Analysis and Synthesis of Signals Using Complex Zeros

M.H. Ackroyd

Submitted for the degree of

Doctor of Philosophy

of

Loughborough University of Technology

1970

SYNOPSIS

This thesis reports studies based on the concept of characterising a bandlimited signal by the locations of its zeros in a "complex time" plane (whose real axis is the real variable, t). The relations between the zeros of a signal and more conventional descriptors such as Nyquist samples, Fourier coefficients or spectral samples are investigated.

Added noise or linear filtering changes the zeros of a signal in a complicated way. The problems which arise are discussed.

Signals can be synthesised directly from a specified zero pattern. The properties of signals having a certain class of zero pattern ("angle coded signals") are studied. A method is presented by which zero patterns can be produced which yield signals (suitable for radar use) having desired properties.

The practical use of angle coded signals would, in some circumstances, call for the use of a Hilbert transform network - a wideband 90° phase shifter. The difficulties of constructing such networks are discussed and ways of overcoming the limitations of existing networks are suggested.

The theory of the distribution of the energy of a signal in the time-frequency plane is given in an appendix. The relations between this t-f distribution and the instantaneous frequency of a signal and 'short-time' spectra are given.

Another appendix applies the theory developed for the design of radar signals described by their zeros in a complex time plane to the design of Huffman sequences, which are described by their zero patterns in the complex frequency plane.

Finally, a new scheme for computing the coefficients of a polynomial from its roots is presented. This scheme, which is based on the discrete Fourier transform, is not unduly affected by round-off error even when used with polynomials of very high order.

ACKNOWLEDGMENTS

The author wishes to thank his supervisor, Professor J.W.R. Griffiths, for his enthusiastic guidance and encouragement. In addition, he would like to thank Professor H. Buckingham for the use of the facilities of the Department of Electronic and Electrical Engineering.

Acknowledgment is made to Professor H.B. Voelcker for the interest he has shown in this work.

CONTENTS

1. Introduction

1.1 Representing a Signal by its Zeros 1

1.2 Organisation of the Work 3

2. Zero Location

2.1 Introduction 5

2.2 Periodic Bandlimited Signals 6

2.3 Finite Bandlimited Signals 10

2.4 Approximate Methods 15

2.5 Zero Location by Axis Shifting 17

3. Effects of Noise and Filtering on the Zeros of Signals

3.1 Introduction 19

3.2 Effects of Linear Filtering on Zeros 20

3.3 Effects of Added Noise 21

4. Angle Coded Signals and their Properties

4.1 Introduction 28

4.2 General Properties of Angle Coded Signals 31

4.3 Spectra of Angle Coded Signals - Theoretical 39

4.4 Spectra of Angle Coded Signals - Experimental 48

5. Angle Coded Signals for Radar and Sonar

5.1 Introduction 51

5.2 Synthesis of Zero Patterns 52

5.3 Example of the Method 54

5.4 Discussion 57

6. Hilbert Transform Networks

6.1 Introduction 60

6.2 A Simplified Quadrature Network 62

6.3 Realisation of Hilbert Transform Networks 63

7. Conclusions 64

Appendix A. The Time-Frequency Energy Distribution of Signals 70

Appendix B. The Design of Huffman Sequences 84

Appendix C. Computing the Coefficients of Polynomials of High Order 85

CHAPTER I

INTRODUCTION

1.1 Representing a Signal by Its Zeros

It is a familiar result of linear system theory that a (rational) transfer function can be specified either by its magnitude and phase (or by its real and imaginary parts) given as a function of the real variable, frequency or by the locations of its poles and zeros in the complex frequency plane.

By considering the reciprocal roles of time and frequency in the Fourier transform relations it might be supposed that something similar would apply to time functions (signals). In fact a bandlimited signal (which, as a function of a complex variable, is an entire function - no singularities in the finite plane) can be specified essentially by the locations of its zeros in the complex time plane in addition to more conventional representations such as its envelope and instantaneous phase given as time functions. A bandlimited signal is analogous to a filter whose impulse response is duration-limited and whose transfer function consequently has only zeros in the finite complex frequency plane.

The possibility of representing a bandlimited signal (i.e. an entire function of exponential type) was shown in 1926 by Titchmarsh[1] and by Paley and Wiener[2].

One theorem of Titchmarsh's paper states that if

$$s(z) = \int_{f_1}^{f_2} S(f)\, e^{j2\pi fz}\, df \qquad (z = t + j\sigma)$$

(i.e. if s(t) is a bandlimited signal), then s(z) can be expressed as an infinite product:

$$s(z) = s(o)\, e^{j\pi(f_1+f_2)z} \prod_{n=1}^{\infty} (1 - z/z_n)$$

(The notation of Titchmarsh's paper is altered here).
Thus, except for a scale factor and a frequency shift, a bandlimited signal can be specified by the locations of its zeros in the complex time plane. Titchmarsh also proved that the zeros tend to be clustered along the real (t) axis of the z-plane and occur on average at the Nyquist rate (in engineering parlance).

A paper was published by Bond and Cahn in 1957[3] which applied the mathematicians' results to signals. Bond and Cahn suggested that as an alternative to specifying a bandlimited signal by its Nyquist samples it was possible to "sample the zeros". They also suggested that signals could be synthesised so that the information was coded explicitly in the zero pattern.

The work of Bond and Cahn appears to have received little attention until Voelcker, in 1966, published his paper entitled "Toward a Unified Theory of Modulation"[4]. In this paper he interpreted modulation processes as being equivalent to manipulations performed on the zero patterns of signals. In addition to his interpretive work, Voelcker suggested means of generating two types of signal in which the information is carried explicitly by the zero pattern. With one type of signal ("real zero" signals), the zeros are restricted to lying on the real (t) axis of the complex-time plane. With the other type of signals (which he termed "angle-coded" signals) each zero is constrained to lie on either of two conjugate lines parallel to the t-axis and at regular intervals in the direction of the t-axis. The work of this thesis is largely concerned with investigating the properties and applications of "angle-coded" signals.

As Voelcker has shown, the zero-based description of signals is evidently a "natural" one to use in studying modulation processes which are essentially multiplicative in nature. Because zero-based signal theory has proved to be such an effective tool in the study of modulation it would be natural to seek other situations where it would be useful. Such situations are suggested by considering why pole-zero methods play such a central part in linear system theory. There seem to be two principal reasons for the widespread use of pole-zero methods.

(i) When systems are cascaded and their transfer functions are multiplied, the resulting pole-zero pattern is the superposition of the individual pole-zero patterns.

(ii) There are simple approximate (qualitative) relationships between the pole-zero pattern of a system and both its time behaviour (for example, its step response) and its frequency response characteristics.

In addition to the study of situations where signals are multiplied it would seem that zero based methods might be useful in studying the properties of signals where both the time <u>and</u> the frequency behaviour are simultaneously of interest. Zero-based ideas are in fact used in this thesis to design signals having desired time-frequency properties.

1.2 Organisation of the Work

As a basis for later work, the formal relationships between the zeros, the Nyquist samples and the spectral samples (or the Fourier coefficients) of band-limited signals are developed in Chapter 2. This is done for periodic signals, using the Fourier series factorisation methods outlined by Voelcker and also for "finite bandlimited" signals (that is, for signals whose Nyquist samples outside some finite interval are all zero). Methods are developed by which the zeros of a bandlimited signal can be located from knowledge of its waveform.

In chapter 3 the important but difficult questions as to how the zeros of a signal are affected by linear filtering or by added noise are posed.

The properties of angle coded signals are studied in chapter 4. This work includes studies, both theoretical and computational, of the spectra of these signals. In connection with the computational work, a new scheme for determining the coefficients of a polynomial from its roots was devised which is described in appendix C.

In chapter 5 a procedure is developed by which a zero pattern can be chosen which yields a signal whose energy distribution in the time-frequency plane approximates some desired form. This provides a synthesis procedure for radar signals. Appendix A provides a background for chapter 5; it clarifies the relationships between ambiguity functions, short-time spectra and other functions of time and frequency which are used in signal theory. Appendix B is, in effect, a translation of chapter 5 from time language into frequency language. The work is, perhaps, of greater practical utility when cast in this form.

Much of the work of this thesis deals with analytic signals. In many applications of analytic signal theory a Hilbert transform network is necessary. Chapter 6 clarifies the reasons for the difficulty in realising satisfactory practical Hilbert transform networks and suggests ways in which they may be constructed so as to work over a wide frequency range.

CHAPTER 2

ZERO LOCATION

2.1 Introduction

A central problem in zero-based signal theory is that of locating the zeros of the signal. In other words, given the waveform of a bandlimited signal, how are its zeros to be found?

As pointed out by Voelcker[4], the real zeros of a signal can be located readily by the use of an axis crossing detector. But complex zeros cannot be located by simple observations on the waveform of a signal. Voelcker suggested that the zeros of a periodic bandlimited signal can be found by numerical factorisation of the polynomial which represents it. He also suggested that, as a non-periodic signal can be represented over a finite interval by a Fourier series, the zeros of a non-periodic signal could also be located to a close approximation by factorisation of a Fourier polynomial.

Bond and Cahn[3], by using the results of Titchmarsh[1], succeeded in expressing the Nyquist samples of a finite bandlimited* (FBL) signal in terms of its real and complex zero locations. However, they were unable to suggest a procedure for locating the complex zeros.

In section 2.2 of this chapter the relations between the Nyquist samples, the Fourier coefficients and the zeros of a periodic bandlimited (PBL) signal are developed. The Nyquist samples and the Fourier coefficients are related (as is well known) by the discrete Fourier transform (DFT),[5] which is a linear reversible transformation. The Fourier coefficients and the zeros are related by a nonlinear transformation which is also reversible.

*The term finite bandlimited signal is applied to a signal whose Nyquist samples outside a finite interval are all zero.

In section 2.3 a procedure is suggested which in principle can be used to determine the zeros of a FBL signal from its Nyquist samples. Theory is developed relating the Nyquist samples, the spectral samples and the zeros of a FBL signal. The Nyquist samples are related to the spectral samples by the DFT, while the Nyquist samples and the zeros are related by a nonlinear transformation.

These procedures are exact in the sense that if the Nyquist samples which specify the signals are known exactly, then in principle the zero locations can be found to any required degree of accuracy. In section 2.4 the application of these methods to the approximate location of zeros is discussed. Approximate zero location methods must be used either when the whole history of the signal is not known (e.g. in real-time operations) or when the dimensionality of the signal is too large for exact methods to be used.

In section 2.5, a rather different approach to the zero location problem is presented. An "axis shifting" filter (which is a nonphysically realisable complex linear filter) converts complex zeros on some line in the complex time plane running parallel to the t-axis into real zeros. The real zeros can then be observed directly. This method does not provide a very practical way to compute zero locations; however, it is a useful conceptual device which is used in chapter 3 in the study of the effects of noise on the zeros of a signal.

2.2 Periodic Bandlimited Signals.

The study of the zeros of periodic signals is the simplest; problems of zero location become (as shown by Voelcker) equivalent to polynomial factorisation.

In this section (and in most subsequent work) only analytic signals whose spectra are zero outside some range (0,W) are considered. (This is no great restriction, as any PBL signal can be brought into this class by a suitable frequency translation). Such a signal can be represented by a finite Fourier series of N terms

$$m(t) = C_0 + C_1 e^{j2\pi t/T} + \ldots + C_{N-1} e^{j2\pi(N-1)t/T}$$

$$= \sum_{k=0}^{N-1} C_k e^{j2\pi kt/T} \qquad 2.2.1$$

where T is the period of the signal and N = TW+1.

The sample values of the signal are thus given by

$$m(n/W) = \sum_{k=0}^{N-1} C_k e^{j2\pi kn/N} \qquad n=0,1,\ldots,N-1 \qquad 2.2.2$$

or, in vector form,

$$\underline{m} = \underline{Y}^* \underline{C} \qquad 2.2.3$$

when the N element vectors $\underline{m}$ and $\underline{C}$ are given by

$$\underline{m} = \begin{bmatrix} m(0/W) \\ m(1/W) \\ \cdot \\ \cdot \\ \cdot \\ m[(N-1)/W] \end{bmatrix} \qquad \underline{C} = \begin{bmatrix} C_0 \\ C_1 \\ \cdot \\ \cdot \\ \cdot \\ C_{N-1} \end{bmatrix} \qquad 2.2.4$$

and the N x N matrix $\underline{Y}^*$ is the conjugate of the DFT matrix $\underline{Y}$, given by

$$\underline{Y} = \begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & Y & Y^2 & \ldots & Y^{N-1} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 1 & Y^{N-1} & Y^{2(N-1)} & \ldots & Y^{(N-1)^2} \end{bmatrix} \qquad 2.2.5$$

where $Y = e^{-j2\pi/N}$.

As equation 2.2.3 has the form of a discrete Fourier transform, its inverse is given by

$$\underline{C} = \frac{1}{N} \cdot \underline{Y}\,\underline{m} \qquad 2.2.6$$

The transformation from the Fourier coefficients of a PBL signal to its Nyquist samples is linear and reversible. As it is a form of the DFT, the fast Fourier transform (FFT) algorithm[5] can be used for computational work.

By substituting the symbol Z for $e^{j2\pi t/T}$, the finite Fourier series 2.2.1 can be rewritten as

$$M(Z) = C_o + C_1 Z + \ldots + C_{N-1} Z^{N-1}$$

$$= \sum_{K=o}^{N-1} C_k Z^k \qquad 2.2.7$$

By the fundamental theorem of algebra, this polynomial in Z has N-1 roots, $Z_1, \ldots, Z_{N-1}$ and can be written in the factored form

$$M(Z) = C_o\ (1 - Z/Z_1) \ldots (1-Z/Z_{N-1})$$

$$= C_o \prod_{i=1}^{N-1} (1-Z/Z_i) \qquad 2.2.8$$

The signal is specified by its Fourier coefficients which can be recovered by multiplying out the right hand side of equation 2.2.8. Thus, except for a multiplying scale factor, the roots of M(Z) specify the signal. The Fourier coefficients are related to the roots by the nonlinear equations.

C_1/C_o = - (sum of reciprocals of all roots)

$= -\ (1/Z_1 + 1/Z_2 + \ldots + 1/Z_{N-1})$

C_2/C_o = + (sum of products of reciprocals of all roots taken two at a time)

$= \ (1/Z_1Z_2 + 1/Z_1Z_3 + \ldots)$

$$C_3/C_0 = -\text{(sum of products of reciprocals of all roots taken three at a time)}$$

$$= -(1/Z_1Z_2Z_3 + 1/Z_1Z_2Z_4 + \ldots)$$

....

$$C_{N-1}/C_0 = (-1)^{N-1} \text{ (product of reciprocals of all roots)}$$

$$= (-1)^{N-1} / (Z_1Z_2 \ldots Z_{N-1}). \qquad 2.2.9$$

To specify the signal completely one datum is needed in addition to the roots of the Fourier polynomial. This could be the zero frequency (or any other) Fourier coefficient.

. . . .

The roots of $M(Z)$ are related to the real and complex zeros of the signal in a simple way4. $M(Z)$ was obtained by replacing $e^{j2\pi t/T}$ by Z in equation 2.2.1. This determines a conformal mapping from the z- (complex time) plane to the Z-plane. To each point in the Z-plane at which $M(Z)$ is zero (i.e. to each root of $M(Z)$) there correspond points in the z-plane at which $m(z)$ is zero (the zeros of the signal). These zeros are given by the solutions of the equations.

$$Z_i = e^{j2\pi z_i/T} \qquad i = 1,2,\ldots,N-1 \qquad 2.2.10$$

Because of the periodicity of the exponential function, the solutions of the equation set 2.2.10 (the signal zeros) form a configuration in the z-plane which is periodic in the t-direction, with period T.

Taking the principal value logarithm of each side of equation 2.2.8, the locations are found of the zeros which lie in the vertical strip in the complex time plane defined by $-T/2<t\leq T/2$ (the principle strip):

$$z_i = T/2\pi \left[\arg Z_i - j\ln|Z_i| \right] \qquad i=1,2,\ldots N-1 \qquad 2.2.11$$

The N-1 zeros in the principal strip define the roots of $M(Z)$ via equation 2.2.10 and, as shown above, these roots define the

Fourier coefficients of the signal except for a scale factor.

The above shows that the Fourier coefficients of a PBL signal are obtained from its zeros except for an arbitrary scale factor, k.

This transformation can be symbolised;

$$\underline{C} = V(\underline{z}), \qquad 2.2.12$$

where the N-1 element vector $\underline{z}$ is defined by

$$\underline{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_{N-1} \end{bmatrix}$$

This transformation is shown in block diagram form in fig.2.2.1. The inverse of this transformation is the scale factor-destroying transformation depicted in fig. 2.2.2 which can be symbolised

$$\underline{z} = V^{-1}(\underline{C}) \qquad 2.2.13$$

The relations between the Fourier coefficients, Nyquist samples and zeros of a PBL signal are shown in fig. 2.2.3. These transformations are conveniently programmed for digital computer, making use of existing FFT routines and polynomial factorising and multiplying-out routines*.

2.3 Finite Bandlimited Signals

A finite bandlimited (FBL) whose Nyquist samples are zero outside the range (-T/2,T/2) and having spectral extent W has N Nyquist instants within the interval (where N+1 is the smallest even integer exceeding TW : a Nyquist sample is assumed to occur at t=o). Fig. 2.3.1 shows such a signal. The signal is assumed

* A scheme for computing the coefficients of a high order polynomial from its roots is presented in appendix C.

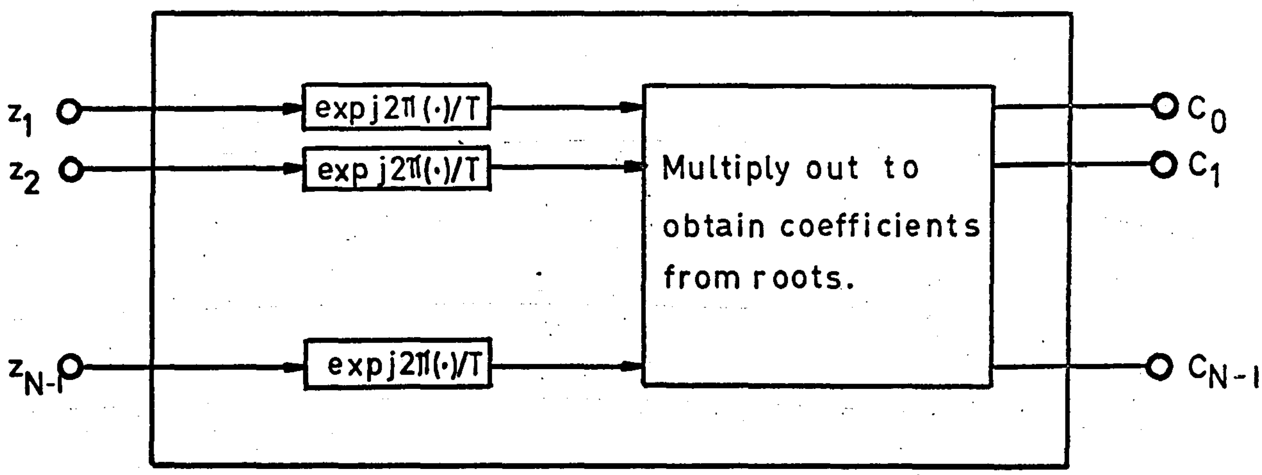


Fig. 2.2.1. Transformation from zeros to Fourier coefficients.

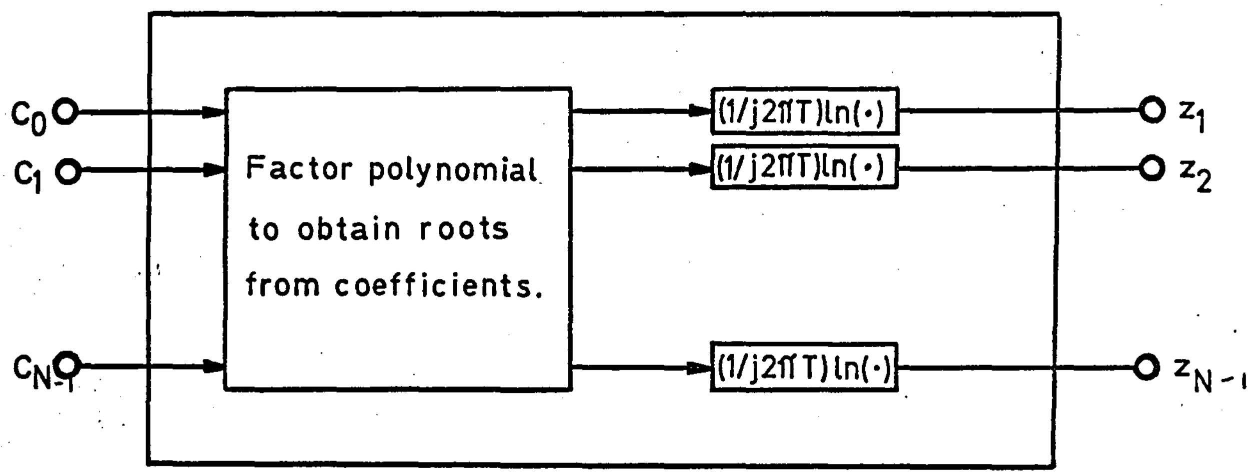


Fig. 2.2.2. Inverse transformation from Fourier coefficients to zeros.

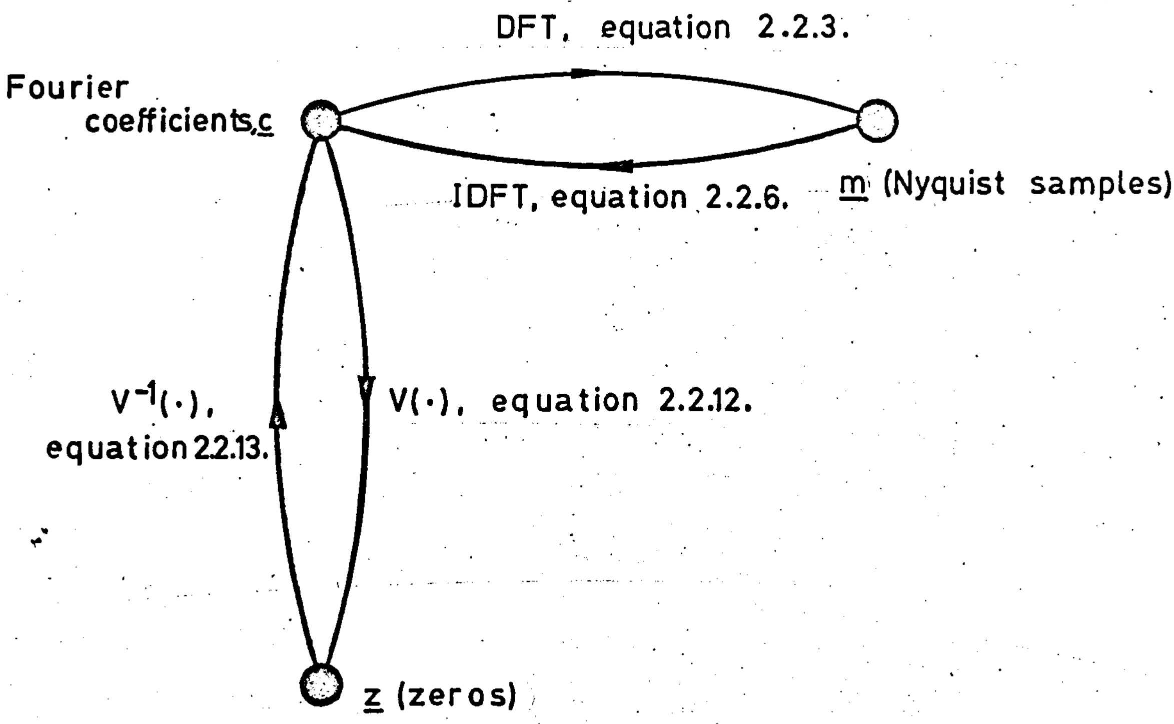


Fig.2.2.3. Relations between the zeros, the Fourier coefficients and the Nyquist samples of a PBL signal.

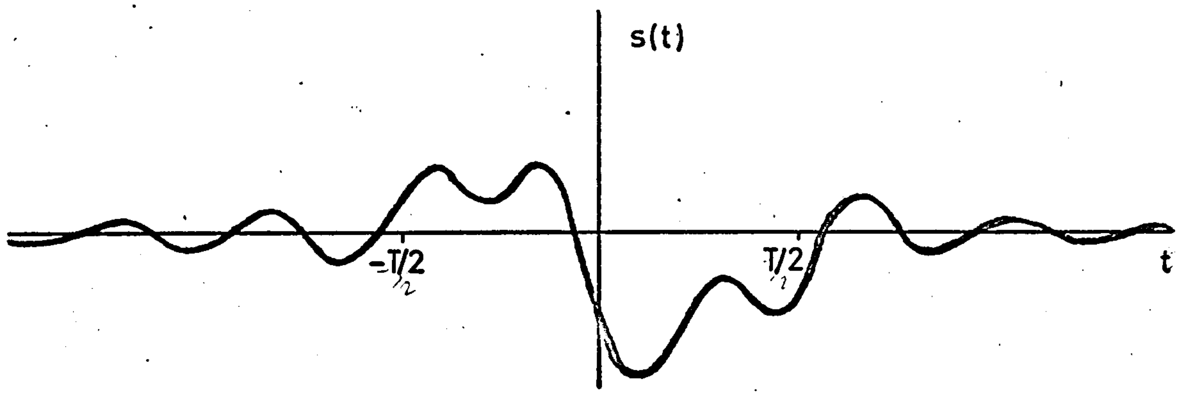


Fig.2.3.1. Finite bandlimited (FBL) signal waveform.

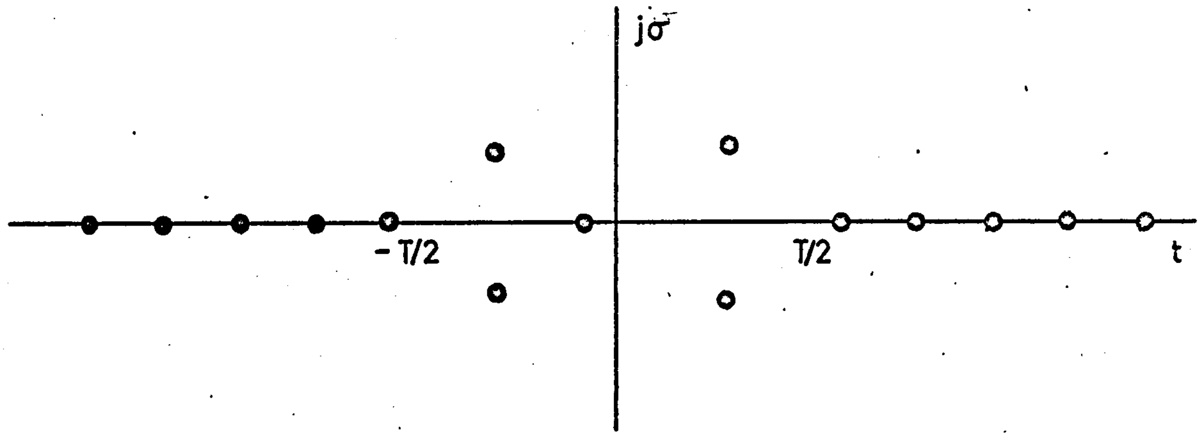


Fig.2.3.2. Zero pattern of FBL signal.

to have N-1 real and complex zeros within the interval (-T/2,T/2) which may occur anywhere except at the origin of the complex time plane*. Outside the interval the zeros occur at the Nyquist instants, i.e., at

$$z = \pm n/W + jo \qquad n = (N-1)/2+1, (N-1)/2+2, (N-1)/2+3, \ldots\ldots$$

as illustrated in fig. 2.3.2.

By Titchmarsh's work, such a signal can be expressed

$$s(t) = s(o) \prod_{k=1}^{N-1} (1-t/z_k) \prod_{n=(N-1)/2+1}^{\infty} 1-(tW/n)^2 \qquad 2.3.1$$

The left product is taken over N-1 terms involving the zeros which may lie anywhere (except the origin). The right product is taken over the terms involving the real zeros which lie at the Nyquist instants outside the interval (-T/2,T/2).

Equation 2.3.1 gives

$$s(t)/s(o) = \prod_{k=1}^{N-1} (1-t/z_k) \quad p(t) \qquad 2.3.2$$

where the function p(t) is defined by

$$p(t) = \prod_{n=(N+1)/2}^{\infty} 1-(tW/n)^2 \qquad 2.3.3$$

Using the product expansion for sin πWt, p(t) can be written

$$p(t) = \frac{\sin \pi Wt}{\pi Wt \prod_{m=1}^{(N-1)/2} 1-(tW/m)^2} \qquad 2.3.4$$

*This assumption is made by Bond and Cahn[3]. While it seems dimensionally correct that N-1 zeros should specify N Nyquist samples to within a multiplicative constant, the assumption seems difficult to prove. Bond and Cahn also assume that the zeros lie within the strip $z=t+j\sigma$ where $|t|<T/2$. Although they tend to cluster within the region there is no reason (or need) for this assumption to hold in general.

At the Nyquist instants, where $t = \pm k/W$, k=o,1,2,...

p(t) has the values[36]

$$p(\pm k/W) = \begin{cases} \dfrac{[(N-1)/2]\,!}{[(N-1)/2 + k]\,!}\,\dfrac{[(N-1)/2]\,!}{[(N-1)/2 - k]\,!} & ,\text{for } k \leq (N-1)/2 \\ o & ,\text{for } k > (N-1)/2 \end{cases} \qquad 2.3.5$$

p(t) is shown graphically in fig. 2.3.3 (for N=9).

The Nyquist samples of s(t) can be obtained from the locations of the N-1 non-Nyquist instant zeros (to within a multiplicative scale factor) by the evaluation of equation 2.3.2, substituting the values given by equation 2.3.5 for p(t). This nonlinear transformation from the zeros to the Nyquist samples can be expressed in vector form

$$\underline{s} = B(\underline{z}) \qquad 2.3.6$$

where the N-1 element vectors $\underline{s}$ and $\underline{z}$ are defined by

$$\underline{s} = \begin{bmatrix} s\,[(N-1)/2W]\,/s(o) \\ \cdot \\ \cdot \\ s\,(1/W)\,/\,s(o) \\ s\,(-1/W)/\,s(o) \\ \cdot \\ \cdot \\ \cdot \\ s\,[-(N-1)/2W]\,/\,s(o) \end{bmatrix}, \quad \underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_{N-1} \end{bmatrix} \qquad 2.3.7$$

The foregoing shows how the Nyquist samples of a FBL signal can be expressed in terms of its zeros. Although details of the present treatment differ, this is basically what was done by Bond and Cahn. In the following, a method is outlined by which the opposite transformation can be made in which the locations of the N-1 non-Nyquist instant zeros of a FBL signal can be obtained from the N Nyquist samples within the interval (-T/2,T/2) (or, for that matter, from any set of N distinct samples within the interval).

If each side of equation 2.3.2 is divided by p(t), the right hand side becomes a polynomial:

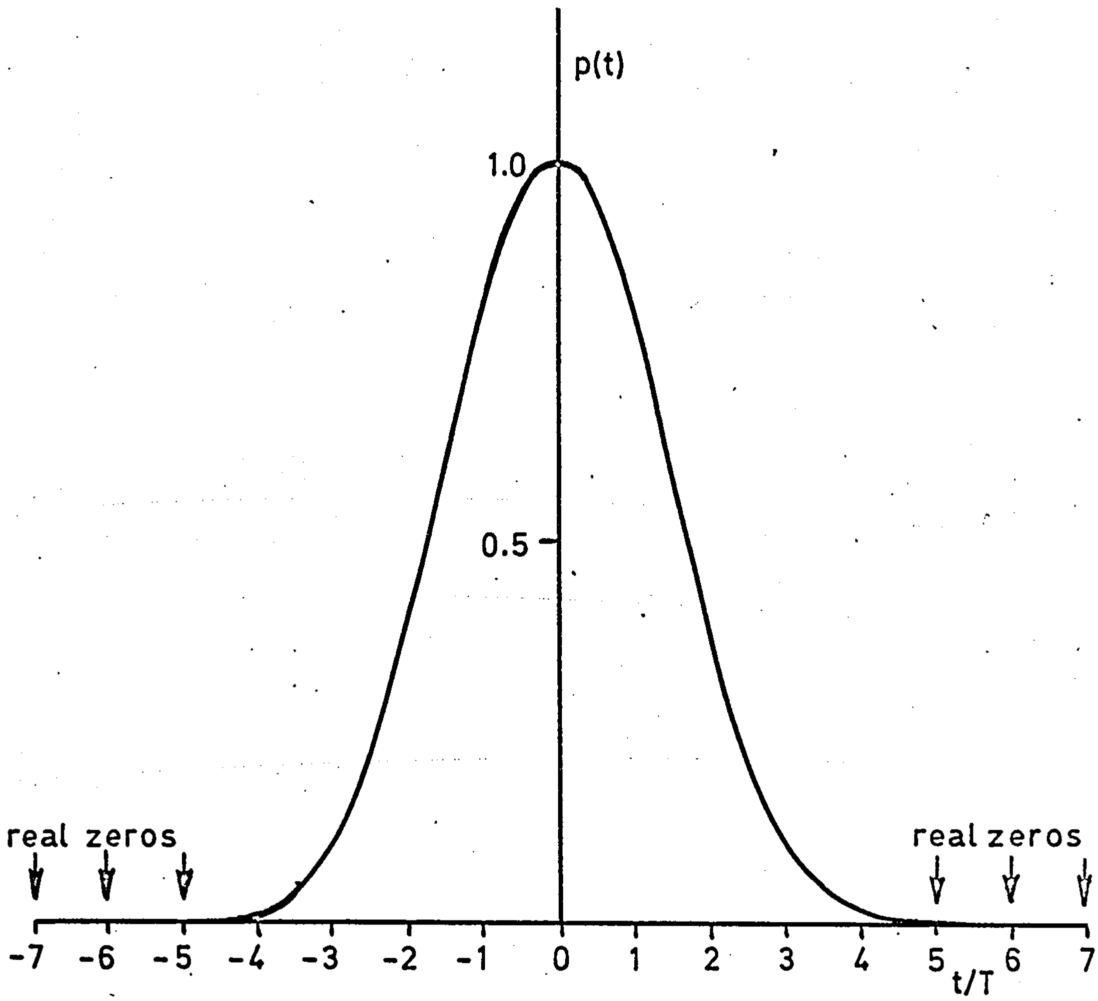


Fig.2.3.3. p(t) for N = 9 (9 non-Nyquist sample instant zeros).

$$s(t) \;/\; s(o)\, p(t) \;=\; a_0 + a_1 t + \ldots + a_{N-1} t^{N-1} \qquad 2.3.8$$

The roots of this polynomial are the N-1 non-Nyquist instant zeros of the signal.

In general, it is sufficient to know the value of an Mth order polynomial at M+1 points in order to be able to determine its coefficients, which is done by the process of solving a set of M+1 linear equations. From the N non zero Nyquist samples, the N coefficients of the right hand side of equation 2.2.7 can be found. The polynomial can then be factored numerically to yield its roots, which are the zeros of the signal. The zeros of a FBL signal can thus, in principle, be found directly from its non-zero Nyquist samples. The inverse of the transformation 2.3.6 thus exists and can be written

$$\underline{z} = B^{-1}\,(\underline{s}) \qquad 2.3.9$$

While the above shows that the transformation from the Nyquist samples to the zeros can be made in principle, nothing has been said about the practical feasibility of this process. The procedure involves the solution of a set of N linear equations and the factorisation of a polynomial of order N-1. The computational difficulty of both of these operations increase very rapidly as N increases. Nevertheless it is of interest that there exists a process by which the zeros of a FBL signal can be found exactly from its Nyquist samples.

The relation between the Nyquist samples and the spectral samples of a FBL signal is straightforward. The Fourier transform, S(f), of a bandlimited signal can be expressed in terms of its Nyquist samples[5], of which only N are non-zero in the case of a FBL signal.

$$S(f) = 1/W \sum_{n=-(N-1)/2}^{(N-1)/2} s(n/W)\, e^{-j2\pi fn/W}, |f|<W/2 \qquad 2.3.10$$

The spectral samples of S(f) are given by

$$S(k/T) = 1/W \sum_{n=-(N-1)/2}^{(N-1)/2} s(n/W)\, e^{-j2\pi kn/N},\ k= -(N-1)/2,\ldots,(N-1)/2 \qquad 2.3.11$$

This is a discrete Fourier transform (DFT) which can be expressed in matrix form as

$$\underline{S} = 1/W\ \underline{W}\ \underline{s} \qquad 2.3.12$$

where the N element vectors $\underline{S}$ and $\underline{s}$ are given by

$$\underline{S} = \begin{bmatrix} S(o) \\ S(1/T) \\ \vdots \\ S\left[(N-1)/2T\right] \\ S\left[-(N-1)/2T\right] \\ \vdots \\ S(-2/T) \\ S(-1/T) \end{bmatrix} \quad \underline{s} = \begin{bmatrix} s(o) \\ s(-/w) \\ \cdot \\ \cdot \\ \cdot \\ s\left[(N-1)/2w\right] \\ s\left[-(N-1)/2w\right] \\ \vdots \\ s(-2/w) \\ s(-1/w) \end{bmatrix} \qquad 2.3.13$$

and $\underline{W}$ is the DFT matrix, given by

$$\underline{W} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & W & W^2 & \cdots & W^{(N-1)} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & W^{(N-1)} & W^{2(N-1)} & \cdots & W^{(N-1)^2} \end{bmatrix} \qquad 2.3.14$$

where $W = e^{-j2\pi/N}$.

The inverse of this transform, giving the Nyquist samples in terms of the spectral samples is thus given by

$$\underline{s} = 1/T\ \underline{W}^*\ \underline{S} \qquad 2.3.15$$

The relations between the spectral samples, the Nyquist samples and the zeros of a FBL signal are shown in fig.2.3.4. It is interesting to compare this figure with fig. 2.2.3. In the case of a PBL signal the zeros are found from the Fourier coefficients (frequency description) of the signal while in the case of a FBL signal, the zeros are found from its Nyquist samples (time description).

2.4 Approximate Methods

The preceeding sections of this chapter presented methods by which the zeros of a FBL or a PBL signal could be found, as accurately as required, from its Nyquist samples. These schemes are practicable when the signals are of moderate dimensionality ($N<60$, say) but for signals of very large dimensionality, such as an information bearing signal in a communication system they are not feasible. If the zero locations of a signal are to be found in "real-time", only the past values of the signal are available: its whole history is not known. In such circumstances procedures are required by which the locations of the zeros can be found approximately without making use of the whole history of the signal. It is possible for such approximate procedures to exist as the relation between the Nyquist samples of a signal and its zeros is a 'localised' one. That is, a zero which is very distant in the complex-time plane from a particular Nyquist instant has very little effect on the sample value associated with that instant (although in principle all the Nyquist samples depend to some extent on all the zeros).

Suppose that the scheme of section 2.2 for locating the zeros of a PBL signal were available. How could this scheme be used to locate (approximately) the zeros of a signal with inconveniently large dimensionality? One approach would be to

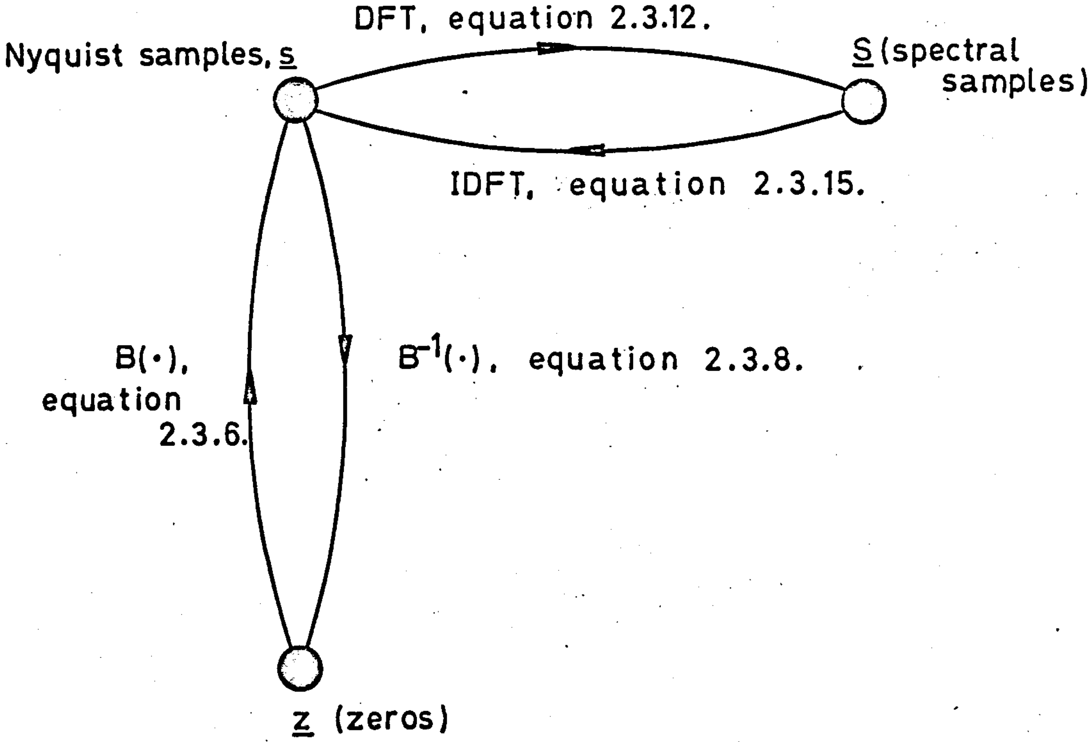


Fig.2.3.4. Relations between the zeros, the Nyquist samples and the spectral samples of a finite bandlimited (FBL) signal.

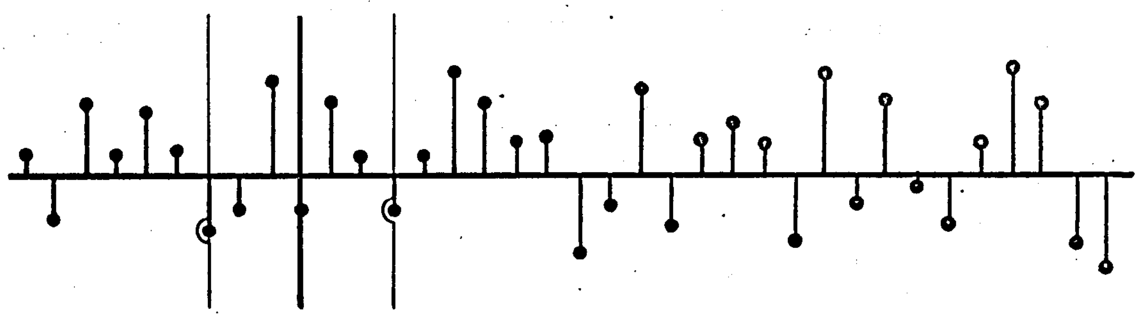
take N adjacent Nyquist samples and to form their periodic repetition (fig.2.4.1.). The periodic sample sequence can be regarded as constituting the samples of a PBL signal whose zeros can be found by the method of section 2.2. However, the zero pattern found by this method is not simply the periodic repetition of a section of the zero pattern of the original signal as the zero locations within the fundamental period of the PBL signal are influenced by the "incorrect" Nyquist samples outside the fundamental period. However, if N is large enough, the zeros near the centre of the fundamental period will not be influenced appreciably by the "incorrect" Nyquist samples outside the period. To ensure accurate zero location then:

(i) N must be large enough.

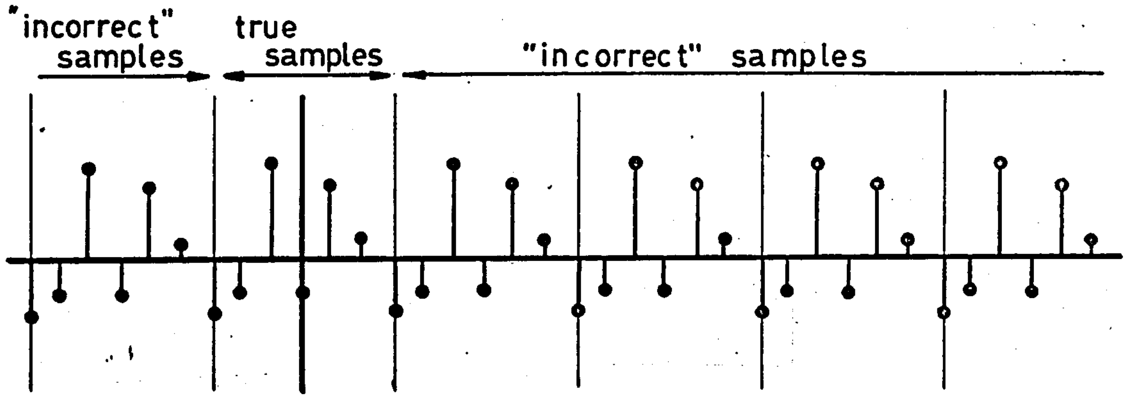
(ii) Zero locations near the centre of the fundamental period must be used.

All the zeros of a signal can be found by repeating the process for successive (overlapping) sections of the original waveform.

