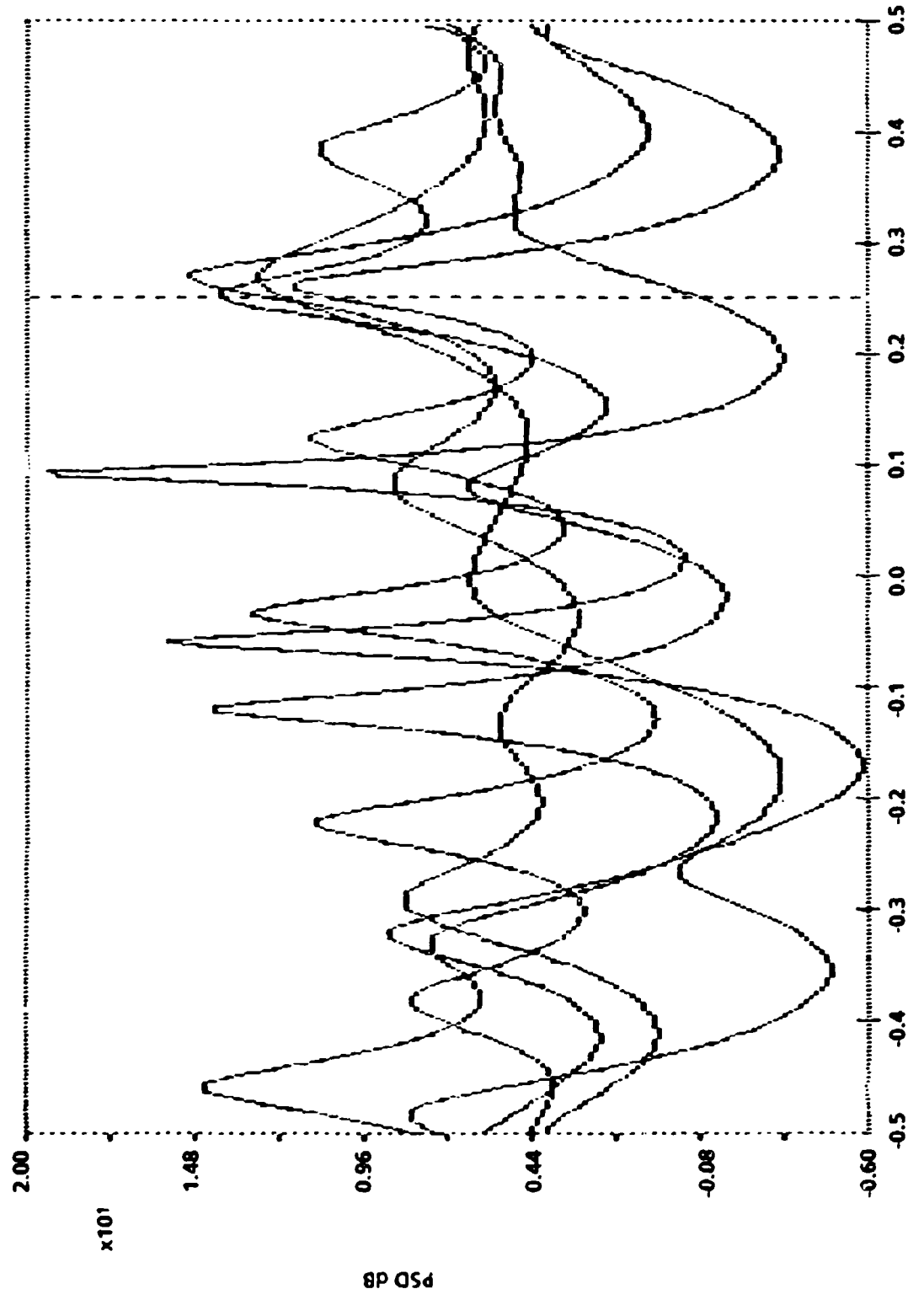


FIGURE 63. BURG SPECTRA (P = 6) OF A SINUSOID (F = 0.25) AND WHITE NOISE, 5 RECORDS, 16 SAMPLES EACH, SNR = -6 dB

REFERENCES

1. Kay, S. M., Modern Spectral Estimation—Theory and Application, Prentice Hall, 1988.

2. Marple, L. S., Digital Spectral Analysis with Applications, Prentice Hall, 1987.

3. Modern Interactive Spectral Analysis (MISA), Spectrasoft, 190 Eighth Street, Providence, RI 02906, 1987.

4. Lee, Y. W., Statistical Theory of Communication, John Wiley & Sons, Inc., New York, NY, 1961.

5. Bergland, G. D., "A Guided Tour of the Fast Fourier Transform," IEEE Spectrum, Jul 1969, p. 41.

6. Rabiner, L. R. and Gold, B., Theory and Application of Digital Signal Processing, Prentice Hall, 1975.

7. Oppenheimer, A. V. and Schafer, R. W., Discrete-Time Signal Processing, Prentice Hall, 1989.

8. Childers, D. G., (Ed.), Modern Spectrum Analysis, IEEE Press, 1978.

9. Kesler, S. B., (Ed.), Modern Spectrum Analysis II, IEEE Press, 1986.

10. Kay, S. M. and Marple, L. S., "Spectrum Analysis—A Modern Perspective," Proc. IEEE, Vol. 69, Nov 1981, pp. 1380-1419.

11. Makhoul, J., "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, Apr 1975, pp 561-580.

12. Burg, J. P., "Maximum Entropy Spectral Analysis," Reprint: Modern Spectrum Analysis, IEEE Press, 1978.

APPENDIX A

POLE-ZERO FORMAT OF FILTER TRANSFER FUNCTION

This is to show the equivalence of the transfer function H(z) as a ratio of polynomials in z as used in the report:

$$H(z) = \frac{\sum_{i=0}^{q} b_i z^{-i}}{\sum_{i=0}^{p} a_i z^{-i}} \qquad (A-1)$$

and as a ratio of products in which factors of the form $(z - z_i)$ and $(z - p_i)$ represent zeroes and poles of the transfer function. From Equation (A-1) one gets

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + \cdots\cdots + b_q z^{-q}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + \cdots\cdots + a_p z^{-p}}$$

If $0 < q < p$, $p - q = N > 0$, then

$$H(z) = \frac{b_0 z^p + b_1 z^{p-1} + b_2 z^{p-2} + \cdots\cdots + b_q z^{p-q}}{a_0 z^p + a_1 z^{p-1} + a_2 z^{p-2} + \cdots + a_p}$$

$$= (z^{p-q}) \cdot \frac{b_0 z^q + b_1 z^{q-1} + b_2 z^{q-2} + \cdots + b_q}{a_0 z^p + a_1 z^{p-1} + a_2 z^{p-2} + \cdots + a_p}$$

$$= z^N \cdot \frac{b_0}{a_0} \cdot \frac{z^q + \frac{b_1}{b_0} z^{q-1} + \frac{b_2}{b_0} z^{q-2} + \cdots + \frac{b_q}{b_0}}{z^p + \frac{a_1}{a_0} z^{p-1} + \frac{a_2}{a_0} z^{p-2} + \cdots + \frac{a_p}{a_0}}$$

$$= z^N \cdot G \cdot \frac{z^q + B_1 z^{q-1} + B_2 z^{q-2} + \cdots + B_{q-1} z + B_q}{z^p + A_1 z^{p-1} + A_2 z^{p-2} + \cdots + A_{p-1} z + A_p}$$

where the numerator and denominator are ordinary polynomials in z. This can be written as

$$H(z) = G \cdot z^N \cdot \frac{(z-z_1)(z-z_2)(z-z_3)\cdots(z-z_q)}{(z-p_1)(z-p_2)(z-p_3)\cdots(z-p_p)}$$

$$= G \cdot z^N \cdot \frac{\prod_{i=1}^{q} (z-z_i)}{\prod_{j=1}^{p} (z-p_j)}$$

where G is a gain factor and z^N implies N zeroes at the origin of the z- plane.

If $p < q$, $q - p = M > 0$, then

$$H(z) = G \cdot \frac{1}{z^M} \cdot \frac{\prod_{i=1}^{q} (z-z_i)}{\prod_{j=1}^{p} (z-p_j)}$$

with M poles at the origin of the z- plane.