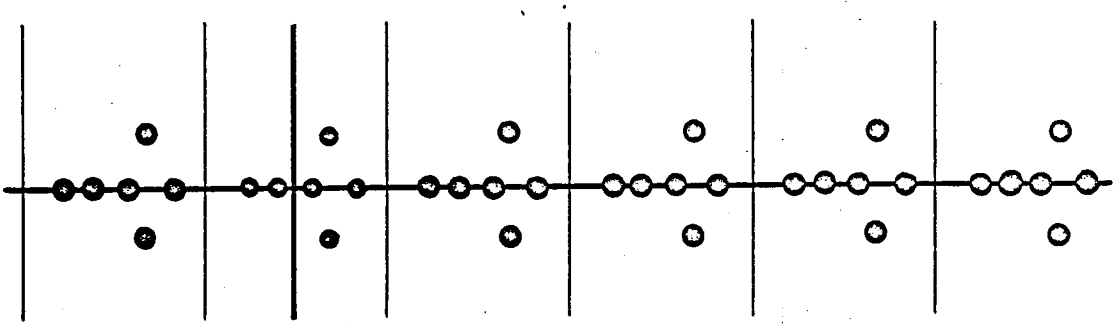
This method can be improved by adding zero Nyquist samples to the ends of the segment which is made periodic (fig.2.4.2.). The point of doing this is that a zero Nyquist sample will generally be "less incorrect" than the periodically repeated Nyquist samples of the original signal. This increases the number of Nyquist samples per period, and hence, apparently, the order of the Fourier polynomial which must be factored. But this is not so, as the extra zeros which are added are known. Thus when the coefficients of the Fourier polynomial have been found from the Nyquist samples (via the DFT), the extra roots which have been added can be divided out before factorisation.



True Nyquist samples of signal.

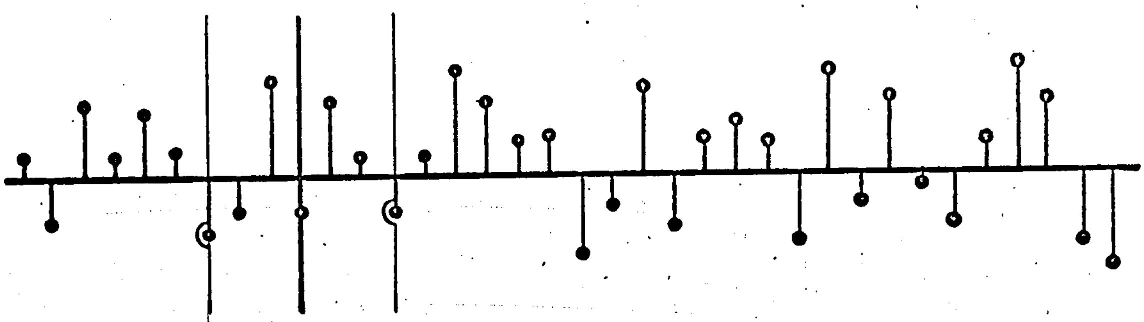


Periodic repetition of central period of true Nyquist samples.

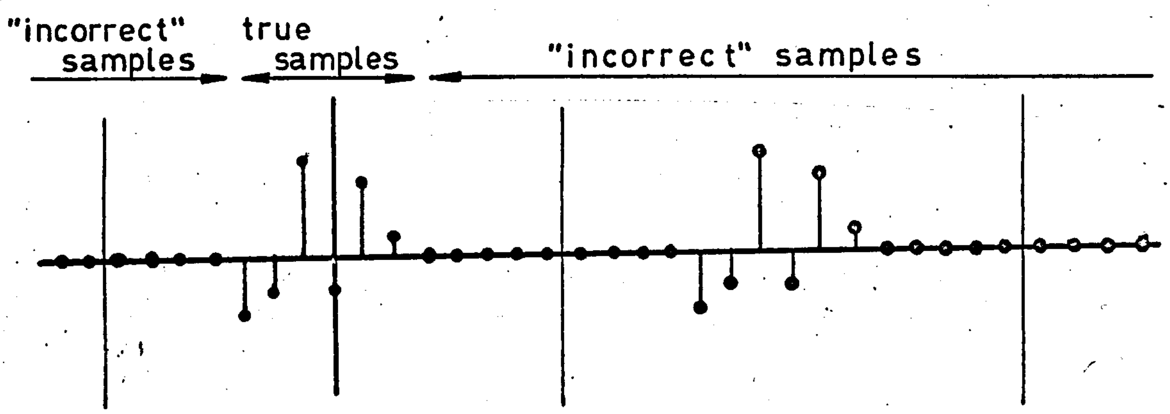


Zero pattern found by Fourier series factorisation.

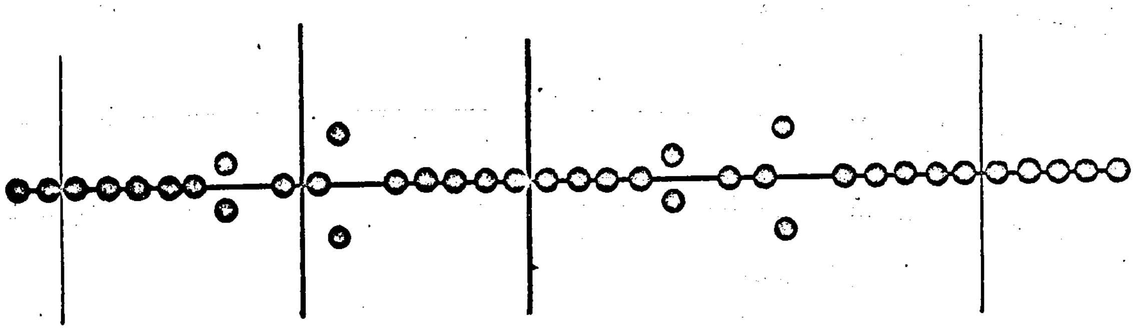
Fig.2.4.1. Periodic repetition of Nyquist samples.



True Nyquist samples of signal.



Periodic repetition of central period of true Nyquist samples with intervening zero-value samples.



Zero pattern found by Fourier series factorisation.

Fig.2.4.2. Provision of zero Nyquist samples results in more accurate zero location; "incorrect" samples are generally less incorrect.

2.5 Zero Location by Axis Shifting

The methods for locating the zeros of a signal which are presented in the previous sections are numerical methods; they depend upon sampling the signal and operating numerically on the samples to produce the zero locations. The scheme presented in this section is different in character. A linear filter is used to operate on the signal so that zeros which lie on some line parallel to the t-axis of the complex time plane, are converted into real zeros. Real zeros are readily detected. By using a sufficient number of such filters the zeros can, in principle, be located to any desired degree of accuracy.

The use of an "axis-shifting" filter provides a simple method of zero location. However, in the present work, the principle of the axis shifting filter is found useful as a conceptual device in studying the effects of added noise on the zeros of signals.

A bandlimited signal lying in the frequency range (f_l, f_l+W) can be written, using the inverse Fourier transform,(as a function of the complex-time variable, z).

$$m(z) = \int_{f_l}^{f_l+W} M_{(f)}\, e^{j2\pi fz}\, df \qquad 2.5.1$$

where $z = t+j\sigma$. Giving σ a particular value, σ_a, a new time function $m_a(t)$ can be formed:

$$\begin{aligned} m_a(t) &= m(t+j\sigma_a) \\ &= \int_{f_l}^{f_l+W} M_{(f)}\, e^{-2\pi f\sigma_a}\, e^{j2\pi ft}\, df \end{aligned} \qquad 2.5.2$$

This equation shows that $m_a(t)$ can be obtained from $m(t)$ by filtering using a transfer function

$$H_a(f) = e^{-2\pi f \sigma_a} \quad , \quad f_l<f<f_l+W$$

$$= \text{anything} \quad , \quad \text{otherwise}. \qquad 2.5.3$$

If $m(z)$ has a zero at $(t_l + j\sigma_a)$, then $m_a(t)$ must have a (real) zero at t_l. At a real zero of a complex or an analytic signal both the real and the imaginary parts of the signal must be zero so that the envelope must be zero there, too. The zeros of a bandlimited signal can thus be located (in principle) by filtering, using transfer functions of the form given above with all possible values of σ_a, and observing the real zeros of the envelopes of the filtered signals.

As an illustration, fig. 2.5.1(a) shows the waveform of the analytic signal lying in the frequency range $(0,W)$ which has the zero pattern shown in fig. 2.5.2. Fig. 2.5.1(b) shows the waveform of the signal after filtering with the appropriate transfer function to convert the lower half plane zeros into real zeros.

Fig. 2.5.3. shows how the scheme might be implemented in a practical way, to provide approximate zero location using a limited number of axis shifting filters. In this application the real zero detectors would be replaced by threshold detectors which would give an indication whenever the envelope of a filtered signal fell below some small value. This scheme would be particularly simple to implement if the zeros of a signal were known to lie nowhere except on two lines parallel to the t-axis (for example, in the case of the angle coded signals which are studied in subsequent chapters). Only two axis shifting filters would then be required.

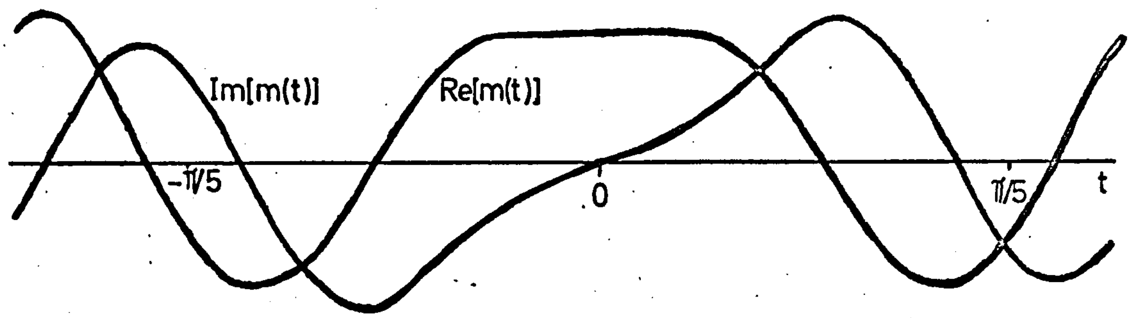


Fig.2.5.1(a). Waveform at input to axis shifting filter.

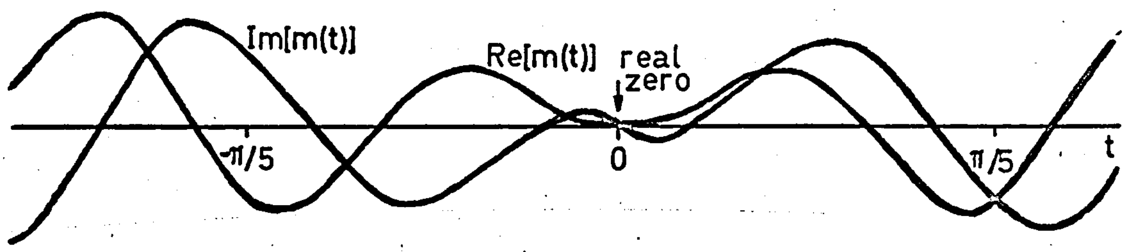


Fig. 2.5.1(b). Waveform at output of axis shifting filter.

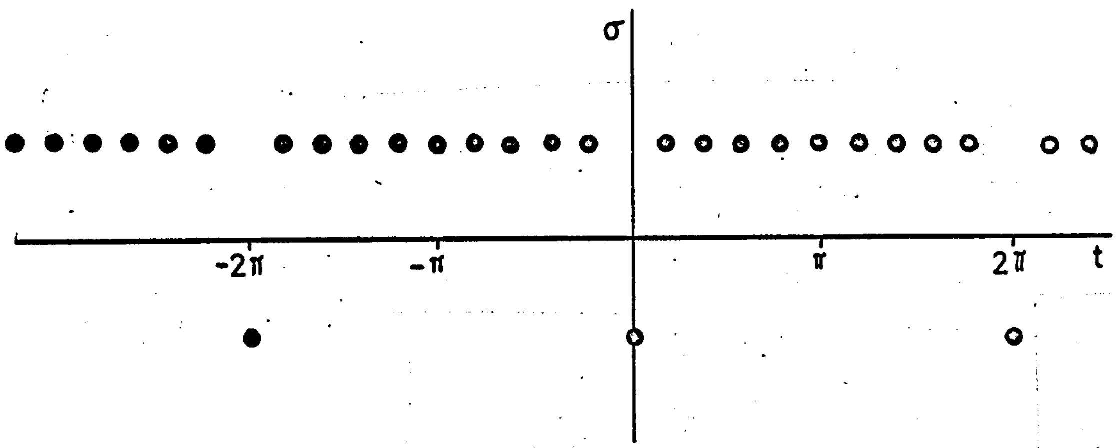


Fig. 2.5.2. Zero pattern of signal at input to axis shifting filter.

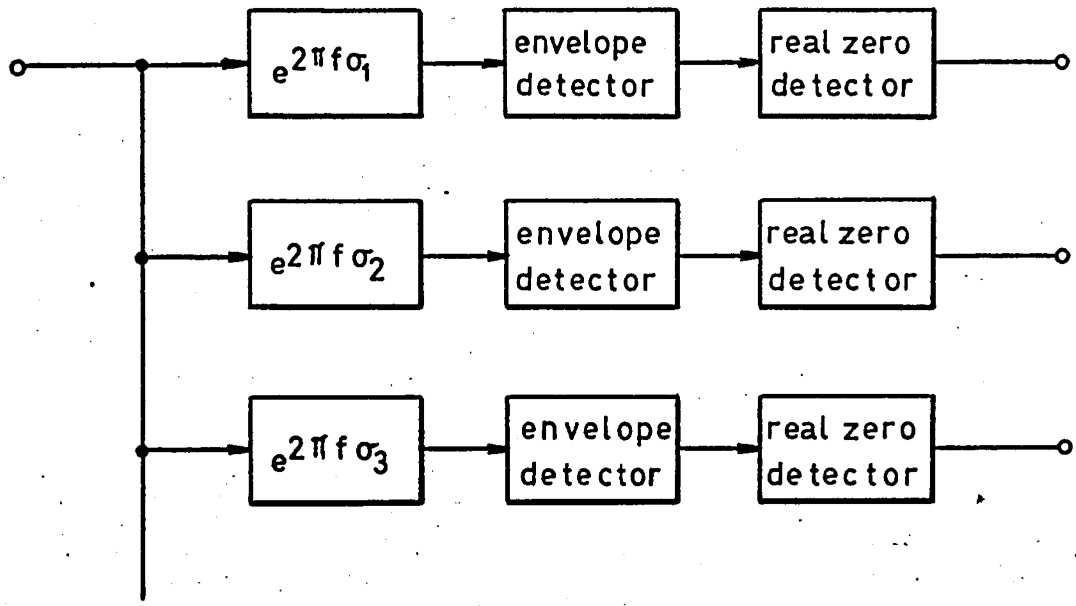


Fig.2.5.3. Scheme for zero location using axis shifting filters.

CHAPTER 3.

EFFECTS OF FILTERING AND NOISE ON THE ZEROS OF SIGNALS.

3.1 Introduction

The effects of physical transmission channels on signals can often be represented as the addition of random noise to the signal or as the operation of a linear filter (or both). Conventional 'linear' signal theory deals very successfully with these situations. For a zero based theory of signals to be of general use, it too should be capable of dealing with such problems. In particular, it should be possible to predict the effects of adding noise or filtering on the locations of the zeros of a signal.

In the case of a PBL signal, which is represented by a Fourier polynomial;

$$M(Z) = C_0 + C_1 Z + \ldots + C_N Z^N \qquad 3.1.1.$$

(where $Z = \exp j2\pi f_0 t$, and where f_0 is the fundamental frequency of the signal), such problems are partly equivalent to studying how the roots of the polynomials

$$M_n(Z) = (C_0+n_0) + (C_1+n_1) Z + \ldots + (C_N + n_N)Z^N \qquad 3.1.2.$$

and

$$M_f(Z) = C_0 H_{(0)} + C_1 H(f_0) Z + \ldots + C_N H_{(Nf_0)} Z^N \qquad 3.1.3.$$

differ from those of M(Z). Here, $n_0, n_1, \ldots, n_N$ are the Fourier coefficients of a periodic noise function which is added to the signal and has the same period and bandwidth. (Note that it may be no less realistic to represent noise as a periodic function than it is to represent the signal as such a function). $H_{(f)}$ is the transfer function of the filter whose effect on the zeros is to be studied. Thus (for PBL signals) problems of the effects of noise and filtering on the zeros of a signal are partly equivalent to questions of how the roots of polynomials are related to their coefficients. This problem is difficult; it is not well understood even though it has been of interest to mathematicians for many years. (When the problems are formulated for non periodic band-limited signals they are even more intractable).

3.2 Effects of Linear Filtering

In general it is difficult to make specific statements about the effects a particular filter will have on a signal. However it is possible to make broad statements about the effects of certain types of filter. For example, as pointed out by Tetarev[6] differentiating a real PBL a sufficient number of times will cause all of its zeros to become real*, and further differentiations regularise the spacings of the zeros.

General statements can also be made about the effect of filters which reduce the bandwidth of signals. Titchmarsh's results show that the average rate of occurrence of zeros is equal to the spectral extent of the signal. Thus a filter which reduces the total spectral extent of the signal (such as a lowpass filter whose bandwidth is less than that of the signal) results in a new zero pattern which has a reduced number of zeros.

* A rigorous proof is given by Szego[7].

The 'pre-envelope' of a real signal is formed from the real signal by filtering with an impulse response $h(t) = (\delta(t) + j/\pi t)$. The spectrum of the pre-envelope is that of the real signal with the negative frequency components eliminated (and the positive frequency components doubled in amplitude) so that the filter reduces the total spectral extent of the signal. In the case of a real signal whose bandwidth is small compared with its centre frequency the reduction in the zero rate can be very large. (Also the zeros of the pre envelope of a real signal are very seldom real or in conjugate pairs - unlike the zeros of the real signal itself which are always thus).

The one type of linear filter whose effect on the zeros of a signal <u>is</u> simple is the axis shifting filter, which was discussed in section 2.5 of the previous chapter. In the following section the axis shifting filter is used as a conceptual device in studying the effects of added noise on the zeros of a signal.

3.3 <u>Effects of Added Noise.</u>

It is possible to set general bounds on the magnitudes of the changes in the roots of a polynomial[8] which result when the coefficients are perturbed by the addition of error terms, as expressed by equation 3.1.2. However, these bounds are very weak for polynomials of only moderately high order; for the noise magnitudes which exist in even the most precise engineering situations they are not relevant.

The root-locus methods of servo theory[9] provide a set of rules by which the loci (as k is varied) of the polynomial $P(z) = Q(z) + k\,R(z)$ can be sketched. In the signal theory context

Q(z) might represent a wanted signal and R(z) might represent an added noise whose power is set by the value of k. Reference 9 treats sampled-data servos and gives examples. Although root locus methods can be applied to the study of the movements of the zeros of signals, it is felt that the method is not very useful. In signal theory it is the general properties of signals and their transformations which is of interest, rather than the study of a unique signal. In any case, with the availability of digital computers it is perhaps more straightforward to plot the root loci by factoring P(z) for a succession of values of k, rather than use the rules.

The relationship between the zeros of a PBL signal and its F-coefficients can be pictured as a multivariable memoryless nonlinear system (fig.3.3.1.). Studying the effects on the signal zeros of added noise is equivalent to examining the effects on the outputs of the nonlinear system of noise added to its inputs. The theory of multivariable memoryless nonlinear systems[10] is of little help. Because of the difficulty of calculating even such 'simple' statistics as the variance of the output variables given the variances (and covariances) of the input variables, the usual approach is to estimate such statistics experimentally[11].

When the power of the noise is small compared with that of the signal, it should be possible to linearise the Fourier coefficient-zero relations so as to be able to calculate the zero movements resulting from the addition of a (small) noise voltage. However, this will only be possible when the Fourier coefficient-zero relationships are sufficiently continuous; that is, when small changes in the Fourier coefficients produce only small corresponding

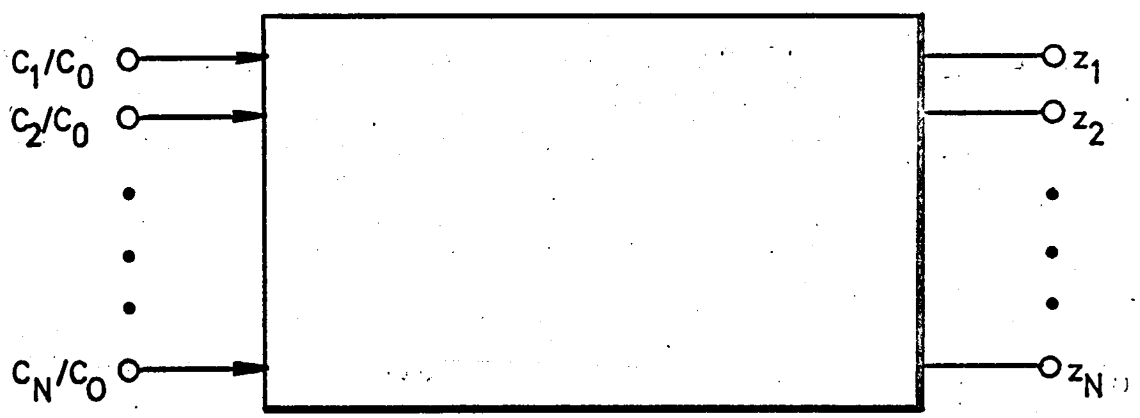


Fig. 3.3.1. Relalations between Fourier coefficients and zeros pictured as a multi-input, multi-output nonlinear system.

changes in the zero positions.

The linearised Fourier coefficient-zero relations are specified by the sensitivity matrix $\underline{S}$;

$$\underline{S} = \begin{bmatrix} \frac{\partial z_1}{\partial C_0} & \frac{\partial z_1}{\partial C_1} & \cdots & \frac{\partial z_1}{\partial C_N} \\ \vdots & & & \\ \frac{\partial z_N}{\partial C_0} & \frac{\partial z_N}{\partial C_1} & \cdots & \frac{\partial z_N}{\partial C_N} \end{bmatrix} \qquad 3.3.1$$

The elements of this matrix are given by

$$\frac{\partial z_k}{\partial C_\ell} = \frac{dz_k}{dZ_k} \frac{\partial Z_k}{\partial C_\ell}$$

Now

$$\frac{dZ_k}{dz_k} = \frac{j2\pi}{T} \exp(j2\pi z_k/T)$$

$$= j2\pi \, Z_K/T$$

and

$$\frac{\partial Z_k}{\partial C_\ell} = -Z_k^{\ell} \Big/ \left. \frac{dM(Z)}{dZ} \right|_{Z = Z_k}$$

Thus

$$\frac{\partial z_k}{\partial C_\ell} = \frac{jT}{2\pi} \; Z_k^{\ell-1} \Big/ \left. \frac{dM(Z)}{dZ} \right|_{Z=Z_k} \qquad 3.3.2$$

When the linearisation procedure is valid, the change, $\underline{\Delta Z}$, in the vector whose elements are the zero positions produced by a change $\underline{\Delta C}$ in the Fourier coefficient vector is approximately given by

$$\underline{\Delta Z} = \underline{S} \; \underline{\Delta C} \qquad 3.3.3$$

The linearisation is only valid when the elements of $\underline{S}$ are sufficiently small. The remainder of this section is a verbal discussion of the circumstances when this is so. The 'axis-shifting' filter of section 2.5 is used as a conceptual device. Real zeros at the output of an axis-shifting filter are pictured as being produced by cancellation of the frequency components of the signal. The effect of added noise is to upset this complete cancellation.

In section 2.5 a filter having the transfer function

$$H(f) = e^{-j2\pi f\sigma_A} \qquad 3.3.4$$

(over the frequency range of the signal) was termed an <u>axis-shifting filter</u>. If its input signal is s(t), its output signal $s_A(t)$, takes the values of s(z) on a line in the z-plane running parallel to the t-axis and cutting the imaginary axis at σ_A. Thus if the input signal to the axis shifting filter has a complex zero at $(t_1+j\sigma_A)$, the output signal from the filter has a real zero at t_1.

A physical understanding of the action of an axis shifting filter can be gained by considering a signal having a zero pattern such as that of fig. 3.3.2. Such a signal may be represented by a Fourier series :

$$m(t) = 1 - a \; e^{j2\pi t/T} \qquad 3.3.5$$

(where $a = e^{2\pi\sigma/T}$). The vector representation of this signal is shown in fig. 3.3.3. The component vectors are of unequal length and so they do not cancel for any real value of time, t.

Suppose that m(t) is passed through the axis-shifting filter which converts its complex zeros into real zeros. The transfer function of this filter is given by equation 3.3.4. m(t) is transformed into the signal

$$m_1(t) = 1 - e^{-2\pi\sigma/T} \; a \; e^{j2\pi t/T}$$

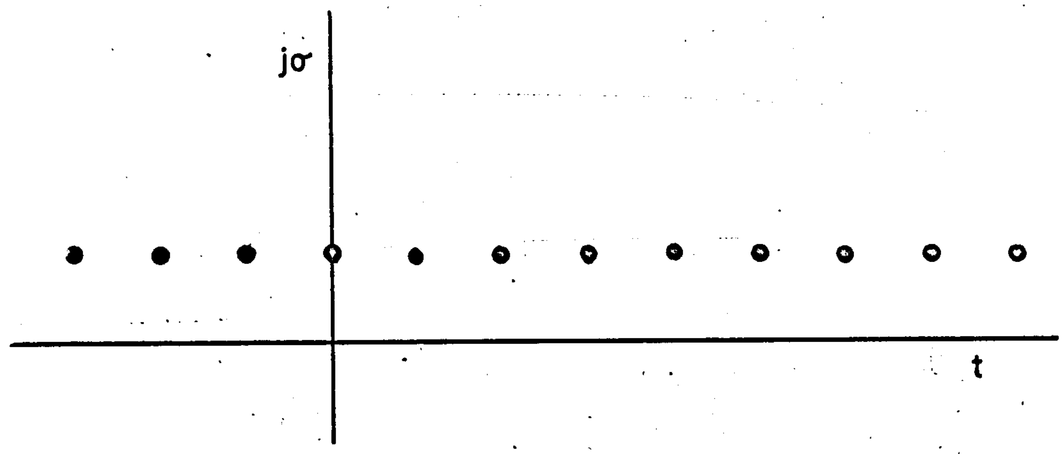


Fig.3.3.2. Zero pattern of signal $1 - a\,e^{j2\pi t/T}$.

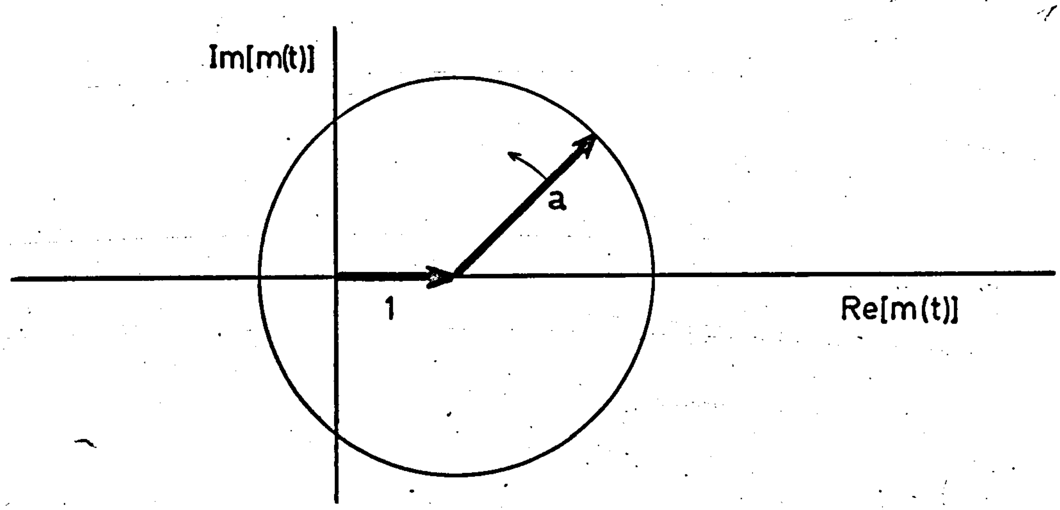


Fig.3.3.3. Vector representation of $1 - ae^{j2\pi t/T}$.

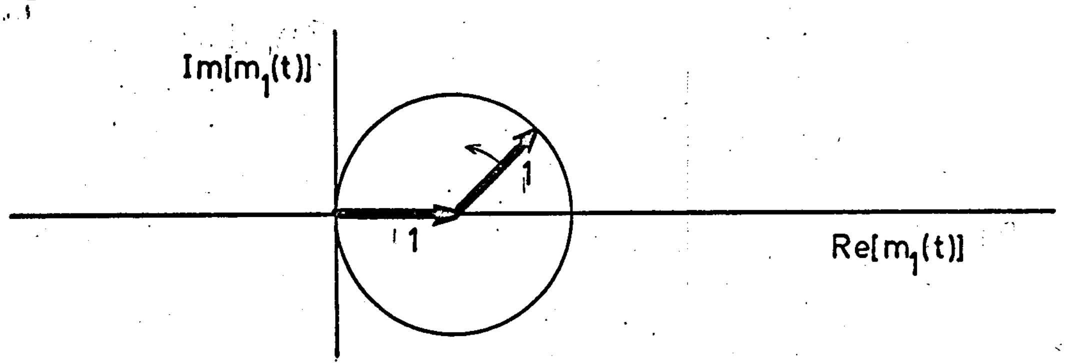


Fig.3.3.4. Vector representation of $1 - e^{j2\pi t/T}$.

$$= 1 - e^{j2\pi t/T}$$

which has the vector diagram shown in fig. 3.3.4. The component vectors are of equal length and so they cancel exactly when they lie in opposite directions. The action of an axis shifting filter in producing real zeros can thus be visualised as being to equalise the magnitudes of the frequency components so that they cancel at certain instants of time (the resulting real zeros)

The effect of an axis shifting filter on a signal with a more complicated zero pattern can be explained in a similar way. The zero pattern of the signal, m(t), will be supposed to be arbitrary except that it contains a 'periodic zero', that is, a zero at positions such as those shown by fig. 3.3.5. It is the action of the axis shifting filter on these periodic zeros that is to be discussed. The signal can be factored into the form

$$m(t) = \left[1 - a\, e^{j2\pi t/T}\right] r(t)$$

The Fourier transform of m(t) can thus be written

$$M(f) = R(f) - a\; R(f-1/T)$$

Suppose that if the 'remainder signal' r(t) is applied to an axis shifting filter the output which results is q(t). Then the output, $m_l(t)$, of the axis shifting filter, in response to m(t) has the F-transform

$$M_l(f) = H(f) \quad M(f)$$

$$= e^{-2\pi f\sigma_1}\left[R(f) - a\; R(f-1/T)\right]. \qquad 3.3.6$$

Equation 3.3.6 becomes

$$M_l(f) = Q(f) - Q(f-1/T)$$

and $m_l(t)$ is thus given by

$$m_l(t) = q(t) - e^{j2\pi t/T}\, q(t)$$

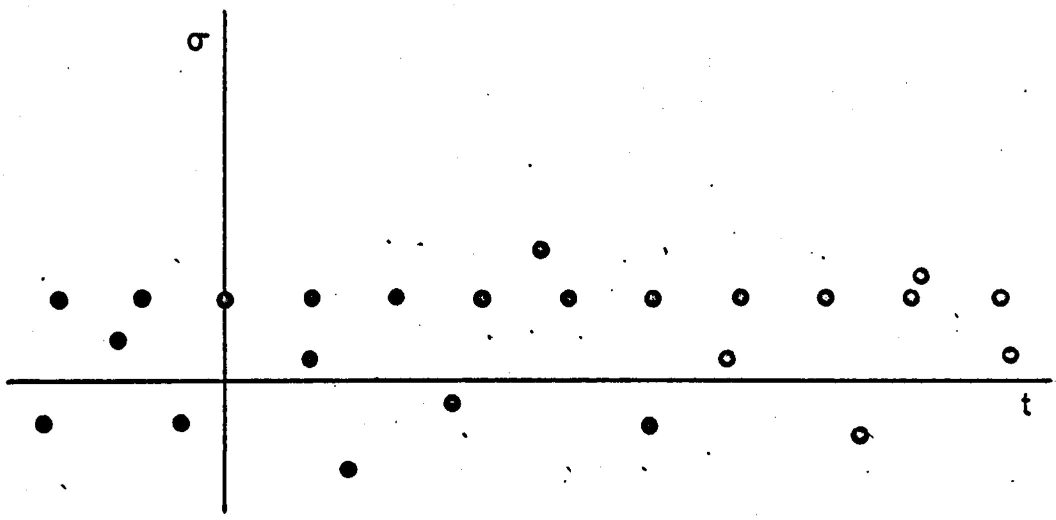


Fig. 3.3.5. Zero pattern with "periodic zero".

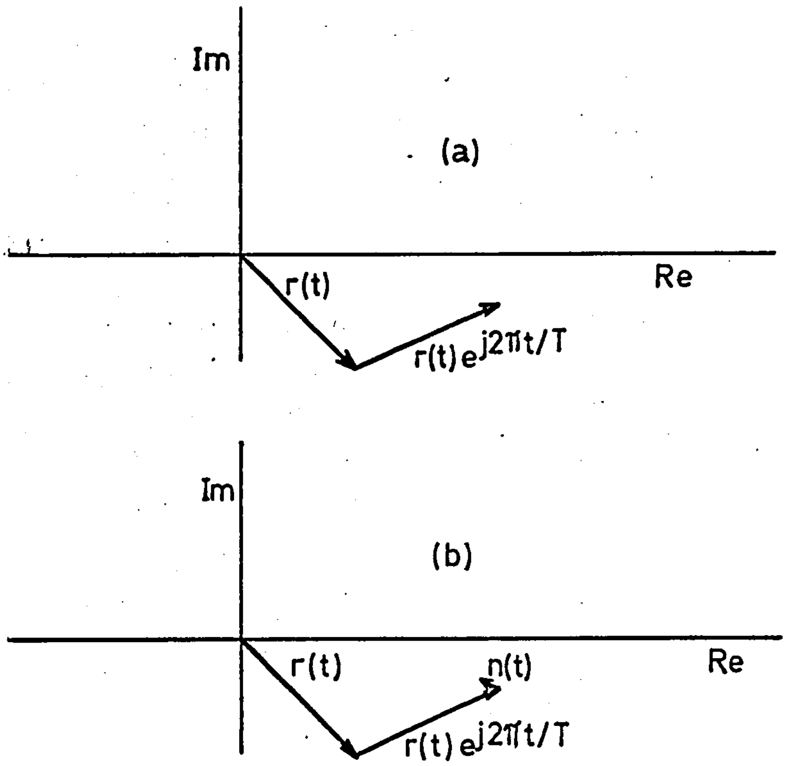


Fig. 3.3.6. Vector representation of (a) signal having a periodic zero and (b) this signal plus noise.

This signal is shown in vector form in fig. 3.3.6a. Like the signal discussed previously, it consists of the sum of two vectors whose lengths are equal but whose instantaneous frequencies differ by $1/T$. The difference between this case and the previous one is that the lengths and angular velocities of the individual vectors are no longer fixed.

In general, when a proportion of noise is added to a bandlimited signal the location of every one of the zeros is changed. The change in the position of each zero depends on the waveform of the noise, the original location of the zero and also in the original locations of all the other zeros of the signal. However, when the change in each zero location is small, the change depends predominantly on the original location of the zero itself and little on the original locations of the other zeros. In this section, because the main concern is to study the conditions under which the change in a zero's location is small, it is permissible to consider only the influence of the location of the zero itself on its sensitivity to noise and to disregard the effect of the other zeros. Only signals having simple zero patterns such as that shown in fig. 3.3.2. are considered here.

If the zeros of the signal are close to the t-axis (that is, the ratio $|\sigma_1|/T$ is small, T being the spacing between zeros in the direction of the t-axis) the signal may be represented by a two-term Fourier series

$$m(t) = 1 + a\, e^{j2\pi t/T} \qquad 3.3.7$$

in which the magnitude of a is close to unity. If random noise, $n(t)$, of small rms value and occupying the same frequency band as the signal is added to $m(t)$, the resulting sum can be represented by a vector diagram such as that of fig. 3.3.6(b). The noise component is represented by a vector of randomly varying length and direction.

The axis shifting filter which converts the zeros of m(t) into real zeros has a gain-frequency characteristic which does not vary greatly over the frequency band of the signal. This is because to produce real zeros the filter has to equalise the amplitudes of the frequency components of m(t) and these do not differ greatly to start with. Fig. 3.3.7. shows the vector diagram of the signal at the output of the axis shifting filter when there would be a real zero in the absence of noise. The output of the filter at this instant consists of noise alone. The magnitude of the noise being small, a small change in the value of the parameter σ_A of the axis shifting filter produces the situation represented in fig. 3.3.8. where there would be a real zero if the instantaneous phase of one frequency component of the signal were changed slightly. Provided that the change is small compared with a Nyquist interval such a change can be effected by advancing or delaying the signal plus noise without appreciable change in the magnitude or phase of the noise vector.

The foregoing verbal argument suggests that when a small proportion of noise is added to a signal whose zeros are close to the t-axis the resulting change in the position of each zero is small. With such signals linearisation of the Fourier coefficient-zero relations might be used to study the effects of added noise on the zeros.

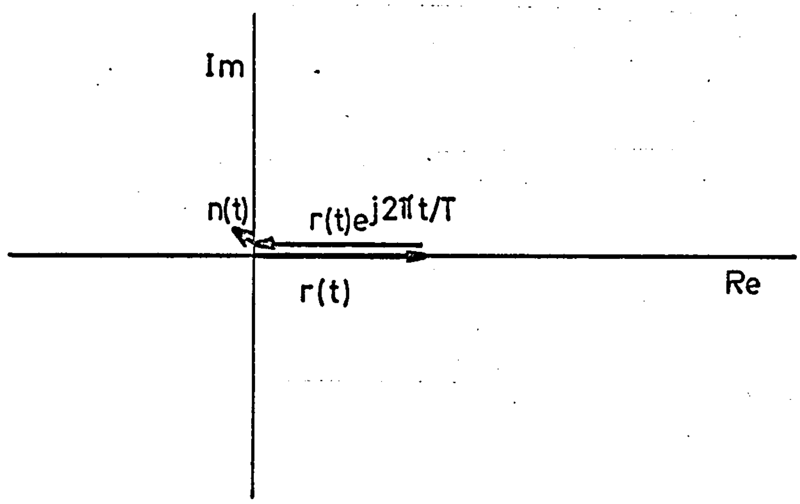


Fig.3.3.7. Signal plus noise at the instant of a real signal zero.

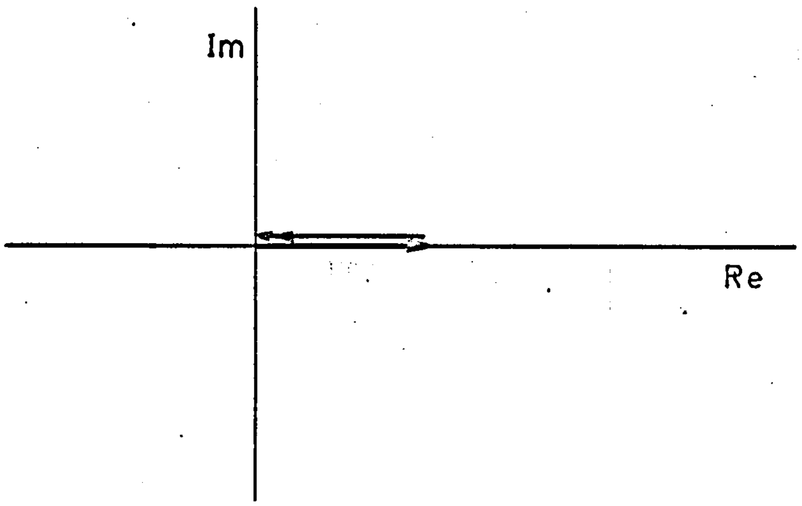


Fig. 3.3.8. Small change in σ_A will produce a real zero.

CHAPTER 4

ANGLE CODED SIGNALS AND THEIR PROPERTIES

4.1 Introduction

As an application of zero based signal theory, Voelcker suggested that signals could be generated so that the zero pattern of a signal would represent binary data explicitly[4]. As an example, he generated periodic signals having zero patterns in which the appearance of a particular zero in the upper half of the complex-time plane would denote a binary '1' and its appearance in the lower half plane, a binary 'o'. Thus by systematically conjugating the N zeros (complex, and not occurring in conjugate pairs) of the fundamental period, a total of 2^N different signals can be generated. As the envelope of a signal remains unaltered when a complex zero is replaced by its conjugate, these signals are members of a "common envelope set" and they differ solely in having different instantaneous phase functions. Accordingly, Voelcker termed signals generated in this way <u>angle coded signals</u>.

In the remaining chapters of this thesis attention is confined principally to the study of angle coded signals having zero patterns like that shown in fig. 4.1.1., in which the zeros occur at regular intervals of T in the direction of the t-axis and at either +s or -s in the direction of the σ-axis.

Fig. 4.1.2 shows the real part, the imaginary part, the envelope and the instantaneous frequency of the signal whose zero pattern is shown in fig. 4.1.3. and which lies in the frequency range (0,W). The <u>aspect ratio</u>, p(=s/T), of this signal is 0.6. Note that the behaviour of the envelope is the same in the vicinity of both the upper half plane (UHP) and the lower half plane (LHP) zeros, while the instantaneous frequency peaks in the vicinity of

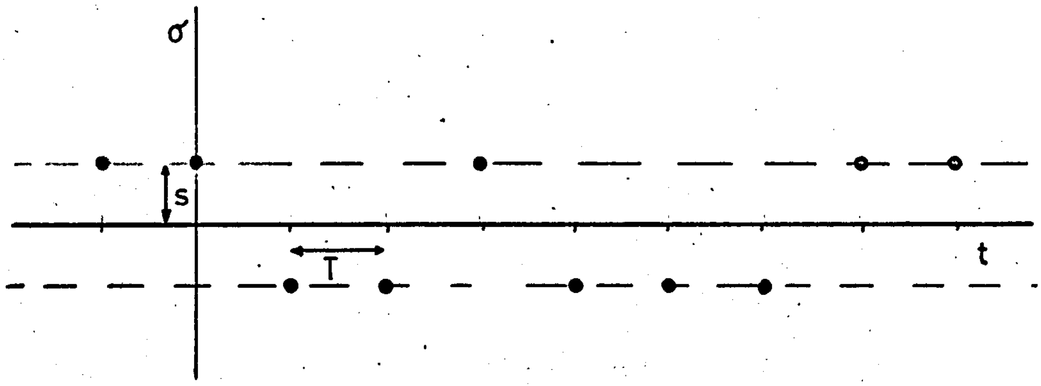


Fig.4.1.1. Typical angle coded signal zero pattern.

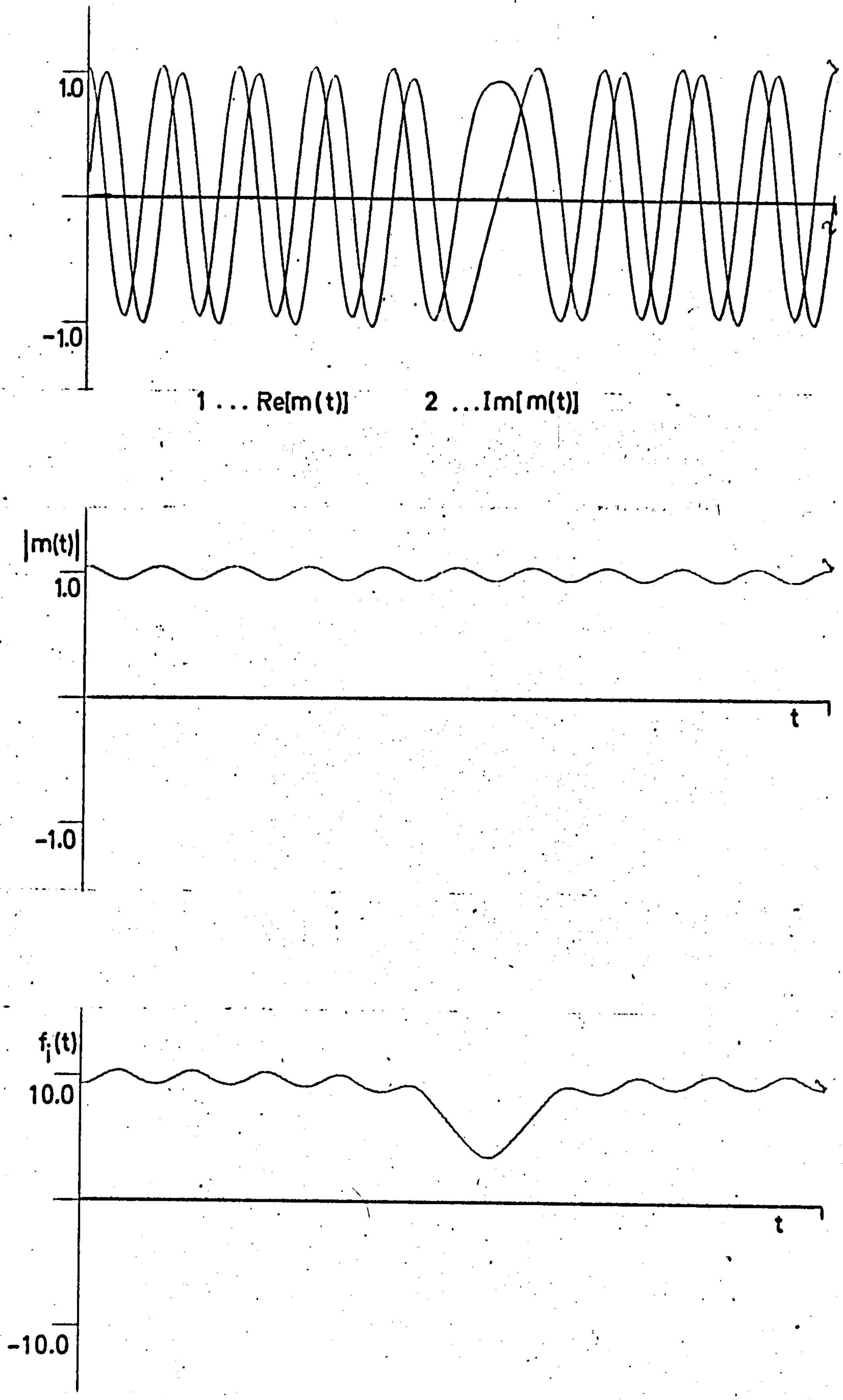


Fig.4.1.2. Waveforms associated with the signal having the zero pattern of fig.4.1.3: p = 0.6.