APPENDIX B

DERIVATION OF THE YULE-WALKER EQUATIONS

The autoregressive (AR) filter models the present output y(n) through a linear combination of the past p outputs y(n-i) for i = 1, 2, 3, · · · · · p, and the present input x(n) where x(n) is a white noise source.

$$y(n) = -\sum_{i=1}^{p} a_i \cdot y(n-i) + x(n)$$

Multiplying by y(n-k) gives

$$y(n) \cdot y(n-k) = -\sum_{i=1}^{p} a_i \cdot y(n-i) \cdot y\,(n-k) + x(n) \cdot y(n-k)$$

Ensemble averaging E{· · ·} on both sides and interchanging the E-operator and Σ-operator (ergodic process assumed) results in:

$$E\left\{y(n) \cdot y(n-k)\right\} = -\sum_{i=1}^{p} a_i \cdot E\left\{y(n-i) \cdot y(n-k)\right\} + E\left\{x(n) \cdot y(n-k)\right\}$$

$$r_{yy}(k) = -\sum_{i=1}^{p} a_i \cdot r_{yy}(k-i) + r_{xy}(k)$$

where $r_{yy}(k)$ is the output autocorrelation function, and $r_{xy}(k)$ is the input-output cross correlation. From the fundamental input-output relationship in the time domain by which the input is convoluted with the impulse response

$$y(n-k) = \sum_{i=-\infty}^{\infty} h(i) \cdot x(n-k-i)$$

one gets

$$r_{xy}(k) = E\left\{x(n) \cdot y(n-k)\right\} = E\left\{\sum_{i=-\infty}^{\infty} x(n) \cdot h(i) \cdot x(n-k-i)\right\}$$

$$= \sum_{i=-\infty}^{\infty} h(i) \cdot E\left\{x(n) \cdot x(n-k-i)\right\} = \sum_{i=-\infty}^{\infty} h(i) \cdot r_{xx}(k+i)$$

$$r_{xx}(k+i) = \sigma_x^2 \cdot \delta(k+i) \quad ,$$

since the autocorrelation function (ACF) of white noise x(n) is the delta function of strength σ_x^2. Therefore,

$$r_{xy}(k) = \sigma_x^2 \sum_{i=-\infty}^{\infty} h(i) \cdot \delta(k+i) \quad .$$

$\delta(k+i)$ as a function of i is equal to zero for all i except $i = -k$, for which

$$\sum_{i=-k-\varepsilon}^{-k+\varepsilon} \delta(k+i) = 1 \quad , \quad \text{with } \varepsilon \to 0 \quad ,$$

so that

$$r_{xy}(k) = \sigma_x^2 \cdot h(-k) \cdot \sum_{i=-k-c}^{-k+c} \delta(k+i) = \sigma_x^2 \cdot h(-k)$$

The value for h(-k) can be found from the basic definition of the z-transform H(z) for h(n):

$$H(z) = \sum_{n=-\infty}^{\infty} h(n) \cdot z^{-n} = \sum_{n=0}^{\infty} h(n) \cdot z^{-n}$$

since for a realizable, causal filter $h(n<0) = 0$, i.e., $h(-k) = 0$ for $k > 0$.

$$H(z) = h(0) \cdot 1 + h(1) \cdot z^{-1} + h(2) \cdot z^{-2} + \cdots\cdots$$

$$\lim_{z \to \infty} H(z) = h(0)$$

Also:

$$\lim_{z \to \infty} H(z) = \lim_{z \to \infty} \frac{1}{1 + \sum_{i=1}^{p} a_i \cdot z^{-i}} = \lim_{z \to \infty} \frac{1}{\left\{1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2} + \cdots\right\}} = 1$$

so that

$$h(0) = 1, \quad r_{xy}(k) = \sigma_x^2 \qquad \text{for } k = 0$$

and

$$h(-k) = 0, \quad r_{xy}(k) = 0 \qquad \text{for } k > 0$$

and finally

$$r_{yy}(k) = -\sum_{i=1}^{p} a_i \cdot r_{yy}(k-i) \qquad k = 1, 2, \cdots p$$

$$r_{yy}(k) = -\sum_{i=1}^{p} a_i \cdot r_{yy}(k-i) + \sigma_x^2 \quad \text{for } k = 0$$

This is the set of Yule-Walker equations, the first of which results in a group of p equations to solve for the p unknowns $a_1, a_2, \cdots, a_p$ given the values of the desired ACF. The second expression for $k = 0$ is just one equation to solve for σ_x^2.

Example of equation for $k = 1$:

$$r_{yy}(1) = -\left| a_1 \cdot r_{yy}(0) + a_2 \cdot r_{yy}(-1) + a_3 \cdot r_{yy}(-2) + \cdots + a_p \cdot r_{yy}(1-p) \right|$$

which can also be written as

$$r_{yy}(1) + a_1 \cdot r_{yy}(0) + a_2 \cdot r_{yy}(-1) + a_3 \cdot r_{yy}(-2) + \cdots + a_p \cdot r_{yy}(1-p) = 0$$

and similarly for $k = 2, 3, \cdots, p$.

The entire set of (p + 1) equations can be written in the short form of matrices and vectors as

$$\begin{bmatrix} r_{yy}(0) & r_{yy}(-1) & r_{yy}(-2) & \cdots & r_{yy}(-p) \\ r_{yy}(1) & r_{yy}(0) & r_{yy}(-1) & \cdots & r_{yy}(1-p) \\ \vdots & & & & r_{yy}(-1) \\ r_{yy}(p) & r_{yy}(p-1) & r_{yy}(p-2) & \cdots & r_{yy}(1) \quad r_{yy}(0) \end{bmatrix} \times \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_x^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This is the form in which the Yule-Walker equations are frequently found in the literature. One can see that the terms of the diagonals in the r_{yy}-matrix are identical. This indicates a so-called Toeplitz structure of the matrix which allows one to solve such a system of equations with a method known as the Levinson-Durbin recursive algorithm.

APPENDIX C

EXAMPLE OF THE LEVINSON RECURSION FOR P = 4

In order to calculate the four coefficients a_{41}, a_{42}, a_{43}, and a_{44} of the fourth order AR filter, one starts out (after initialization which gives a_{11} and σ_1^2) with order $p = 2$:

$$a_{22} = - \frac{r_{ss}(2) + a_{11} \cdot r_{ss}(1)}{\sigma_1^2}$$

$$a_{21} = a_{11} + a_{22} \cdot a_{11}^* \qquad \sigma_2^2 = \left(1 - \left| a_{22} \right|^2\right) \cdot \sigma_1^2$$

These are all parameters of model order 2. Now proceeding to model order $p = 3$:

$$a_{33} = - \frac{r_{ss}(3) + \left(a_{21} \cdot r_{ss}(2) + a_{22} \cdot r_{ss}(1)\right)}{\sigma_2^2}$$

$$a_{31} = a_{21} + a_{33} \cdot a_{22}^*$$

$$a_{32} = a_{22} + a_{33} \cdot a_{21}^*$$

$$\sigma_3^2 = \left(1 - \left| a_{33} \right|^2\right) \cdot \sigma_2^2$$

Model order $p = 4$ (with known parameters for $p = 3$):

$$a_{44} = - \frac{r_{ss}(4) + \left(a_{31} \cdot r_{ss}(3) + a_{32} \cdot r_{ss}(2) + a_{33} \cdot r_{ss}(1) \right)}{\sigma_3^2}$$

$$a_{41} = a_{31} + a_{44} \cdot a_{33}^*$$

$$a_{42} = a_{32} + a_{44} \cdot a_{32}^*$$

$$a_{43} = a_{33} + a_{44} \cdot a_{31}^*$$

$$\sigma_4^2 = \left(1 - \left| a_{44} \right|^2 \right) \cdot \sigma_3^2$$

These last five parameters completely describe the desired AR filter to model the data. The autocorrelation values $r_{ss}(k)$ must be known.