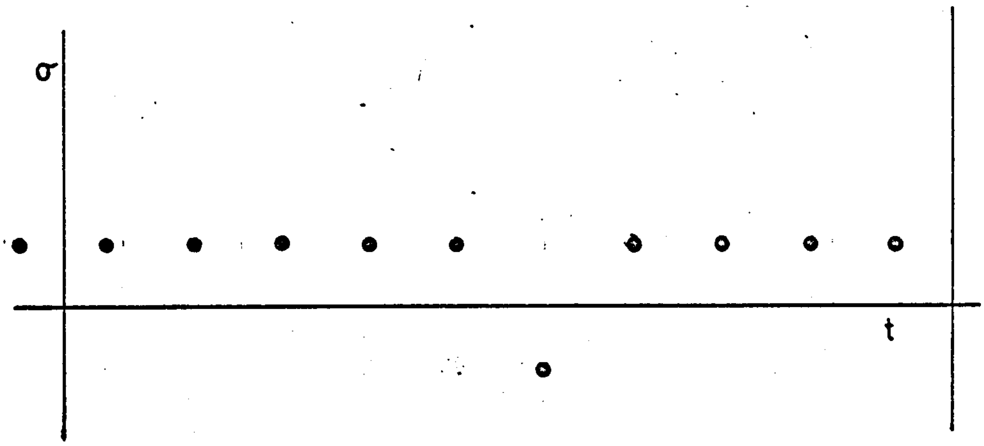


Fig.4.1.3. Fundamental period of zero pattern of example.

an UHP zero and it dips near a LHP zero. Fig. 4.1.4. shows the locus of this analytic signal[12]. Examination of the motion of the analytic signal vector as time progresses gives insight into 'why' the instantaneous frequency and envelope behave as they do in the vicinity of UHP and LHP zeros. Fig. 4.1.5. shows the analytic signal locus for the "minimum phase"[4] member of the common envelope set (all zeros in the LHP), and fig. 4.1.6., for the "maximum phase" member. These loci illustrate the physical implications of <u>minimum</u> and <u>maximum</u> phase. The analytic signal locus encircles the t-axis once for each UHP zero. When all the zeros are UHP the average rate of increase of the instantaneous phase (which is the average of the instantaneous frequency) is W revolutions per second and when all zeros lie in the LHP, it is zero*.

Fig. 4.1.7. shows the waveforms associated with the zero pattern of fig. 4.1.3. when the aspect ratio, p, has the value 0.2 (a small value). The peaks of the instantaneous frequency are large and the dips of the real part and the envelope come close to the t-axis. Fig. 4.1.8. shows the waveforms which result when p has the value 4.0 (a large value). The fluctuations of the envelope and the instantaneous frequency are much reduced in magnitude; so much so that for most practical purposes the signal can be regarded as being a pure frequency modulation signal.

* Reference 13 contains a discussion of the mean value of the instantaneous frequency of a signal which could be cast in terms of zero-based theory. See also ref. 14.

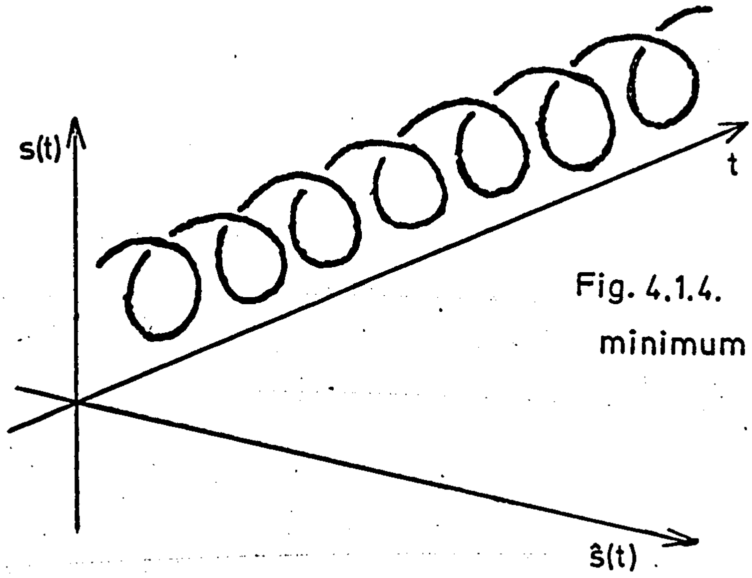


Fig. 4.1.4. Locus of minimum phase signal.

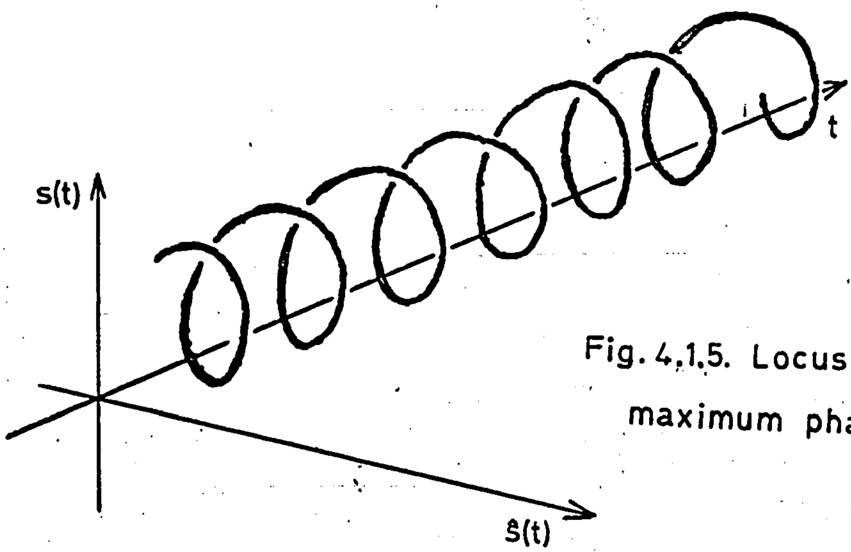


Fig. 4.1.5. Locus of maximum phase signal.

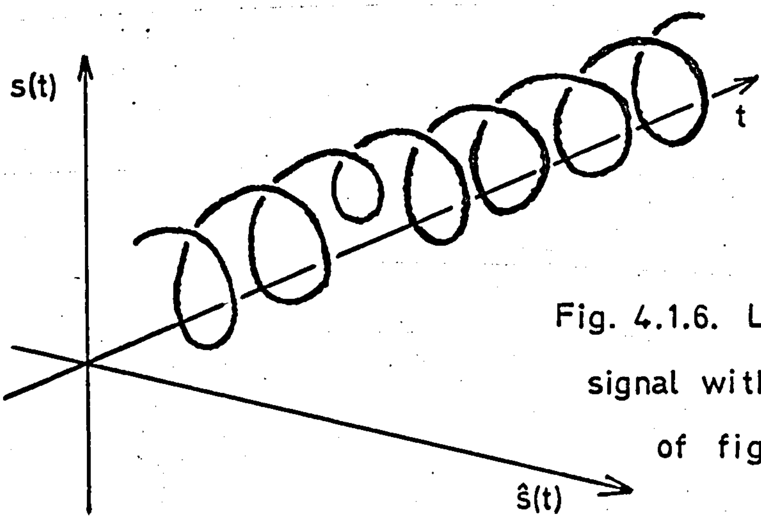


Fig. 4.1.6. Locus of signal with zero pattern of fig. 4.1.3.

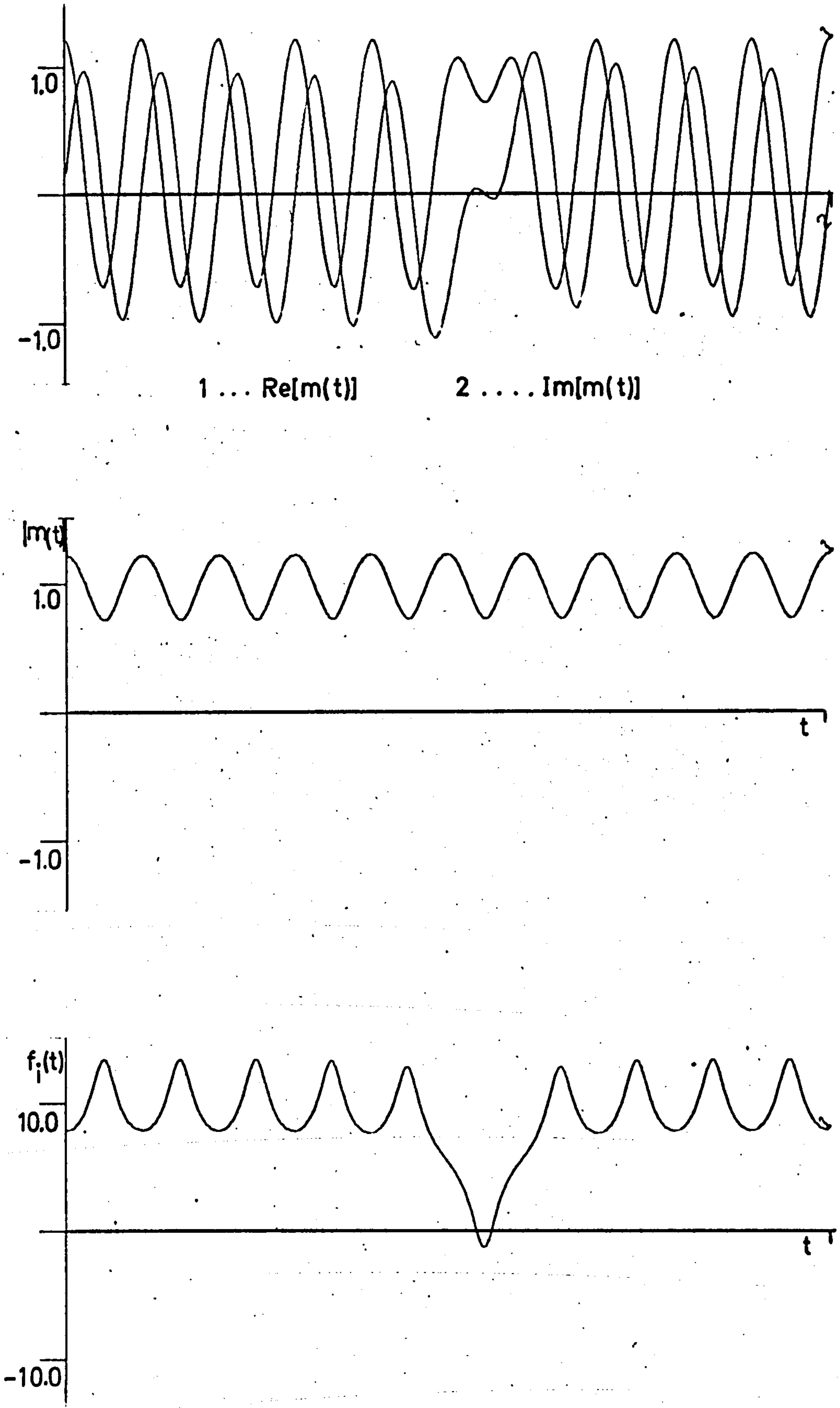


Fig.4.1.7. Waveforms associated with the signal having the zero pattern of fig.4.1.3.: p = 0.2.

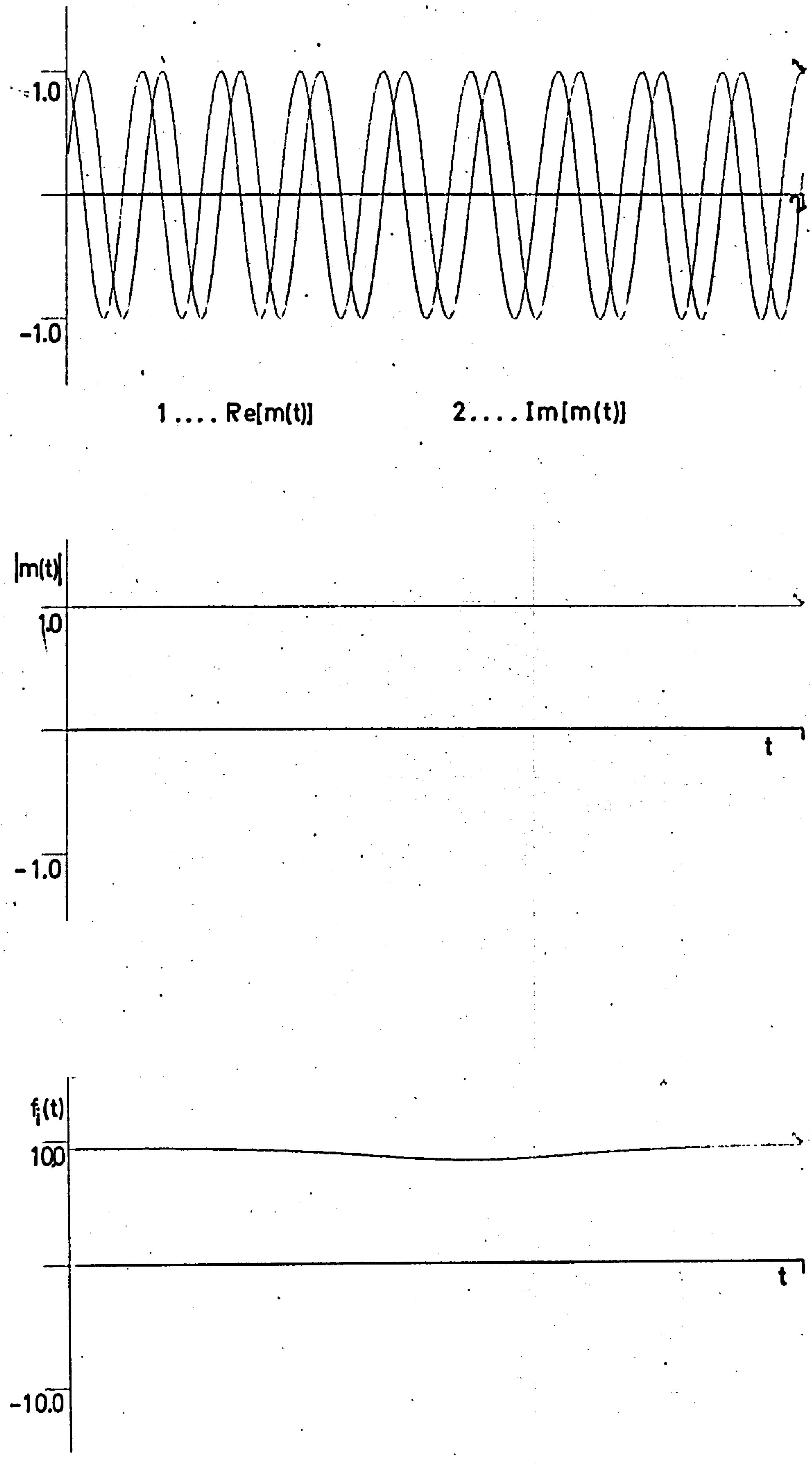
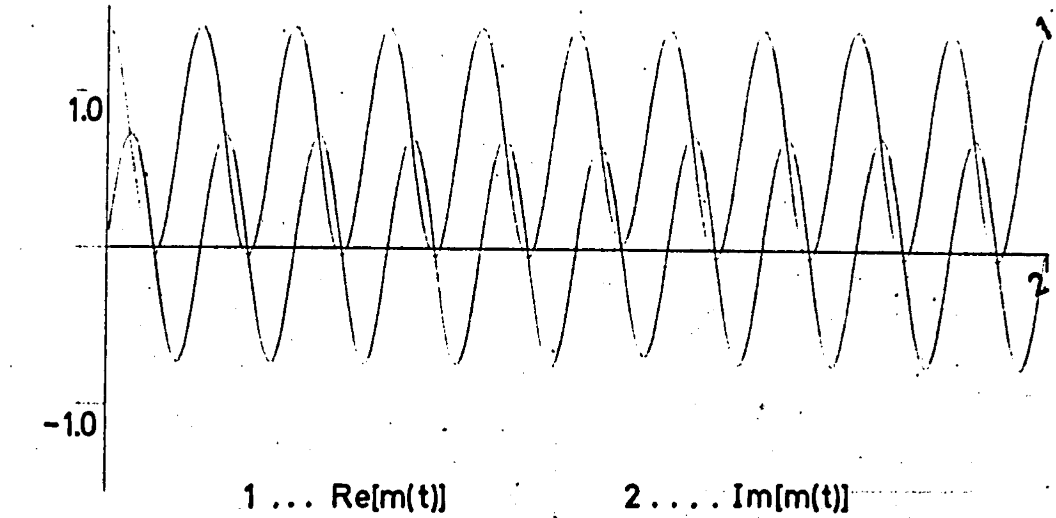


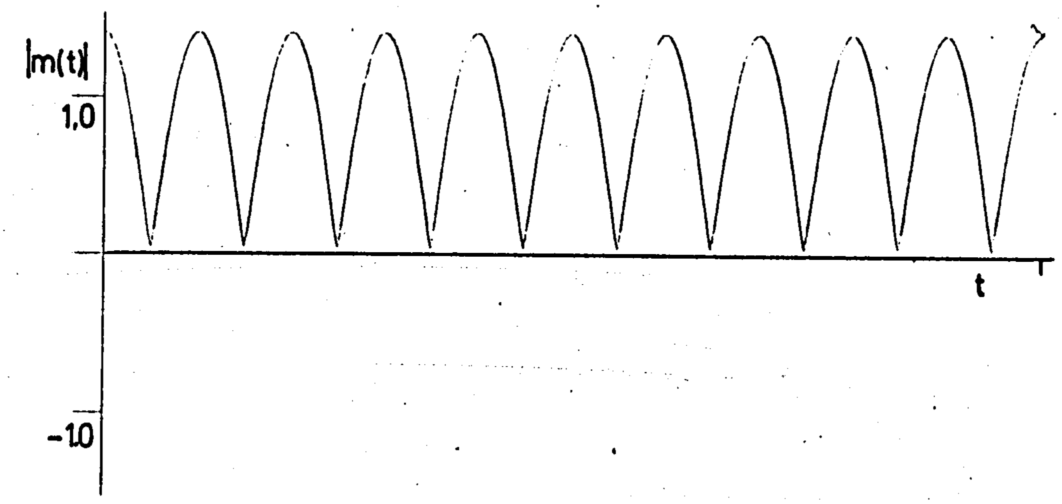
Fig. 4.1.8. Waveforms for p = 4.0.

Fig. 4.1.9. shows the waveforms associated with the same zero pattern for other values of p. The characteristics of the signals undergo a qualitative change at a value of p of about 0.6. For example, a dip in the real part of the signal near a LHP zero occurs only for p less than about 0.6. This qualitative change in nature at a value of p of about 0.6 occurs with other signal characteristics such as the form of the power spectrum (as will be seen later).

As a result of the multiplicative nature of the relationships between the zeros and the waveform of a signal, zero based methods seem to be best suited for discussing situations where signals are multiplied together (as Voelcker has illustrated in his treatment of modulation) or where logarithmic properties of signals are to be studied. However, if angle coded signals are to be used in conventional communication system applications it is important to discuss 'linear' properties of the signals such as the forms of their spectra.

Chapter 2 showed that the relations between the Nyquist samples of a signal and its zeros are generally complicated (i.e. multivariable and nonlinear) and so it is likely that other 'linear' properties of angle coded signals cannot be exactly expressed in terms of the characteristics of their zero patterns in a simple way. Nevertheless, it might be hoped that simple approximate (asymptotic, qualitative) relations might be found between 'linear' properties of the signal and its zero pattern characteristics rather in the way that while the exact relationships between the modulating function and the spectra of an FM signal may be complicated, simple approximate relations often exist.





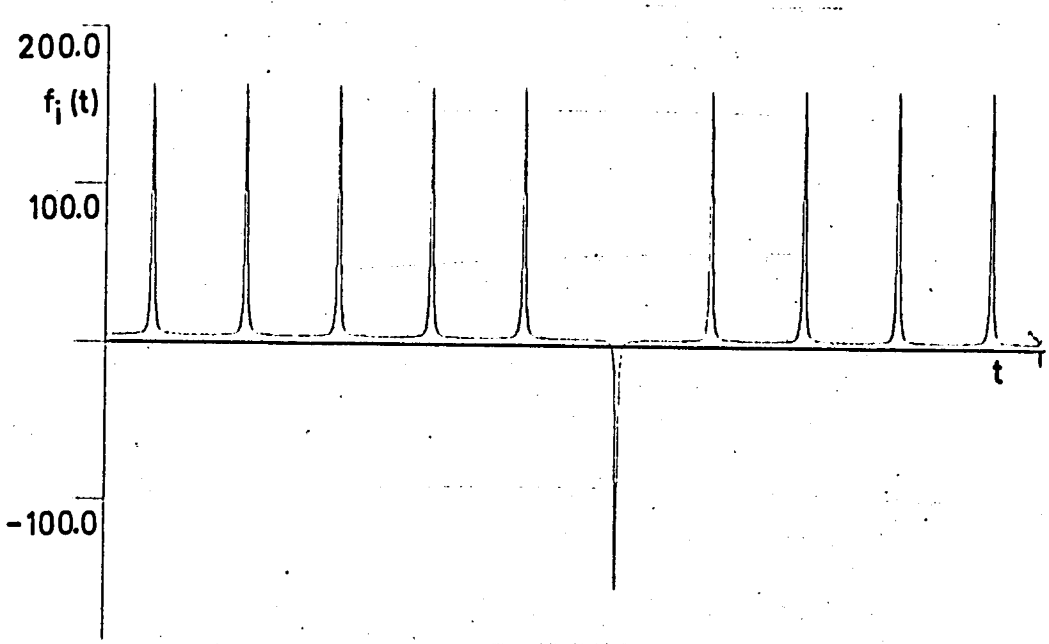


Fig. 4.1.9.a. Waveforms for p = 0.01.

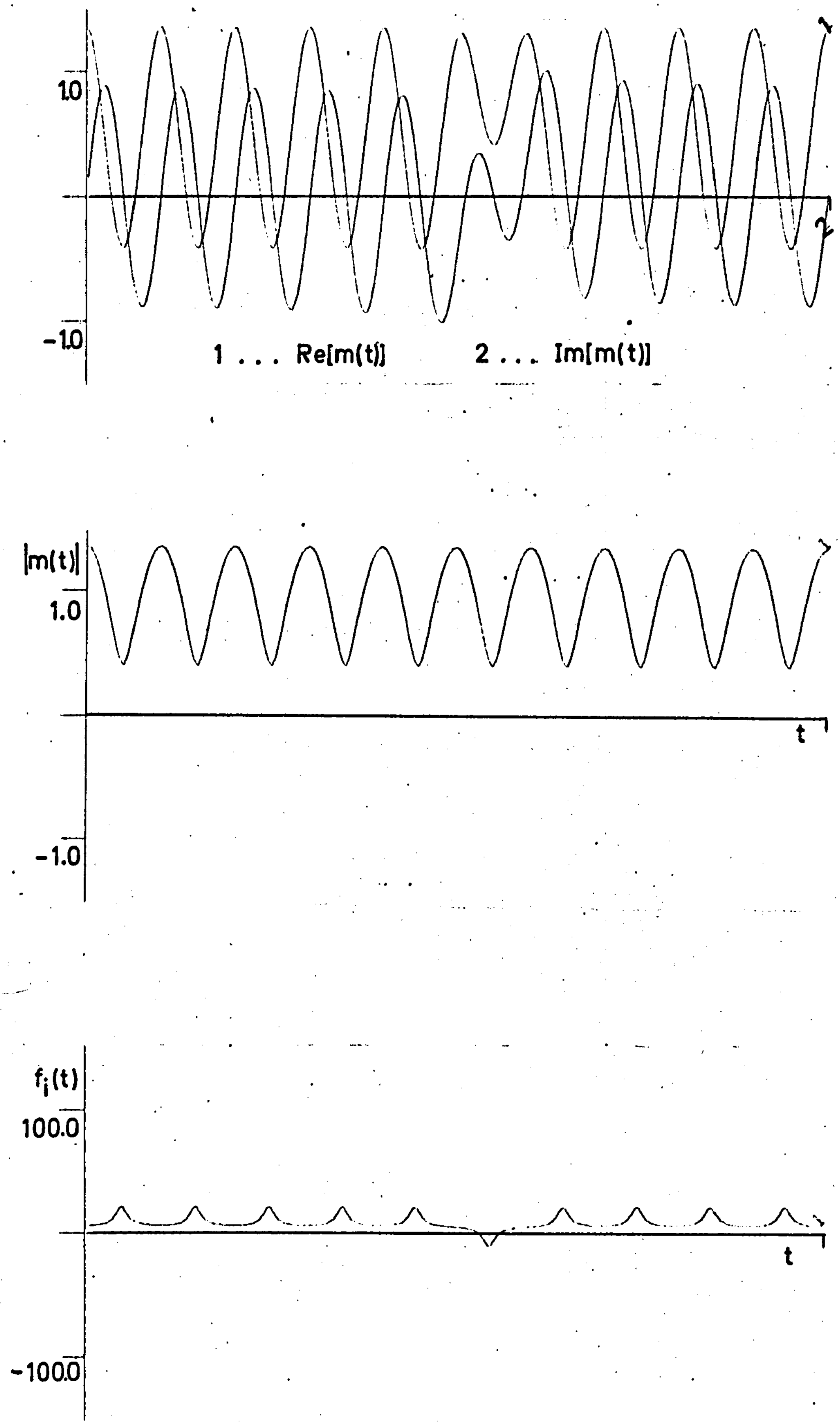


Fig. 4.1.9.b. Waveforms for p = 0.1.

This chapter investigates the properties of angle coded signals whose spectra lie in the range (0,W). It is a trivial matter to extend the results to signals lying in some other frequency range.

4.2 General Properties of Angle Coded Signals

As mentioned in the introduction to this chapter, the zeros of the signals considered have occurred at regular intervals T in the direction of the t-axis and at either +s or -s in the direction of the σ-axis of the complex time plane. The zero locations can thus be written

$$z_n = nT + j\alpha_n s, \; n= \ldots -2,-1,0,1,2, \ldots \qquad 4.2.1$$

where

$$\alpha_n = \begin{cases} +1 \text{ , if the nth zero lies in the UHP} \\ -1 \text{ , if the nth zero lies in the LHP} \end{cases}$$

Fourier Coefficients

A periodic angle coded signal having N zeros per period can be represented (from chapter 2) :

$$m(t) = C_o(1-e^{j2\pi t/NT} / e^{j2\pi z_1/NT}) \ldots (1-e^{j2\pi t/NT} / e^{j2\pi z_N/NT})$$

$$= C_o \prod_{i=1}^{N} (1-e^{j2\pi(t-z_i)/NT}) \qquad 4.2.2$$

On expansion this gives the Fourier polynomial :

$$m(t) = C_o + C_1 e^{j2\pi t/NT} + \ldots . \; C_N e^{j2\pi t/T}$$

$$= \sum_{k=o}^{N} C_k e^{j2\pi kt/NT} \qquad 4.2.3$$

The zero frequency Fourier coefficient (the "d.c." component) of a periodic angle coded signal having a certain particular zero pattern is determined by the power (i.e. the mean square), P.

Expressing P in terms of the Fourier coefficients

$$P = \sum_{k=0}^{N} |c_k|^2$$

gives, for C_o

$$C_o = (P - \sum_{k=1}^{N} |c_k|^2)^{\frac{1}{2}} \qquad 4.2.4.$$

(C_o is assumed to be positive and real without loss of generality). This expression is used below.

Envelope

The properties of the envelope of signals from a common envelope set can be studied by considering any member of the set. It is most simple, however, to consider either the maximum phase or the minimum phase member as these members have only two frequency components in their Fourier series[4]. The Fourier series of the minimum phase signal can be written at once from equation 4.2.3.

$$m(t) = C_o \left[1 + (-1)^N e^{j2\pi t/T} e^{-2\pi s/T}\right]$$

$$= C_o \left[1 + (-1)^N A \; e^{j2\pi t/T} \right] \qquad 4.2.5.$$

where $A = e^{-2\pi s/T} = e^{-2\pi p}$

The squared envelope is given by

$$|m_{(t)}|^2 = m(t)m^*(t)$$

$$= C_o^2 \left[1 + A^2 - 2A \cos(2\pi t/T)\right] \qquad 4.2.6.$$

The spectrum of the squared envelope is zero outside the frequency range (-W,W) - as indeed it must be, by the squared envelope theorem[4,15]. For the minimum phase signal the zero-frequency Fourier coefficient is, from 4.2.4. and 4.2.5., given by

$$c_o^2 = P/(1+A^2) \qquad 4.2.7.$$

and so the envelope is given by

$$|m(t)| \quad = \quad P\left\{\left[1+A^2 - 2A\cos(2\pi t/T)\right] / (1+A^2)\right\}^{\frac{1}{2}} \qquad 4.2.8.$$

The illustrations of the waveforms of angle coded signals shown in the introduction (figs.4.1.7.- 4.1.9) showed that, in a qualitative way, the fluctuations of the envelope of an angle coded signal diminish as p, the aspect ratio, is increased. From equation 4.2.8 the maximum, $|m_{MAX}|^2$, and minimum, $|m_{MIN}|^2$, values of the squared envelope are given by

$$|m|^2_{MAX} \quad = \quad P\ (1+A)^2 / (1+A^2) \qquad 4.2.9.$$

and

$$|m|^2_{MIN} \quad = \quad P\ (1-A)^2 / (1+A^2) \qquad 4.2.10.$$

respectively. The peak-to-peak envelope fluctuation $|m|pp$ is thus given by

$$|m|pp \quad = \quad 2\left[PA / (1+A^2)\right]^{\frac{1}{2}} \qquad 4.2.11.$$

Fig. 4.2.1 shows the maximum and minimum envelope values and the peak-to-peak envelope fluctuation plotted against p (for a signal of unit mean square). When p is large (the zeros of the signal lie far from the t-axis) the envelope fluctuations become diminutive. In this case angle coded signals of the type considered here can be considered, for practical purposes, to be purely frequency modulated signals. This observation is exploited several times in the rest of this work.

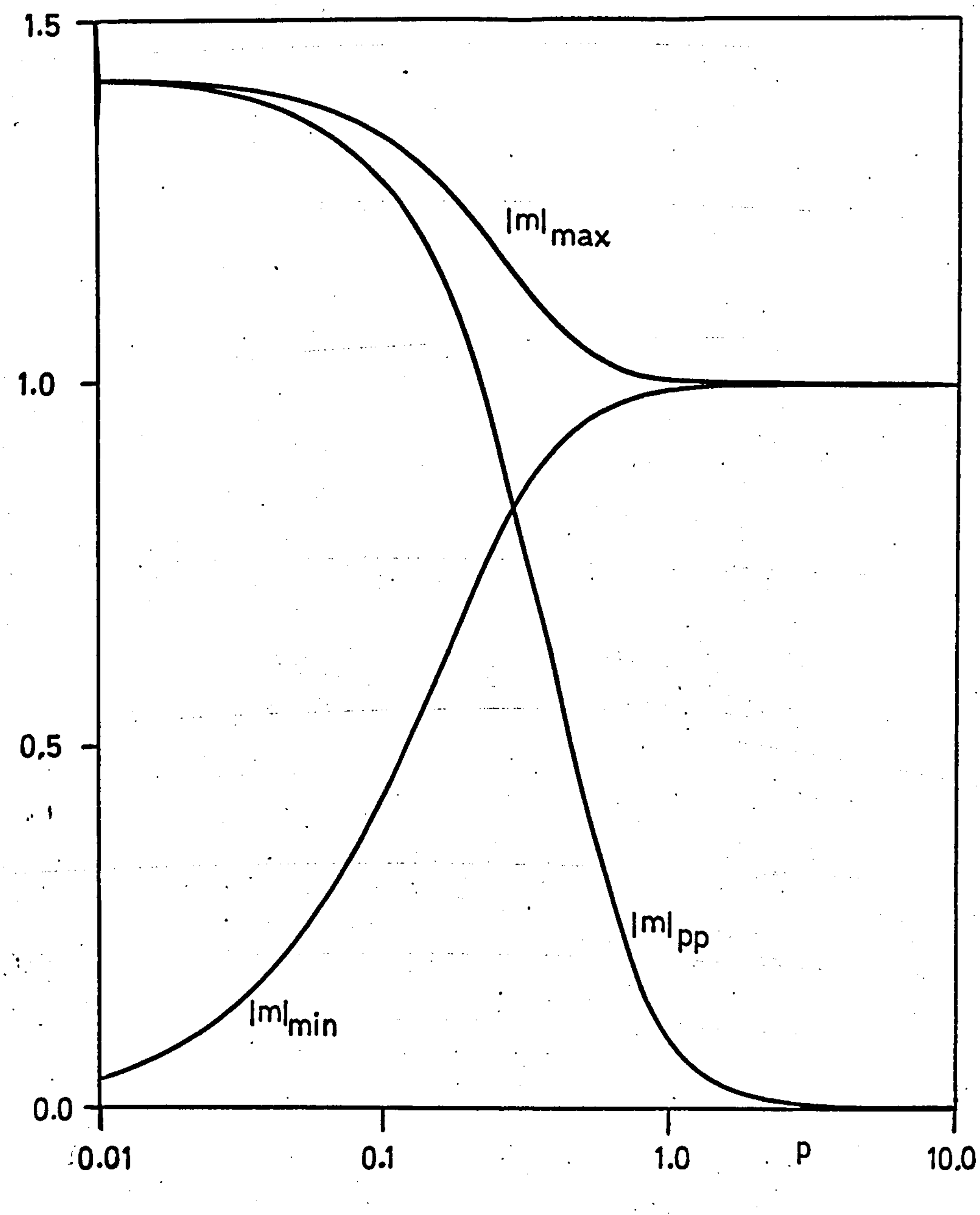


Fig.4.2.1. Variation of envelope properties with zero pattern aspect ratio, p.

Instantaneous Frequency

The instantaneous frequency of an angle coded signal is of central importance in these studies partly because, as mentioned above, an angle coded signal of large p can be approximately represented as purely FM and also because the instantaneous frequency of a signal can be expressed in terms of its zeros in a simple way. As Voelcker has shown[4], the instantaneous phase derivative can be expressed*

$$\dot{\phi}(t) = \sum_{n=-\infty}^{\infty} \frac{\sigma_n}{\sigma_n^{\,2} + (t-\tau_n)^2} + K \text{ (radians/sec)} \qquad 4.2.12$$

where $z_n (= \tau_n + j\sigma_n)$ is the location of the nth zero of the signal. In the case of an angle coded signal, where $|\sigma_n| = s$ and $\tau_n = nT$, the instantaneous phase derivative consists of a series of pulses of identical shape:

$$\dot{\phi}(t) = K + \sum_{n=-\infty}^{\infty} \alpha_n \; V(t-nT) \qquad 4.2.13$$

where $\alpha_n = \text{sgn}\ \sigma_n$

and where

$$V(t) = \frac{|\sigma_n|}{\sigma_n^{\,2} + t^2} \qquad 4.2.14$$

The constant K can be evaluated most simply by physical reasoning. In the case of angle coded signals lying in the frequency range (O,W) the average rate of change of phase (i.e. the mean of $\dot{\phi}(t)$) of the minimum phase signal is zero. For this

$$K = \pi/T \, . \qquad 4.2.15.$$

$\dot{\phi}(t)$ can be pictured as being the output of a linear filter whose impulse response is V(t) and whose input consists of a sequence of positive and negative unit impulse functions.

* The instantaneous phase derivative is 2π times the instantaneous frequency.

That is

$$\dot{\phi}(t) = V(t) * \sum_{n=-\infty}^{\infty} \alpha_n \, \delta(t-n\tau) + K \qquad 4.2.16$$

where * denotes the operation of conv olution. This is illustrated in fig. 4.2.2. Regarding the instantaneous frequency function as being the output of a linear filter is conceptually helpful. It suggests how the instantaneous frequency function can be synthesised electronically as a preliminary stage in the generation of angle coded signals. It also makes calculation of such properties of the instantaneous frequency function as its power spectrum and its mean square conceptually simple - at least to an electrical engineer.

The impulse response of the equivalent linear system is shown in fig. 4.2.3., plotted for various values of p. When p is small, the width of the impulse response (defined in some appropriate way) is small compared with the Nyquist interval, T, and its maximum magnitude is very large. As p tends to zero, V(t) becomes an impulse function of strength π. The physical reason for this behaviour of $\dot{\phi}(t)$ can be seen by considering the vector diagram of the min.phase angle coded signal (for example) as the value of p is reduced (fig. 4.2.4.) p approaches zero, the rate of change of phase as the signal goes through its minimum magnitude becomes more and more rapid; when the zeros are real (p=o) it makes a jump in phase of π radians at each envelope minimum. If p is increased the width of V(t) becomes larger (and its maximum value is reduced). When p has a value of about 0.6, the width of V(t) is comparable with T, the zero spacing in the direction of the t-axis. When p is small, UHP and LHP zeros can be distinguished by observation of the peaks and dips of $\dot{\phi}(t)$

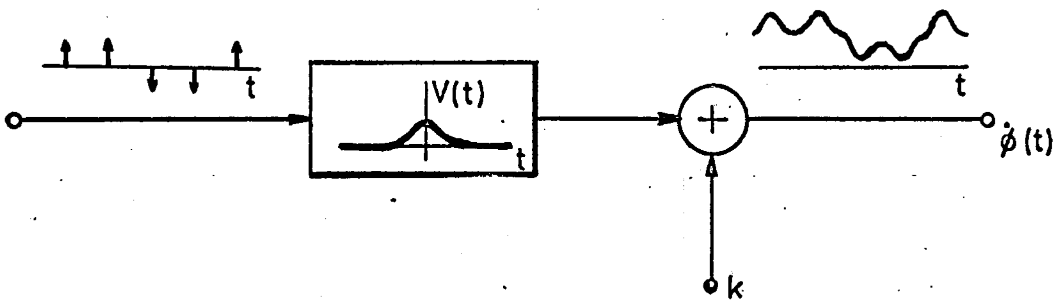


Fig.4.2.2. Instantaneous phase derivative generated as the response of a linear filter to an impulse train.

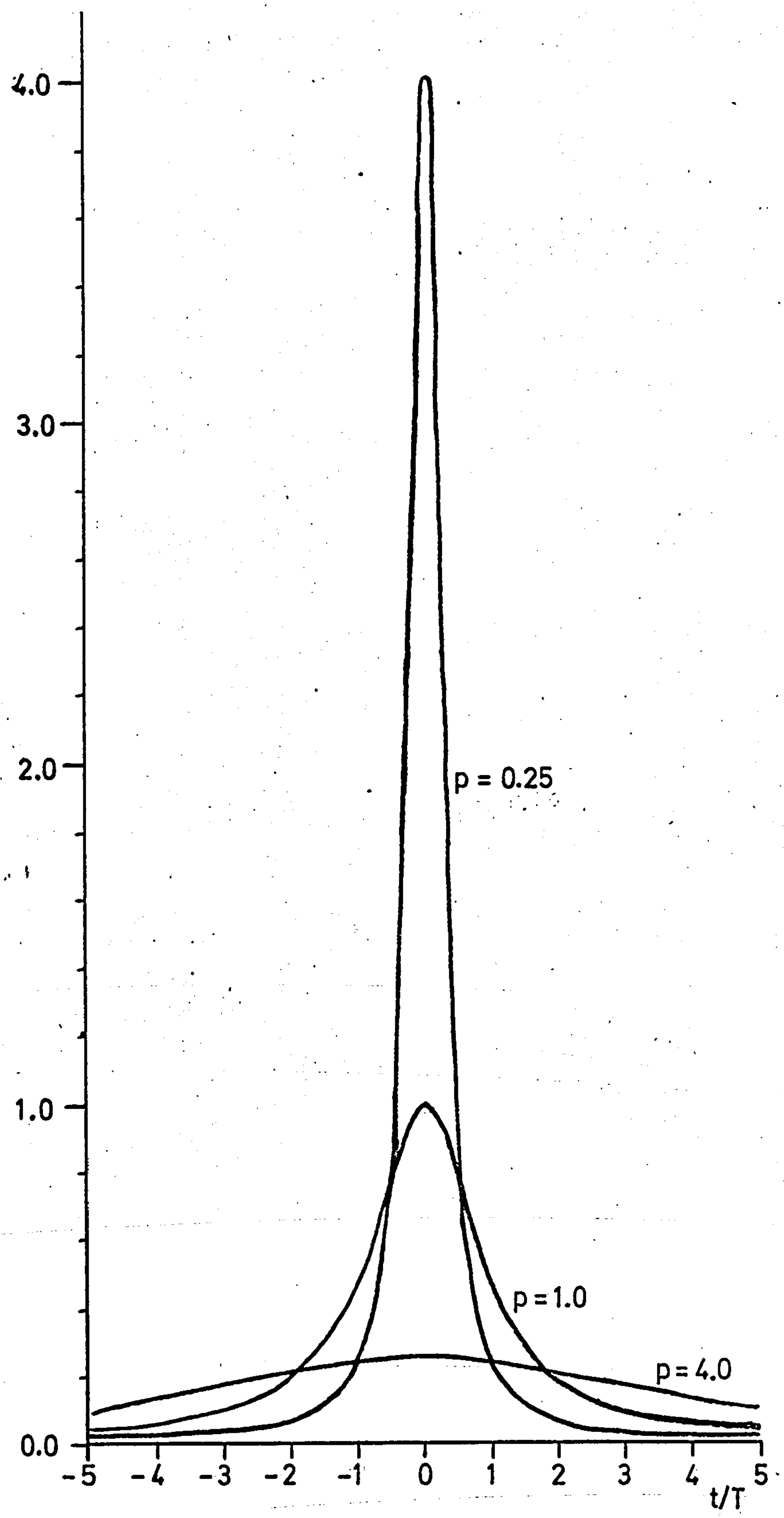


Fig.4.2.3. V(t) plotted for various values of p.

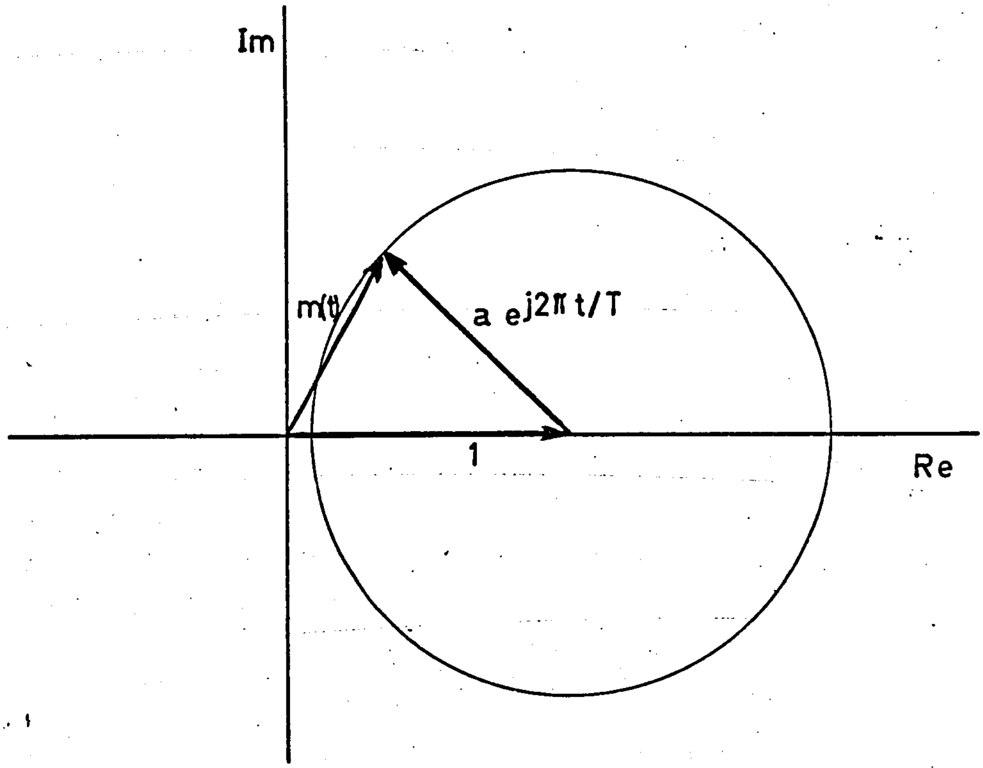


Fig.4.2.4. Vector diagram of the analytic signal
$m(t) = 1 + a e^{j2\pi t/T}$.

but as p is made larger over-lapping of the individual terms of the series of equation 4.2.14 occurs. This is discussed in detail later: Note here however, that the effect can be pictured as intersymbol interference of a sequence of impulses spaced at intervals of T is transmitted through a linear channel whose impulse response is V(t).

Spectral Moments

The normalised moments of the spectrum of a signal[16,17] can be used to provide measures of the centre frequency, the bandwidth, the skewness, etc. of the spectrum. For a finite energy signal, the first moment of the energy density spectrum is defined

$$f_1 = \int_{-\infty}^{\infty} |M(f)|^2 \, f \, df \Big/ \int_{-\infty}^{\infty} |M(f)|^2 \, df \qquad 4.2.17$$

This expression can be manipulated into the form

$$f_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |m(t)|^2 \quad \dot{\phi}(t) \, dt \Big/ \int_{-\infty}^{\infty} |m(t)|^2 dt, \qquad 4.2.18$$

which, for present purposes is a more useful form. For a signal of finite power, the first moment can be defined by the limit

$$f_1 = \lim_{T\to\infty} \quad \frac{1}{2\pi} \int_{-T}^{T} |m(t)|^2 \quad \dot{\phi}(t) \, dt \Big/ \int_{-T}^{T} |m(t)|^2 \, dt \qquad 4.2.19$$

For a periodic signal, this definition is equivalent to

$$f_1 = \frac{1}{2\pi} \int_{0}^{T_p} |m(t)|^2 \quad \dot{\phi}(t) \, dt \Big/ \int_{0}^{T_p} |m(t)|^2 \, dt \qquad 4.2.20$$

where T_p is the period of the signal.