APPENDIX D

DERIVATION OF THE LEAST SQUARE ERROR OF THE PREDICTION FILTER

The error e(n) between the actual signal sample s(n) and its prediction $\hat{s}(n)$ is

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^{P} a_i s(n-i)$$

The totaled squared error over a given time interval of n samples is

$$\sum_n e^2(n) = \sum_n \left(s(n) + \sum_{i=1}^{P} a_i \cdot s(n-i) \right)^2$$

$$= \sum_n s^2(n) + 2 \sum_n s(n) \sum_{i=1}^{P} a_i \cdot s(n-i) + \sum_n \left(\sum_{i=1}^{P} a_i \cdot s(n-i) \right)^2$$

In order to find the parameter set $\{a_i\}$ which minimizes this expression, one calculates the partial derivatives with respect to all a's and sets them to zero. In order not to confuse the derivative index and the delay index i, one chooses another index for calculating the derivative, say k. The derivative of the first term is zero, since it is independent of all a. The derivative of the second term can be written as

$$2 \sum_n s(n) \frac{\partial}{\partial a_k} \left\{ a_1 s(n-1) + a_2 s(n-2) + \cdots + a_k \cdot s(n-k) + \cdots + a_p \cdot s(n-p) \right\}$$

$$= 2 \sum_n s(n) \cdot s(n-k) \qquad \text{for } k = 1, 2, \ldots, p$$

The third term has the form $z = y^2(x)$ for which

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x} = 2y \cdot \frac{\partial y}{\partial x}$$

The derivative of this term can then be written as

$$2 \sum_{n} \left(\sum_{i=1}^{p} a_i \cdot s(n-i) \right) \cdot \frac{\partial}{\partial a_k} \left\{ \sum_{k=1}^{p} a_k \cdot s(n-k) \right\}$$

where

$$\frac{\partial}{\partial a_k} \left\{ \sum_{k=1}^{p} a_k \cdot s(n-k) \right\} = \frac{\partial}{\partial a_k} \left\{ a_1 \cdot s(n-1) + \cdots + a_k \cdot s(n-k) + \cdots + a_p \cdot s(n-p) \right\}$$

$$= s(n-k)$$

so that the third term is

$$2 \sum_{n} \sum_{i=1}^{p} a_i \cdot s(n-i) \cdot s(n-k) \qquad \text{for } k = 1, 2, \cdots p$$

The total derivative of the squared error is then set to zero:

$$2 \sum_{n} s(n) \cdot s(n-k) + 2 \sum_{n} \sum_{i=1}^{p} a_i \cdot s(n-i) \cdot s(n-k) = 0$$

The ensemble average of this expression leads to the requirement for the minimum mean square error:

$$E\left\{\sum_{n} s(n)\cdot s(n-k)\right\} = -\sum_{i=1}^{p} a_i \cdot E\left\{\sum_{n} s(n-i)\cdot s(n-k)\right\}$$

which is

$$r_{ss}(k) = -\sum_{i=1}^{p} a_i \cdot r_{ss}(k-i) \qquad \text{for } k = 1, 2, 3, \cdots p$$

APPENDIX E

SET OF ERROR EQUATIONS FOR N DATA SAMPLES

Expanding the error equation for a prediction filter of order p

$$e(n) = s(n) + \sum_{i=1}^{p} a_i \cdot s(n-i) = \sum_{i=0}^{p} a_i\, s(n-i)$$

one gets the set of (N + p) error samples possible for an observed set of N data, s(0) through s(N-1). All unknown data for (N-1) < n < 0 are set to zero.

$$e(0) = a_0 s(0) + a_1 \cancel{s(-1)} + a_2 \cancel{s(-2)} + \ldots\ldots + a_{p-1} \cancel{s(1-p)} + a_p \cancel{s(-p)}$$
$$e(1) = a_0 s(1) + a_1 s(0) + a_2 \cancel{s(-1)} + \ldots\ldots + a_{p-1} \cancel{s(2-p)} + a_p \cancel{s(1-p)}$$
$$e(2) = a_0 s(2) + a_1 s(1) + a_2 s(0) + \ldots\ldots + a_{p-1} \cancel{s(3-p)} + a_p \cancel{s(2-p)}$$
$$\vdots$$
$$e(p-1) = a_0 s(p-1) + a_1 s(p-2) + a_2 s(p-3) + \ldots\ldots + a_{p-1} s(0) + a_p \cancel{s(-1)}$$

$$e(p) = a_0 s(p) + a_1 s(p-1) + a_2 s(p-2) + \ldots\ldots + a_{p-1} s(1) + a_p s(0)$$
$$e(p+1) = a_0 s(p+1) + a_1 s(p) + a_2 s(p-1) + \ldots\ldots + a_{p-1} s(2) + a_p s(1)$$
$$\vdots$$

Covariance Method

$$e(N-2) = a_0 s(N-2) + a_1 s(N-3) + a_2 s(N-4) + \ldots\ldots + a_{p-1} s(N-1-p) + a_p s(N-2-p)$$
$$e(N-1) = a_0 s(N-1) + a_1 s(N-2) + a_2 s(N-3) + \ldots\ldots + a_{p-1} s(N-p) + a_p s(N-1-p)$$

$$e(N) = a_0 \cancel{s(N)} + a_1 s(N-1) + a_2 s(N-2) + \ldots\ldots + a_{p-1} s(N-p+1) + a_p s(N-p)$$
$$e(N+1) = a_0 \cancel{s(N+1)} + a_1 \cancel{s(N)} + a_2 s(N-1) + \ldots\ldots + a_{p-1} s(N-p+2) + a_p s(N-p+1)$$
$$e(N+2) = a_0 \cancel{s(N+2)} + a_1 \cancel{s(N+1)} + a_2 \cancel{s(N)} + \ldots + a_{p-1} s(N-p+3) + a_p s(N-p+2)$$
$$\vdots$$
$$e(N+p-3) = a_0 \cancel{s(N+p-3)} + a_1 \cancel{s(N+p-4)} + \ldots\ldots\ldots\ldots + a_{p-1} s(N-2) + a_p s(N-3)$$
$$e(N+p-2) = a_0 \cancel{s(N+p-2)} + \ldots\ldots\ldots\ldots + a_{p-2} \cancel{s(N)} + a_{p-1} s(N-1) + a_p s(N-2)$$
$$e(N+p-1) = a_0 \cancel{s(N+p-1)} + \ldots\ldots\ldots\ldots + a_{p-2} \cancel{s(N+1)} + a_{p-1} \cancel{S(N)} + a_p s(N-1)$$

Autocorrelation Method

APPENDIX F

DERIVATION OF BURG'S ERROR RECURSION

The errors of the forward and backward predictors are (with $a_{pp} = k_p$)

$$e_p^f(n) = s(n) + \sum_{i=1}^{p} a_{pi} \cdot s(n-i) = s(n) + \sum_{i=1}^{p-1} a_{pi} \cdot s(n-i) + k_p \cdot s(n-p)$$

$$e_p^b(n) = s(n-p) + \sum_{i=1}^{p} a_{pi}^* \cdot s(n-p+i)$$

Substitution of a_{pi} by Levinson's recursion

$$a_{pi} = a_{p-1,i} + k_p \cdot a_{p-1,p-i}^*$$

gives

$$e_p^f(n) = s(n) + \sum_{i=1}^{p-1} \left(a_{p-1,i} + k_p \cdot a_{p-1,p-i}^* \right) \cdot s(n-i) + k_p \cdot s(n-p)$$

$$= s(n) + \sum_{i=1}^{p-1} a_{p-1,i} \cdot s(n-i) + k_p \left(\sum_{i=1}^{p-1} a_{p-1,p-i}^* \cdot s(n-i) + s(n-p) \right)$$

where

$$s(n) + \sum_{i=1}^{p-1} a_{p-1,i} \cdot s(n-i) = e_{p-1}^f(n)$$

Now substituting p-i=k, n-i=n-(p-k)

and with i=1→k=p-1, i=(p-1)→k=1

$$\sum_{i=1}^{p-1} a^{*}_{p-1,\,p-i} \cdot s(n-i) + s(n-p) \rightarrow \sum_{k=p-1}^{1} a^{*}_{p-1,k} \cdot s(n-p+k) + s(n-p)$$