For an angle coded signal, $|m(t)|^2$ is given by equation 4.2.8 and $\dot{\phi}(t)$ by the series of equation 4.2.12. Substituting these expressions for $|m(t)|^2$ and $\dot{\phi}(t)$ in equation 4.2.20

$$f_1 = K/2\pi + \lim_{M\to\infty} \frac{1}{2\pi} \sum_{n=-M}^{M}{}' a_n / \int_{-MT}^{MT} p\left[1+A^2-2A\cos(2\pi t/T)\right]/(1+A^2)dt$$

$$= K/2\pi + \lim_{M\to\infty} \frac{1}{2\pi} \sum_{n=-M}^{M} a_n / 2MTp$$

where $a_n = \int_{-\infty}^{\infty} \frac{\sigma_n}{\sigma_n^2+t^2} p(1+A^2-2A\cos(2\pi t/T)/(1+A^2)\,dt$

$$= p\pi \left[1 - e^{-2\pi s/T}\right] \operatorname{sgn} \sigma_n$$

Thus,

$$f_1 = W/2 + \frac{W}{2}\left[1 - e^{-2\pi p}\right](n_u - n_L) \qquad 4.2.21$$

where n_u is the ratio of the arrange number of UHP zeros to the total number of zeros per unit time and n_L is the corresponding measure of the number of LHP zeros.

Fig. 4.2.4.1.shows f_1/W plotted as a function of p, in the cases of minimum phase and maximum phase signals. When p is very small, f_1 is hardly affected by whether the zeros are UHP or LHP as,in this case, the magnitudes of the frequency components are nearly equal. When p is very large, one frequency component predominates, lying at either extreme of the frequency band according to whether the signal is minimum or maximum phase.

The second moment of the energy density spectrum of a finite energy signal can be defined by

$$f_2 = \int_{-\infty}^{\infty} f^2 \,|M(f)|^2\, df \Big/ \int_{-\infty}^{\infty} |M(f)|^2\, df \qquad 4.2.22$$

$$= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} \left(\frac{d}{dt}|m(t)|\right)^2 + \dot{\phi}(t)^2\,|m(t)|^2\,dt \Big/ \int_{-\infty}^{\infty} |m(t)|^2\,dt \qquad 4.2.23$$

In the case of a finite power signal f_2 can be defined by

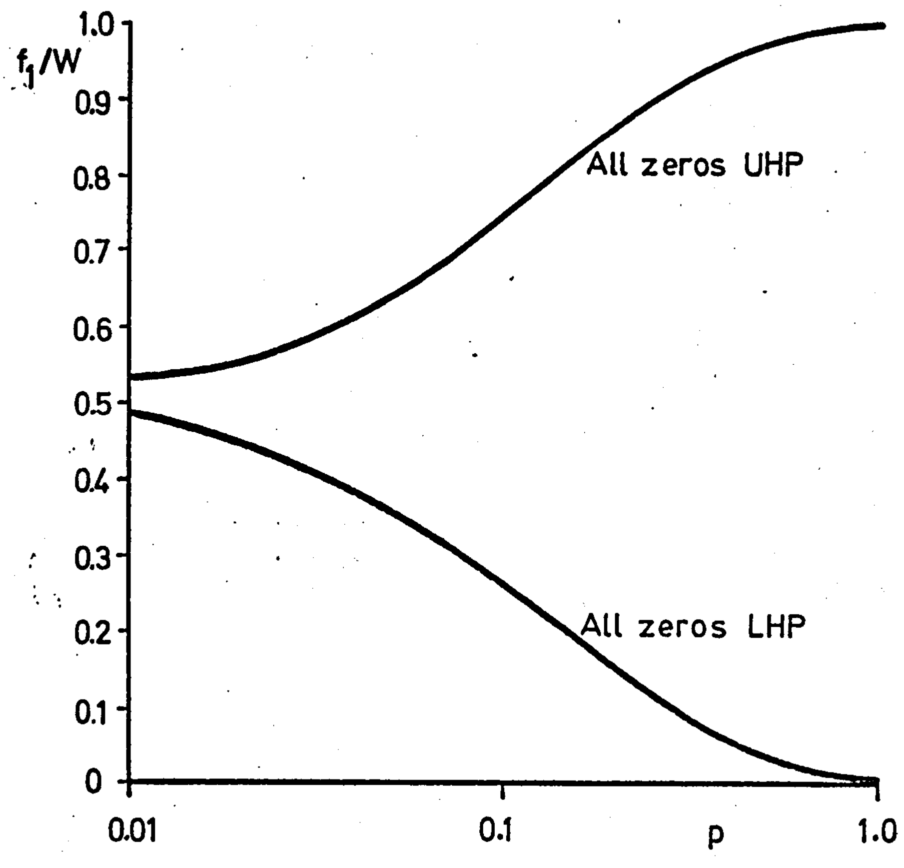


Fig.4.2.4.1. Normalised first spectral moment of minimum and maximum phase angle coded signals as a function of p, the zero pattern aspect ratio.

$$f_2 = \lim_{T\to\infty} \frac{1}{4\pi^2} \int_{-T}^{T} (\frac{d}{dt}|m(t)|)^2 + \dot{\phi}(t)^2 \; |m(t)|^2 \; dt \Big/ \int_{-T}^{T} |m(t)|^2 \; dt \qquad 4.2.24$$

f_2 has two components; one involves the envelope alone of the signal and is thus the same for all members of the common envelope set. The other component involves the instantaneous frequency function and is thus generally different for the various members of the common envelope set. It is not possible to express f_2 exactly in terms of the zeros of the signal in a simple way (as could be done for f_1). The reason for this is that f_2 is not a <u>linear</u> functional of $\dot{\phi}(t)$ - unlike f_1. However, when p is large (that is, greater than about 2) the envelope fluctuations become very small and equation 4.2.24 can be approximated by

$$f_2 \simeq \lim_{T\to\infty} \frac{1}{(2T)(4\pi^2)} \int_{-T}^{T} \dot{\phi}^2(t) \; dt \qquad 4.2.25$$

Thus f_2 (for large p) is approximately given by the mean square of $\frac{1}{2\pi} \; \dot{\phi}(t)$, which can be calculated in a straightforward way if the relevant statistics of the zero locations are known.

If the probability that a particular zero lies in the upper half plane is 0.5, and is independent of the locations of the other zeros of the angle coded signal (this is a reasonable assumption if binary data is to be transmitted with an UHP zero representing binary '1' and a LHP zero representing binary '0') then $\dot{\phi}(t)$ is a train of uncorrelated V(t) pulses. Thus

$$\overline{\dot{\phi}(t)^2} = E/T + k^2$$

where E is the 'energy' of one pulse given by the expression

$$E = \int_{-\infty}^{\infty} V(t)^2 \; dt$$

$$= \pi/2s \quad sec^{-1}$$

Thus, for a random angle coded signal (as one with zeros randomly in the upper and lower half planes will be termed), f_2 is given by

$$f_2 = 1/(16\pi sT) + 1/4T^2 \text{ (cycles/sec)}^2$$
$$= W^2/(16\pi p) + W^2/4 \qquad 4.2.26.$$

The 'standard deviation', f_{SD}, of the spectrum about its first moment or "centroid" can be used as a measure of the bandwidth of a signal. For an angle coded signal (with p large) this is, from equations 4.2.21 and 4.2.26,

$$f_{SD} = \sqrt{f_2 - f_1^2}$$
$$= 1/(4\pi\sqrt{sT})$$
$$= W/(4\pi\sqrt{p}) \qquad 4.2.27.$$

The <u>effective</u> bandwidth of a random angle coded signal (as measured by the standard deviation of the spectrum) is thus a fraction of the 'strict' bandwidth W (outside which the spectrum is zero). As p is increased and the zeros move away from the real axis in the complex plane, the effective bandwidth of a large-p angle coded signal decreases. This effect is mentioned again in the next section in which the power spectra of random angle coded signals are discussed.

4.3 Spectra of Random Angle Coded Signals - Theoretical

In the statistical approach to communication system analysis, information carrying signals are represented as being sample functions from a random process. The power spectrum of the random process is of interest for perhaps two principal reasons. It enables the effect on the signal (on average) of linear filtering operations to be calculated. It also provides an indication of the effectiveness of the signal process for communicating in the presence of noise. Shannon[19] derived the ideal form of power spectrum for the signal process for signalling in the presence of additive Gaussian noise of a given spectral form : by comparing the spectra of different signal processes (for example, the angle coded signal process) with the ideal, their relative effectiveness can be gauged.

A <u>continuing random angle coded signal process</u> ('continuing process', for short) is a random process whose sample functions are angle coded signals (all of equal power, p) whose zeros occur at regular instants separated by T in the direction of the t-axis and randomly +js or -js in the direction of the σ-axis of the complex time plane. A CP is not a stationary process; for example, its ensemble-arranged magnitude (which is simply the envelope) depends explicitly on time according to equation 4.2.8

The autocorrelation function $R(t,\tau)$, of the continuing process is given by

$$R\,(t,\tau) \;=\; E\left[m(t)\; m^*(t+\tau)\right] \qquad 4.3.1$$

By considering the nature of the zero pattern it can be seen that the autocorrelation function must be periodic in t, with period T. That is

$$R\,(t,\tau) \;=\; R\,(t+nT,\tau) \quad \text{for} \quad n = 0, \pm 1, \pm 2, \ldots$$

As the autocorrelation function of the CP is time dependent, the power spectrum is, too. However, for many purposes it is sufficient to consider the time average[20], $S(f)$, of the power spectrum, $S(t,f)$.

$$S(f) = \frac{1}{T}\int_0^T S\,(t,f)\; dt \qquad 4.3.2.$$

where S(t,f) is the Fourier transform of R(t,τ)

$$S(t,f) = \int_{-\infty}^{\infty} R(t,\tau)\, e^{-j2\pi f\tau}\, d\tau \qquad 4.3.3$$

For purposes of calculating the power passed by filters and so on, the time averaged power spectrum can be treated like the power spectrum of a stationary random process. It is the time averaged power spectrum that is treated in this chapter.

In the experimental estimation of spectra described later in this chapter, the idea of a <u>periodic random angle coded signal process</u> (periodic process, in short) is used. This is defined to be a random process whose sample functions are periodic angle coded signals (all of equal power, P) having N zeros per period. The zeros in the fundamental period of each sample function are chosen randomly so that each zero lies in either the UHP or the LHP with equal probability. The autocorrelation function, Rp(t,τ), of the PP, in addition to being periodic in t with period T, is also periodic in τ with period NT:

$$Rp(t,\tau) = R_p(t+mT,\ \tau+nNT),\quad m = 0,\pm 1,\pm 2,\pm 3,\ldots\ldots$$
$$n = 0,\pm 1,\pm 2,\pm 3,\ldots\ldots \qquad 4.3.4$$

Here it is the time averaged autocorrelation function (referred to simply as the autocorrelation function from here on), defined by

$$R_p(\tau) = \frac{1}{T}\int_0^T R_p(t,\tau)\, dt, \qquad 4.3.5$$

which is of interest.

If $m_p(t)$ is a sample function from the periodic process, its Fourier series expansion[21] is

$$m_p(t) = \sum_{n=0}^{N} a_n\, e^{j2\pi nt/NT} \qquad 4.3.6$$

(The series contains only N+1 terms, as the sample functions are band limited to the frequency range (0,1/T)).

The Fourier coefficients a_n, given by

$$a_n = \frac{1}{NT} \int_0^{NT} m_p(t)\, e^{-j2\pi nt/NT}\, dt, \qquad 4.3.7$$

are uncorrelated random variables such that

$$\begin{aligned} E\left[a_n\right] &= E\left[m_p(t)\right] \quad , n=o \\ E\left[a_n\right] &= o \qquad\qquad\quad , n\neq o \end{aligned} \qquad 4.3.8$$

and

$$E\left[a_k\, a_n^*\right] = \begin{cases} r_k \;, & k=n \\ o \;, & k\neq n \end{cases} \qquad 4.3.9$$

where the r_k are the coefficients of the Fourier series expansion of $R_p(\tau)$.[21] The autocorrelation function can thus be expressed

$$R_p(\tau) = \sum_{k=o}^{N} r_k\, e^{j2\pi k\tau/NT} \qquad 4.3.10$$

The time averaged power spectrum, the Fourier transform of $R_p(\tau)$, is thus

$$S_p(f) = \sum_{k=o}^{N} r_k\, \delta(f-k/NT) \qquad 4.3.11$$

If the period, NT, of the periodic process is great enough it may be assumed that any signal drawn from the periodic process will have similar statistical properties (within its fundamental period) to those of a signal from the continuing process with the same characterising parameters (power, bandwidth and 'aspect ratio', p). Then the autocorrelation function, $R_p(\tau)$, of the periodic process will provide (within its fundamental period) a good approximation to that of the continuing process, $R(\tau)$ (fig.4.3.1.)

$$R(\tau) \approx R_p(\tau)\ \mathrm{rect}(\tau/NT)$$

Taking the Fourier transform of this approximate equation gives

$$S(f) \approx S_p(f) * NT\ \mathrm{sinc}\ (fNT) \qquad 4.3.12$$

Thus the spectrum of the continuing process is given approximately by interpolating the line spectrum of the periodic process with the sinc function of equation 4.3.12 or indeed any reasonable interpolating function of similar width and height.

The remainder of this section derives theoretical expressions for the power spectrum of the continuing process which become valid in the extreme cases of very large and very small aspect ratio p.

<u>Case (i) - Small p</u>

Instead of studying the analytic angle coded signal process, it is expedient here to consider its real part. There is no problem, of course, in finding the spectrum of an analytic process if the spectrum of its real part is known. Fig.4.3.2. shows a section of the waveform of the real part of an angle coded signal whose zeros lie close to the t-axis (i.e., p is small). This figure also displays the real part of the analytic signal

$$m_1(t) = a\left[1 + e^{j2\pi t/T}\right], \qquad 4.3.13$$

where a is a real number such that the two signals are of equal power). The two waveforms are very similar in form, except that the real part of the angle coded signal dips below the t-axis in the vicinity of an UHP zero and passes above it near a LHP zero.

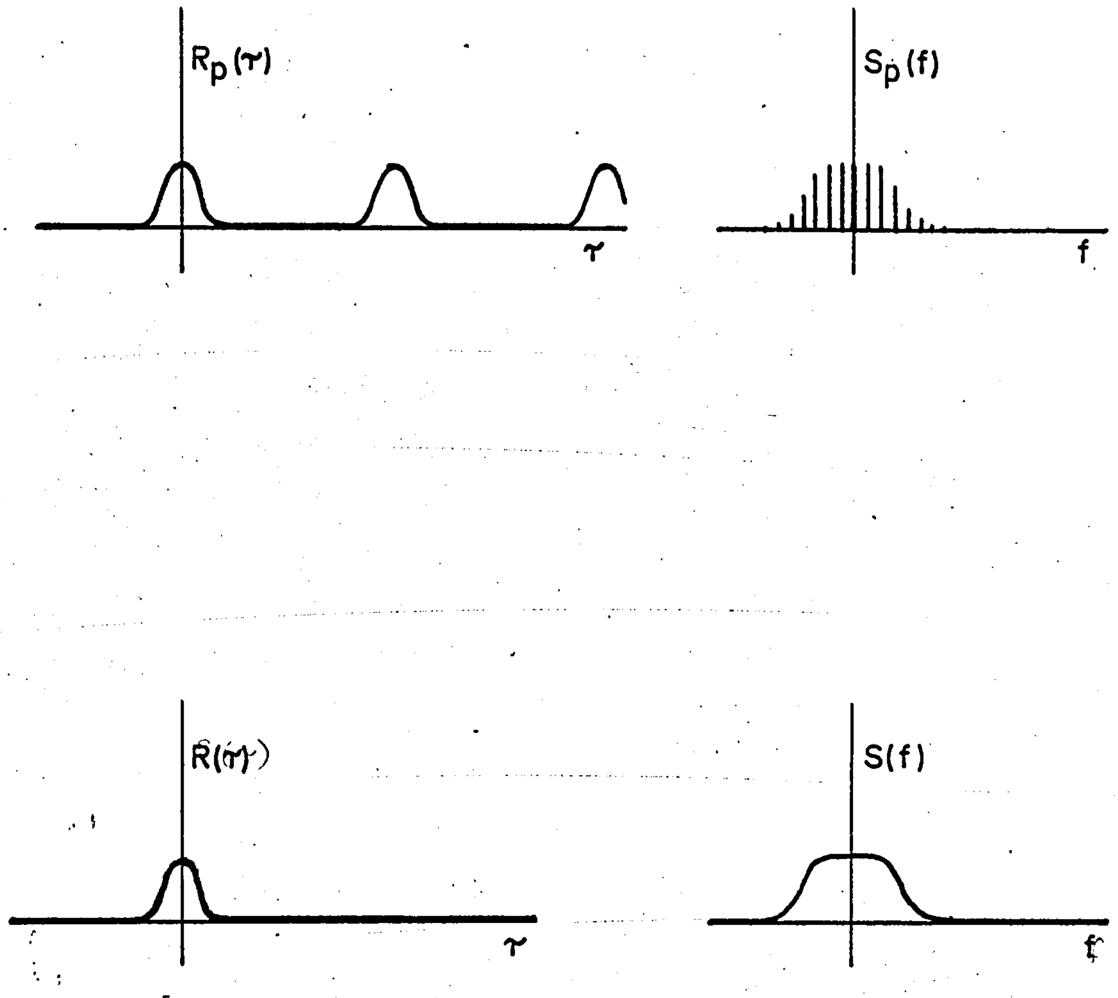


Fig.4.3.1. Autocorrelation functions and power spectra of periodic and continuing processes.

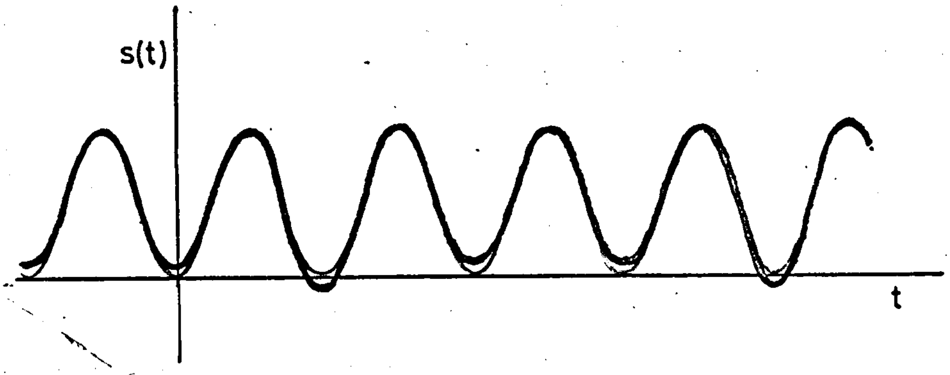


Fig.4.3.2. Waveform of the real part of an angle coded signal with its zeros close to the real axis.

Let the real part of the angle coded signal (which is a sample function from the continuing process) be represented by

$$s(t) = \frac{1}{\sqrt{2}} \left[1 + \cos(2\pi t/T) \right] + d(t) \qquad 4.3.14$$

where the analytic signal has been normalised to have unit mean square, so that in equation 4.3.13, a has the value $1/\sqrt{2}$.

s(t) is bandlimited to (-W,W) and therefore d(t) must be too, since the other term on the RHS of equation 4.3.2 is also bandlimited to (-W,W). Thus, by the sampling theorem, d(t) can be represented by the series

$$d(t) = \sum_{m=-\infty}^{\infty} d(m/2W)\ \mathrm{sinc}\ (2Wt-m) \qquad 4.3.15$$

Now at the t-coordinate of the zeros, $t = nT$, $n = 0; \pm 1, \pm 2, \ldots$ the expression $(1+\cos 2\pi t/T)$ is zero and so at these points

$$d(nT) = s(nT), \qquad n=0,\pm 1,\pm 2,\ldots \quad .$$

The instantaneous phase of the signal makes a sudden increase of π radians in the vicinity of each UHP zero and a sudden decrease of π in the vicinity of each LHP zero. The phase change is sudden because of the impulse-like character of the instantaneous frequency function for signals of small p. In addition, there is a steady increase of phase at the rate of π radians in T seconds, because the centre frequency of the process lies at W/2. The result of this is that the instantaneous phase of the small p signal at the instants nT $(n=0,\pm 1,\pm 2,\ldots)$ is approximately an integer multiple of π which results in the signal being real at these instants. Thus d(nT) is given by

$$d(nT) = \alpha_n \quad |m|min \qquad n=o,\pm 1,\pm 2,\ldots$$

(where α_n is unity if the nth zero lies in the UHP and -1 if it lies in the LHP - sect.4.2) At the other sampling instants, $t = nT/2$, $n = \pm 1, \pm 3, \pm 5, \ldots$, it is evident that s(t) takes (very nearly, for p small) the same values as the real part of $m_I(t)$ and thus

$$d(nT/2) = o, \qquad n = \pm 1, \pm 3, \pm 5, \ldots$$

Thus equation 4.3.15 becomes, approximately

$$d(t) \simeq \quad |m|min \sum_{n=-\infty}^{\infty} \alpha_n \text{ sinc } (2Wt-2n) \qquad 4.3.16$$

d(t) can be thought of as being the output of an ideal low pass filter whose impulse response is $\text{sinc}(2Wt)$ and which is excited with randomly positive and negative impulses of area $|m|_{min}$ and occurring at regular intervals of T. Thus the power spectrum, $S_d(f)$, of d(t) is that of bandlimited white noise.

$$S_d(f) = W \text{ rect } (f/W) \qquad 4.3.17$$

where w, the spectral density is given by

$$W = |m|^2_{min}/2 \qquad \text{volt}^2 \text{ /Hz} \qquad 4.3.18$$

Thus, from equations 4.2.10, 4.3.14 and 4.3.17 the power density spectrum of an analytic angle coded signal of small p is approximately given by*

*The power spectrum $S_A(f)$ an analytic signal is related to the power spectrum of its real part, $S_R(f)$, by

$$S_A(f) = \begin{cases} 4S_R(f), & f > o \\ S_R(f), & f = o \\ o & f < o \end{cases}$$

$$S(f) \simeq P \left\{ \alpha \left[\delta(f)+\delta(f-W) \right] + \beta \,\mathrm{rect}(f/W-\tfrac{1}{2}) \right\} \qquad 4.3.19$$

when

$$\alpha = 2A/(1+A^2)$$

and

$$\beta = (1-A)^2/(1+A^2)$$

This approximate expression for S(f), valid in the case of extremely small p, is shown graphically in fig.4.3.3. and also later in this chapter when it is compared with experimental estimates of the power spectrum.

Case (ii) - Large p

Section 4.2 showed that the envelope fluctuations of angle coded signals, when p is large, are very small. In fact, for practical purposes they can be regarded as purely frequency modulated signals. The instantaneous frequency of a large-p angle coded signal consists of a series of slowly varying V(t) functions and so must itself be a slowly varying function of time. The "purely FM" character of large p angle coded signals and the slow variations of their instantaneous frequency functions suggests that the quasi stationary method of calculating FM spectra is applicable.

The instantaneous frequency of random angle coded signal can be regarded as being produced by the operation of filtering a sequence of randomly positive and negative impulses with a filter whose impulse response is V(t). When p is large, V(t) is of much larger duration than, T, the interval between impulses and so, at any instant, the instantaneous frequency consists of the sum of many quantities which have similar magnitudes but random signs.

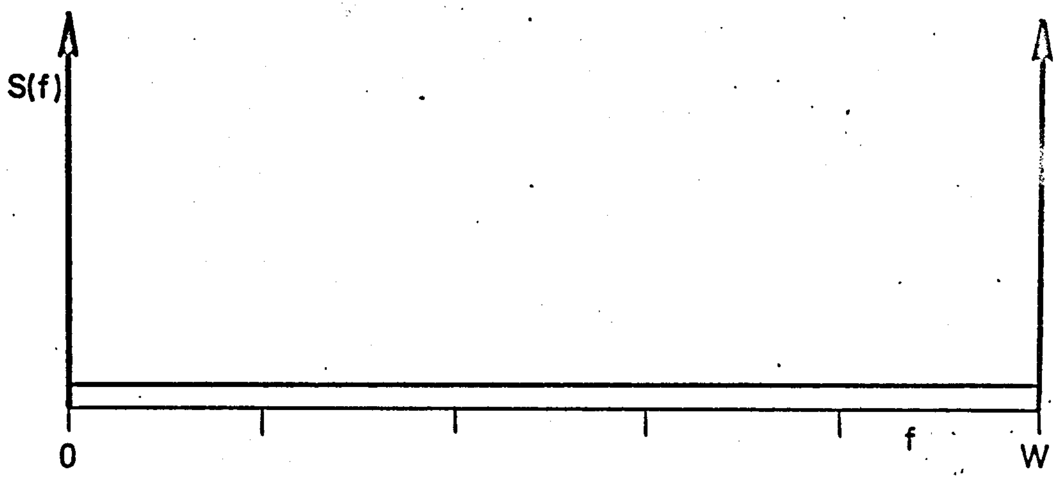


Fig.4.3.3. Approximate spectrum of random angle coded signal (for p small).

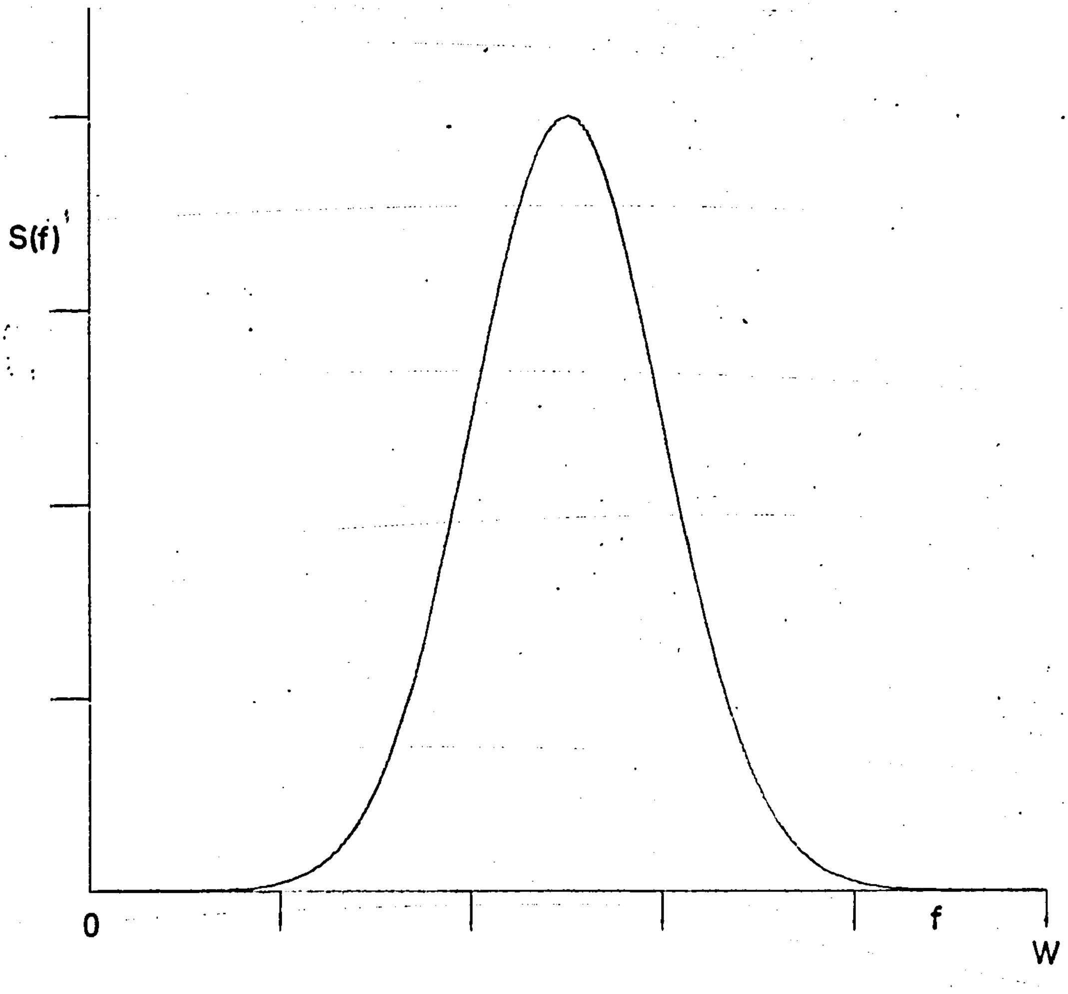


Fig.4.3.4. Approximate spectrum of random angle coded signal (for p large).

Under these conditions the central limit theorem should apply and the instantaneous frequency, $f_i(t)$, should be a random function whose probability distribution $p(f_i)$ is Gaussian. Thus

$$p(f_i) = \frac{1}{f_\sigma^2 \sqrt{2\pi}} e^{-(f_i - f_m)^2/2f_\sigma^2} \qquad 4.3.20$$

The mean, f_m, of the instantaneous frequency for a random angle coded signal lying in the frequency range (0,W) is simply W/2. The variance, f_σ^2, is given by

$$f_\sigma^2 = \frac{W}{(2\pi)^2} \int_{-\infty}^{\infty} V(t)^2 \, dt$$

since the successive V(t) pulses are uncorrelated and occur at the rate of W per second. Thus

$$f_\sigma^2 = W/(8\pi s)$$
$$= W^2/(8\pi p)$$

When the quasi-stationary hypothesis applies, the power of a signal in a frequency band (f_i, f_i+df) is proportional to the proportion of the time that the instantaneous frequency of the signal lies within that band. Thus, the power spectrum, S(f), of a large-p angle coded signal of power P is approximately given by

$$S(f) \approx P/(f_\sigma\sqrt{2\pi}) \; \exp\left[-(f-f_m)^2/2f_\sigma^2\right]$$
$$= (P\,2\sqrt{P}/W) \; \exp\left[-(f-W/2)^2 \; 4\pi p/w^2\right] \qquad 4.3.21.$$

This approximate expression for S(f) is shown graphically in fig.4.3.4. One indication of the approximate nature of this expression is that it incorrectly indicates the spectrum to be non-zero (although very small) at frequencies outside the range (0,W).

4.4 Spectra of Random Angle Coded Signals - Experimental

Computer experiments ("Monte Carlo" calculations) were performed to verify the theory of section 4.3 and to establish the range of validity of the approximations. The procedure used was, briefly, as follows. The power spectrum of a periodic process (sect.4.3) was estimated by repeatedly choosing zero patterns at random, computing Fourier coefficients of the corresponding angle coded signals and averaging the squared magnitude of each Fourier coefficient over the set of generated signals. (An estimate of the standard deviation of each estimate was produced at the same time). The estimated periodic process spectrum was interpolated to provide an estimate of the continuing process spectrum. These estimated spectra are displayed graphically later in this section, together with the theoretically derived spectral forms of the previous section.

The outline of the computer program shown in fig.4.4.1. requires little explanation but the following discusses some particular points.

At stage (1) (fig.4.4.1) the roots Z_n of the Fourier polynomial are chosen to lie at regular angular intervals in the complex plane and randomly on a circle of radius A or A^{-1}. This corresponds to choosing the zeros of the signal to lie at regular intervals T in the direction of the t-axis of the complex time plane and randomly at +s or -s in the direction of the σ-axis. The 'random' choices are made according to the outcome of whether numbers produced by the pseudo random number generating routine UTR1* lie above or below the mean.

*Available at Loughborough University of Technology Computer Centre.

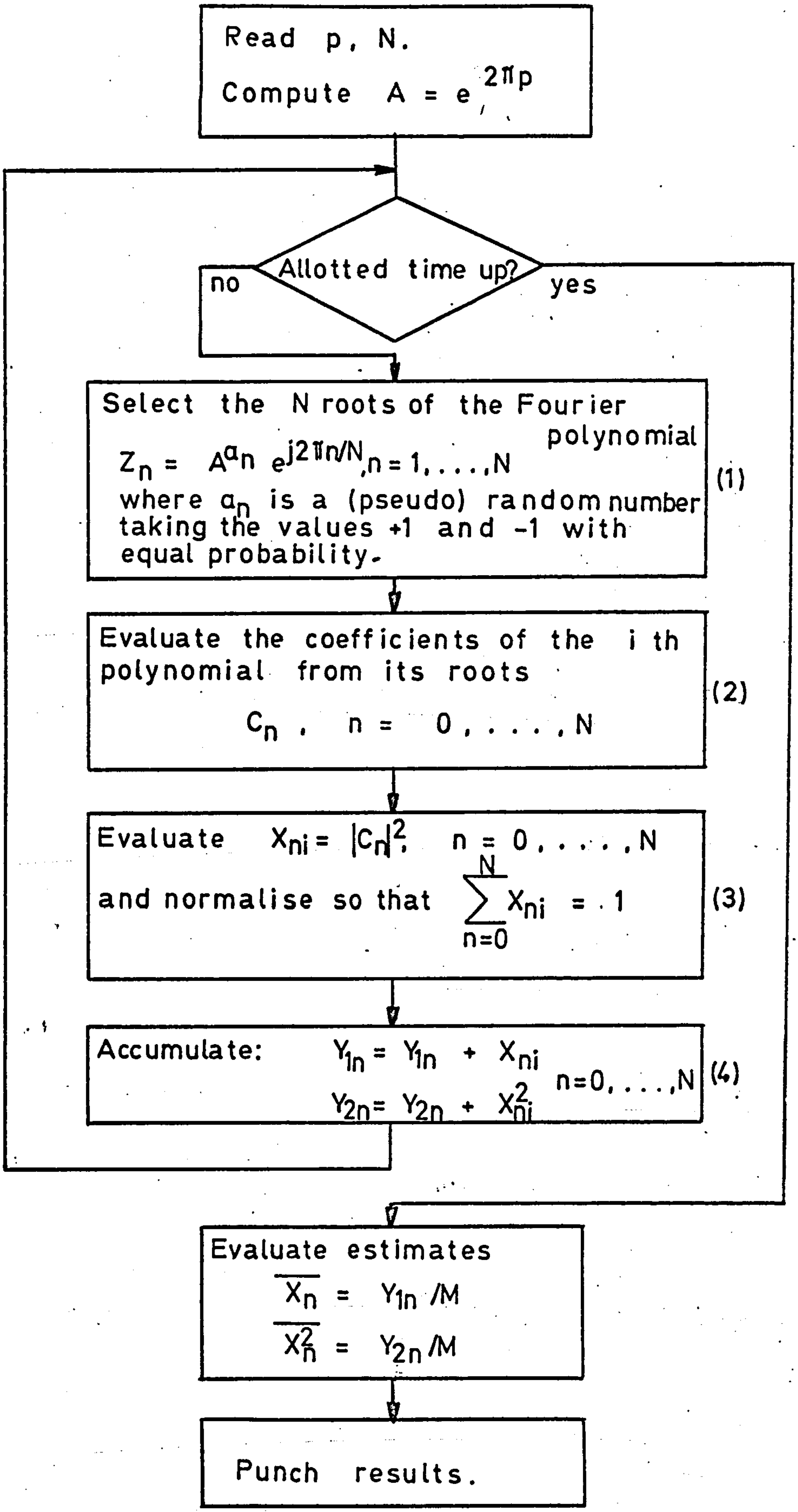


Fig. 4.4.1. Flow diagram of spectrum estimation program.

At stage (2), the standard scheme for computing the coefficients of a polynomial from its roots proved to give totally incorrect results when it was used with polynomials of order greater than about 40, due to the accumulation of round off errors. To overcome this effect, a scheme was devised based on the discrete Fourier transform (DFT) which gave good accuracy with polynomials of order of at least 250. The details of this scheme are given in appendix C.

At stages (3) and (4) the estimate of the power in the nth spectral line $\overline{X}_n$ was computed as

$$\overline{X}_n = \frac{1}{M} \sum_{i=1}^{M} X_{ni} ,$$

where X_{ni} is the squared magnitude of the nth Fourier coefficient of the ith randomly generated Fourier polynomial. An estimate of the standard deviation of this estimate was taken as $\overline{S}_n$, defined by the expression

$$(\overline{S}_n)^2 = \frac{1}{M} \left[\overline{X_n^2} - (\overline{X}_n)^2 \right]$$

where $\overline{X_n^2}$ is computed according to the expression

$$\overline{X_n^2} = \frac{1}{M} \sum_{n=1}^{M} X_{ni}^2.$$

Dividing the estimates of the weights of the spectral lines of the periodic process by N+1 gives estimates of the values of the spectrum of the continuing process at the sample points.

These sample values were interpolated by straight lines in plotting the estimated spectra. A vertical line of length representing twice the estimated standard deviation of each estimated point is drawn with its centre on the point.

The results are displayed in figs. 4.4.2. - 4.4.10. The theoretical form of the spectrum is also shown on each figure; the large p form the plots for the spectra of signals with p greater than 0.5 and the small p form for the spectra of signals with p less than 0.5.

The experimental results show good agreement with the theoretical results. The agreement is quite surprisingly close in the case of signals of only moderately large values of p.

In most of the experiments the signals which were studied had 63 zeros per period. To check that this was a large enough number, the spectra of signals with 127 zeros per period were also estimated (for the large p cases of p=4.0 and p=8.0). By comparing figures 4.4.7. and 4.4.8. with 4.4.9. and 4.4.10. it can be concluded that 63 zeros per period is a sufficiently large number for the periodic process to provide a good model of the continuing process, except when p is very large (greater than about 4).

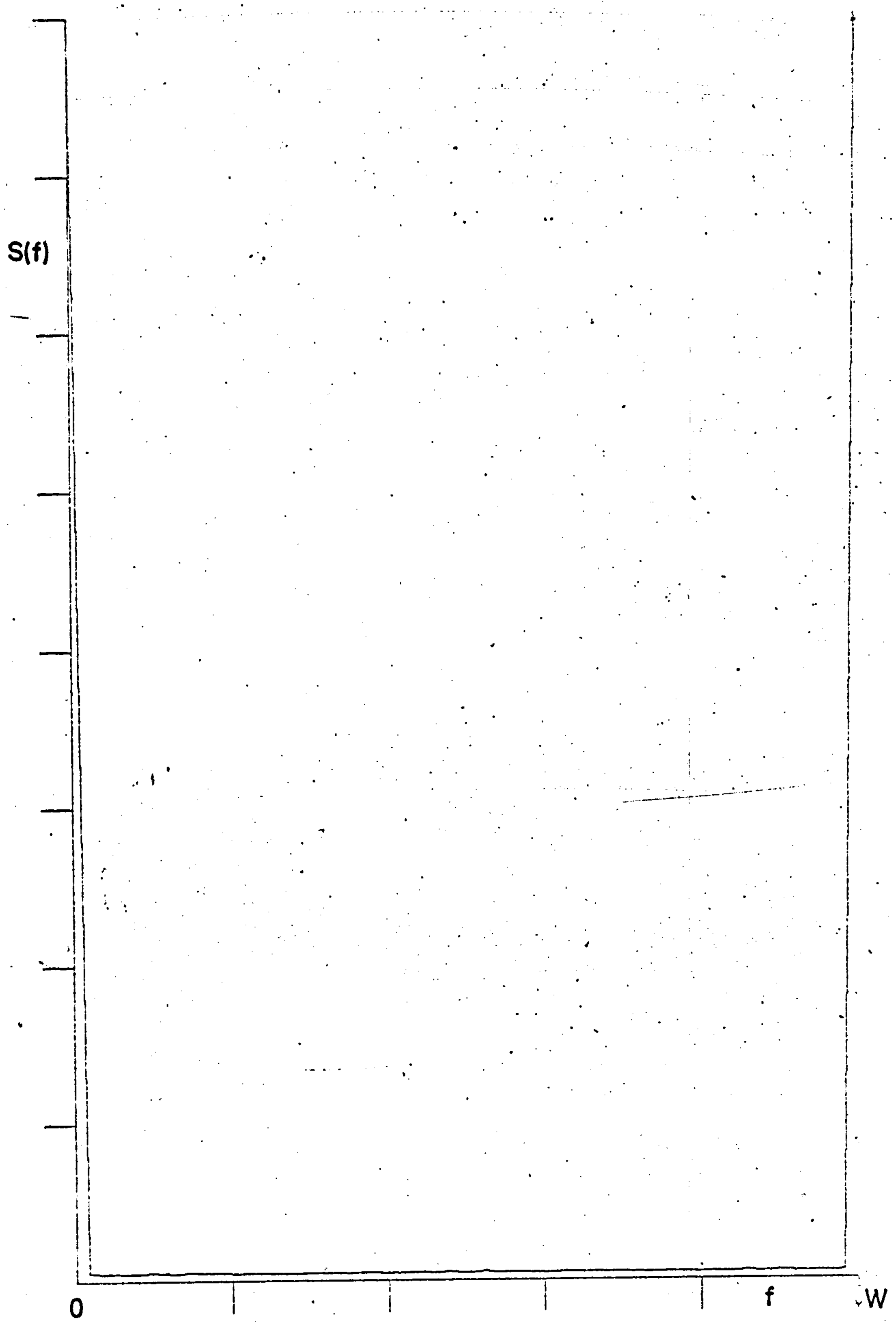


Fig.4.4.2. Estimated spectrum for p = 0.05. Theoretical small - p spectrum differs imperceptibly and so is not shown.

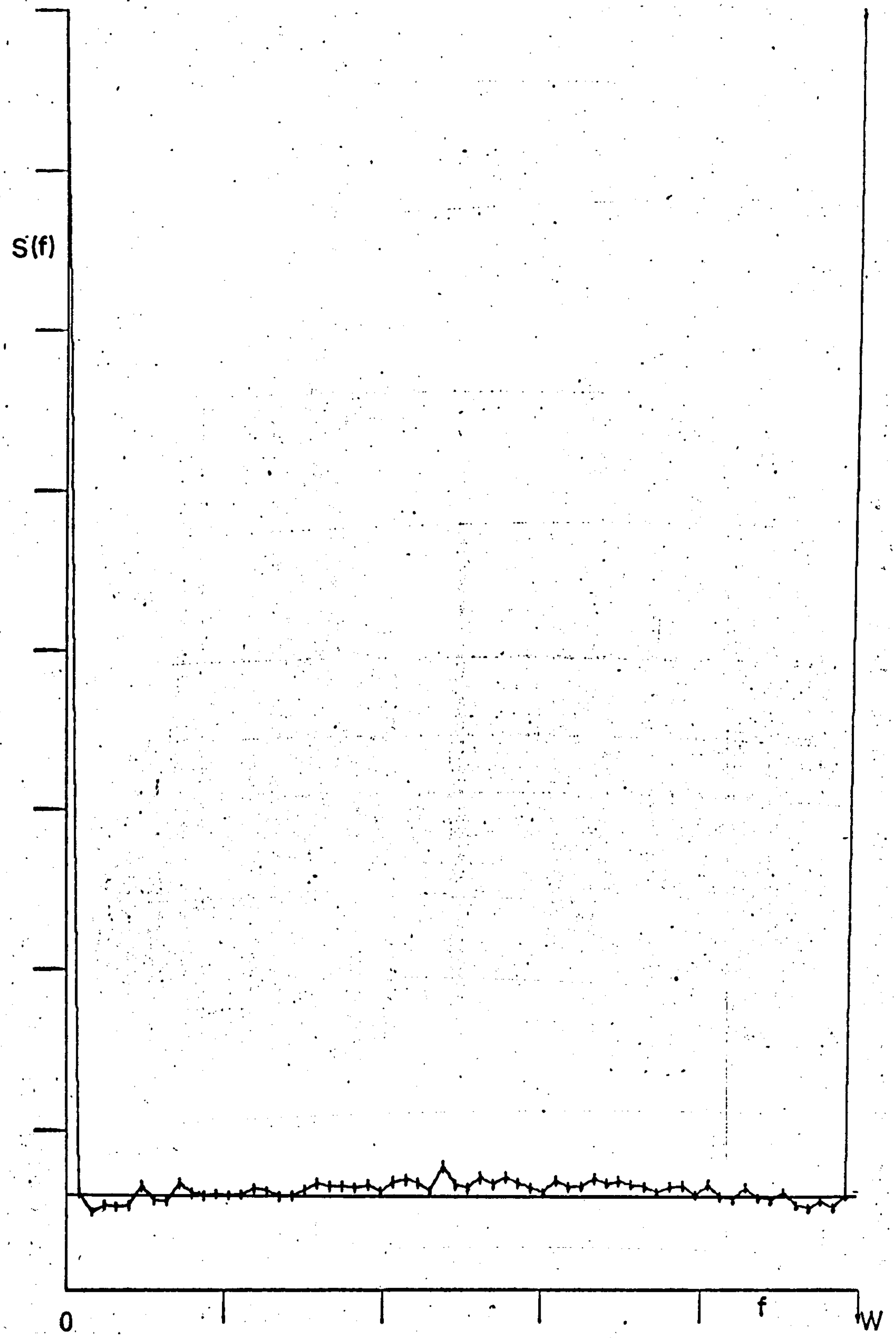


Fig. 4.4.3. Estimated and theoretical small-p spectra for p = 0.25.

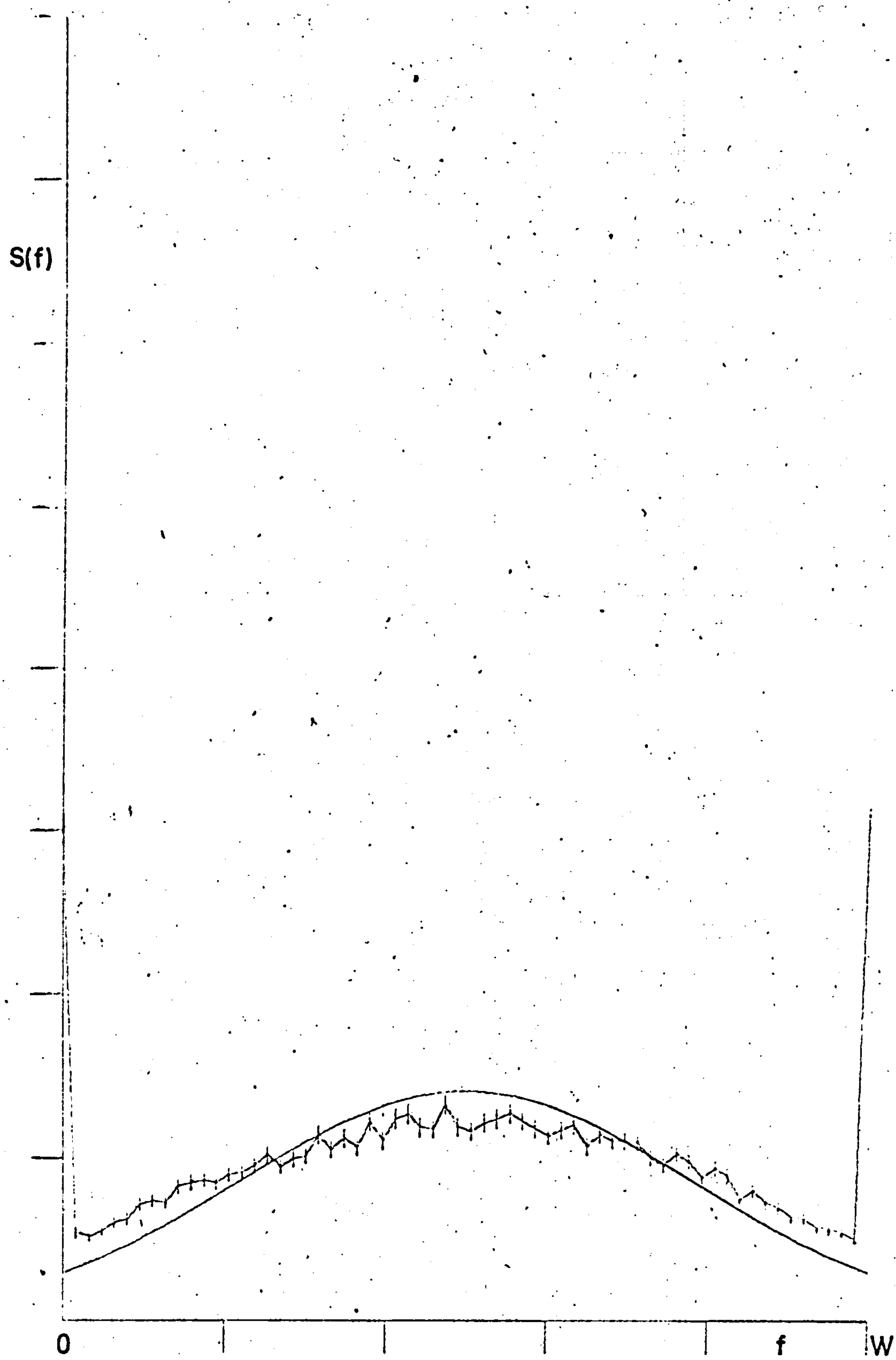


Fig.4.4.4. Estimated and theoretical large-p spectra for p = 0.5.

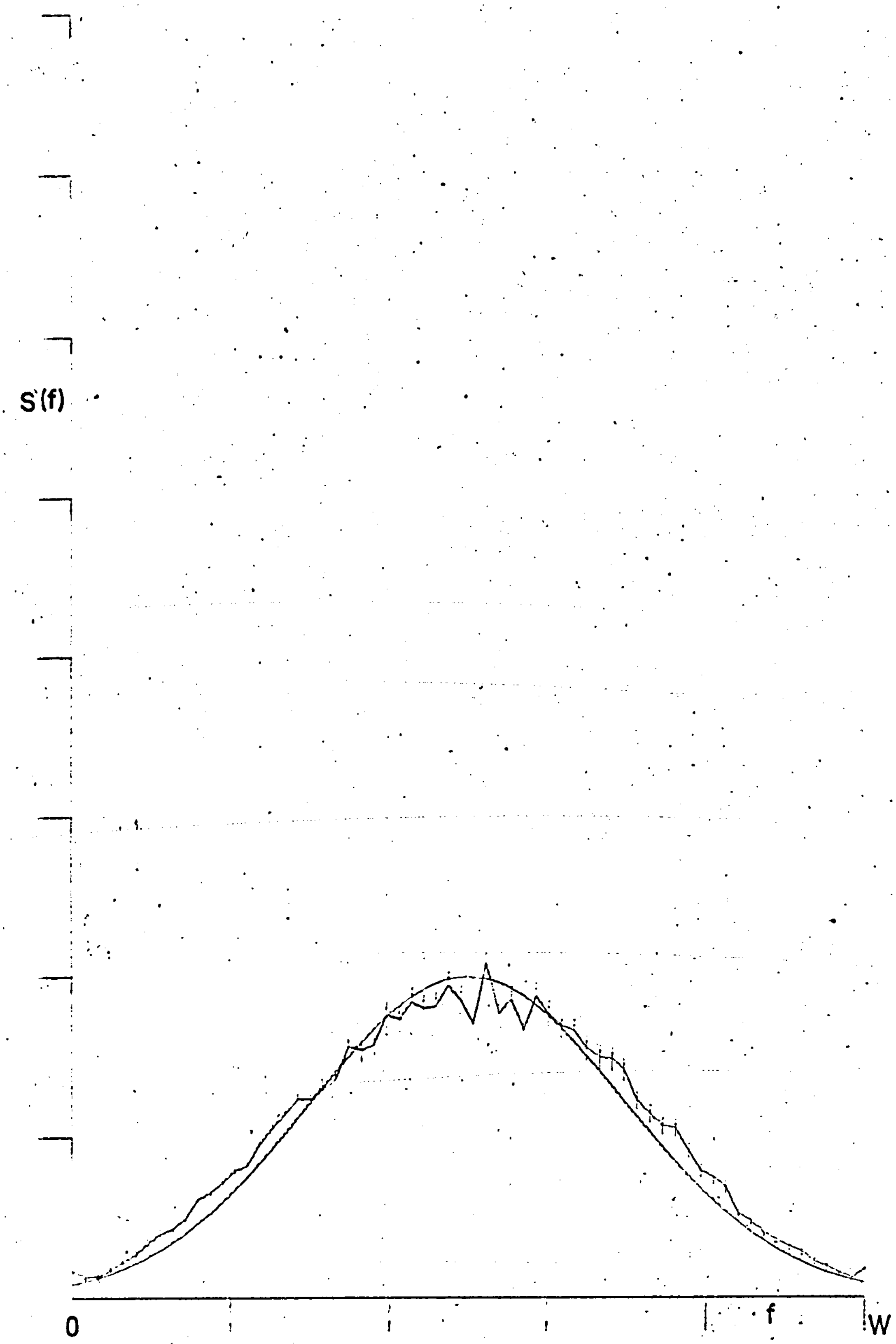


Fig. 4.4.5. Estimated and theoretical large-p spectra for p = 1.0.

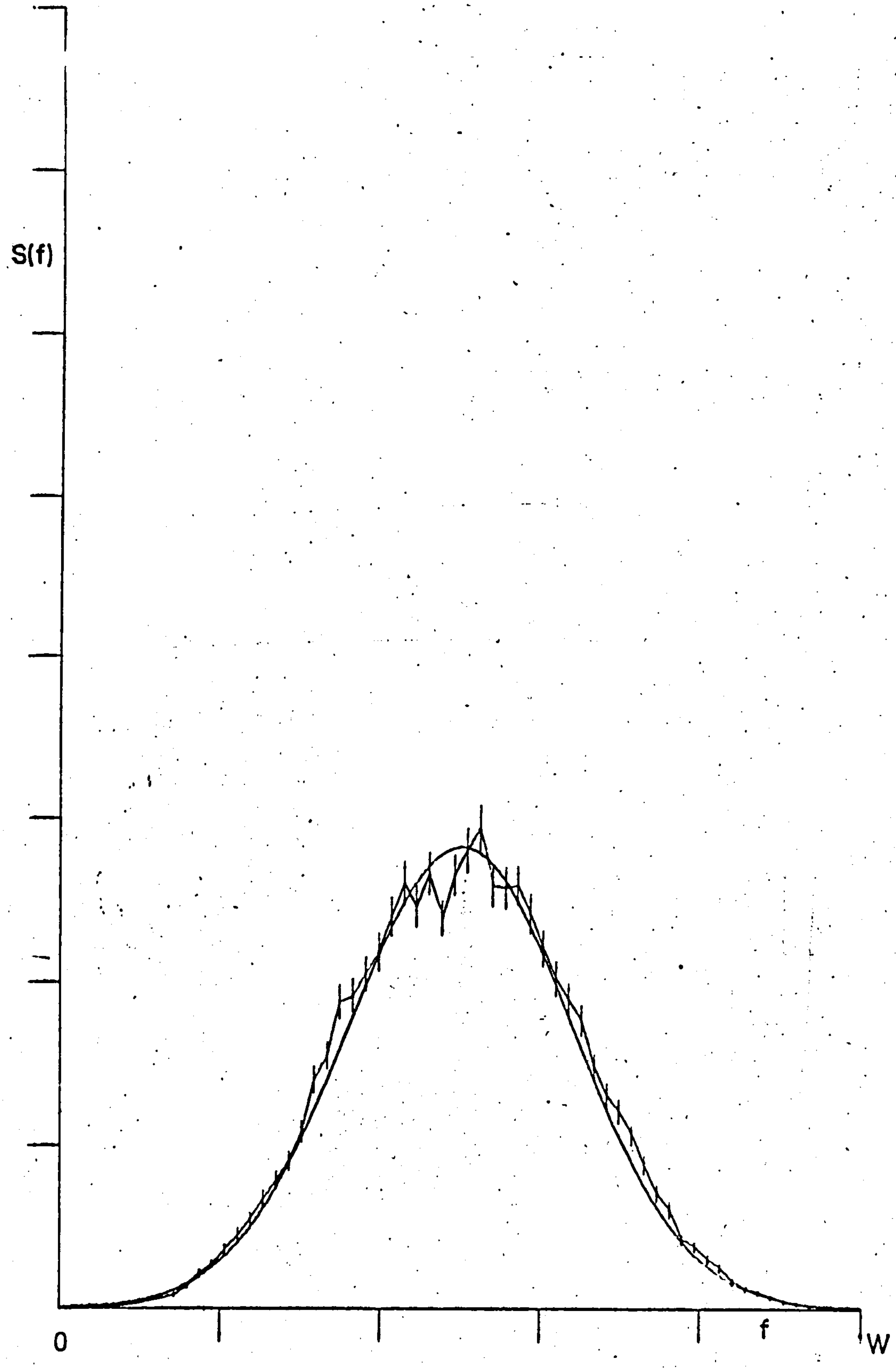


Fig. 4.4.6. Estimated and theoretical large-p spectra for p = 2.0.

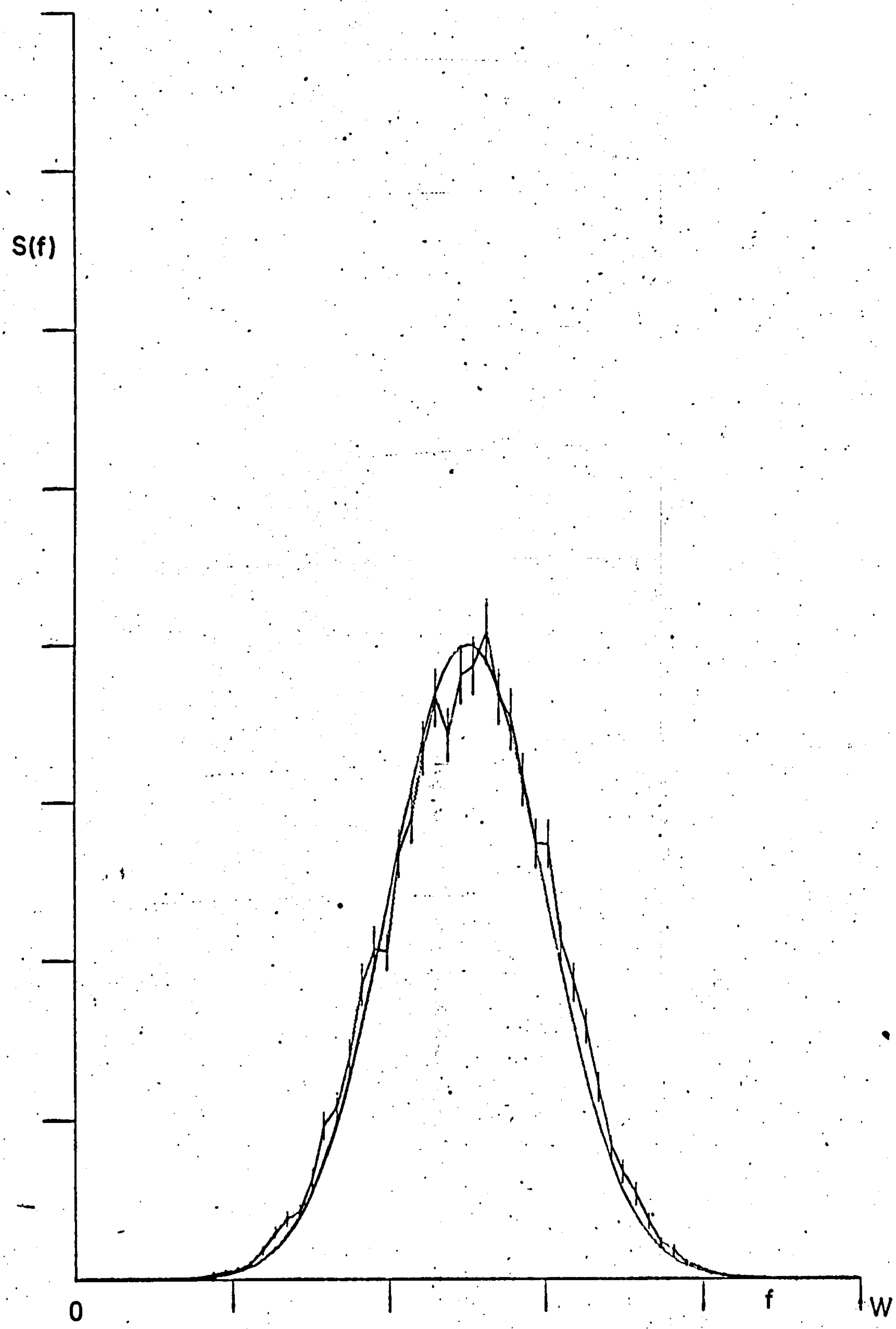


Fig.4.4.7. Estimated and theoretical large-p spectra for p = 4.0. Estimate made using 63 zeros per period.

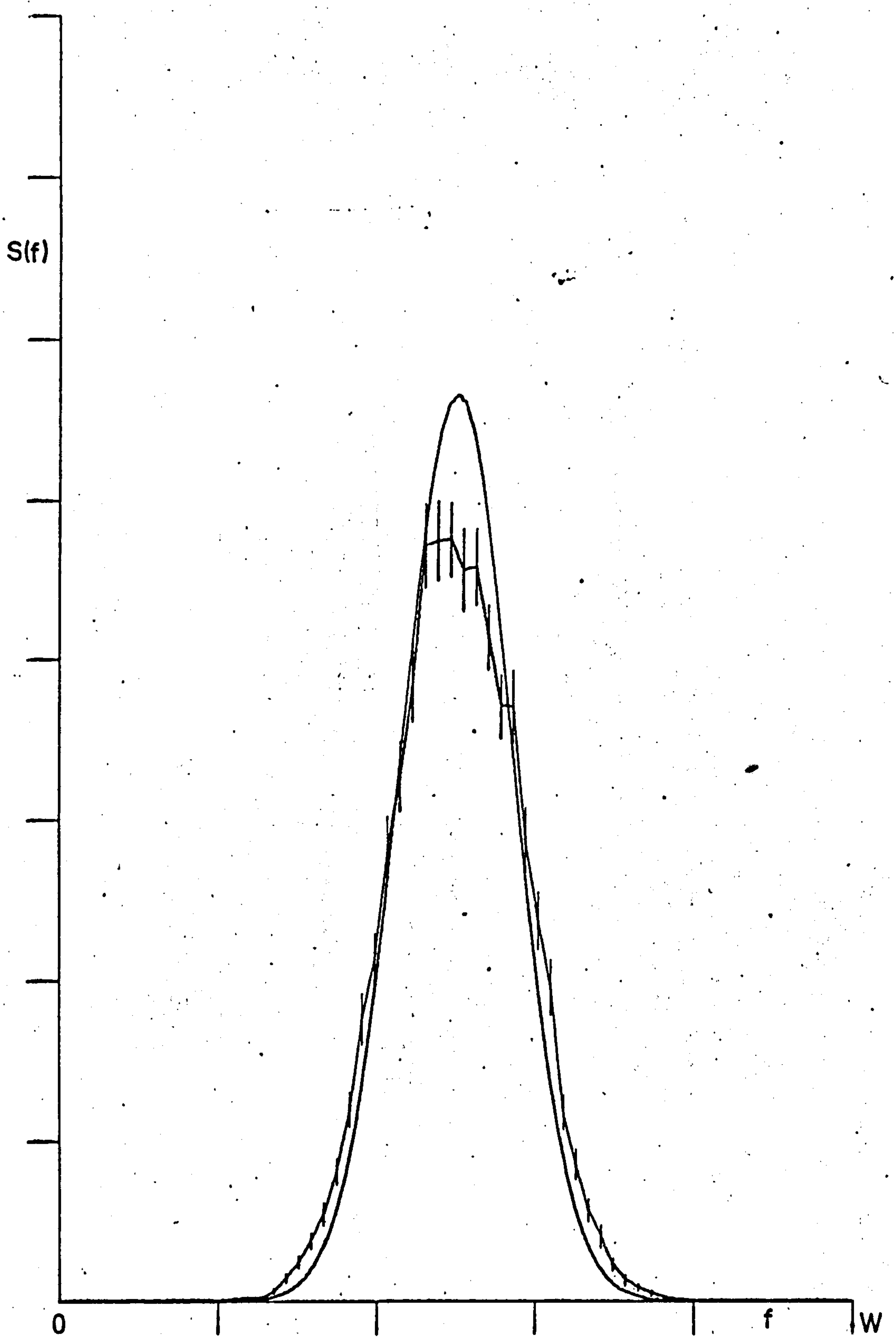


Fig. 4.4.8. Estimated and theoretical large-p spectra for p = 8.0. Estimate made using 63 zeros per period.

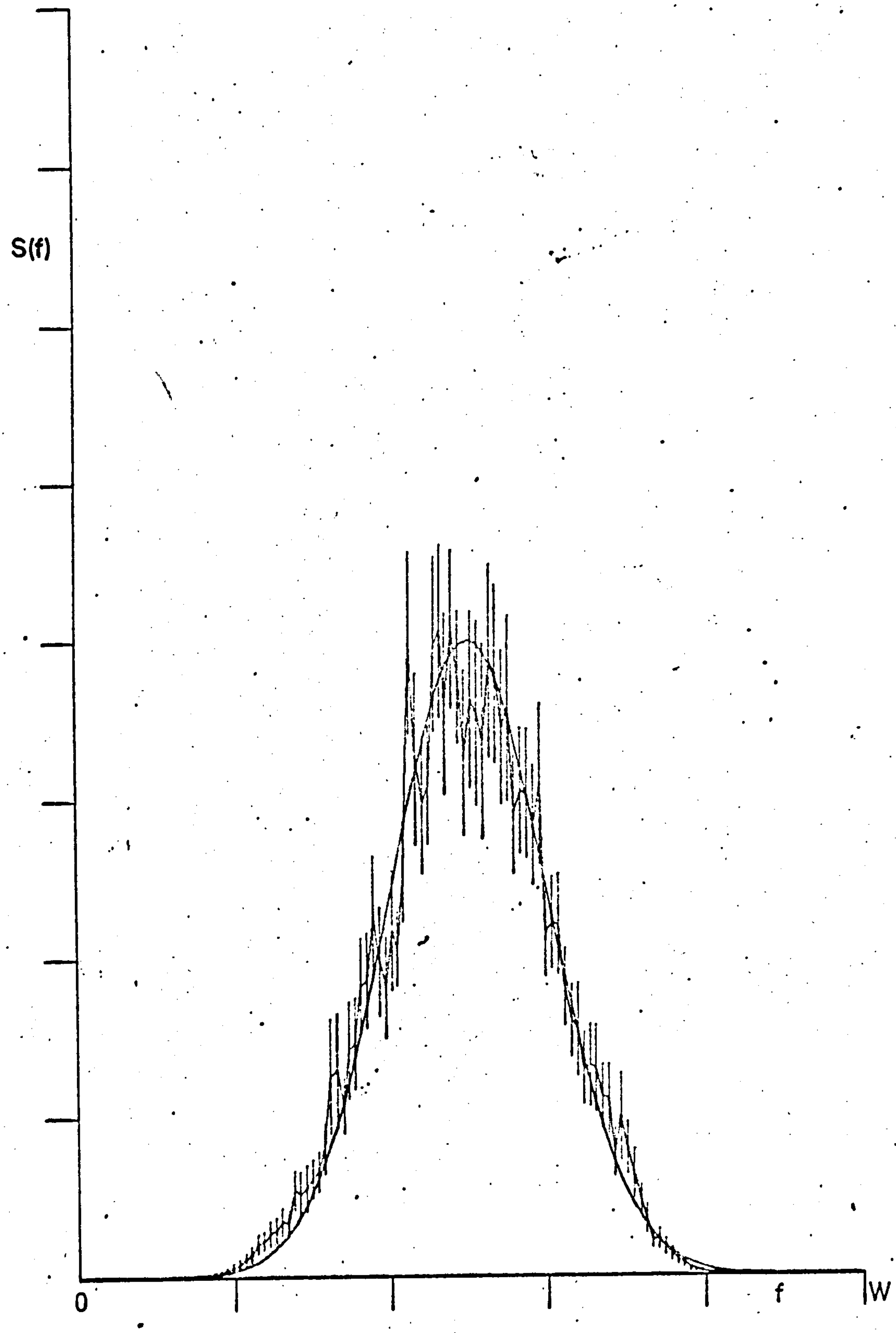


Fig. 4.4.9. Estimated and theoretical large-p spectra for p = 4.0. Estimate made using 127 zeros per period.

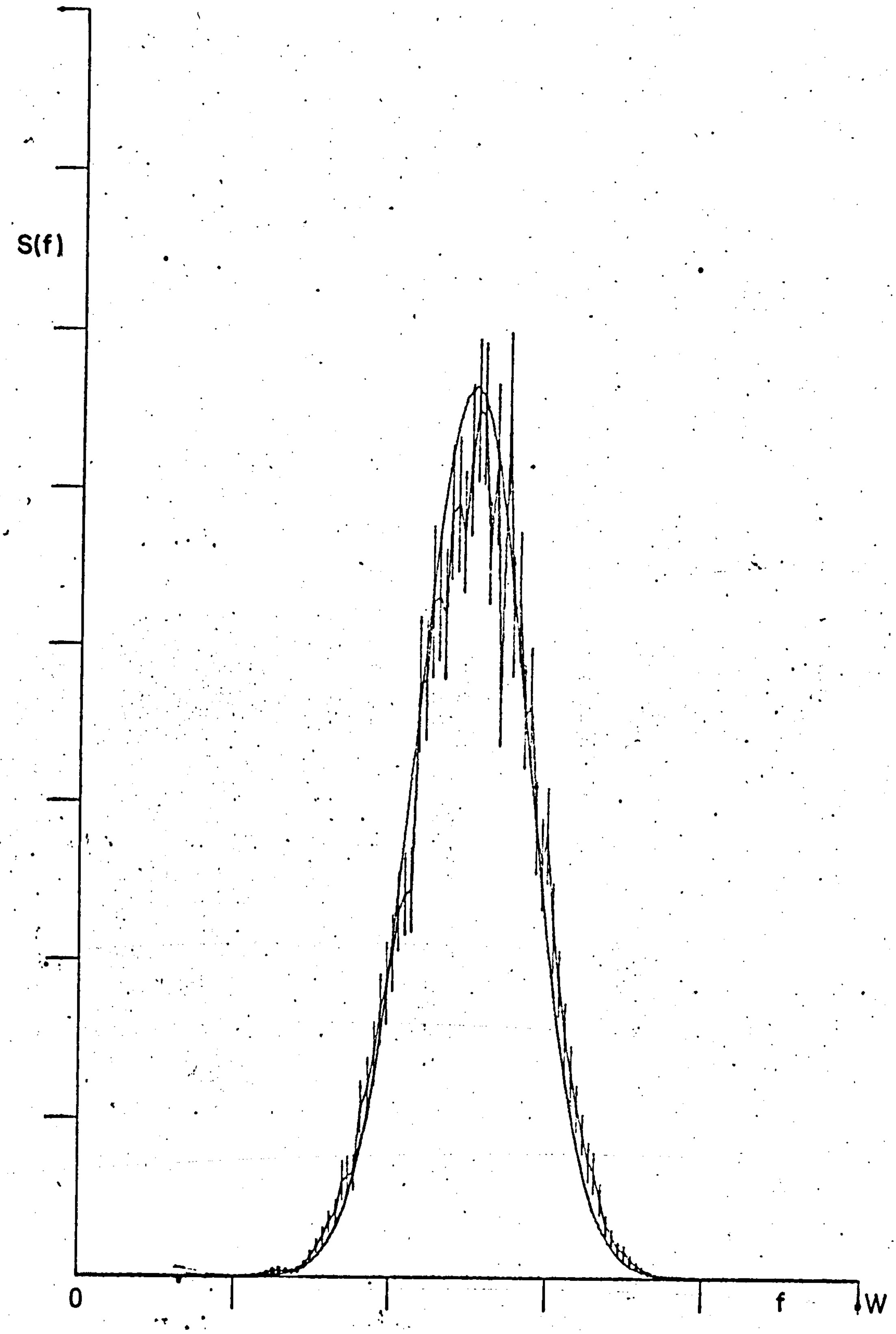


Fig.4.4.10. -Estimated and theoretical large-p spectra for p= 8.0. Estimate made using 127 zeros per period.

CHAPTER 5

ANGLE CODED SIGNALS FOR RADAR AND SONAR

5.1 Introduction

In a radar or a sonar system the transmitted signal returns delayed in time according to the range of the target and shifted in frequency according to its velocity* (together with an amplitude change and various disturbances). The classic problem of radar signal design consists of choosing a waveform which permits targets at various specified portions of the delay-doppler shift plane to be distinguished (in the presence of noise).

The ambiguity function[17,22] provides a means of comparing the resolution capabilities of signals. Ideally it would be possible to synthesise a signal directly to yield a specified ambiguity function but there are great theoretical difficulties in this approach. Instead, what is commonly done is to choose a suitable waveform for a given application from those signals whose ambiguity functions have been catalogued[22]. Phase modulated signals are often used because they require only simple and efficient transmitter circuits. Because angle coded signals, for p large, behave for practical purposes as purely frequency modulated signals it might be expected (as Voelcker has suggested) that they would be suitable for use as radar signals.

In radar signal design one is concerned simultaneously with the temporal and the spectral behaviour of signals. As mentioned in chapter 1, by the analogy with root locus methods (which are found

* To a close approximation

useful when the time and frequency behaviour of linear systems are of interest), it might be expected that zero based signal theory would be useful in the design of radar signals.

This chapter shows that zero based methods can be useful in radar signal design. In particular, a method is presented by which an angle coded signal can be synthesised whose time-frequency energy density distribution (appendix A) approximates some desired form. The resulting signal is a dual of a Huffman "impulse equivalent" pulse sequence*[23]. Appendix C, which represents a translation of this chapter from time language into frequency language, applies this theory to the design of Huffman sequences.

5.2 Synthesis of Zero Patterns

The signal design problem here principally consists of choosing the zero locations of an angle coded signal so that the energy density of the signal is concentrated about some line in the time-frequency (t-f) plane. Reference 24 discusses the problem of choosing suitable forms of t-f energy density distribution.

If the energy density of the signal is concentrated about a line in the t-f plane which forms a single valued function of time, then it is meaningful to speak of "the frequency at which the power of the signal acts" at that time. A measure of this quantity, $f_c(t)$, is given by the normalised first moment of the real part of the complex energy density function[38], $e(t,f)$, (taken with respect to frequency at a given instant of time). Appendix A shows that this quantity is given by the instantaneous frequency of the signal whether or not this is a slowly varying function of time. Thus the

*Pointed out, according to Prof. Voelcker, by Prof. Titlebaum instantaneous frequency at a given instant of time represents the "centre of gravity" of a cross section of $e(t,f)$ taken in the frequency direction at that instant (fig.5.2.1.).

The instantaneous frequency of an angle coded signal can be expressed very simply in terms of the zero locations (sect.4.2). The instantaneous phase derivative, $\dot{\phi}(t)$, can be thought of as being produced by the convolution of the function

$$V(t) = \frac{s}{s^2 + t^2} \tag{5.2.1}$$

with the impulse series

$$I(t) = \sum_{n=-\infty}^{\infty} \alpha_n \, \delta(t-nT) \tag{5.2.2}$$

together with the addition of the constant, k. α_n takes the values +1 and -1 according to whether the nth zero lies in the upper or lower half planes. Symbolically,

$$\dot{\phi}(t) = V(t) * I(t) + K \tag{5.2.3}$$

To produce an angle coded signal whose energy is concentrated about some line, $f_d(t)$, in the t-f plane, it suffices to choose a zero pattern which results in the instantaneous frequency function of the signal following this line. It can be seen from equations 5.2.2 and 5.2.3 that this is equivalent to choosing a sequence of regularly spaced positive and negative unit impulses which, after being smoothed by convolution with V(t) and being added to K, approximate $f_d(t)$.

A delta-sigma modulator[25] (fig. 5.2.2.) is a system whose output consists of a sequence of equally spaced positive and negative impulses the weights of which are all of the same magnitude. It works in such a way that its output, when smoothed by a lowpass filter, is

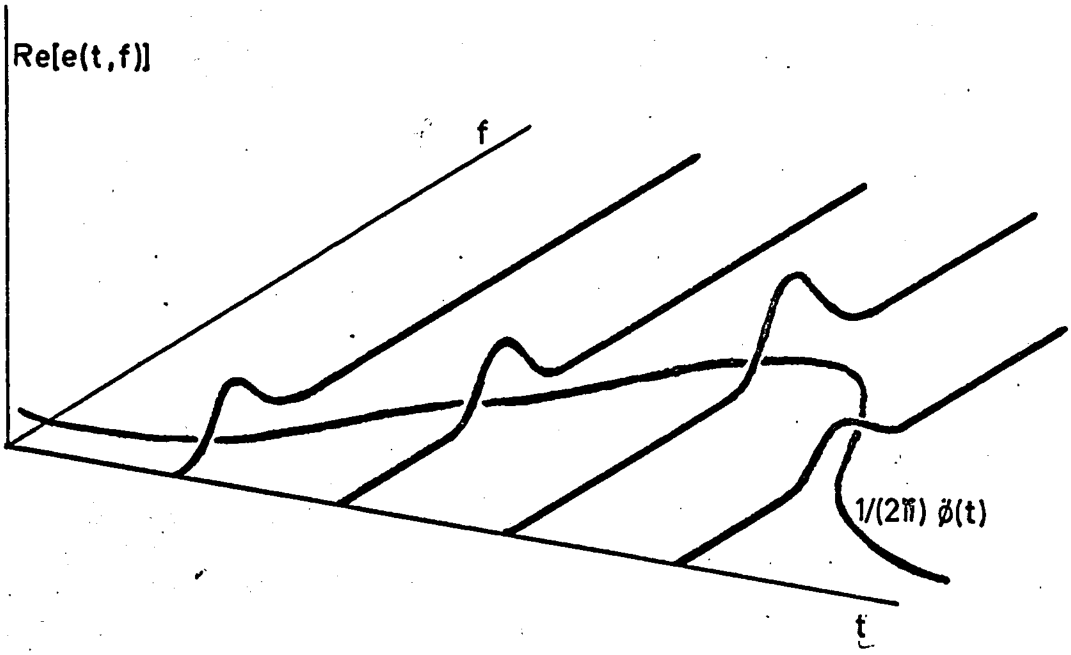


Fig.5.2.1. Instantaneous frequency as normalised first moment of the time-frequency energy distribution.

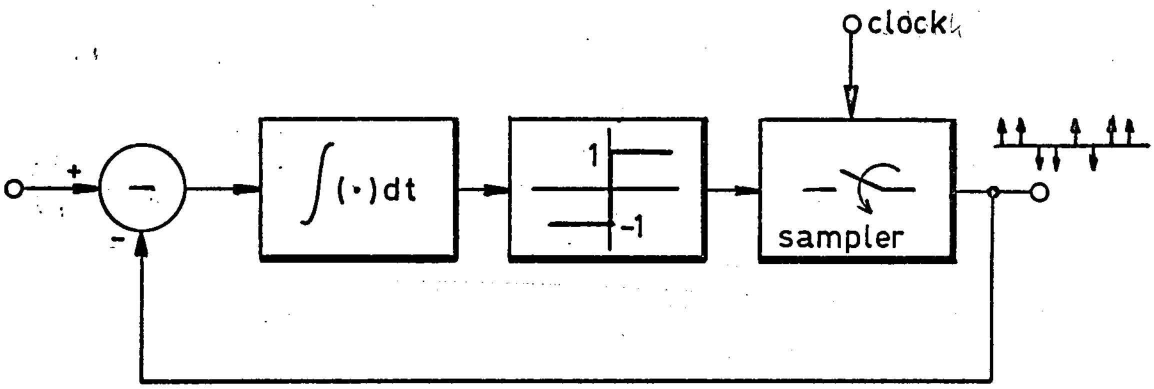


Fig.5.2.2. Delta-sigma modulator.

transformed into a signal which presents an approximation to its input waveform. Thus if it is required to approximate a function by the convolution of a sequence of impulses with a function having the nature of the impulse response of a lowpass filter (as in the present situation) the impulse sequence can be found by applying the desired function to the input of a delta sigma modulator and recording its output.

The impulse response of the lowpass filter in this case is V(t) so that the zero pattern can be chosen by applying $\frac{1}{2\pi} f_d(t)$ (with $k/2\pi$ subtracted) to the input of a delta-sigma modulator whose sampling period is T and whose output impulses have weights of unit magnitude. A positive impulse from the delta modulator at the nth sampling instant indicates that the nth zero should lie in the UHP; a negative impulse indicates that it should lie in the LHP.

5.3 Example of the Method

To illustrate the use of the method the form of t-f energy distribution sketched in fig. 5.3.1 was chosen as the desired energy distribution. This energy distribution resembles that of the periodic repetition of a linear FM pulse (which is a commonly used radar signal).

The duration-bandwidth product of one period of this energy distribution is 34 which implies that a periodic zero pattern with 34 zeros per period should be used. (when T, the spacing between zeros, will be one thirty-fourth of the period of the energy density function).

The desired instantaneous frequency function $f_d(t)$, corresponding to this desired t-f energy distribution is shown in fig. 5.3.2.

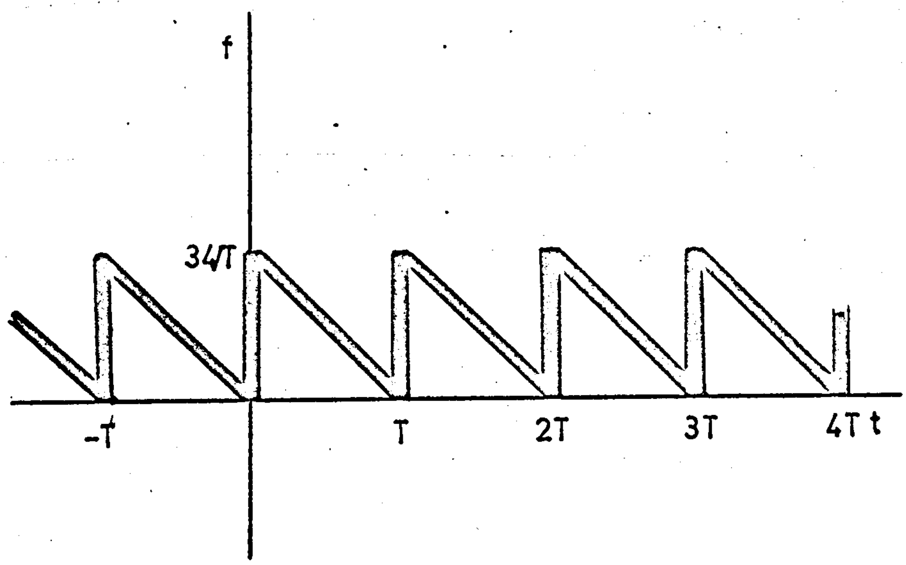


Fig. 5.3.1. Desired t-f energy distribution.

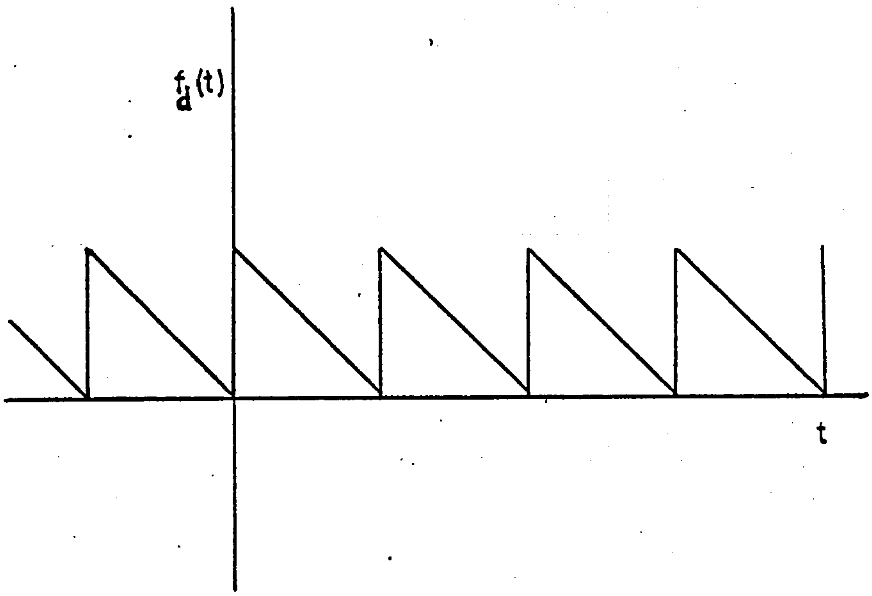


Fig. 5.3.2. Desired instantaneous frequency function.

A system similar to a delta-sigma modulator was simulated (by hand) with $f_d(t)$ (less its mean value) as input. One period of the resulting impulse sequence was

(+++++++++-++++-+-+-+-+---+---------).

The Fourier coefficients of the resulting periodic angle coded signal for various values of p are displayed in table 5.1 and plotted in fig. 5.3.3. They were computed from the zero locations specified by the above sequence making use of the methods of section 2.3.

The complex energy density function of an analytic signal m(t) is defined by

$$e(t,f) = m(t)\, M^*(f)\, e^{-j2\pi ft} \qquad 5.3.1.$$

If m(t) is periodic, with period T and is represented by a Fourier series

$$m(t) = \sum_{k=0}^{N} C_k\, e^{j2\pi kt/T} \qquad 5.3.2$$

its Fourier transform is a series of impulses

$$M(f) = \sum_{k=0}^{N} C_k\, \delta(f-k/T) \qquad 5.3.3.$$

Thus e(t,f) is given by

$$e(t,f) = \sum_{k=0}^{N} \sum_{\ell=0}^{N} C_k C_\ell^*\, e^{-j2\pi(f-k/T)t}\, \delta(f-\ell/T). \qquad 5.3.4.$$

e(t,f) consists of lines running in the direction of the t-axis, and spaced at integer multiples of the fundamental frequency in the direction of the f-axis.

To provide a display of e(f,t), the weight of each line (a time function) was plotted by evaluating the expression

n	p = 0.5		p = 1.0		p = 2.0	
	Re[C_n]	Im[C_n]	Re[C_n]	Im[C_n]	Re[C_n]	Im[C_n]
1	2.08	0.00	0.43	0.00	0.02	0.00
2	0.00	2.74	0.00	1.14	0.00	0.10
3	-1.80	-1.43	-1.51	-0.60	-0.27	-0.06
4	1.83	-0.01	1.60	-1.00	0.30	-0.45
5	-1.26	0.65	-0.43	1.85	0.39	0.78
6	0.77	-1.55	-0.70	-1.97	-1.19	-0.25
7	0.55	0.17	2.18	0.65	1.43	-0.97
8	0.52	0.58	-1.06	1.23	-0.19	2.42
9	-1.36	1.67	-0.84	0.23	-2.37	-1.64
10	-0.89	-0.91	-1.42	-1.26	2.01	-1.38
11	0.26	-0.99	1.79	-1.46	0.72	0.75
12	1.20	-1.04	0.96	0.70	0.98	1.25
13	0.33	1.12	0.75	1.02	-1.97	1.56
14	-0.09	0.48	-1.17	1.32	-1.01	-1.06
15	-0.30	0.64	-0.52	-0.20	-1.21	-0.58
16	-0.67	0.04	-1.59	0.21	0.08	-2.30
17	-1.67	1.93	-1.73	-0.50	-0.33	-1.15
18	-2.25	1.89	-2.97	0.35	0.24	-2.68
19	-1.67	1.93	-1.73	-0.50	-0.33	-1.15
20	-0.67	0.04	-1.59	0.21	0.08	-2.30
21	-0.30	0.64	-0.52	-0.20	-1.21	-0.58
22	-0.09	0.48	-1.17	1.32	-1.01	-1.06
23	0.54	1.12	0.75	1.02	-1.97	1.56
24	1.20	-1.04	0.96	0.70	0.98	1.25
25	0.26	-0.99	1.79	-1.46	0.72	0.75
26	-0.89	-0.91	-1.42	-1.26	2.01	-1.38
27	-1.36	1.67	-0.84	0.23	-2.37	-1.64
28	0.52	0.58	-1.06	1.23	-0.19	2.42
29	0.55	0.17	2.18	0.65	1.43	-0.97
30	0.77	-1.55	-0.70	-1.97	-1.19	-0.25
31	-1.26	0.65	-0.43	1.85	0.39	0.78
32	1.88	-0.01	1.60	-1.00	0.30	-0.45
33	-1.80	-1.43	-1.51	-0.60	-0.27	-0.06
34	0.00	2.74	0.00	1.14	0.00	0.10
35	2.08	0.00	0.43	0.00	0.02	0.00

Table 5.1. Fourier coefficients of synthesised signals.

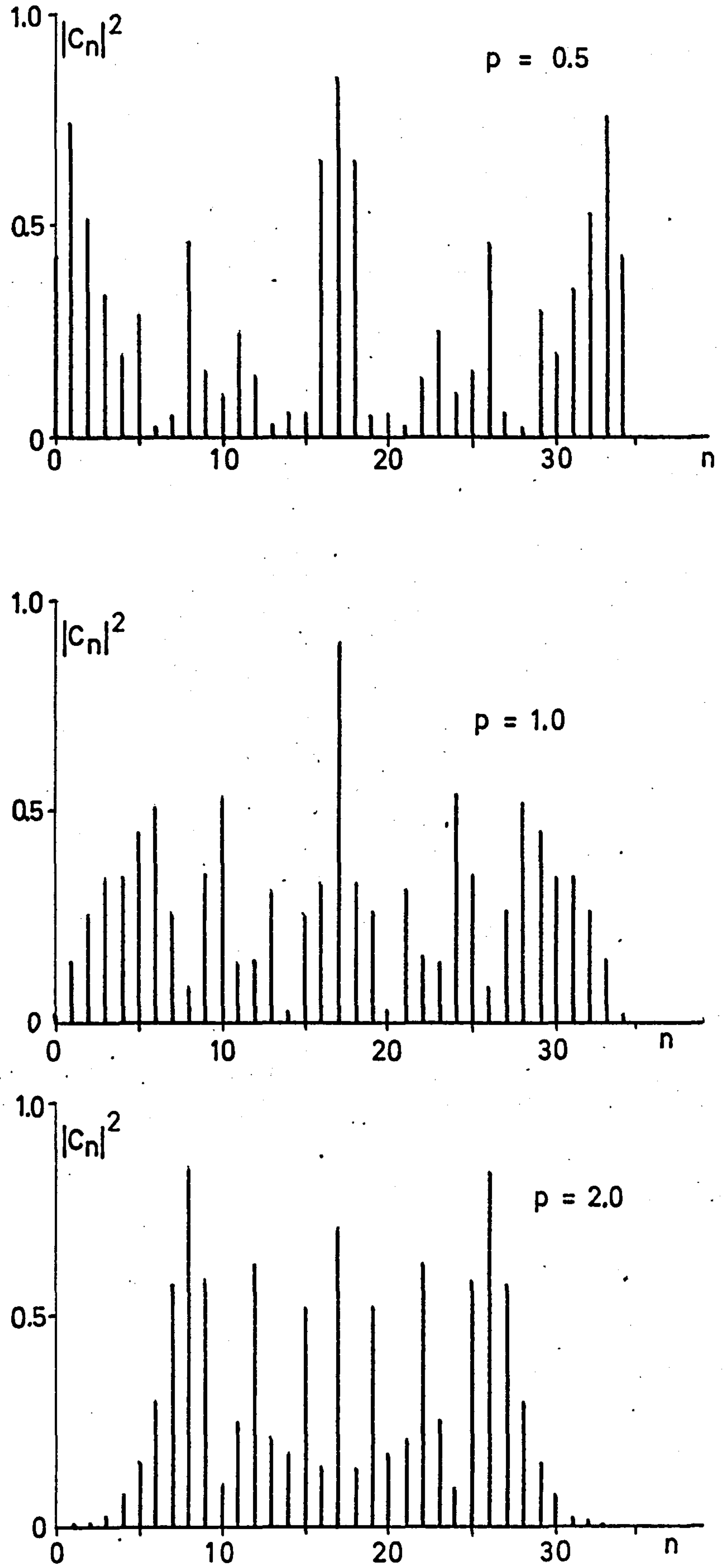


Fig.5.3.3. Fourier coefficient magnitudes of synthesised signal for various values of p.

$$e_\ell(t) = C_\ell^* \sum_{k=0}^{N} C_k \, e^{-j2\pi(f-k/T)t} \qquad 5.3.5$$

at 400 points and interpolating with straight lines (which are too small to be perceived in the display). The form of display adopted is similar to that used by Singleton and Poulter[26]. Each line weight function, $e_\ell(t)$, has a constant added to it which represents the frequency of the line. The line is then displayed as a conventional x-y plot - with the exception that any portion of a trace which lies below a previously plotted trace is suppressed to enhance the "3-D" appearance. The results are shown in figures 5.3.4 - 5.3.6.