Renaming k → i and noting that

$$\sum_{k=p-1}^{1} = \sum_{k=1}^{p-1} ,$$

this expression becomes

$$s(n-p) + \sum_{i=1}^{p-1} a^{*}_{p-1,i} \cdot s(n-p+i)$$

$$= s\Big((n-1)-(p-1)\Big) + \sum_{i=1}^{p-1} a^{*}_{p-1,i} \cdot s\Big((n-1)-(p-1)+i\Big) = e^{b}_{p-1}(n-1)$$

so that the forward error is recursively

$$e^{f}_{p}(n) = e^{f}_{p-1}(n) + k_{p} \cdot e^{b}_{p-1}(n-1)$$

The derivation of the recursion backward error is similar and results in

$$e^{b}_{p}(n) = e^{b}_{p-1}(n-1) + k^{*}_{p} \cdot e^{f}_{p-1}(n)$$

DISTRIBUTION

	Copies
Commander Naval Sea Systems Command Attn: PMS-409 (Mr. J. Neely) Washington, DC 20362	1
Office of Naval Research Attn: Library Arlington, VA 22217	1
Office of Chief of Naval Operations Washington, DC 20350	1
Commanding Officer Naval Underwater Systems Center Attn: Library New London, CT 06320	1
Commanding Officer Naval Underwater Systems Center Attn: Library Newport, RI 02841-5047	1
Commander Naval Ocean Systems Center Attn: Library San Diego, CA 92152	1
Commander Naval Air Development Center Attn: Library Warminster, PA 18974	1
Commanding Officer Naval Coastal Systems Center Attn: Library Panama City, FL 32407	1
Commanding Officer Naval Research Laboratory Attn: Library 4555 Overlook Avenue, S.W. Washington, DC 20390	1
Commander David Taylor Research Center Attn: Library Bethesda, MD 20084	1
Commander Naval Weapons Center Attn: Library China Lake, CA 93555	1
Commanding Officer Naval Avionics Facility Attn: Library 21st and Arlington Avenue Indianapolis, IN 46218	1
Commander Naval Ship Engineering Center Attn: Library Washington, DC 20362	1
Scripps Institute of Oceanography University of California Attn: Library La Jolla, CA 93150	1
NASA Scientific and Technical Information Facility Attn: Library P.O. Box 5700 Bethesda, MD 20546	1

DISTRIBUTION (Cont.)

	Copies
The University of Texas Applied Research Laboratory Attn: L. Mellenbruch Austin, TX 78712	1
The Pennsylvania State University Applied Research Laboratory Attn: Library P.O. Box 30 State College, PA 16801	1
Library of Congress Attn: Gift & Exchange Division Washington, DC 20540	4
Defense Technical Information Center Cameron Station Alexandria, VA 22304-5145	12

Internal Distribution:		Copies
E231		2
E232		3
E52	(J. Gallagher)	1
U02	(J. Goeller)	1
U20	(J. Sherman)	1
U202	(T Ballard)	1
U24	(J. Price)	1
U25	(B. Otte)	1
U25	(W. Beatty)	1
U25	(S. Chang)	1
U25	(D. Claybaugh)	1
U25	(B. Davis)	1
U25	(A. Delagrange)	1
U25	(S. Hebbert)	1
U25	(S. Le)	1
U25	(M. Pruessner)	2
U25	(L. Weeks)	1
U25	(M. Williams)	1
U32	(P. D. Jackins)	1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	29 June 1990	

4. TITLE AND SUBTITLE
Spectral Estimation Techniques with Emphasis on Short Data Records—A Primer

5. FUNDING NUMBERS

6. AUTHOR(S)
Manfred K. Pruessner

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Surface Warfare Center
White Oak Laboratory (Code U25)
10901 New Hampshire Avenue
Silver Spring, MD 20903-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
NAVSWC TR 90-236

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report reviews the development of the spectrum from Schuster's periodogram and the Blackman-Tukey approach via autocorrelation to the Fast Fourier Transform by Cooley-Tukey as routinely used today. It gives the results of theses techniques when applied to some transient data. The report proceeds to parametric autoregressive methods developed more recently because of their promise of higher resolution. The analysis results of transient data using these methods are compared to those obtained with the classical approach.

14. SUBJECT TERMS
Spectral Analysis; Parametric Analysis; Autoregressive Modeling; Linear Prediction; Burg's Algorithm

15. NUMBER OF PAGES
148

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and its title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

BLOCK 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If Known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank.
NTIS	- Leave blank.

Block 13. Abstract. Include a brief ***(Maximum 200 words)*** factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code ***(NTIS only)***

Blocks 17.-19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.