The instantaneous phase derivative was evaluated, also at 400 points, using the expression

$$\dot{\phi}(t) = \text{Im}\left[\dot{m}(t) \,/\, m(t)\right]$$

$$= \text{Im}\left[(j2\pi/T) \sum_{k=0}^{N} k\, C_k\, e^{j2\pi kt/T} \,/\, \sum_{k=0}^{N} C_k e^{j2\pi kt/T}\right] \qquad 5.3.6.$$

In the computations, T was assigned the value unity; this normalises the scales of the resulting graphs.

The characteristics of the angle coded signals having the synthesised zero pattern (for various values of p) are shown plotted in figures 5.3.7 - 5.3.9. These plots show the envelope, the real and the imaginary parts of the waveform, the instantaneous frequency and the real part of the t-f energy distribution. One period of each function is shown, except for the t-f energy distribution, where for clarity two periods are shown.

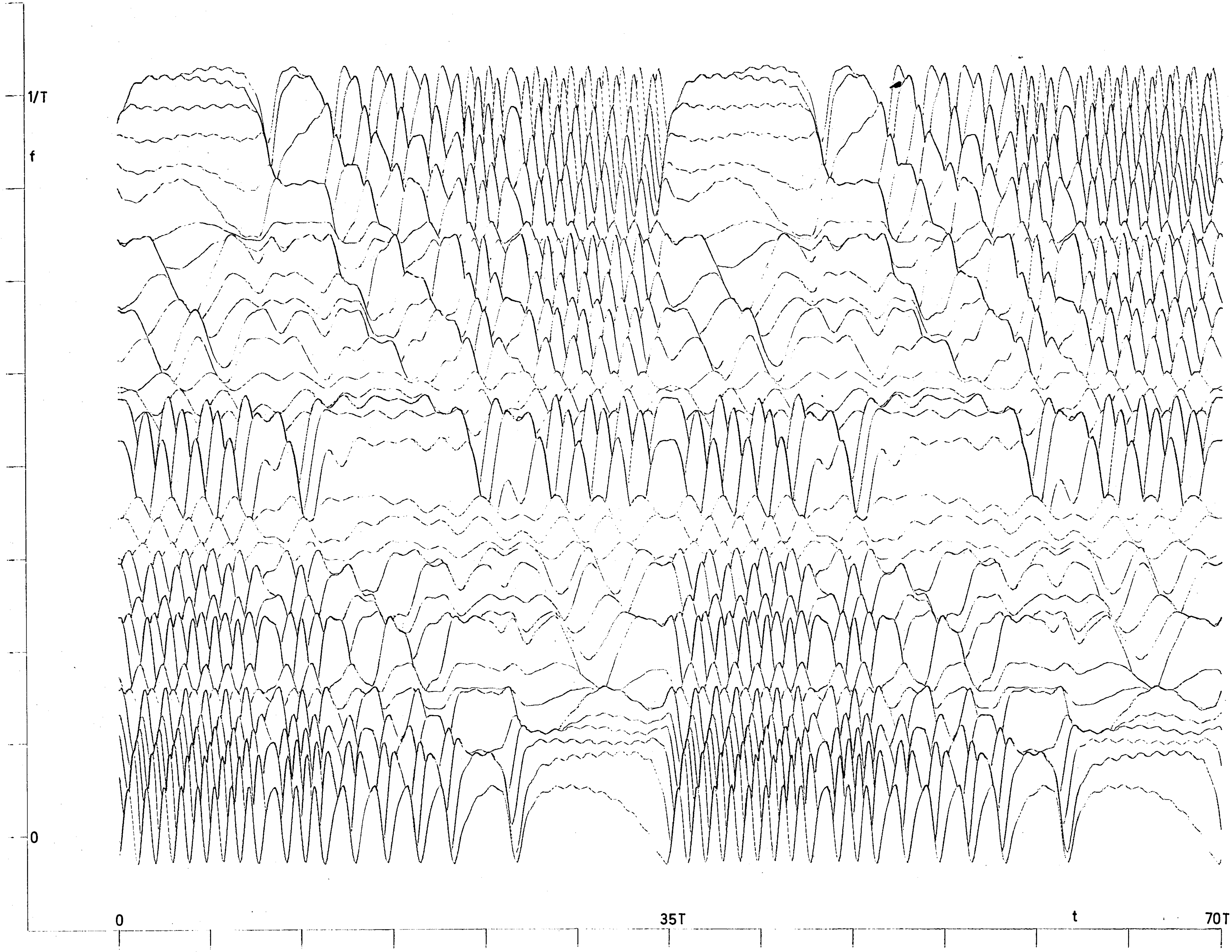


Fig.5.3.4. t - f energy density distribution of synthesised signal: p = 0.5.

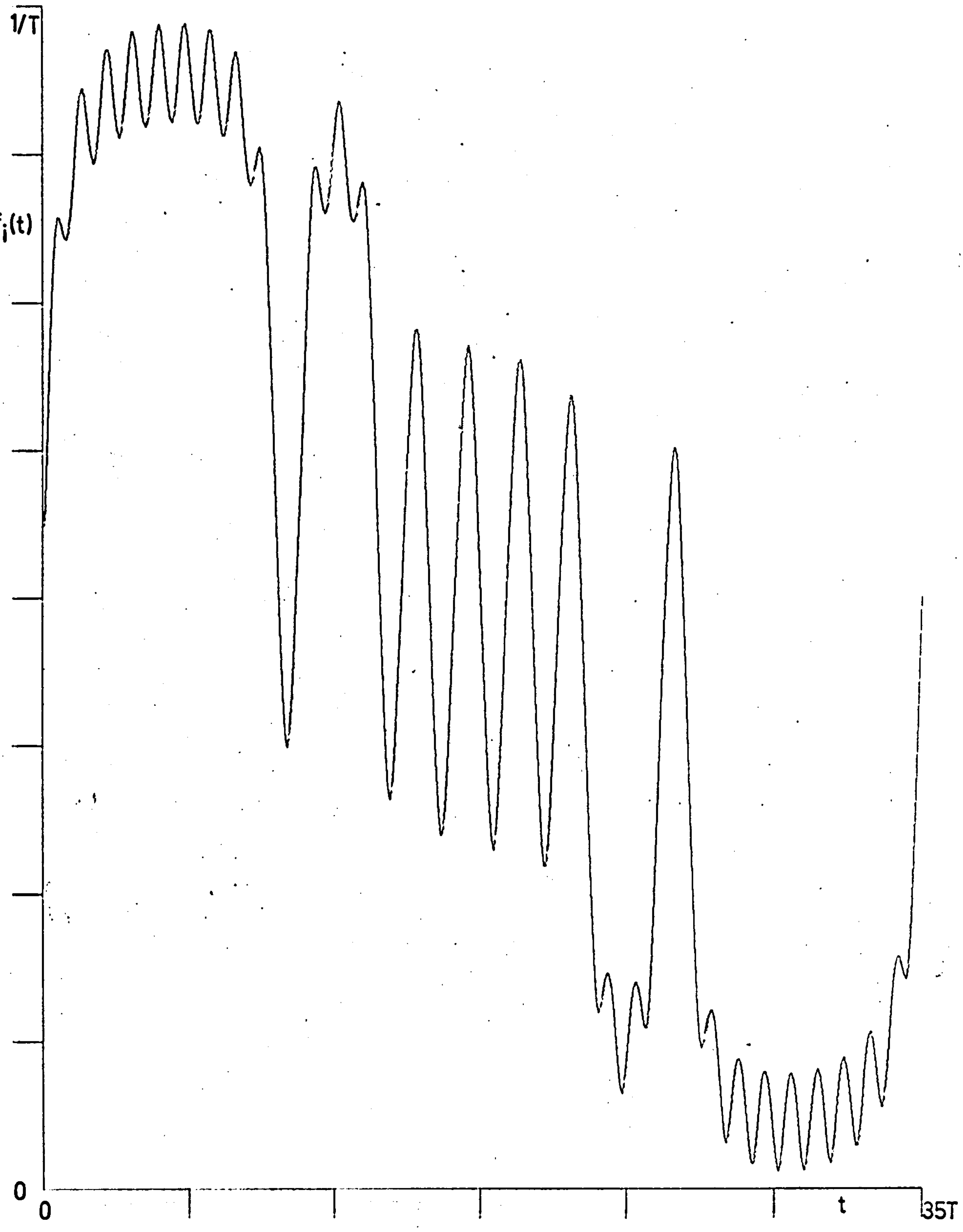
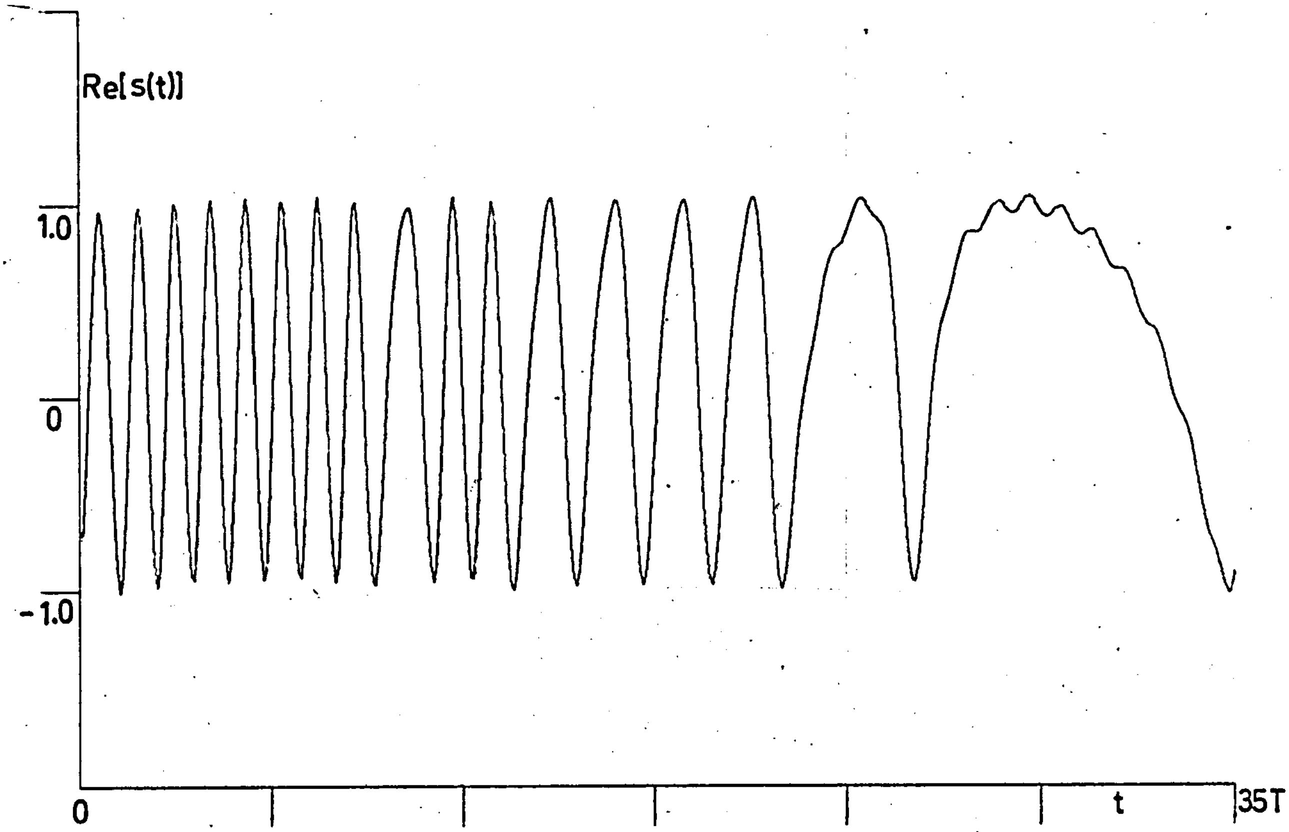


Fig.5.3.5.a. Instantantaneous frequency of synthesised signal: p = 0.5 .



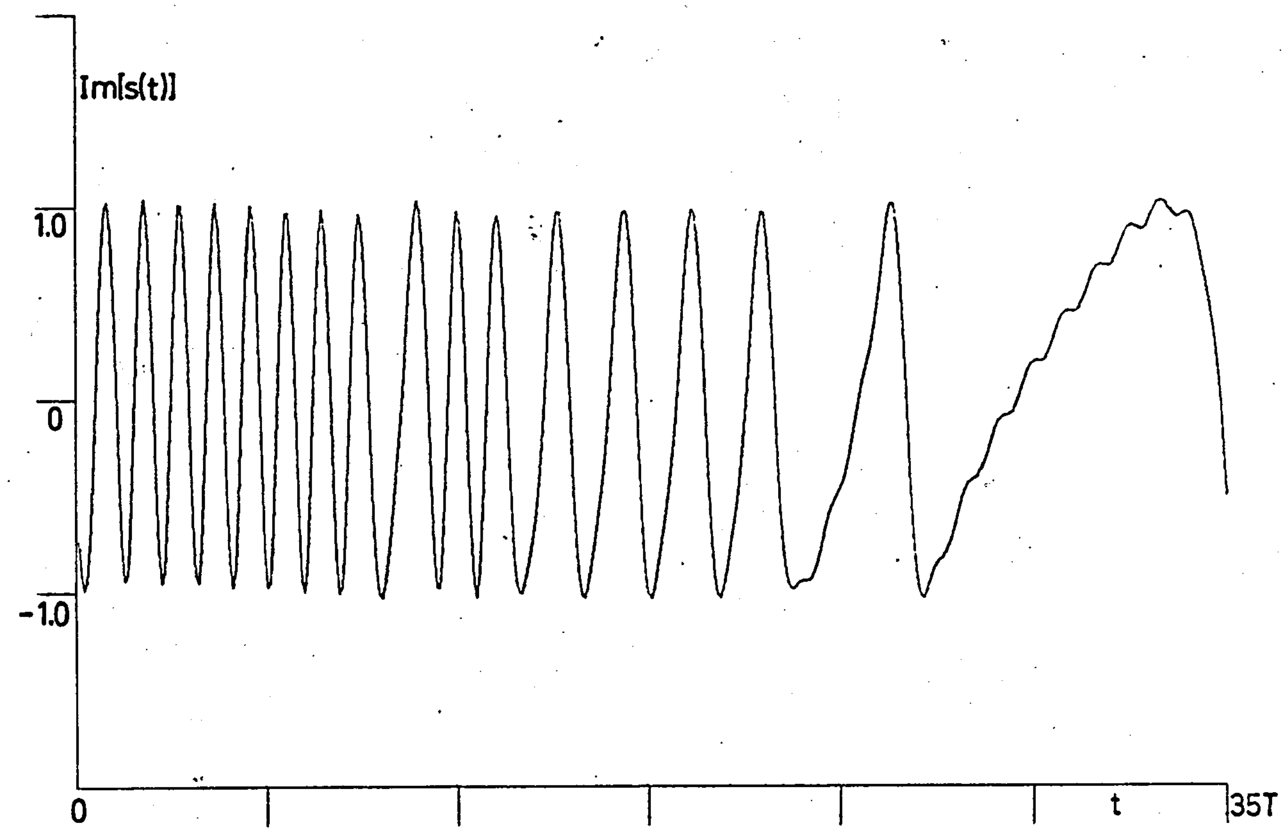


Fig.5.3.5.b. Real and imaginary parts of synthesised signal: p = 0.5.

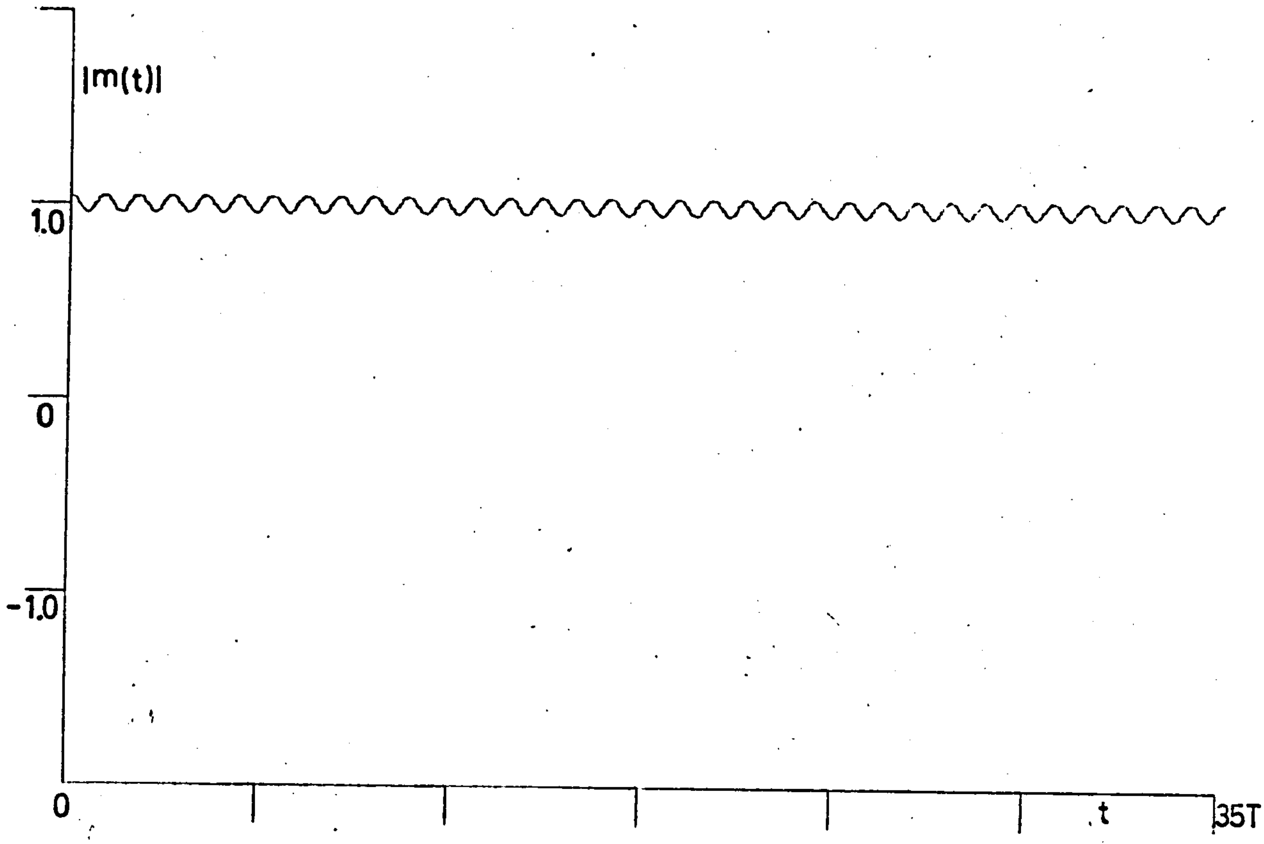


Fig. 5.3.5.c. Envelope of synthesised signal: p = 0.5 .

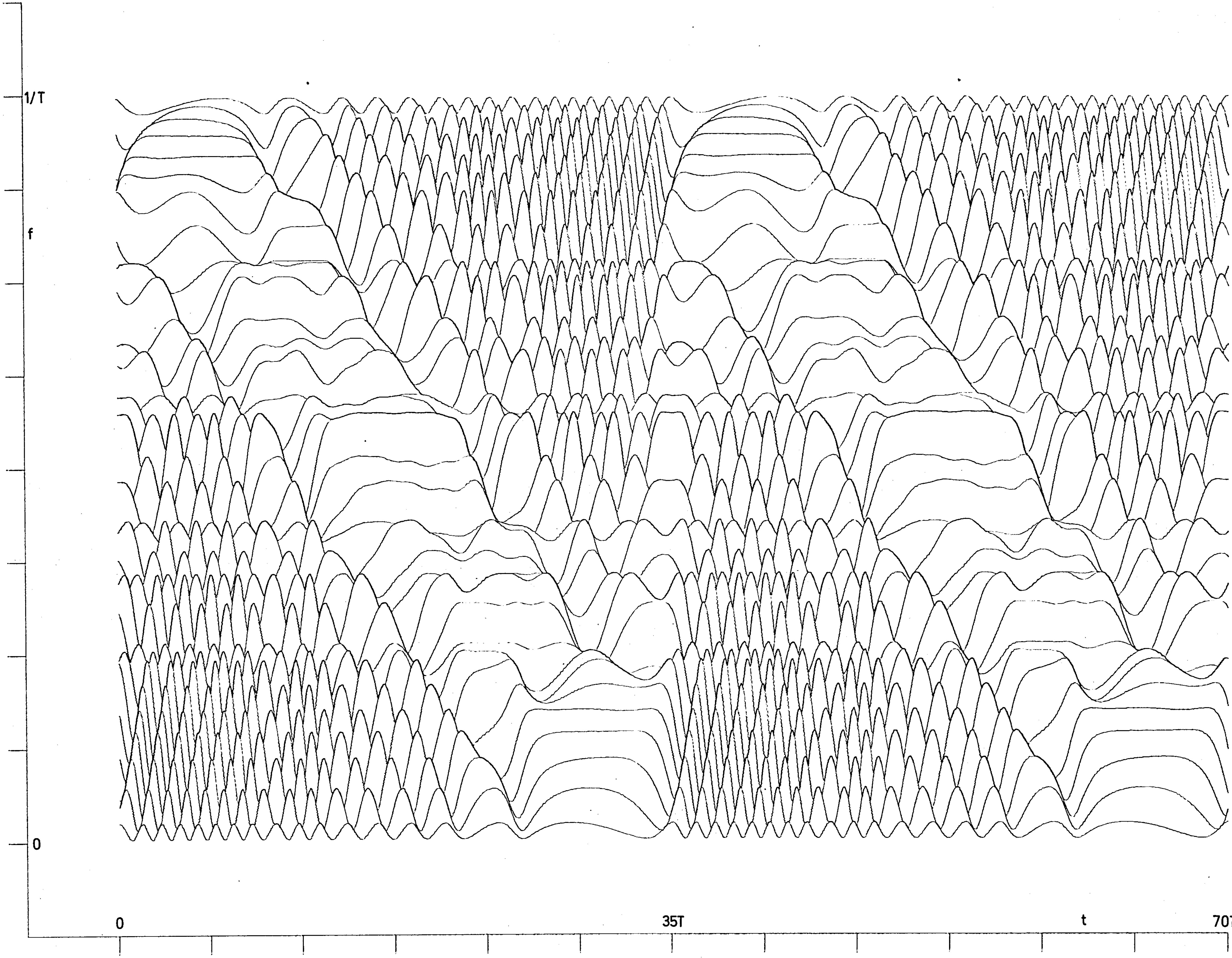


Fig. 5.3.6. t - f energy distribution of synthesised signal: p = 1.0.

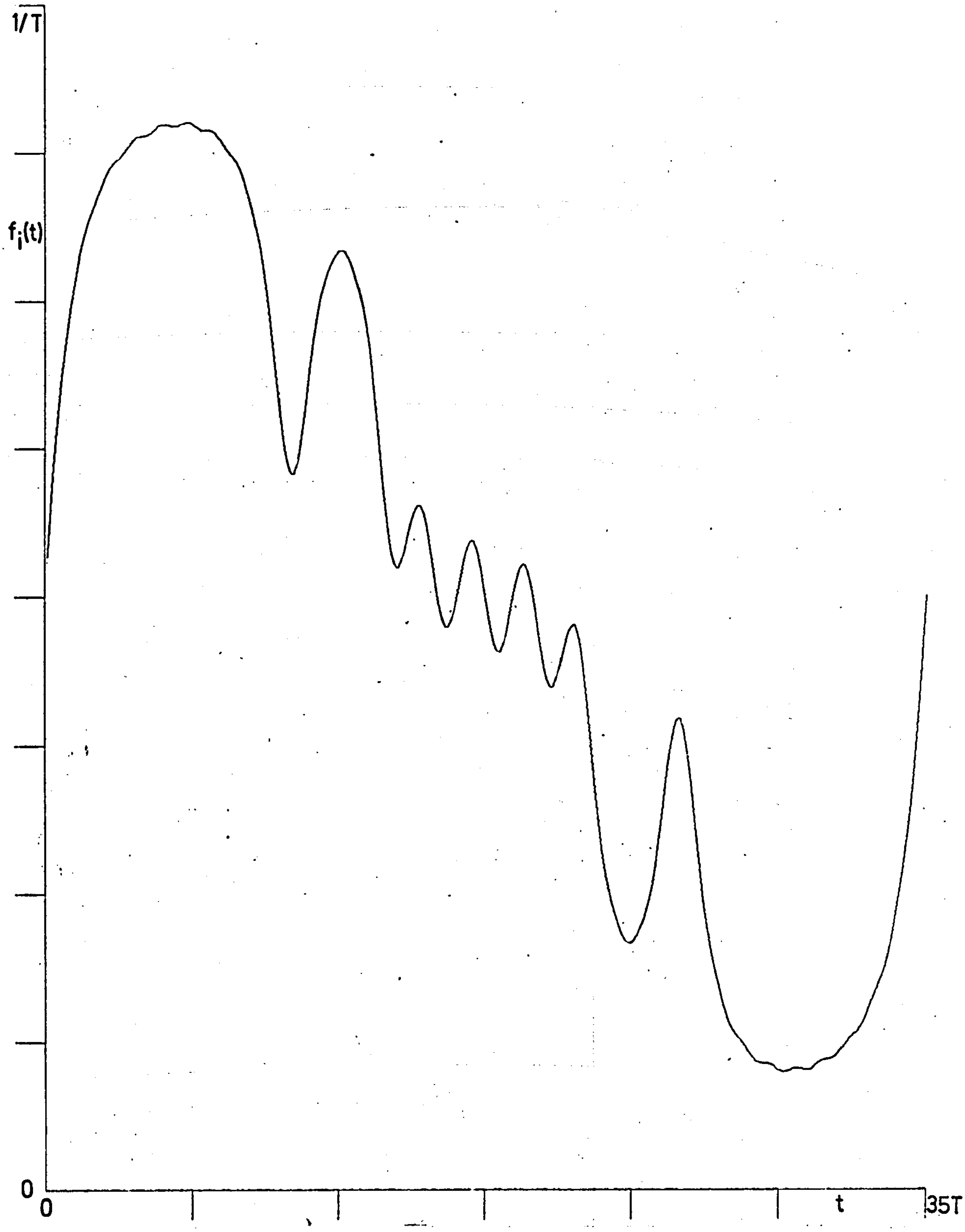


Fig. 5.3.7.a. Instantaneous frequency of synthesised signal: p = 1.0.

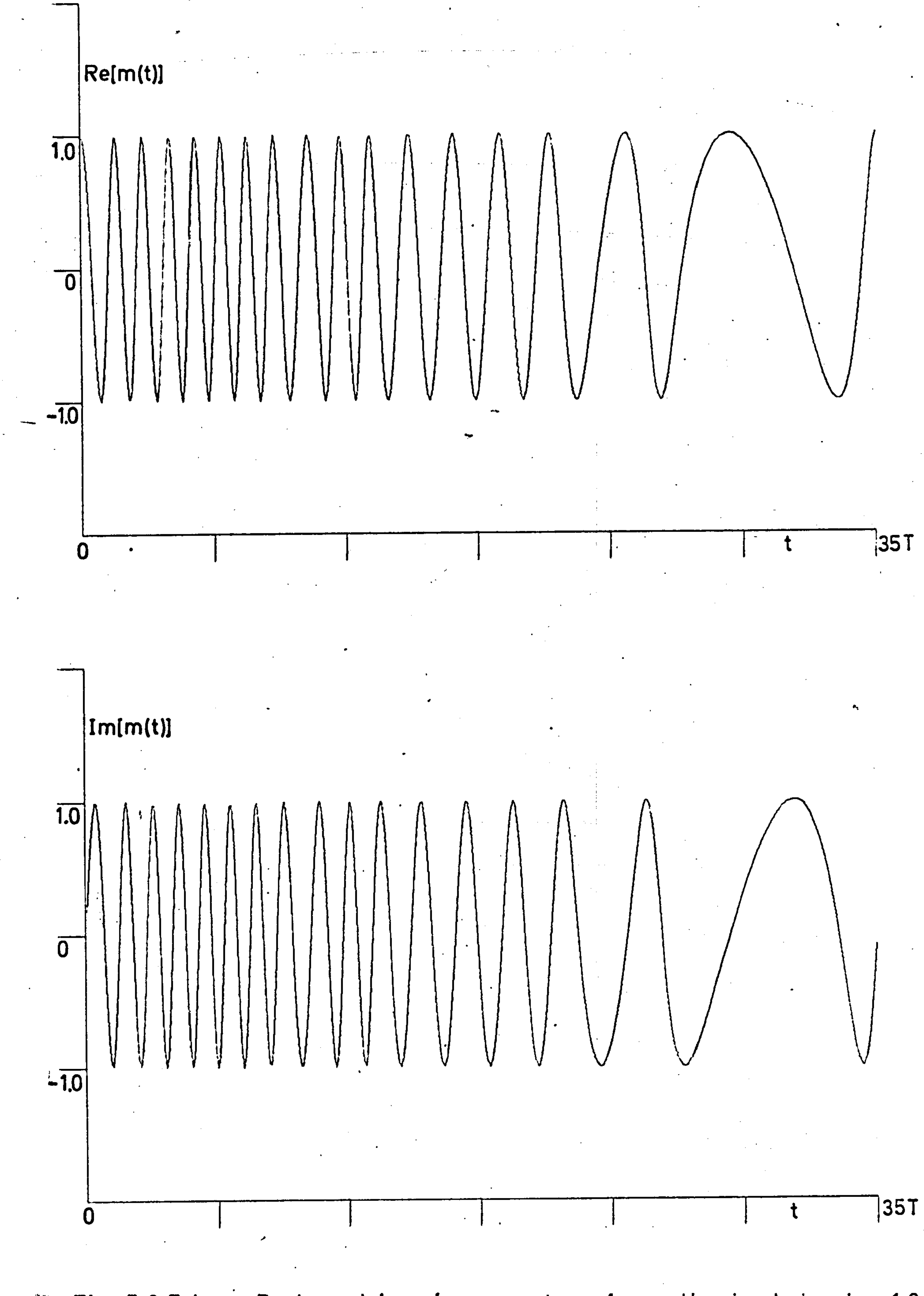


Fig. 5.3.7. b.   Real and imaginary parts of synthesised signal: p=1.0.

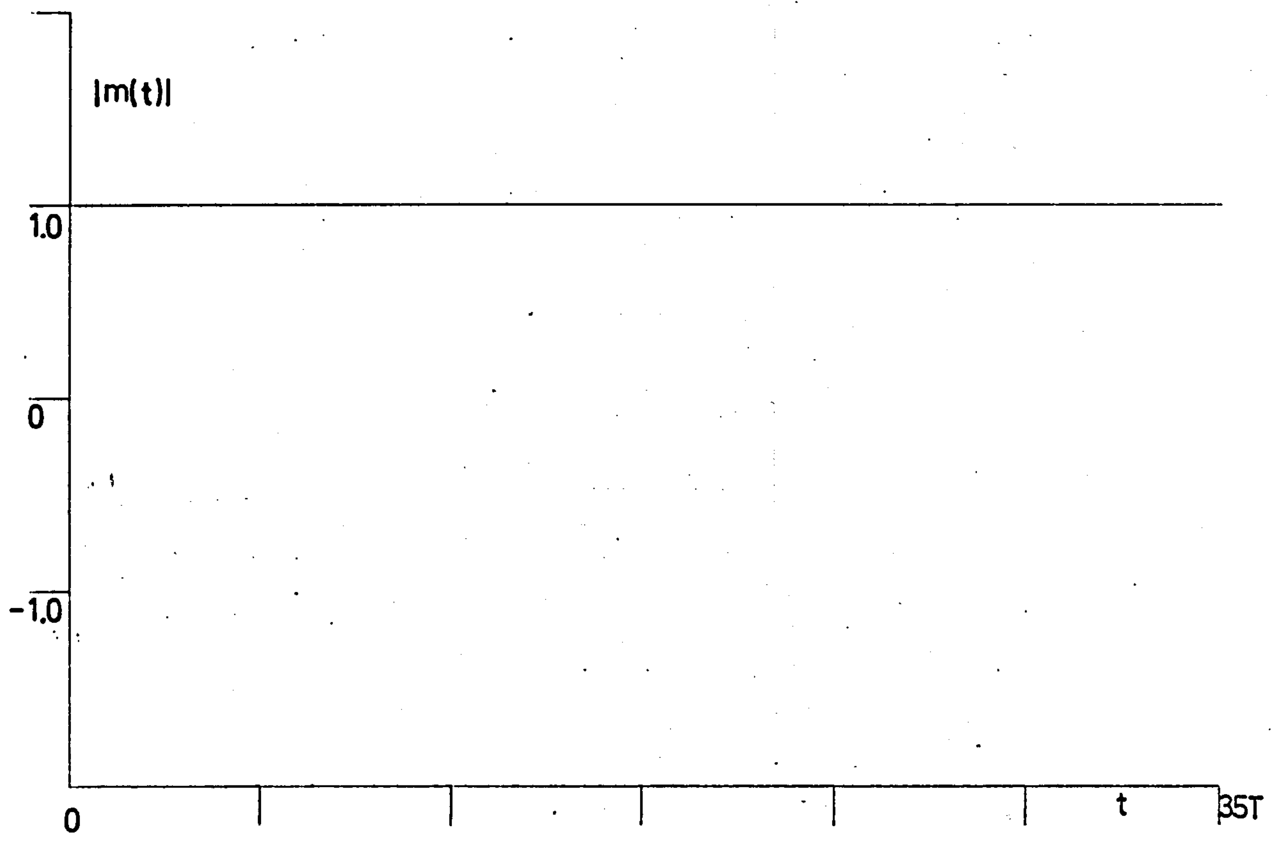


Fig. 5.3.7.c. Envelope of synthesised signal: p = 1.0. Envelopes of larger-p signals differ imperceptibly.

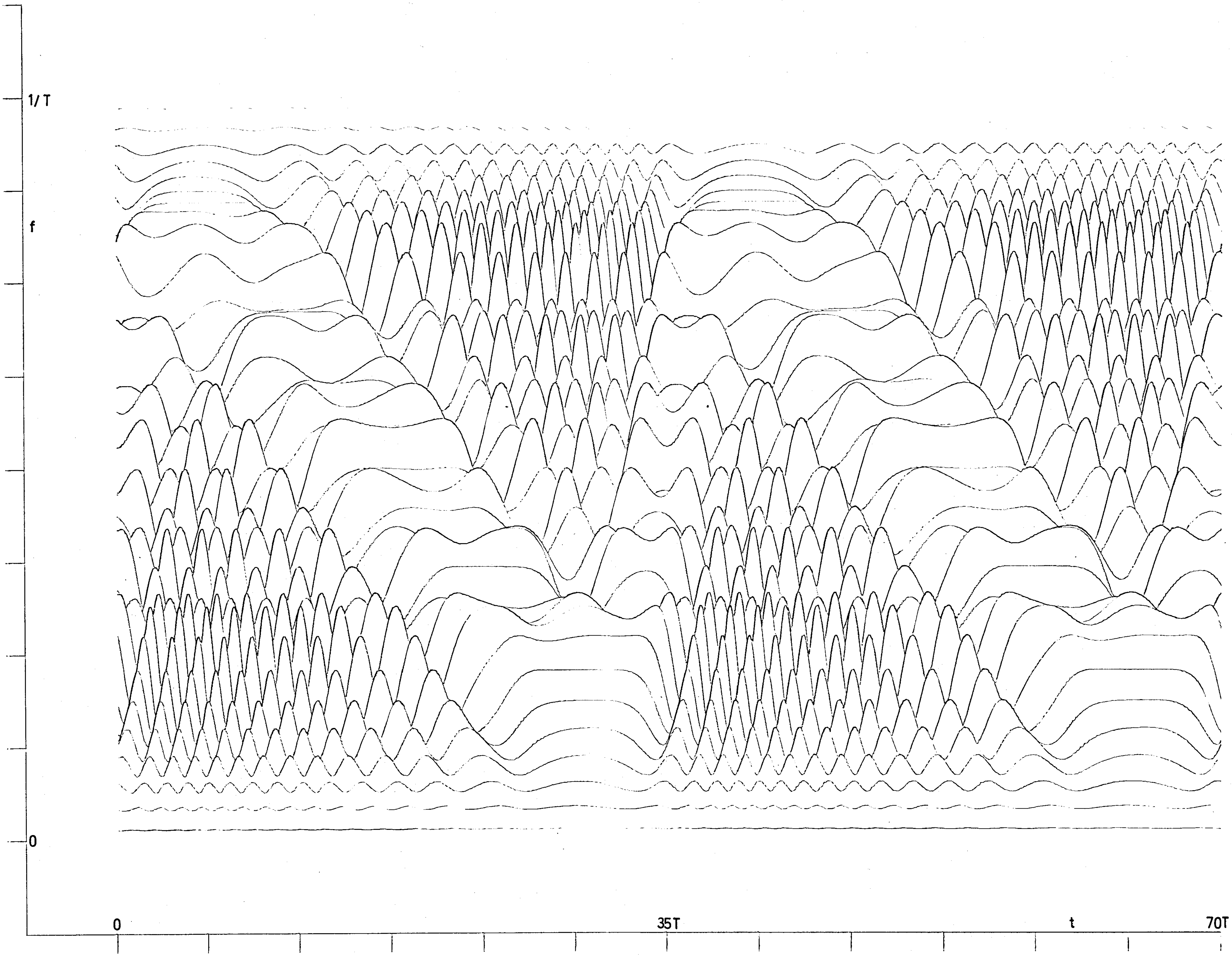


Fig. 5.3.8. t - f energy distribution of synthesised signal: p = 2.0.

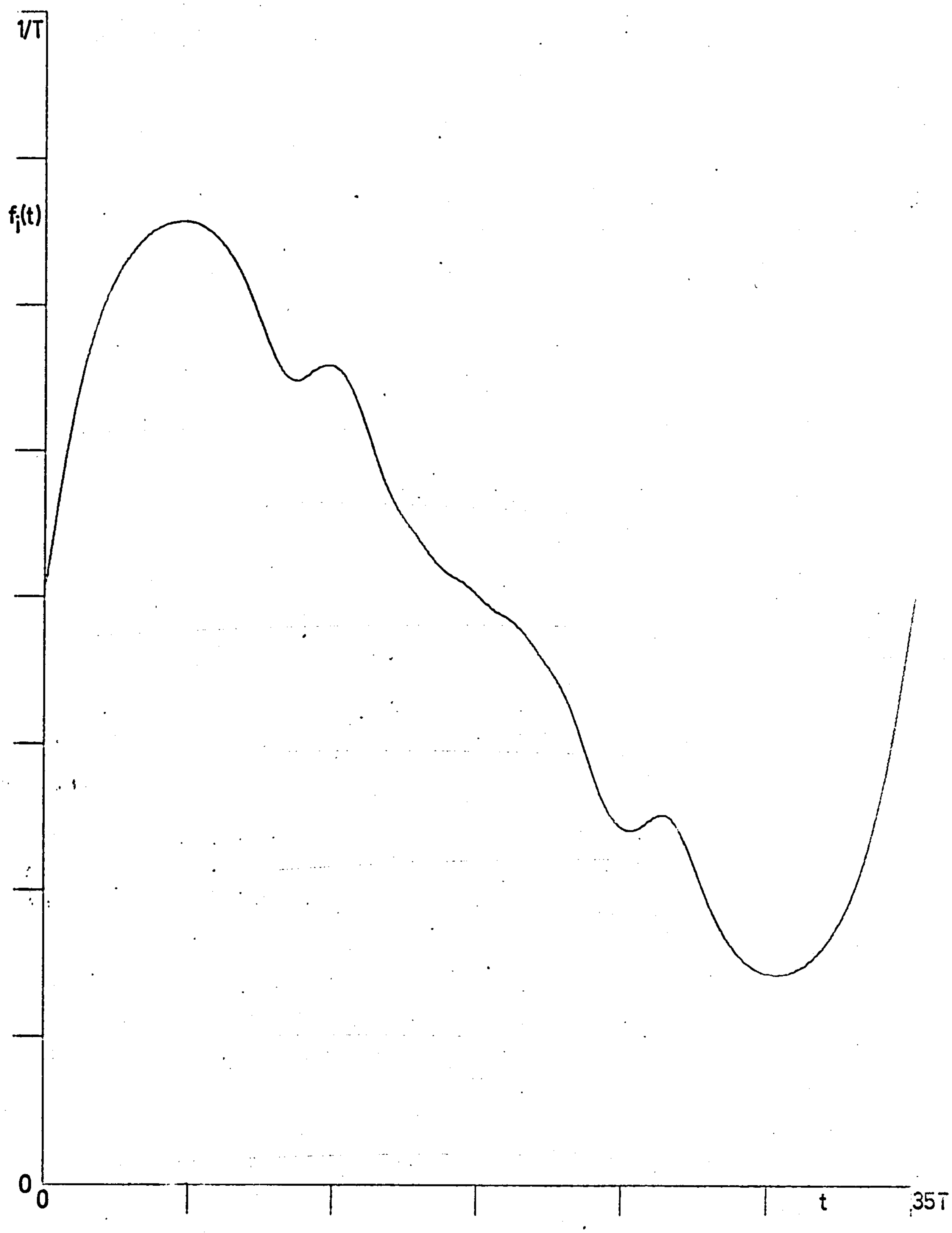
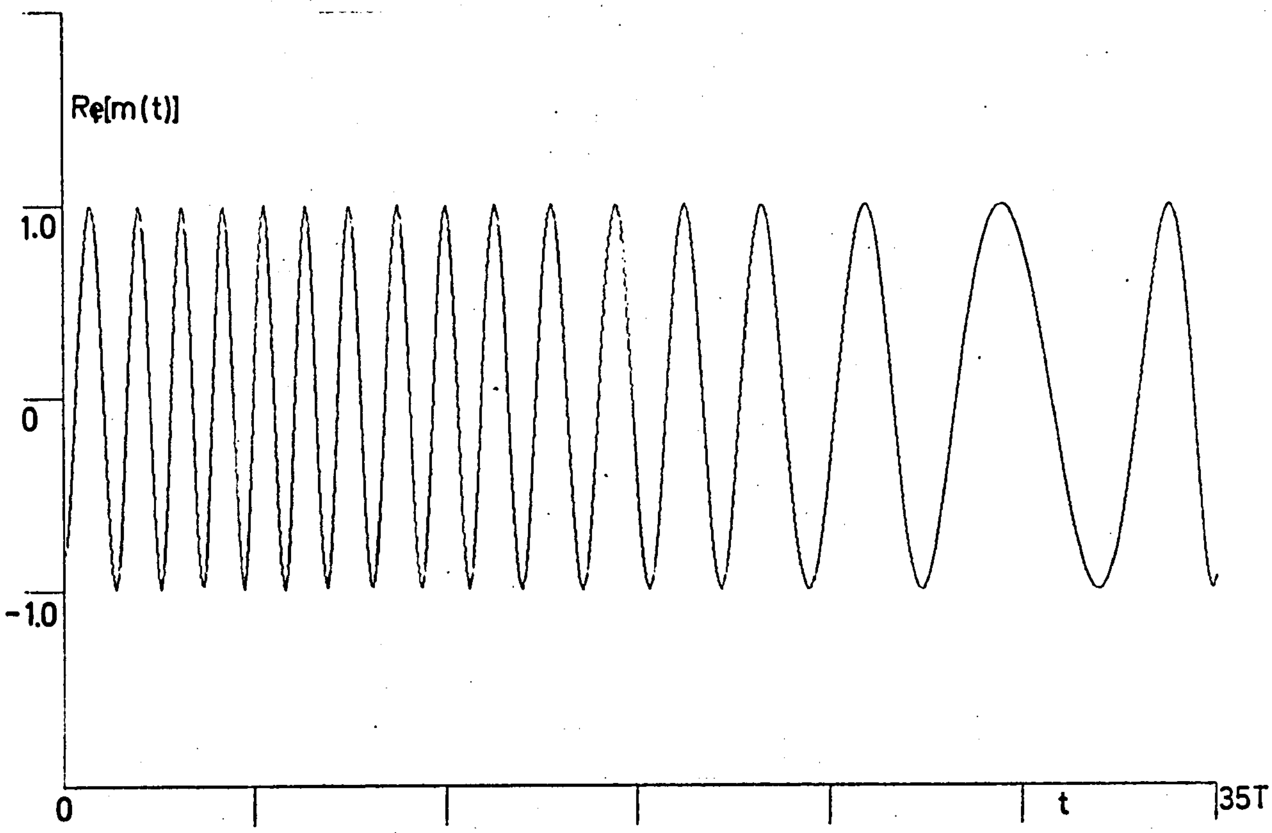


Fig. 5.3.9.a. Instantaneous frequency of synthesised signal: p = 2.0.



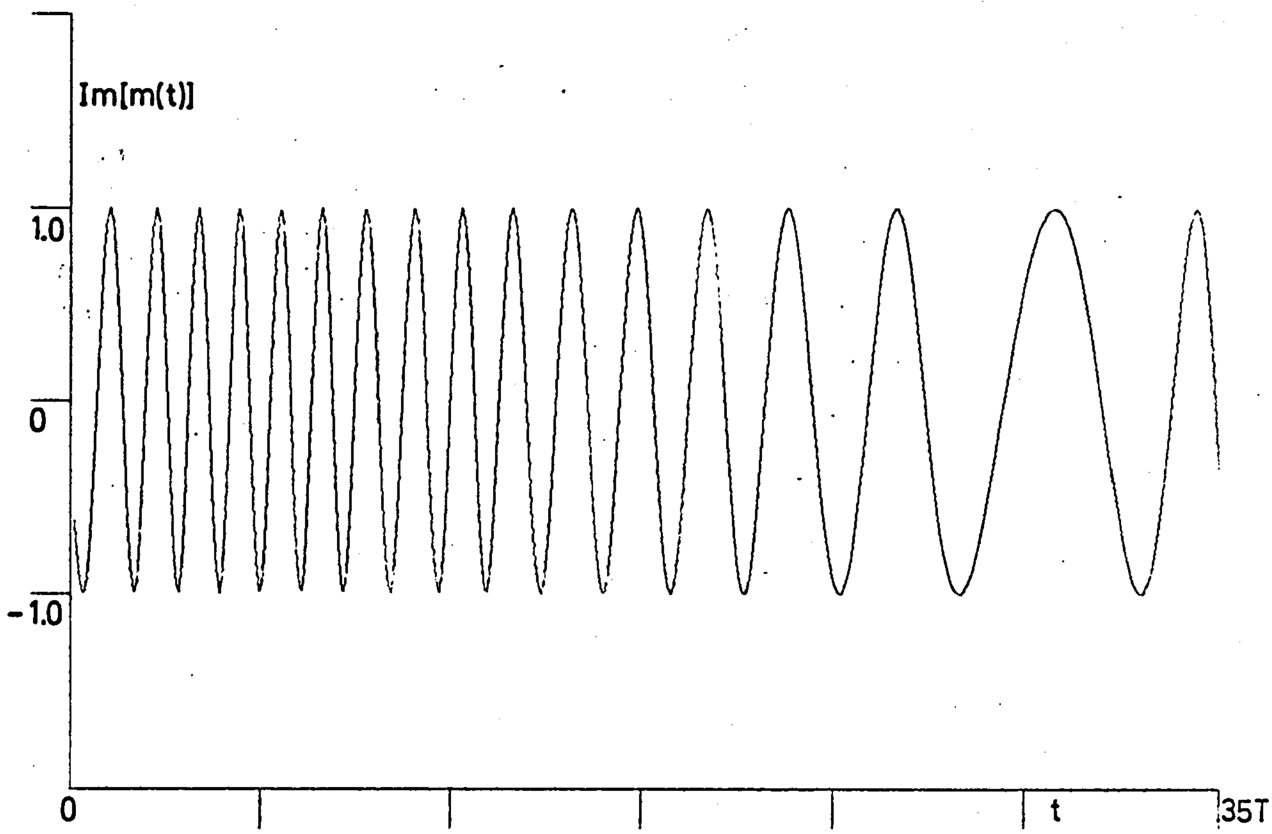


Fig.5.3.9.b. Real and imaginary parts of synthesised signal: p = 2.0.

5.4 Discussion

Inspection of the plotted t-f energy density distributions and the corresponding waveforms places in evidence the following points:

(i) For signals having a value of p of about 1.0, the instantaneous frequency function provides a good approximation to the required form.

(ii) For p small, $f_i(t)$ presents a poor approximation to the desired form. This is also true when p is large.

(iii) The t-f energy density distribution of the large p signals is concentrated in the region of the line formed by the instantaneous frequency function.

(iv) The envelope fluctuations of the large p signals are very small in magnitude. For practical purposes they can be viewed as purely frequency-modulated signals.

From (i) above, it can be concluded that the zero pattern synthesis scheme put forward in the previous section provides a means by which angle coded signals can be produced having a t-f energy density distribution which approximates a desired form.

The failure of $f_i(t)$ to provide a good approximation noted in (ii) when p is either too small or too large is due to V(t) failing in the first case to provide sufficient smoothing of the impulse train. In the second case, the width of V(t) results in excessive smoothing, so that $f_i(t)$ cannot follow the rapid changes in the desired instantaneous frequency form.

The "FM-like" nature of the angle coded signals with p large, noted in (iv) (as in section 4.3) implies that a transmitter for such signals can be simple and efficient; a class-C final amplifier can be used in the output stage.

The time-frequency dual of an angle coded signal is a Huffman pulse sequence. Huffman sequences have the property that their autocorrelation functions are zero for all time shifts greater than the length of one pulse of the sequence and less than one pulse length less than the sequence duration. Thus Huffman sequences (which because of this property are called "impulse equivalent") can provide excellent resolution between targets of known velocity and differing ranges. An angle coded signal, by the dual of this statement, should provide excellent resolution between targets of known ranges and differing velocities. This can be seen by considering the ambiguity function of an angle coded signal for zero time shift (corresponding to a target of known range):

$$\chi(o,\zeta) = \int_{-\infty}^{\infty} m(t) \quad m^*(t+o)e^{-j2\pi\zeta t} \quad dt \qquad 5.4.1$$

This is simply the Fourier transform of the squared envelope of the signal, which from section 4.2 (equation 4.2.6. and 7) is given by

$$\chi(o,\zeta) = P\left[\frac{-A}{1+A^2} \quad \delta(f-1/T) + \delta(f) - \frac{A}{1+A^2} \quad \delta(f+1/T)\right] \qquad 5.4.2$$

The ambiguity function, for zero time shift is thus zero for any frequency shift less than the bandwidth of the signal (apart from zero frequency shift). By analogy with the term "impulse equivalent" angle coded signals might be termed "spectral line equivalent". An angle coded signal is thus a good signal for resolving targets of known range and differing velocities. This is perhaps an academic observation, as it is difficult to visualise a situation in which the velocity of a target is unknown while its range remains known.

To conclude, a method has been presented by which angle coded signals can be synthesised so as to have a desired t-f energy density distribution. It must be admitted, however, that it is not obvious that an angle coded signal is greatly preferable to a signal produced (for example) by directly frequency modulating a carrier with the desired instantaneous frequency function. Attributes of the angle coded signal which might conceivably be useful are its strict bandlimitedness and its "spectral line equivalent" property. The author believes that the results of this chapter are of more general usefulness when translated from time language into frequency language and applied to the design of Huffman pulse sequences. This is done in appendix B.

CHAPTER 6

HILBERT TRANSFORM NETWORKS

6.1 Introduction

The work of chapters 4 and 5 of this thesis deal with analytic angle coded signals. To make practical application of analytic signal theory some form of Hilbert transform network is necessary. In addition to applications of the present work, Hilbert transform networks are of importance in :

(i) single-sideband data transmission

(ii) vestigial-sideband transmission of television signals[27]

(iii) minimum phase modulation[28]

(iv) real-zero interpolation[4]

As Gabor explained[16], a physical Hilbert transform network must inevitably involve some form of signal storage and delay. The difficulty of constructing a Hilbert transform network becomes evident when the "quantity of information" (in the Nyquist sense) that must be stored is considered.

The minimum bound on the delay which must be provided by a Hilbert transform network is set by the need for it to give a 90° phase shift at all the frequencies within its working range. This implies that the delay period cannot be less than one quarter period of the lowest frequency, f_L at which the network is to operate.

By the sampling theorem, to specify a lowpass signal whose upper frequency is f_H, the signal must be sampled at a rate of at least $2f_H$. The total equivalent number of signal samples, N_S, which must be stored by the network must exceed the product of the minimum time delay and the sampling frequency :

$$N_s > (2f_H)/(4f_L)$$

that is,

$$N_s > f_H / (2f_L) \qquad 6.1.1.$$

This formula illustrates at once the difficulties which are involved in making Hilbert transform networks to handle signals of large "frequency ratio" (f_H/f_L). For example, to form the Hilbert transform of signals covering the frequency range 50 Hz to 10 kHz requires at a minimum the storage of the equivalent of 100 signal samples. It is to be concluded that Hilbert transform networks to cope with large frequency ratio signals must inevitably be complicated, having many degrees of freedom.

The observation that the Hilbert transform networks of Gouriet[27] and Lyannoy[29] used an equivalent number of stored samples more than twice the minimum number set by the above theory led to the development of the "simplified quadrature network" which is reported in the reprinted letter which follows.

eprinted from *Electronics Letters*, Vol. 4, No. 6, 22nd March 1968

SIMPLIFIED QUADRATURE NETWORK

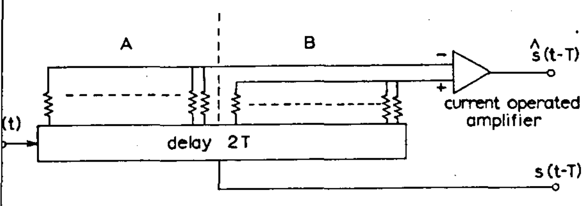
The number of sections in the delay line of a quadrature network can be halved by replacing one half of the line by an *RC* network. Measured characteristics of an experimental network are presented.

n ideal Hilbert transformer, or quadrature network, would ave constant gain and $\pi/2$ phase shift at all frequencies. The npulse response of such a network would be[1]

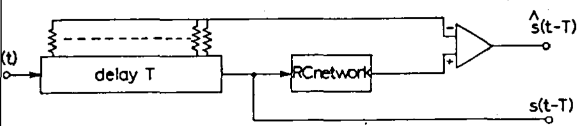
$$h(t) = \frac{1}{\pi t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

that its output signal would depend on both the entire ture and past of its input signal. Practical approximations the ideal quadrature network incorporate some form of elay to 'convert the future into the past'.[2] Thus a quadrature etwork has two outputs: a delayed replica of the input signal he in-phase output) and the quadrature version of this out ut. The time that the in-phase signal must be delayed epends on the lowest frequency that the quadrature net ork must handle.

An artificial delay line provides a practical means of roducing the required delay. Quadrature networks have been ade[1,3] which use the delay line in a transversal filter, here an approximation to the ideal impulse response is



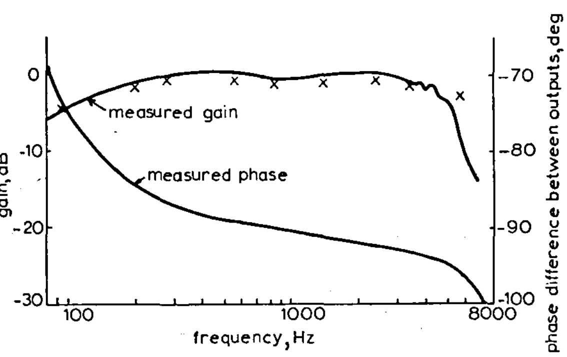
ig. 1A *Quadrature network using delay line*



ig. 1B *Simplified quadrature network*

ade straightforwardly by adjustment of the weighting esistors connected to delay-line taps (Fig. 1A). However, sed in this way, the total time delay provided by the line ust be twice the period T that the in-phase signal is delayed.

The number of sections required in the delay line can be alved by terminating it with a short circuit,[4] so that reflection ccurs and its length is effectively doubled. In practice, when his technique is used, attenuation in the delay line causes the mpulse response of the quadrature network to be not truly dd, which results in phase error. Using this method, the



ig. 2 *Frequency response of quadrature network*
× × × calculated gain

ttenuation cannot be compensated by adjustment of the eighting-resistor values.

The method proposed here allows compensation of ttenuation in the delay line. It can also be applied to digital Hilbert transform networks[5] and analogue quadrature net orks, in which the necessary time delay is obtained by means

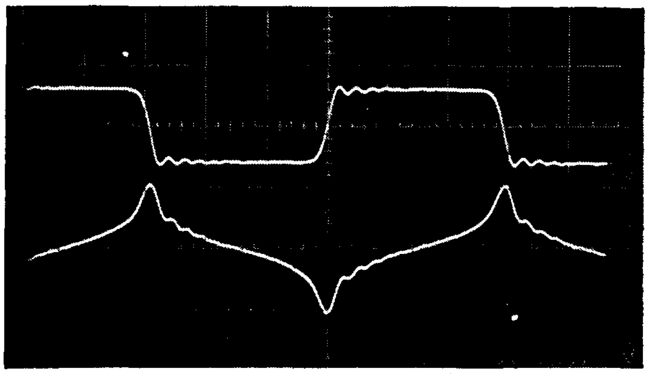


Fig. 3 *Response to square-wave input*

other than a (theoretically) lossless delay line (e.g. the transistor delay network of Krause[6]).

The quadrature output signal $\hat{s}(t-T)$ has two components: one depends on the future of the in-phase output signal $s(t-T)$, and one depends on its past. In Fig. 1A, the half A of the transversal filter forms the component which is dependent on the future of the in-phase output signal. The use of some form of storage such as an artificial delay line is unavoidable here. The half B forms the component dependent on the past of the in-phase signal. In principle, this half of the transversal filter can be replaced by a simpler network (Fig. 1B).

For the phase difference between the delayed input signal (i.e. the in-phase output) and the quadrature output to be $\pi/2$, it is sufficient that the impulse response at the quadrature output is odd about the ordinate $t = T$. However, unless this has the form described earlier, which, of course, requires an infinite delay, the gain will vary with frequency. Nevertheless, it is possible to find an impulse response of which the part corresponding to positive time ($t > 0$) can be realised by a simple *RC* network and yet which yields a quadrature network whose gain is sensibly constant over some quite wide frequency range, even though its impulse response is quite a crude approximation to the ideal.

As an example, for the construction of an experimental network, this part of the impulse response was chosen to be of the form

$$\left.\begin{aligned} g(t) &= A e^{-\omega_a t} + B e^{-\omega_b t} && \text{for } t \geqslant 0 \\ &= 0 && \text{for } t < 0 \end{aligned}\right\} \quad . \quad . \quad . \quad (2)$$

which can be realised without difficulty using a simple active *RC* network to replace the second half of the transversal filter. The transfer function of a network with the odd impulse response of which eqn. 2 is the positive time component is

$$H(j\omega) = K \frac{j\omega(\omega^2 + \omega_c^2)}{(\omega^2 + \omega_a^2)(\omega^2 + \omega_b^2)} \quad . \quad . \quad . \quad . \quad (3)$$

The operating-frequency range of the quadrature network was set by the delay line used; the lower limit was set by the limited delay of the line and the upper limit by its cutoff frequency. ω_a and ω_b were chosen to lie within the working angular-frequency range, close to its limits. ω_c was made the geometric mean of ω_a and ω_b, so that the gain/frequency characteristic had two peaks of equal height.

Fig. 2 shows the calculated variation of gain with frequency of a quadrature network whose impulse response is

$$h(t) = k(e^{-1780|t|} + 10e^{-17800|t|}) \operatorname{sgn}(t) \quad . \quad . \quad . \quad (4)$$

this being the inverse transform of eqn. 3 with the chosen values of ω_a, ω_b and ω_c.

Also shown are the measured gain and phase characteristics of an experimental quadrature network designed to have this impulse response. Fig. 3 shows the in-phase and the quadrature output waveforms in response to a square-wave input signal.

The experimental quadrature network uses an active *RC* network in place of the second half of the transversal filter. The first half uses a constant-*k* artificial delay line (used because suitable inductors were available), giving a total time delay of 1·85 ms. The weighting resistors were adjusted to compensate for the increase in attenuation along the line.

The delayed in-phase signal is differentiated and added to the output signal.[3] This improves the frequency response and the closeness of the phase difference between the outputs to $\pi/2$.

Truncation of one of the components of the impulse response due to the finite length of the delay line produces a deviation in the phase difference between outputs from $\pi/2$. Calculation shows that the phase error due to this cause is small, in the working frequency range, compared with that due to the imperfect characteristics of the delay time. Phase error is also introduced by the time-discrete approximation to the continuous impulse response made in the transversal filter.

The writer wishes to thank J. W. R. Griffiths for helpful discussions.

M. H. ACKROYD

15th February 1968

Department of Electrical Engineering
University of Technology,
Loughborough, Leics., England

References

1 GOURIET, G. G., and NEWELL, G. F.: 'A quadrature network for generating vestigial-sideband signals', *Proc. IEE*, 1960, **107B**, pp. 253–260

2 GABOR, D.: 'Theory of communication', *J. IEE*, 1946, **93**, Part III, pp. 429–457

3 LYANNOY, V. P.: 'A wide-band phase shifter containing an artificial delay line in an OAPM system', *Izv. VUZ Radiotekhnika*, 1960, **3**, p. 247

4 VON URFF, C. A., and ZONIS, F. I.: 'The square-law single-sideband system', *IRE Trans.*, 1962, **CS-11**, pp. 257–267

5 STOREY, D. J., and DONNE, R. F.: 'Synthesis of the Hilbert transform of a train of rectangular pulses', *Electronics Letters*, 1967, **3**, pp. 126–128.

6 KRAUSE, G.: 'Analog-Speicherkette: eine neuartige Schaltung zum Speichern und Verzögern von Signalen', *ibid.*, 1967, **3**, pp. 544–546

6.3 Realisation of Hilbert Transform Networks.

The following approaches to the construction of Hilbert transform networks (as opposed to 90° phase-difference networks) have been suggested or used :

(i) Transversal filters using artificial delay lines[27,28,29]

(ii) Time varying networks[30] (in essence, combined SSB modulator and demodulator)

(iii) Expansion of the impulse response in orthonomal functions[31]

(iv) Nonlinear networks with shift registers[32] (for use with clocked binary inputs only)

(v) 90° phase difference network with transversal filter delay equaliser[33].

(vi) Shift register transversal filters[34] (for use with clocked binary inputs).

(vii) On-line digital computer.

In the future, the last two methods are likely to be of importance. Unlike an analogue artificial delay line, there is no limit to the time delay that can be obtained using a shift register (while maintaining a given time resolution). In conjunction with a delta-modulator[35] or a conventional analogue-digital converter[34] the shift register transversal filter scheme can be adapted to handle analogue signals. This scheme also has the attraction that it could be realised in micro-miniature form.

CHAPTER 7

Conclusions

This thesis reports investigations into the problems of specifying bandlimited signals in terms of their complex-time plane zeros and into the properties and uses of signals synthesised in terms of their zeros.

Zero-based signal theory is evidently no panacea. In particular, the relations between linear effects (filtering and the addition of noise) seem to be particularly intractable. This is perhaps to be expected; the polynomial root-coefficient relationship problem has interested mathematicians for many years and yet remains little understood. The problem of the effects of linear filtering on frequency modulated signals, too, although much studied, is not well understood. In view of the complexity of noise effects and the non-rectangularity of their spectra, there is no clear reason why the angle coded signals of chapter 4 should be used in a communication system, rather than more conventional signals which are simpler to generate.

Zero based signal theory seems most likely to be of use in the study of situations where signals are multiplied or where linear operations are performed on the logarithms of signals. One such situation is the study of homomorphic filtering[37] where the logarithm of a signal is applied to a linear filter. The homomorphically filtered signal is produced by forming the exponential of the output of the linear filter. However it is not obvious how to apply zero based signal theory as the output of a homomorphic filter is generally not bandlimited even if its input is.

The relationship between the zero pattern of a signal and its t-f energy density distribution might be capable of development in greater detail. One can speculate that this might be useful in studying speech problems.

The theory of short-time spectra, which is of practical use in the experimental study of time varying systems, is far from being complete. In particular, procedures for computing "best" estimates of the spectra of time varying processes from experimental measurements need to be developed.

REFERENCES

1. E.C. Titchmarsh, "The zeros of certain integral functions", Proc.Lond.Math.Soc., 25, pp 283-302 (1926).

2. R.E.A.C. Paley and N. Wiener, Fourier Transforms in the Complex Domain (1934)

3. F.E. Bond and C.R. Cahn, "On Sampling the Zeros of Bandwidth Limited Signals", I.R.E. Trans., IT-4, pp 110-113 (1958)

4. H.B. Voelcker, "Toward a Unified Theory of Modulation", Proc.I.E.E.E., 54, pp 340-353 and pp 735-755 (1966).

5. I.E.E.E. Trans. on Audio and Electroacoustics, AU-15, No.2 (1967). (Special issue on the fast Fourier transform).

6. V.N. Tetarev, "The connection between the distribution function for the zeros of a harmonic signal and its frequency characteristic in the case of extreme amplitude limiting", Telecommunications and Radio Engineering, Pt.I, 19, pp 62-64 (1965)

7. G. Szego, Orthogonal Polynomials (American Mathematical Society, 1959).

8. A.M. Ostrowski, Solution of Equations and Systems of Equations , (Academic Press, 1966).

9. J.G. Truxal, Automatic Control System Synthesis , (McGraw Hill, 1955).

10. I.E. Kazakov, "Generalisation of the method of statistical linearisation to multidimensional systems", Automat. Remote Control, 26, pp 1201-1206 (1965).

11. J.C. West, "The equivalent gain matrix of a multivariable non-linearity", Measurement and Control, 2, pp T1-T5(1969)

12. E.C. Cherry, "Quelques remarques sur le temps considéré comme variable complexe", L'Onde Electrique, pp 7-13 (1954)

13. E.C. Cherry and V.J. Phillips, "Some possible uses of single sideband signals in formant tracking systems", J. Acous. Soc. Am., 33, pp 1067-1077 (1961)

14. K. Hiramatsu, T. Matsuno, C.L. Coates, "Analysis of the phase function in single sideband signal and its application to speech", Electronics Laboratory, University of Texas, Technical Report No. 24.

15. J. Dugundji, "Envelopes and pre-envelopes of real waveforms", Trans. I.R.E., IT-4, pp 53-57 (1958)

16. D. Gabor, "Theory of communication", J.I.E.E., 93, pt. 3, pp 429-457 (1946)

17. P.M. Woodward, Probability and information theory with applications to radar, (Pergamon, 1953)

18. J. Ville, "Theorie et applications de la notion de signal analytique", Cables et Transmission, 2, pp 61-74(1948)

19. C.E. Shannon and W. Weaver, The Mathematical theory of communication , (Illinois U.P., 1949)

20. A.A. Kharkevich, Spectra and Analysis (Translated from the Russian), (Consultants Bureau, 1960).

21. A. Papoulis, Probability, Random Variables and Stochastic Processes , (McGraw-Hill, 1965).

22. C.E. Cook and M. Bernfield, Radar Signals (Academic Press, 1967).

23. D.A. Huffman, "The generation of impulse equivalent pulse trains", Trans. I.R.E., IT-8, pp 510-516 (1962).

24. D.E. Weston, "New calculation methods for echo-ranging ambiuity", Trans. I.E.E.E., IT-13, pp 432-441 (1967).

25. H. Inose and Y. Yasuda, "A unity bit coding method by negative feedback", Proc. I.E.E.E., 51, pp 1524-1535 (1963)

26. R.C. Singleton and T.C. Poulter, "Spectral analysis of the call of the male killer whale", Trans. I.E.E.E., AU-15, pp 104-113 (1967).

27. G.G. Gourlet and G.F. Newell, "A quadrature network for generating vestigial sideband signals", Proc. I.E.E., 107 B, pp 253-260 (1960).

28. K.H. Powers, "The compatibility problem in single sideband transmission, Proc. I.R.E., 48, pp 1431-1435 (1960.

29. V.P. Lyannoy, "A wideband phase shifter containing an artificial delay line in an OAPM system", Izv.VUZ Radiotekhnika, 3, p 247 (1960).

30. S. Deutsch, "A method of obtaining 90° phase shift of constant amplitude", Proc. I.R.E., 41, p 797 (1953).

31. E.G. Gilbert and J. Otterman, "The synthesis of linear filters with real or imaginary transfer functions", Trans. A.I.E.E., . ,pp 323-330, (1960).

32. D.J. Storey and R.F. Donne, "Synthesis of the Hilbert Transform of a train of rectangular pulses", Electronics Letters, 3, pp 126-128 (1967).

33. R.K.P. Galpin, Private Communication , 1968.

34 H.B. Voelcker, "Generation of digital signalling waveforms", I.E.E.E. Trans., COM-16, pp 81-93, (1968).

35. G.K. McAuliffe, "ADEM - An adaptively data equalised high speed MODEM", Proc. Mil. E. Conv., pp 332-337 (1964).

36. D.E. Vakman, Sophisticated Signals and the uncertainty principle in radar, (Translated from the Russian), (Springer-Verlag, 1968).

37. A.V. Oppenheim, R.W. Schafer, T.G. Stockham, "Nonlinear filtering of multiplied and convolved signals", Proc.I.E.E.E., 56, pp 1264-1291, (1968).

38. A.W. Rihaczek, "Signal Energy Distribution in time and frequency", I.E.E.E. Trans., IT-14, pp 369-374, (1968).

39. C.W. Helmstrom, Statistical theory of signal detection, (Pergamon, 1960).

40. D. Thorn, An introduction to generalised circuits, p 114 et seq., (Wadstock, 1963).

41. M.J. Levin, "Instantaneous spectra and ambiguity functions", I.E.E.E. Trans., IT-10, pp 95-97, (1964).

42. V.L. Larrowe, "Analog computer measurement of time-varying power spectra", (Ph.D. Thesis, University of Michigan, 1964).

APPENDIX 'A'

The Time-Frequency Energy Distribution of Signals

A.1. Introduction

The concept of the time-frequency (t-f) energy density distribution of a signal was found useful in chapter 5. Section A2 of this appendix is a reprint of a paper which provides an introduction to the notion of the t-f energy density distribution of a real signal. The complex t-f energy density distribution of a complex signal has been discussed by Rihaczek[38] but without emphasis of its physical interpretation. In order to give a physical interpretation, section A3 discusses the significance of power when analytic signals are involved. Finally, section A4 presents some properties of the complex t-f energy density distribution of a complex signal.

U.D.C. 621.095.21

Instantaneous and Time-Varying Spectra—An Introduction

By

M. H. ACKROYD, B.Sc.†

Several definitions of 'time-varying', 'short-time' and 'instantaneous' spectra exist. The paper relates these to the time-frequency energy distribution of a signal and to the time-varying power spectrum of a non-stationary random process. The treatment emphasizes physical interpretation, rather than mathematical rigour.

List of Principal Symbols

$e(t,f)$	time-frequency energy density distribution of a signal
$e_s(t,f)$	version of $e(t,f)$ smoothed in the t-direction.
E	energy
E_T	total energy
f	frequency
Δf	frequency interval
$G_t(f)$	'short-time' spectrum according to Fano.
$h(t)$	impulse response
$H(f)$	transfer function corresponding to $h(t)$
$i(t)$	current
$p(t,f)$	'instantaneous spectrum' according to Page
$P(t,f)$	time dependent power spectrum of a random process
$q(t)$	power
$R(t,\tau)$	time dependent autocorrelation function
$\mathrm{Re}(\cdot)$	'real part of $(\cdot)$'
$s(t)$	signal as a time function
$S(f)$	Fourier transform of $s(t)$
t	time
Δt	time interval
T	effective duration of a signal or an impulse response
$u(t)$	unit step function
W	effective bandwidth of a signal or filter
$X(f)$	reactance
$Y(f)$	admittance
$\overline{(\cdot)}$	'ensemble average of $(\cdot)$'

1. Introduction

For the usual purposes of time-invariant system analysis it is sufficient to consider a signal as a function of time alone or as a function of frequency alone. However, in certain situations (which are usually associated with the study of time-*varying* linear systems) one is concerned with both the time and the frequency characteristics of a signal at once. One such situation is in the study of speech and the vocal mechanism.[1,2] Another is in radar system theory where signals suffer both a time delay and a frequency shift between being transmitted and subsequently being received after reflection by a moving target.[3,4]

† Department of Electronic and Electrical Engineering, University of Technology, Loughborough, Leicestershire.

Intuitively, it is evident that the energy of a signal does have a distribution in time and frequency in some sense. For example, the acoustic energy of a short blast on a whistle is 'obviously' located both at the frequency of the note which is blown *and* at the epoch in time when the whistle is sounded.

Many definitions of 'short-time', 'instantaneous' and time-varying power and energy spectra can be found in the literature.[5–9] In addition various instruments have been constructed to measure short-term spectra.[10,11] Thus further implicit definitions of short-term spectra couched in the mechanisms and parameters of these instruments have been introduced. The questions arise: How are these definitions related? Can they be regarded as approximations to some 'exact' or 'true' instantaneous spectrum? How are they to be interpreted in physical terms?

In what follows it is shown that an exact definition of $e(t,f)$, the energy density distribution in time and frequency of a signal, can be made. This definition applies whether the signal is deterministic, such as a pulse of specified shape, or whether it is random. If signals originate in a random process, then each individual signal from the process has its own energy distribution in time and frequency. The time-varying power spectrum of the random process, $P(t,f)$, is found by averaging all the possible energy density functions in accordance with their probability of occurrence.[12–16] In other words, $P(t,f)$ is the average of $e(t,f)$ over the ensemble. $P(t,f)$ itself, for a well defined random process is not a random function any more than is, for example, the mean of the process.

The short-term spectrum measured according to a particular definition can be interpreted in two ways: it provides an approximation to the time-frequency energy density distribution of the signal and it can also provide an estimate of the time varying power spectrum if the signal is a random one.

D

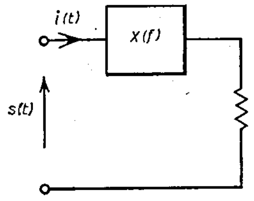


Fig. 1. Idealized circuit.

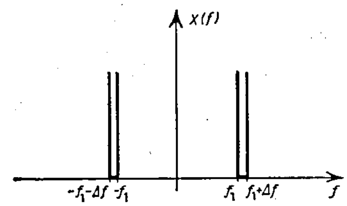


Fig. 2. Reactance of idealized circuit element.

2. Energy Density of a Signal in Time and Frequency

This section derives an expression for the distribution of the energy of a signal in the time-frequency plane.[18]

Replacing the real signal by the analytic signal† would have allowed the mathematics to have been made slightly more elegant and would have given substantially similar results. However, the conceptual difficulties involved in assigning physical meanings to the various results would have been increased (although this can be done in a satisfactory way). To minimize the difficulties of visualizing the results, the analysis is presented in terms of the real signal.

The definition of the energy density function can be understood by considering the circuit shown in Fig. 1. $X(f)$ is supposed to be a purely reactive circuit element having infinite reactance at all frequencies except over a narrow band where it has zero reactance. The reactance of this element is shown diagrammatically in Fig. 2. The admittance between the terminals of the circuit is

$$\begin{aligned} Y(f) &= 1, \quad f_1 \leqslant |f| \leqslant f_1 + \Delta f \\ &= 0, \quad \text{otherwise.} \end{aligned} \qquad \ldots\ldots(1)$$

Such an idealized admittance function is not physically realizable. However, it is nevertheless possible to *calculate* the current that would flow in the circuit in response to a given voltage waveform.

From the waveform of a signal, $s(t)$, the distribution of its voltage as a function of frequency is given by the Fourier transform

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft)dt \qquad \ldots\ldots(2)$$

Conversely, the waveform is given by the inverse transform

$$s(t) = \int_{-\infty}^{\infty} S(f) \exp(j2\pi ft)df \qquad \ldots\ldots(3)$$

Applying the signal $s(t)$ as a voltage to the terminals of the circuit, the current that flows only has frequency components lying in the frequency range f_1 to $f_1+\Delta f$. The waveform of the current is given by the inverse Fourier transform

$$\begin{aligned} i(t) &= \int_{f_1}^{f_1+\Delta f} S(f) \exp(j2\pi ft)df + \int_{-f_1-\Delta f}^{-f_1} S(f) \exp(j2\pi ft)df \\ &= 2 \operatorname{Re}\left[\int_{f_1}^{f_1+\Delta f} S(f) \exp(j2\pi ft)df\right] \end{aligned} \qquad \ldots\ldots(4)$$

The power entering the terminals of the circuit is given by the product of the applied voltage and the current which flows:

$$q(t) = s(t)i(t) \qquad \ldots\ldots(5)$$

Thus the energy which enters the circuit between a time t_1 and $t_1+\Delta t$ is given by

$$E = \int_{t_1}^{t_1+\Delta t} s(t)i(t)dt \qquad \ldots\ldots(6)$$

This quantity is *defined* as twice the part of the energy of the signal contained in the frequency range f_1 to $f_1+\Delta f$ and in the time interval t_1 to $t_1+\Delta t$ in the t–f plane. Thus the energy density at t_1 and f_1 is given by

$$e(t_1, f_1) = \lim_{\substack{\Delta t \to 0 \\ \Delta f \to 0}} \tfrac{1}{2}E/\Delta t \Delta f \qquad \ldots\ldots(7)$$

Substituting equations (4) and (6) into (7) and writing t and f in place of t_1 and f_1 gives

$$e(t, f) = s(t) \operatorname{Re}[S(f) \exp(j2\pi ft)] \qquad \ldots\ldots(8)$$

This is the required expression for the energy density distribution in time and frequency.†

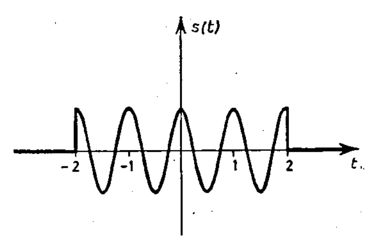


Fig. 3. Pulsed carrier signal of example.
(Shown for $T = 4.0$, $f_0 = 1{\cdot}0$.)

† The analytic signal[3,5,19] is a complex function of time. Its relation to the real signal is simple; the analytic signal is derived by doubling the amplitudes of the positive frequency components and eliminating the negative frequency components of the real signal. As an example, the analytic signal corresponding to the real signal $A \cos \Omega t$ [$= A/2(\exp j\Omega t + \exp -j\Omega t)$] is $A \exp j\Omega t$.

† Reference 18 derives the complex energy density function of the analytic signal. The present definition appears in reference 8.

Example

The pulsed-carrier signal shown in Fig. 3 can be written (using Woodward's *rect* function[3]†)

$$s(t) = A \operatorname{rect}(t/T) \cos 2\pi f_0 t.$$

The corresponding energy density function is

$$e(t,f) = \frac{A^2}{4} \operatorname{rect}(t/T) \times$$
$$[\cos 2\pi(f+f_0)t + \cos 2\pi(f-f_0)t] \times$$
$$\left[\frac{\sin \pi(f+f_0)T}{\pi(f+f_0)} + \frac{\sin \pi(f-f_0)T}{\pi(f-f_0)}\right]$$

which is shown graphically in Fig. 4. Note that as T, the duration of the pulse, is increased the spread of the energy density on each side of the carrier is reduced. Eventually, as T becomes infinite and the signal becomes a pure cosine wave, the whole of the energy density becomes concentrated at the frequency of the carrier.

The example illustrates that, for values of t where the signal is zero, $e(t,f)$ *is also zero*. If the waveform of the example were applied to a bank of bandpass filters and square-law envelope detectors in the arrangement traditionally used to measure 'short-term' spectra, the output voltage of each filter would not immediately fall to zero at the end of the input pulse. Instead, it would die away gradually due to the response time of the filter. In fact, the function of time and frequency measured by a filter bank is an approximation to $e(t,f)$, as explained in Section 4 of this paper.

The function $e(t,f)$ is symmetrical about the time axis; the negative frequencies merely mirror the positive frequencies:

$$e(t,f) = e(t,-f).$$

This results directly from the symmetry properties of the Fourier transforms of real signals. Thus in plotting $e(t,f)$ it is only necessary to consider positive values of f.

Integrating $e(t,f)$ over all values of f gives the energy density (that is, the power) of the signal at time t. This is readily verified by integrating equation (8):

$$\int_{-\infty}^{\infty} e(t,f)\mathrm{d}f = s(t)^2 \qquad \text{......(9)}$$

Similarly, the energy density spectrum of the signal is got by integrating $e(t,f)$ over all time;

$$\int_{-\infty}^{\infty} e(t,f)\mathrm{d}t = |S(f)|^2 \qquad \text{......(10)}$$

† The *rect* function is defined by

$$\operatorname{rect}(t/T) = \begin{matrix} 1, & |t| < T/2 \\ 0, & |t| > T/2. \end{matrix}$$

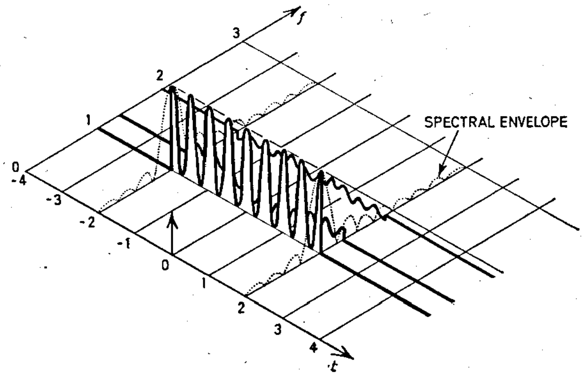


Fig. 4. Sketch of $e(t,f)$ of example. Note: Cross-sections at $f = 1{\cdot}0, 1{\cdot}3, 1{\cdot}9$ are shown.

The total energy of the signal is thus found by integrating $e(t,f)$ over the whole time-frequency plane;

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e(t,f)\mathrm{d}t\,\mathrm{d}f = E_T \qquad \text{......(11)}$$

There exists a misconception that it is not possible to measure exactly the t–f energy density function of a given waveform and that this is a consequence of Gabor's uncertainty relation.[19] However, the uncertainty principle of waveform analysis is not concerned with the measurement of t–f energy density distributions: instead, it states that if the effective bandwidth of a signal is W then its effective duration cannot be less than about $1/W$ (and conversely, the bandwidth of a signal of effective duration T cannot be less than about $1/T$). In fact, given the waveform of a signal, its t–f energy density distribution can in principle be computed *exactly* using the defining relation (8). However, if $e(t,f)$ is measured *approximately* by the use of a bank of bandpass filters each having a bandwidth W, then it is evident that variations of $e(t,f)$ in the f-direction which are finer than W will be obscured. If the effective duration of the impulse response of each filter is T, then details of $e(t,f)$ in the t-direction which are finer than T will be obscured. The only sense in which the uncertainty principle applies to the measurement of t–f energy distributions is that it prohibits filters from having both short impulse responses *and* narrow bandwidths (as the impulse response of a filter, like any other signal, is subject to the $TW \geqslant 1$ uncertainty relation). Thus a bank of fixed bandwidth filters cannot provide both good spectral and good temporal resolution (although the resolution may be entirely adequate for practical purposes, of course).

Exactly parallel limitations on resolution apply when short term spectral analysis is performed by the process of multiplying the time waveform by contiguous 'time window' functions to produce a succession of short waveforms which are each subjected to

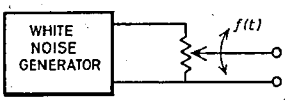


Fig. 5. Non-stationary (time varying) noise source.

Fourier analysis to produce a succession of cross-sections of a short-term spectrum. (The use of this procedure is described in detail in reference 20.)

3. Power Spectra of Non-stationary Processes[12–17]

A non-stationary process is one whose statistics vary with time. Here, one is interested in processes which are non-stationary in the wide sense, that is, with processes whose mean or whose autocorrelation vary with time. Figure 5 shows a model of a particular non-stationary process. Stationary white noise (from an ideal noise diode held at constant temperature, for example) is applied to a time-varying attenuator. Intuitively, one feels that the output signal has a time-varying spectral density. Many non-stationary processes can be represented as a time-varying network excited by a stationary random process.

The autocorrelation of a real process is defined

$$R(t,\tau) = \overline{s(t)s(t+\tau)} \qquad \ldots\ldots(12)$$

The line above the product indicates that the ensemble average is to be taken, that is, the product is to be averaged over all possible pairs of values of $s(t)$ and $s(t+\tau)$ in accordance with their probability of occurring jointly. $R(t,\tau)$ exists for signals whose total energy is finite and also for signals whose total energy is infinite but whose average power is finite.

Taking the Fourier transform of equation (12) with respect to τ and denoting it by $P(t,f)$ gives

$$\begin{aligned} P(t,f) &= \int_{-\infty}^{\infty} R(t,\tau)\exp(-\mathrm{j}2\pi f\tau)\,\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} \overline{s(t)s(t+\tau)}\exp(-\mathrm{j}2\pi f\tau)\,\mathrm{d}\tau \qquad \ldots\ldots(13) \end{aligned}$$

Assuming that the order of averaging and integrating can be interchanged, and that the F-transform of $s(t)$ exists,† equation (13) becomes

$$\begin{aligned} P(t,f) &= \overline{s(t)\int_{-\infty}^{\infty} s(t+\tau)\exp(-\mathrm{j}2\pi f\tau)\,\mathrm{d}\tau} \\ &= \overline{s(t)S(f)\exp(\mathrm{j}2\pi ft)} \qquad \ldots\ldots(14) \end{aligned}$$

Equation (13) is the time-varying analogy of the Wiener-Khintchine relation, which states that the power spectrum of a stationary random process is the Fourier transform of its autocorrelation function. $P(t,f)$ is generally a complex function, in contrast to the purely real power spectra of stationary processes. This happens as $R(t,\tau)$, unlike the stationary case, is not generally an even function in τ.

The real part of $P(t,f)$ is the ensemble average value of $e(t,f)$ as given by equation (8). The physical interpretation of this quantity is that Re $[P(t_1,f_1)]\Delta f$ is expected value of the power entering the circuit of Fig. 1 at time t_1 when $s(t)$ is applied as a voltage across its terminals. The imaginary part of $P(t,f)$ is of no direct interest here, but the following brief discussion explains its significance in a different context.

Had the analytic signal been used in the foregoing arguments, then $P(t,f)$ would have played an analogous part to the 'complex power' of a.c. phasor theory.[21] Its real part would have represented the power entering the circuit of Fig. 1 and its magnitude the volt-amperes. As in the case of complex power, the imaginary part has little physical significance but is useful in analysis in that it provides the discrepancy between the real part and the magnitude of $P(t,f)$ in the correct way.

Relations similar to (9), (10) and (11) apply to $P(t,f)$.[16,17] The ensemble mean power of the signal is given by

$$\int_{-\infty}^{\infty} P(t,f)\,\mathrm{d}f = \overline{s(t)^2} \qquad \ldots\ldots(15)$$

The expected energy of the signal, if it is finite, is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(t,f)\,\mathrm{d}f\,\mathrm{d}t = E_{\mathrm{T}} \qquad \ldots\ldots(16)$$

By integrating $P(t,f)$ with respect to time in the case of a signal of finite energy, or averaging over time in the case of a signal of finite power, it is possible to define a mean energy (or power) spectrum of the process. This is discussed in detail in references 16 and 17.

4. Short-term Spectra—Relations between Definitions

Many definitions of short-term spectra exist, as was mentioned in the Introduction. Some were introduced to provide Wiener–Khintchine-like relationships with corresponding short-term autocorrelation functions[7,9] and others to accord with the results of physical measurement, being cast in terms of the past and present values of the signal.[6,9] This Section shows that the various definitions can be regarded as approximations to $e(t,f)$.

The block diagram of the system which produces $e(t,f)$ from $s(t)$ is shown in Fig. 6. The relationship between this block diagram and equation (8) becomes

† These assumptions are always satisfied for physical signals. The requirement for the Fourier transform of $s(t)$ to exist can be relaxed by the use of the generalized harmonic analysis.

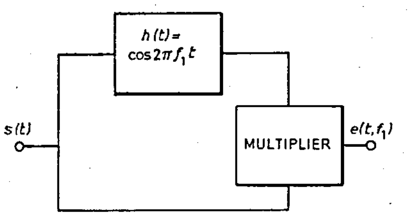


Fig. 6. Scheme for measuring $e(t, f)$.

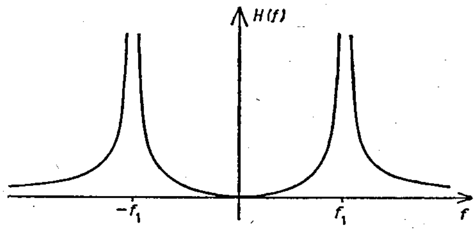


Fig. 8. Transfer function of physically realizable filter of equation (23).

clear when it is realized that Re $[S(f_1)\exp(j2\pi f_1 t)]$ is simply the response to $s(t)$ of the system whose impulse response is

$$h(t) = \cos 2\pi f_1 t \qquad \text{......(17)}$$

This is not a physically realizable system as its impulse response is non-zero for negative t. Nevertheless, its output can be calculated or it can be simulated.[24] The transfer function corresponding to the impulse response $h(t)$ is

$$H(f) = \tfrac{1}{2}[\delta(f+f_1)+\delta(f-f_1)] \qquad \text{......(18)}$$

which represents a filter having an infinitely narrow bandwidth (Fig. 7).

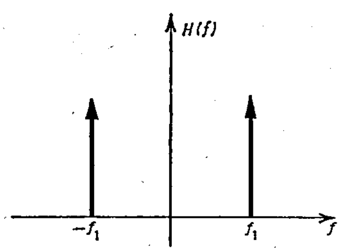


Fig. 7. Transfer function of 'ideal' filter.

4.1. *Page's Definition*

Page[6] defined an 'instantaneous power spectrum' in terms of a 'running transform';

$$S_t(f) = \int_{-\infty}^{t} s(\tau)\exp(-j2\pi f\tau)d\tau \qquad \text{......(19)}$$

The running transform depends only on the past and present values of the signal and not on its future values. His instantaneous spectrum was defined as the rate of change with time of the squared magnitude of $S_t(f)$:

$$p(t,f) = \frac{\partial}{\partial t}|S_t(f)|^2 \qquad \text{......(20)}$$

Equations (19) and (20) can be manipulated into the alternative form

$$p(t,f) = 2s(t)\,\mathrm{Re}\,[S_t(f)\exp(j2\pi ft)] \qquad \text{......(21)}$$

Now Re $[S_t(f_1)\exp(j2\pi f_1 t)]$ can be shown to be the response to $s(t)$ of a filter whose impulse response is

$$h(t) = u(t)\cos 2\pi f_1 t \qquad \text{......(22)}$$

where $u(t)$ is the unit step function. This impulse response is physically realizable in the sense that it is zero for negative t. The transfer function corresponding to this impulse response is

$$H(f) = jf/2\pi(f_1^2-f^2) \qquad \text{......(23)}$$

whose magnitude is shown plotted in Fig. 8.

$p(t, f_1)$ is thus produced (except for a scale factor) by a system having the block diagram of Fig. 6 in which the 'ideal' filter of equation (18) is replaced by the less perfect (but realizable) filter of equation (23). $p(t, f)$ can therefore be thought of as an approximation to $e(t, f)$ with poorer spectral resolution.

4.2. *Fano's Definition*

Fano[7,9] defined a 'short-time' power spectrum by the expression

$$G_t(f) = \left|\int_{-\infty}^{t} s(\tau)\exp[(-\alpha+j2\pi f)(t-\tau)]d\tau\right|^2 \qquad \text{......(24)}$$

(A scale factor which is not of interest here is omitted.) This definition was chosen by Fano as being the Fourier transform of a 'short-time' autocorrelation function. Equation (24) can be rewritten

$$G_t(f) = \left|\int_{-\infty}^{t} s(\tau)\exp[-\alpha(t-\tau)]\cos 2\pi f(t-\tau)d\tau\right|^2 + \left|\int_{-\infty}^{t} s(\tau)\exp[-\alpha(t-\tau)]\sin 2\pi f(t-\tau)d\tau\right|^2 \qquad \text{......(25)}$$

Thus $G_t(f_1)$ is produced by the system whose block diagram is shown in Fig. 9. The signal is applied to two filters whose impulse responses are $u(t)\exp(-\alpha t)\cos 2\pi f_1 t$ and $u(t)\exp(-\alpha t)\sin 2\pi f_1 t$ and the outputs of these filters are squared and added to yield $G_t(f_1)$.

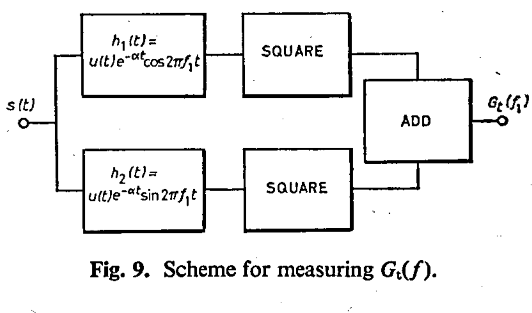


Fig. 9. Scheme for measuring $G_t(f)$.

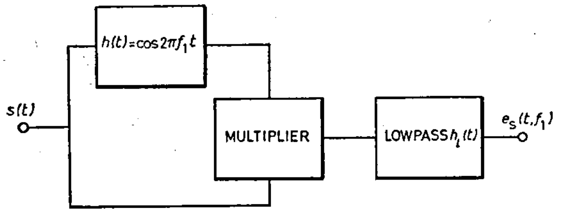


Fig. 10. Scheme for measuring $e_s(t,f)$ ($e(t,f)$ smoothed in the t-direction).

Suppose that in measuring $e(t,f)$ the fine structure in the t-direction is of no interest. This irrelevant detail can be removed by smoothing $e(t,f)$ in the t-direction with a low-pass filter. Figure 10 shows the block diagram of a system to produce $e_s(t,f_1)$, the smoothed version of $e(t,f_1)$. This is simply the system of Fig. 6 which produces $e(t,f_1)$ with the addition of a low-pass filter at its output. The smoothed version of $e(t,f)$ can be expressed by the convolution integral

$$e_s(t,f) = \int_{-\infty}^{\infty} e(\tau,f)h(t-\tau)\mathrm{d}\tau \quad \ldots\ldots(26)$$

where $h(t)$ is the impulse response of the low-pass filter. Equation (26) becomes, on substituting the defining relation for $e(t,f)$,

$$e_s(t,f) = \mathrm{Re}\left[\int_{-\infty}^{\infty} S(f)\exp(\mathrm{j}2\pi f\tau)s(\tau)h(t-\tau)\mathrm{d}\tau\right]$$

$$= \mathrm{Re}\left[S(f)\exp(\mathrm{j}2\pi ft)\times \int_{-\infty}^{\infty} s(\tau)\exp[-\mathrm{j}2\pi f(t-\tau)]h(t-\tau)\mathrm{d}\tau\right] \quad \ldots\ldots.(27)$$

$e_s(t,f)$ can thus also be calculated by the scheme shown in Fig. 11, which is the block diagram representation of equation (27). The similarity between Fig. 9 and Fig. 11 becomes exact if, in Fig. 11,

(i) the filters having impulse responses $\cos 2\pi ft$ and $\sin 2\pi ft$ are replaced by filters having impulse responses $u(t)\exp(-\alpha t)\cos 2\pi ft$ and $u(t)\exp(-\alpha t)\sin 2\pi ft$, respectively; and

(ii) the impulse response of the low-pass filter, $h(t)$, is given by

$$h(t) = u(t)\exp(-\alpha t)$$

From the similarity of these block diagrams it is evident that Fano's definition of the short-term spectrum, $G_t(f)$, can be thought of as an approximation to $e(t,f)$.

4.3. *Other Definitions*

The definition of $e_s(t,f)$ given above (in which $h(t)$ can be any low-pass filter impulse response) was termed by Schroeder and Atal[9] a 'short-time spectrum of the second kind'. This definition was used by them as it is the Fourier transform of a corresponding short-time autocorrelation function.

Their 'short-time spectrum of the first kind' represents a generalization of the short-term spectrum of Fano in which the simple band-pass filters of Fig. 9 are replaced by band-pass filters of a more general form. Their 'short-time spectrum of the third kind' is measured by a system similar to that used in measuring $e_s(t,f)$ in which the ideal band-pass filter is replaced by a physically realizable one. All these definitions thus provide approximations to $e(t,f)$.

The effect of using non-ideal band-pass filters in the various short-term spectrum measuring systems is to give reduced spectral resolution which is equivalent to a smoothing of $e(t,f)$ in the f-direction. The various definitions of short-term spectrum thus correspond to modifications of $e(t,f)$ made by smoothing in the t- and f-directions with various weighting functions.

5. Measurement of Short-term Spectra

Roughly speaking, situations in which a short-term spectrum is to be measured can be placed in two categories, although the division between categories is by no means clear. In one category, the signal itself is the prime object of interest, for example in sonar signal design. In the other category, the principal object of study is not the signal as such, but rather the process from which it originates. Examples here are the spectrographic study of the speech forming process and seismology.

In the first category, the t–f energy distribution is required, perhaps in the finest detail possible. The finest possible detail is provided by computing $e(t,f)$ directly from its defining relation.† However,

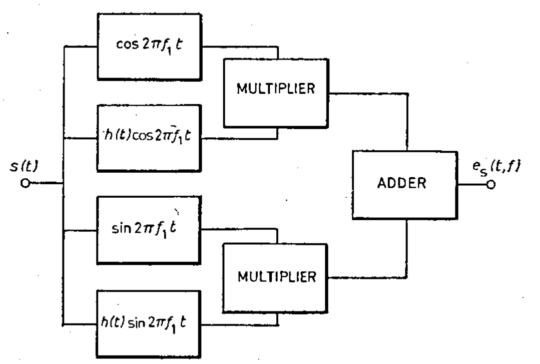


Fig. 11. Alternative scheme for measuring $e_s(t,f)$.

† This can be done economically by using the fast Fourier transform algorithm. Reference 22 is an entire issue devoted to this subject; reference 23 gives a FORTRAN sub-routine.

the fine structure of $e(t,f)$ (such as its fluctuations over a period of one or two cycles of the frequency concerned) may be of no interest. Then a degree of smoothing can be provided in the t-direction by an appropriate choice of $h(t)$ in computing $e_s(t, f)$ as given by equation (26).

In the second category of situations, one often wishes to estimate the power spectrum of a time-varying random process. If $P(t,f)$ changes with time at a rate which is of the same order as the rate at which the signals fluctuate, there is little that can be done with a single signal. With a set of signals from the process, however, $e(t,f)$ can be computed for each member of the set and the average taken to yield an estimate of $P(t,f)$.

If the time variation of $P(t,f)$ is slow compared with the rate of fluctuation of the signal then $P(t,f)$ can be estimated by smoothing $e(t,f)$ calculated for a single signal from the random process. However, at the present time the question of what the best smoothing operation is in a particular situation is not well understood. Ideally, procedures would be available by which the optimum smoothing operation could be chosen on the basis of the available *a priori* knowledge of the process being studied. (For example, in measuring the spectrum of the noise from a vehicle moving past a fixed microphone at a particular speed certain features of the non-stationary noise process such as the rate at which the level changes, the total Doppler shift, etc., are known beforehand.) Developing such procedures is a matter for further research.

In the absence of systematic methods for choosing the smoothing operation to be performed on $e(t,f)$ it is necessary to rely on the intuition of the investigator. In short-term spectrum analysis as it is usually done (using a filter bank or by Fourier analysis of weighted sections of the signal), the smoothing operation is implicitly chosen by the investigator when he chooses filter bandwidths (and so on) so that the measured short-term spectrum appears 'best' according to his subjective judgement.

6. Conclusion

The intention of this paper has been to serve as an introduction to the concepts of the t–f energy density distribution and the time-varying power spectrum and to clarify the physical meaning of the results of practical short-term spectrum measurement.

The idea of the t–f energy density distribution of a signal has been used to relate the various definitions of 'short-time' and 'instantaneous' spectra and to interpret them in physical terms. The t–f energy density distribution of a signal can in principle be calculated exactly if its waveform is known. The uncertainty principle of waveform analysis makes no restriction on the accuracy with which $e(t,f)$ can be computed. The only sense in which it does apply is that it restricts the t–f plane resolution that can be obtained when short-term spectrum measurement is made by a bank of bandpass filters or by an analogous method.

The time-varying power spectrum, $P(t, f)$, of a random process can be regarded as being the average of the t–f energy density functions of the individual signals from the process. A smoothed version of a measured $e(t,f)$ function can provide an estimate of $P(t, f)$, but as yet rules which specify the 'best' smoothing operation for a given application do not exist. Thus the design of a short-term spectrum analysis remains a matter in which heavy reliance must be placed on intuition.

7. References

1. Potter, R. K., Kopp, G. A. and Green, H. C., 'Visible Speech'. (Van Nostrand, New York, 1947.)
2. Flanagan, J. L., 'Speech Analysis, Synthesis and Perception'. (Springer-Verlag, New York, 1965.)
3. Woodward, P. M., 'Probability and Information Theory, with Applications to Radar' (2nd edition). (Pergamon, London, 1964.)
4. Benjamin, R., 'Modulation, Resolution and Signal Processing in Radar, Sonar and Related Systems'. (Pergamon, London, 1966.)
5. Ville, J., 'Théorie et applications de la notion de signal analytique', *Câbles et Transmission*, **2**, pp. 61–74, 1948. Translation available by I. Selin, 'Theory and applications of the notion of complex signal', Report T–92, RAND Corporation, Santa Monica, California, USA.
6. Page, C. H., 'Instantaneous power spectra', *J. Appl. Phys.*, **23**, pp. 103–106, 1952.
7. Fano, R. M., 'Short-time autocorrelation functions and power spectra', *J. Acoust. Soc. Amer.*, **22**, pp. 546–550, 1950.
8. Levin, M. J., 'Instantaneous spectra and ambiguity functions', *Trans. I.E.E.E. on Information Theory*, **IT–10**, pp. 95–97, 1964.
9. Schroeder, M. R. and Atal, B. S., 'Generalised short-time power spectra and autocorrelation functions', *J. Acoust. Soc. Amer.*, **34**, pp. 1679–1683, 1962.
10. Ehrich, W. G. and Steinberg, B. D., 'Panoramic spectrum analysis in real time', *I.R.E. Nat. Conv. Rec.*, **7**, Part 2, pp. 62–69, 1959.
11. Gill, J. S., 'A versatile method for short-term spectrum analysis in "real-time",' *Nature*, **189**, pp. 117–119, 1961.
12. Lampard, D. G., 'A generalisation of the Wiener–Khintchine theorem to non-stationary processes', *J. Appl. Phys.*, **25**, pp. 802–803, 1954.
13. Urkowitz, H., 'Pre-envelopes of non-stationary bandpass processes', *J. Franklin Inst.*, **277**, pp. 31–36, 1964.
14. Bendat, J. S. and Piersol, A. G., 'Measurement and Analysis of Random Data'. (Wiley, New York, 1966.)
15. Silverman, R. A., 'Locally stationary random processes', *Trans. I.R.E., on Information Theory*, **IT–3**, pp. 182–187, 1957.

16. Karkevich, A. A., 'Spectra and Analysis' (Translated from the Russian). (Consultants Bureau, New York, 1960.)
17. Papoulis, A., 'Probability, Random Variables and Random Processes'. (McGraw-Hill, New York, 1965.)
18. Rihaczek, A. W., 'Signal energy distribution in time and frequency', *Trans. I.E.E.E. on Information Theory*, **IT-14**, pp. 369–374, 1968.
19. Gabor, D., 'Theory of communication', *J. Instn Elect. Engrs*, **93**, Part 3, pp. 429–457, 1946.
20. Singleton, R. C. and Poulter, T. C., 'Spectral analysis of the call of the male killer whale', *Trans. I.E.E.E. on Audio*, **AU-15**, pp. 104–113, 1967.
21. Thorn, D., 'An Introduction to Generalised Circuits', p. 114 *et seq.* (Wadsworth, Belmont, Calif., 1963.)
22. Special Issue on Fast Fourier Transform, *I.E.E.E. Trans. on Audio and Electroacoustics*, **AU-15**, pp. 43–117, June 1967. (This issue contains reference 20.)
23. Robinson, E. A., 'Multichannel Time Series Analysis with Digital Computer Programs'. (Holden-Day, San Francisco, 1967.)
24. Storey, L. R. O. and Grierson, J. K., 'Time-symmetric filters' (part 1), *Electronic Engng*, **30**, pp. 586–592, 1958.

Manuscript first received by the Institution on 26th June 1969 and in final form on 29th October 1969. (Paper No.1308/CC67.)

A.3. The t-f Complex Energy Density Distribution

Power and Analytic Signals.

At least three physical interpretations can be assigned to an analytic signal (or any complex signal) :

(i) The real part can be considered to represent a physical signal while the imaginary part is regarded as a purely mathematical fiction which may serve in defining the envelope or the instantaneous frequency of a signal. The analytic signal in this case is often termed the pre-envelope[15] of the real signal.

(ii) Each part of the signal can be represented as a voltage (for example) which is transmitted by a separate wire.

(iii) A complex signal can be defined to be the complex envelope[39] of a high frequency waveform. The real part, $s_r(t)$, represents the amplitude of the in-phase component and the imaginary part, $s_i(t)$, represents the amplitude of the quadrature component of the associated high frequency signal. The real high frequency signal is then

$$v(t) = s_r(t) \cos 2\pi f_c t - s_i(t) \sin 2\pi f_c t \qquad \text{A3.1.}$$

A complex envelope may or may not be an analytic signal.

For most purposes of linear systems analysis the physical interpretation applied to analytic signals is immaterial. The lowpass analytic angle coded signals discussed in this thesis could equally well be given any of the three above meanings. However, when power and energy are involved (or equivalently, when signals are multiplied together) care is needed in interpreting the physical meaning of the complex quantities which result. The interpretation of an analytic signal used here is the complex envelope, which represents a generalisation of the a.c. circuit theory notion of a phasor. In fact, except for the conventional use of rms values instead of peak values, the phasor and the complex envelope representation of a sinusoid of frequency f_c are identical. What follows is a simple generalisation of the theory of power in a.c. circuits.

A narrowband signal, v(t), has a spectrum of the form shown in fig.A1. If this signal is applied as a voltage across the terminals of a time-invarient passive circuit, the current which flows, i(t), is also narrowband and has a spectrum of the same general form as v(t). The power, p(t), entering the circuit is given by the product of the voltage and the current

$$p(t) = v(t)\, i(t)$$

$$= |m(t)| \cos\left[2\pi f_c t + \phi_m(t)\right] |n(t)| \cos\left[2\pi f_c t + \phi_n(t)\right]$$

$$= \frac{|m(t)|\,|n(t)|}{2} \left\{\cos\left[4\pi f_c t + \phi_m(t) + \phi_n(t)\right] + \cos\left[\phi_m(t) - \phi_n(t)\right]\right\} \qquad \text{A3.2}$$

where m(t) and n(t) are the complex envelopes of v(t) and i(t), respectively, so that

$$v(t) = |m(t)| \cos\left[2\pi f_c t + \phi_m(t)\right]$$

$$i(t) = |n(t)| \cos\left[2\pi f_c t + \phi_n(t)\right]$$

The power has a spectrum of the form shown in fig.A2 and can be considered as being the sum of two components; one is a bandpass function whose spectrum is concentrated near the frequencies $\pm\, 2f_c$. The other component is a lowpass function whose functions depend on the complex envelopes of v(t) and i(t) and do not involve f_c. It is usually this quantity which is of interest. In a.c. circuit practice, this is the quantity which

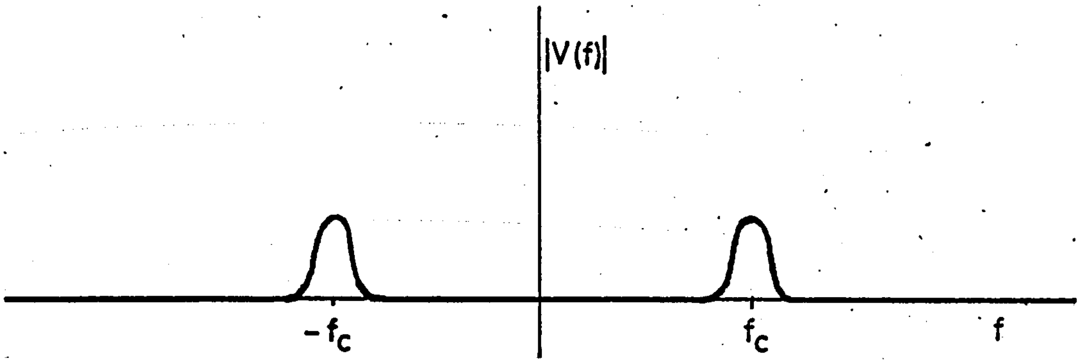


Fig. A1. Spectrum of narrowband voltage waveform.

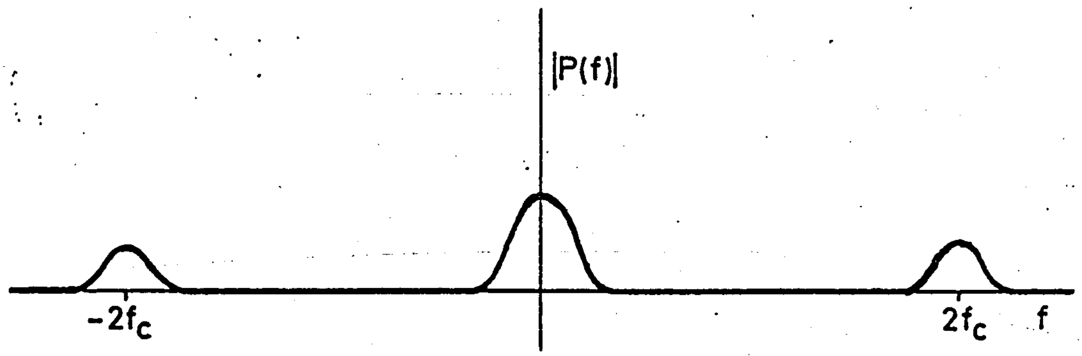


Fig. A2. Spectrum of power waveform.

is measured by a wattmeter (which indicates only the steady component of the power and does not follow the fluctuations at twice the mains frequency).

The cross-power, $p(t)$, of two analytic signals $m(t)$ and $n(t)$ can be defined:

$$p(t) = m(t)\ n^*(t) \tag{A3.3}$$

This is a generalisation of the notion of complex power[40] to non sinusoidal analytic signals. It is a generally complex quantity whose real part is nothing else but twice the slowly varying part of $p(t)$. Its magnitude (if $m(t)$ and $n(t)$ are the complex envelopes of a voltage and a current, respectively) is the <u>volt-amperes</u> in the circuit and is equal to the product of the magnitudes of $m(t)$ and $n(t)$. The imaginary part of $p(t)$ (the 'reactive volt-amperes') has little physical significance[40], but serves to produce the discrepancy between the magnitude and the real part of $p(t)$.

Complex Energy Density

The complex energy density distribution in time and frequency, $e(t,f)$, has been used formally by Rihaczek[38], following Ville[18], and Levin[41] but its physical interpretation has not been emphasised.

If the analytic signal $m(t)$ is applied as a voltage across a (non-realisable) admittance given by

$$Y(f) = \begin{cases} 1, & \text{for } f_o < |f| < f_o + \Delta f \\ 0, & \text{otherwise} \end{cases}$$

then the current, $n(t)$ is given by

$$n(t) = \int_{f_o}^{f_o+\Delta f} M(f)\, e^{j2\pi ft}\, df.$$

The complex power entering the circuit is thus

$$p(t) = m(t)\, n^*(t)$$

$$= m(t) \int_{f_o}^{f_o+\Delta f} M^*(f)\, e^{-j2\pi ft}\, df.$$

The complex energy, E_c, which enters the circuit in the time interval $(t_o, t_o + \Delta t)$ is given by the integral

$$E_c\,(t_o, \Delta t, f_o, \Delta f) = \int_{t_o}^{t_o+\Delta t} p(t)\, dt$$

The complex energy density at time t_1 and frequency f_1 is defined by

$$e(t_1, f_1) = \lim_{\substack{\Delta t \to o \\ \Delta f \to o}} \frac{E_c(t_1, \Delta t, f_1, \Delta f)}{\Delta t\, \Delta f}$$

$$= m(t_1)\, M^*(f_1)\, e^{-j2\pi f_1 t_1} \qquad \text{A3.4}$$

$e(t,f)$ is a complex quantity. The previous section shows that its real part can be interpreted as follows. If a real narrowband signal, of which $m(t)$ is the complex envelope and which has the form given by equation A3.1, is applied to an impedance which acts as a unit resistor over a very narrow frequency range about $(f_I + f_C)$ and which otherwise acts as an open circuit, then $e(t,f_I)$ gives the power entering the impedance at time t (to be exact, the lowpass component of the power).

A.4 Properties of the Complex Energy Density Function

The complex energy density distribution is a function with many interesting properties. Many properties of a signal such as its envelope, energy, etc. can be expressed in terms of $e(t,f)$ in a simple and direct way (see Rihaczek's paper[38]).In this section several more signal properties of $e(t,f)$ are presented.

The following is a reprint of an item of technical correspondence which points out that the instantaneous frequency of a signal is simply related to $e(t,f)$.

Proc. IEEE, 58, p141, 1970.

Instantaneous Spectra and Instantaneous Frequency

Abstract—A simple relation exists between the instantaneous frequency and the instantaneous spectrum of an analytic signal which does not depend on an assumption that the instantaneous frequency varies slowly.

The *instantaneous spectrum* or time-frequency energy density distribution of a signal was introduced by Ville[1] (in a slightly different form from that used here) and has subsequently been used by others.[2-4] A survey of earlier work is contained in the paper by Rihaczek.[4]

A signal whose spectrum is zero for negative frequencies is termed an *analytic signal.*[1] The waveform of an analytic signal is a complex function of time whose real and imaginary parts form a Hilbert transform pair. The complex t-f energy density distribution $e(t, f)$ of an analytic signal $m(t)$ can be defined by the expression

$$e(t, f) = m(t)M^*(f) \exp(-j2\pi ft)$$

where $M(f)$ is the Fourier transform of $m(t)$. The real part of $e(t, f)$ is the quantity of prime interest; it represents the power of the signal per unit bandwidth at frequency f and time t.

In the case of a signal whose instantaneous frequency varies slowly, it has long been realized that the energy density of the signal is concentrated in the t-f plane about a line which follows the instantaneous frequency. This result is supported by physical reasoning and by the principle of stationary phase.[4] The instantaneous frequency of an analytic signal is defined by

$$f_i(t) = (1/2\pi)d/dt[\arg m(t)].$$

However, it does not seem to be widely known that there is a simple relation between $e(t, f)$ and $f_i(t)$ which holds even when $f_i(t)$ varies rapidly.

The normalized first moment of the real part of $e(t, f)$, taken with respect to frequency at a given instant of time, provides a measure of the center frequency of the signal at that time. Expressed as a function of time, this quantity is

$$f_c(t) = \mathrm{Re}\left[\int_{-\infty}^{\infty} fe(t, f)df \Big/ \int_{-\infty}^{\infty} e(t, f)df\right]$$

which, on substituting the definition of $e(t, f)$ and integrating, becomes

$$\begin{aligned} f_c(t) &= \mathrm{Re}\{(j/2\pi)[\dot{m}^*(t)/m^*(t)]\} \\ &= (1/2\pi)d/dt[\arg m(t)]. \end{aligned}$$

This is nothing other than $f_i(t)$, the instantaneous frequency of the signal. Thus, in the sense given above, the instantaneous frequency provides a measure of the "frequency at which the power of the signal acts at a given time" which applies not only to signals for which $f_i(t)$ varies slowly but indeed to any signal.

M. H. ACKROYD
Dept. of Electronic and Elec. Engrg.
Loughborough University of Technology
Loughborough, Leics., England

Manuscript received September 4, 1969; revised September 26, 1969.

[1] J. Ville, "Theorie et applications de la notion de signal analytique," *Cables et Transmission*, vol. 2, pp. 61–74, January 1948.

[2] M. J. Levin, "Instantaneous spectra and ambiguity functions," *IEEE Trans. Information Theory* (Correspondence), vol. IT-10, pp. 95–97, January 1964.

[3] C. A. Stutt, "Some results on real-part/imaginary-part and magnitude/phase relations in ambiguity functions," *IEEE Trans. Information Theory*, vol. IT-10, pp. 321–327, October 1964.

[4] A. W. Rihaczek, "Signal energy distribution in time and frequency," *IEEE Trans. Information Theory*, vol. IT-14, pp. 369–374, May 1968.

Other moments of $e(t,f)$ also provide measures of signal properties. For example the expression*

$$\overline{tf} = \frac{\iint tf\ \ddot{e}(t,f)\ dt\ df}{\iint e(t,f)\ dt\ df} \qquad \text{A4.1}$$

becomes, on substituting the defining relation for $\dot{e}(t,f)$,

$$\overline{tf} = \frac{1}{E} \int t\ \dot{\phi}(t)\ |m(t)|^2\ dt \qquad \text{A4.2}$$

which is Helmstrom's[39] measure of "the amount of FM in a signal". The physical meaning of $\overline{tf}$ becomes, perhaps, clearer when the expression A4.1 is considered.

An important property of $e(t,f)$ is that its two dimensional Fourier transform is the ambiguity function of radar theory[38]

$$\chi(\tau,\xi) = \iint e(t,f)\ e^{-j2\pi(f\tau+\xi t)}\ dt\ df$$

where the ambiguity function is defined by

$$\chi(\tau,\xi) = \int m(t)\ m^*(t+\tau)\ e^{-j2\pi\xi t}\ dt.$$

The two dimensional autocorrelation function of $e(t,f)$ gives the squared magnitude of the ambiguity function[38]. (This is related to the fact that the squared magnitude of an ambiguity function is its own Fourier transform). This result can be generalised to the cross-ambiguity function, $\chi_{12}(\tau,\xi)$, of two signals $m_1(t)$ and $m_2(t)$.

$$|\chi_{12}(\tau,\xi)|^2 = \chi_{12}(\tau,\xi)\ \chi^*_{12}(\tau,\xi)$$

$$= \int m_1(t) m_2^*(t+\tau)\ e^{-j2\pi\xi t}\ dt \int M_2(f-\xi)\ M_1^*(f)\ e^{j2\pi f\tau}\ df$$

*From here onwards all integrals are taken over an infinite interval.

$$= \iint m_1(t)\, M_1^*(f)\, e^{-j2\pi ft} \quad m_2^*(t+\tau)\, M_2(f-\xi)\, e^{j2\pi(f-\xi)(t+\tau)} dt\, df$$

$$= \iint e_1(t,f)\, e_2^*(t+\tau, f-\xi)\; dt\, df \qquad \text{A4.3}$$

where $e_1(t,f)$ and $e_2(t,f)$ are the t-f energy density functions of $m_1(t)$ and $m_2(t)$.

Thus the squared magnitude of the cross-ambiguity function of two signals is given by the two-dimensional cross correlation function of their individual energy density functions. This relation is useful in sketching the squared magnitudes of cross (or auto) ambiguity functions. The author has found it more straightforward to sketch the t-f energy density functions of the individual signals and then to use A4.3 than to attempt to sketch $e(t,f)$ directly from its definition.

The expression A4.3 allows a clarification to be made of the relations between "short term" spectra and the t-f energy density distribution of a signal.

In practice, short term spectra are usually measured by procedures which are equivalent to one of the following :

(i) The signal waveform, $s(t)$, is multiplied by a window function $w(t)$ (shifted so as to be centred at t_1) and the squared magnitude of the Fourier transform of the product is computed. The resulting function of frequency is regarded as constituting a cross section of the short term spectrum, $S_t(t,f)$, taken in the f-direction at time t_1.

(ii) The signal is filtered with a transfer function $H(f)$ (shifted in frequency so as to be centred at f_1). The squared magnitude of the output waveform of the filter is taken to be a cross section of the short term spectrum,

$S_f(t,f)$, in the t-direction at frequency f_1.

The short term spectrum defined by the first process can be written

$$S_t(t_1,f_1) = \left| \int s(t)\, w(t-t_1)\, e^{-j2\pi f_1 t}\, dt \right|^2 \qquad \text{A4.4}$$

From its similarity of form to the squared magnitude of a cross-ambiguity fraction, this expression can be re-written using A4.3 :

$$S_t(t_1,f_1) = \iint e_s(t,f)\, e_{w*}^{*}\, (t-t_1,\ f-f_1)\, dt\, df$$

where $e_s(t,f)$ and $e_{w*}(t,f)$ are the t-f energy density distributions of s(t) and $w^*(t)$, respectively. This becomes

$$S_t(t_1,f_1) = \iint e_s(t,f)\, e_w\, (t-t_1,\ f_1-f)\, dt\, df$$

where $E_w(t,f)$ is the t-f energy density distribution of w(t). Thus, as suggested in the paper of section 2, the short term spectrum is indeed a smoothed version of the t-f energy density of the signal, the smoothing being done by a two-dimensional convolution with the t-f energy density function of the window.

The short term spectrum defined by the second process can be expressed

$$S_f(t_1,f_1) = \left| \int S(f)\, H(f-f_1)\, e^{j2\pi f t_1}\, df \right|^2$$

which, applying Parseval's theorem, becomes

$$S_f(t_1,f_1) = \left| \int s(t)\, h(t_1-t)\, e^{-j2\pi f_1 t}\, dt \right|^2$$

If $h(-t)$, the impulse response of the lowpass prototype of the filter is identified with $w(t)$, the last expression becomes identical to A4.4, which can be rewritten

$$S_f(t_1,f_1) = \iint e_s(t,f)\, e_h(t_1-t, f-f_1)\, dt\, df$$

(where $e_h(t,f)$ can be shown to be identical to $e_w(-t,-f)$).

Thus the short term spectrum, whether measured using a frequency window or a time window can be regarded as being modifications of the t-f energy density distribution of the signal made by a two dimensional convolution with the t-f energy density distribution of the appropriate window. The equivalence of the time window and frequency window approaches was first shown by Larrowe[42].

APPENDIX 'B'

This appendix is similar to chapter 5 but the theory is applied to the design of Huffman sequences instead of angle coded signals. The principal difference between chapter 5 and this appendix is that the words time and frequency are interchanged.

This appendix has been accepted for publication as a paper in I.E.E.E. Transactions on Aerospace and Electronic Systems.

LOUGHBOROUGH UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING

The Design of Huffman Sequences

by M.H. Ackroyd

Abstract

The paper presents a method by which the zeros of the polynomial representing a Huffman (impulse-equivalent) pulse sequence can be chosen so as to exert a degree of control on the form of the energy distribution of the signal in the time-frequency plane. This makes it possible to design Huffman pulse sequences which are suitable for use as radar or sonar signals in situations where significant target velocity occurs.

1. Introduction.

To overcome the effects of noise and interference in a radar or a sonar system the transmitted signal should have large energy which, with a transmitter of limited peak power output, necessitates a pulse of long duration. Provided that the modulation in phase and amplitude of the transmitted signal is chosen properly, the use of long duration pulses does not preclude the attainment of good resolution between targets at different ranges. This is because the limits on range resolution are set by the effective width of the autocorrelation function of the signal, rather than by the duration of the signal itself.

The classic problem of radar signal design is finding a signal whose ambiguity function[1] has a magnitude of some desired form in the τ-ϕ (time delay-doppler shift) plane.

Because of the difficulties of direct synthesis, the commonly used approach is to select a suitable waveform from among those whose ambiguity functions have been catalogued[2].

In cases where significant target velocity is not encountered, it is sufficient to control the form of the autocorrelation function of the signal (which is simply the ambiguity function on the τ axis of the τ-ϕ plane). To obtain good resolution, the autocorrelation function should be narrow, like that of a single short pulse. Huffman[3] showed how to design finite trains of contiguous pulses modulated in amplitude and phase in such a way that their autocorrelation functions would resemble that of a single large pulse. The autocorrelation function of such a pulse sequence, shown sketched in fig. 1, is zero for shifts greater than the duration of one pulse except that for shifts of about the total duration of the pulse sequence the autocorrelation function is unavoidably non-zero.

Such sequences were termed by Huffman "impulse equivalent".

The amplitudes of the complex envelopes of the N+1 transmitted pulses are represented as the coefficients of a polynomial having N roots*.

$$Q = C_0 + C_1 D + \dots + C_N D^N \quad \dots \quad \dots \quad \dots \quad \dots \quad (1)$$

Huffman showed that for the pulse train to be impulse equivalent the roots of Q should lie at equal angular intervals in the complex plane on either of two circles one of which has some radius X while the other has radius X^{-1}.

There are three problems in designing a Huffman sequence:

(i) choosing the number of pulses in the sequence.

(ii) choosing the radius of one or other of the circles on which the roots lie in the complex plane.

(iii) deciding on which circle each root should lie.

These problems are interrelated; in general answer to each problem depends on the answer to the other two and also on exactly what properties the Huffman sequence is required to have.

A desirable property of a Huffman sequence is that its <u>energy ratio</u>, (the ratio of the total energy of the sequence to the energy of the largest pulse) should be large. The energy ratio evidently cannot exceed N+1, for which case all the pulses would be of equal amplitude.

---

*If a transmitted pulse is of amplitude A and phase θ, the corresponding complex envelope value is $A\exp(j\theta)$.

Injeyan[4] has found a nine-pulse sequence with the large energy ratio of 8.125*.

The autocorrelation function of a Huffman sequence (fig. 1) has sidelobes which can easily be shown to be $X^{-N}/(1+X^{-2N})$ of the amplitude of the central lobe. The latter is equal to the energy of the signal, of course. The tolerable sidelobe level sets limits on X and N. What constitutes a tolerable level depends, of course, on the environment in which the signal is to be used.

It is possible to generate Huffman sequences with the property of having purely real pulse amplitudes (i.e. phase 0 or π), with consequent simplicity of implementation, by choosing the complex roots of Q to be conjugate pairs. However purely real Huffman sequences usually have poor energy ratios or large autocorrelation function sidelobes. Reference (5) describes the practical use of a real Huffman sequence.

---

*It is, in fact, possible to improve slightly on Injeyan's sequence: For example the sequence (1.000+j0:000, 0.000+j0.996, - 0.496 - j0.790, 0.787 - j0.618, 0.180 - j0.887, 0.787 - j0.618, - 0.496 - j0.790, 0.000+j0.996, 1.000+j0.000) is a nine pulse sequence having an energy ratio of 8.54. This is the highest energy ratio that can be obtained with this root pattern, which is (--- + - +++), where a minus sign indicates a root on the smaller circle and a plus sign indicates one on the larger. This pulse sequence, like Injeyan's has relatively large autocorrelation function sidelobes; they have a magnitude which is about 7% of the central peak.

In addition to exerting control on the foregoing properties of a Huffman sequence (energy ratio, sidelobe level, purely AM pulses) it may also be desirable to control the form of its ambiguity function <u>off</u> the τ-axis. This paper presents a method for choosing the roots of the polynomial representing a Huffman sequence so that the energy density distribution of the signal in the time frequency (t-f) plane approximates some desired distribution. This permits a degree of control on the form of the ambiguity function in the τ-ϕ plane as a result of the one-to-one relation between the form of the energy distribution of a signal in the t-f plane and the form of its ambiguity function in the τ-ϕ plane[7].

2. Huffman sequences and Zero Patterns.

Instead of dealing directly with a pulse sequence represented by a polynomial such as (1) it is more convenient here to consider a signal m(t) which consists of a series of impulses having complex weights;

$$m(t) = C_o\ \delta(t) + C_1\ \delta(t-T) + \ldots + C_N\ \delta(t-NT). \quad \ldots \quad \ldots \quad (2)$$

where T is the duration of each pulse of the transmitted signal. The complex envelope n(t), of the transmitted signal is given by the convolution of m(t) with a rectangular function of duration T. Symbolically,

$$n(t) = m(t) \ * \ \mathrm{rect}(t/T) \ \ldots \ldots \ \ldots \ \ldots \ \ldots \ \ldots \quad (3)$$

Taking the Laplace transform of (2) gives

$$M(s) = C_o + C_1 \exp(-sT) + \ldots + C_N \exp(-sNT) \quad \ldots \ \ldots \ \ldots \quad (4)$$

and making the substitution

$$z = \exp\ (sT) \quad \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \ \ldots \quad (5)$$

gives the z-transform of the impulse sequence:

$$M(z) = C_o + C_1 z^{-1} + \dots + C_N z^{-N} \dots \quad \dots \quad \dots \quad \dots \quad (6)$$

As mentioned in the introduction, Huffman showed that for the polynomial (1) to represent an impulse equivalent sequence its roots should be chosen so that they lie at equal angular intervals in the complex plane, with each zero being chosen to lie either on a circle of some radius X or else on one of radius X^{-1}. Thus (6) can be written in factored form

$$M(z) = C_o \prod_{n=1}^{N} (1 - z^{-1} z_n) \quad \dots \quad \dots \quad \dots \quad \dots \quad (7)$$

where the zeros of M(z) are given by

$$z_n = \begin{cases} X \exp(j2\pi n/N), \text{ if the nth zero lies on the circle of radius } X \\ X^{-1} \exp(j2\pi n/N), \text{ if the nth zero lies on the circle of radius } X^{-1} \end{cases}, \quad n = 1,2,\dots,N \quad (8)$$

fig.2 shows a typical z-plane zero pattern for a sequence of length 9. Reversing the substitution (5), the zeros of M(s) in the s-plane are given by

$$s_{n+k} = \frac{1}{T} \ln (z_n)$$

$$= \frac{1}{T} \ln |z_n| + j(\arg z_n \pm k2\pi) \quad \begin{matrix} k = 0, 1, 2, \dots \\ n = 1, 2, \dots\dots, N \end{matrix} \quad \dots \quad (9)$$

The zero pattern is periodic in the direction of the $j\omega$ axis and is thus defined by the locations of the zeros in any strip of width $2\pi/T$ such as the one shown shaded in fig. 3. The zeros in this strip are given by

$$s_n = \begin{cases} \frac{1}{T}\ln X + j2\pi n/(NT), & \text{if the nth zero of } M(z) \text{ lies on the circle of radius } X. \\ -\frac{1}{T}\ln X + j2\pi n/(NT), & \text{if the nth zero of } M(z) \text{ lies on the circle of radius } X^{-1}. \end{cases}$$

$$n = 1, 2, \ldots, N \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

(fig. 2 shows the s-plane zero pattern corresponding to the z-plane zero pattern of fig. 1).

In a paper of 1926, Titchmarsh[6] showed that, in engineering language, the Laplace transform of a function which is effectively zero outside the range (t_1,t_2) can be expressed in the form

$$M(s) = M(o)\ e^{-s(t_1+t_2)/2} \prod_{k=-\infty}^{\infty} (1 - s/s_k) \quad \ldots \quad \ldots \quad \ldots \quad (11)$$

where the zeros of $M(s)$ are $s_1, s_2, \ldots$. In the present case $t_1 = 0$ and $t_2 = NT$. This infinite product expansion of $M(s)$ is used in the next section where it is necessary to take logarithms.

3. Complex Energy Density in Time and Frequency.

The time-frequency energy density distribution[(7)] of a complex signal is given by

$$e(t,f) = m(t)\ M^*(j2\pi f)e^{-j2\pi ft} \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (12)$$

where $M(j2\pi f)$ is the Fourier transform of $m(t)$, given by

$$M(j2\pi f) = M(s)\Big|_{s = 0 + j2\pi f}$$

The real part of $e(t,f)$ represents the power of the signal per unit bandwidth at frequency f and at time t. Among many interesting properties of $e(t,f)$ is that its two-dimensional autocorrelation function gives the squared magnitude of the ambiguity function of the signal*[7]. Thus if a form of e(t,f) which leads to a desired ambiguity function is known (or can be found) it is sufficient to work in the t-f plane and control e(t,f) instead of working in the τ-ϕ plane. Finding suitable form of e(t,f) presents a problem, but one which can sometimes be solved by using knowledge of the energy density distribution functions of signals which are known to have ambiguity functions close to the required form but which are not impulse equivalent.

If the bulk of the energy is concentrated about a line in the t-f plane which forms a single-valued function of f, then the normalised first moment $\tau_g(2\pi f)$ of the real part of e(t,f) (taken with respect to t) provides a measure of the time at which the power of the signal is concentrated at a given frequency (fig. 4). Thus

$$\tau_g(2\pi f) = \mathrm{Re}\left[\int_{-\infty}^{\infty} te(t,f)dt \Big/ \int_{-\infty}^{\infty} e(t,f)dt\right] \quad \dots \quad \dots \quad \dots \quad (13)$$

which, on substituting (12) and performing the integrations, becomes

$$\tau_g(\omega) = \mathrm{Re}\left[\frac{j\frac{d}{d\omega}\ M\ (j\omega)}{M(j\omega)}\right] \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (14)$$

$$= \mathrm{Re}\left[j\frac{d}{d\omega} \ln M(j\omega)\right] \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (15)$$

where $\omega = 2\pi f$. $\tau_g(\omega)$ is, in fact, the group delay of the signal. Substituting the infinite product expansion (11) into (15) gives $\tau_g(\omega)$ in terms of the zero locations of M(s):

---

*The author has found this fact a useful aid in sketching ambiguity functions of signals.

$$\tau_g(\omega) = NT/2 + \sum_{k=-\infty}^{\infty} \frac{\sigma_k}{\sigma_k^2 + (\omega-\omega_k)^2} \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (16)$$

where $s_k = \sigma_k + j\omega_k$.

As was shown in the previous section, the s-plane zero patterns of Huffman sequences have zeros occurring at regular intervals of $2\pi/(NT)$ in the direction of the $j\omega$ axis, and either at $(1/T)\ln X$ in the RHP or at $(-1/T)\ln X$ in the LHP, according to whether the corresponding zero of $M(z)$ lies on the X-radius circle or the X^{-1} radius circle, respectively. Thus by considering equation (16) it can be seen that the group delay $\tau_g(\omega)$ can be represented as the sum of a constant term and a series of functions of identical shape displaced by various amounts of a positive or negative sign:

$$\tau_g(\omega) = NT/2 + \sum_{k=-\infty}^{\infty} \operatorname{sgn}\sigma_k\, \nu(\omega-2\pi k/NT) \quad \dots \quad \dots \quad \dots \quad \dots \quad (17)$$

$$\text{where } \nu(\omega) = \frac{|\sigma_k|}{\sigma_k^2 + \omega^2} \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (18)$$

From equation (17) it is evident that $t_g(\omega)$ can be pictured as being produced by the convolution of $\nu(\omega)$ with the series of regularly spaced positive and negative unit impulses given by

$$I(\omega) = \sum_{k=-\infty}^{\infty} \operatorname{sgn}\sigma_k\, \delta(\omega-2\pi k/NT) \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (19)$$

together with the addition of the constant term NT/2. This impulse series is periodic as a result of the periodicity of the zero pattern in the direction of the $j\omega$ axis.

The problem of finding a zero pattern which yields a Huffman sequence whose t-f energy concentration follows some desired line thus becomes equivalent to finding an impulse series having the form given by equation (17) which when smoothed by convolution with $\nu(\omega)$ gives an approximation to the required group delay function.

A delta-sigma modulator[8] (fig. 5) is a system whose output consists of a sequence of equally spaced positive and negative impulses (the weights of which are all of the same magnitude). It works in such a way that its output, when smoothed by a lowpass filter, is transformed into a signal which presents an approximation to its input waveform.

A sequence of impulses which, when smoothed by convolution with $\nu(\omega)$, approximates a desired group delay function can thus be found by simulating a delta-sigma modulator whose input is the desired group delay function (less the mean value, $NT/2$).

When the zero pattern has been chosen, it remains only to multiply out (7) to obtain the complex amplitudes of the pulse sequence.

4. Example.

As an illustration of the use of the method a required t-f energy density distribution similar to that of a 'linear FM' signal having a duration-bandwidth (TB) product of 34 was chosen. The appropriate sequence length is about 35 pulses; each zero gives about one unit of TB product.

Fig. 6 shows the corresponding $\tau_g(\omega)$ function which, when used as the input to a simulated delta-sigma modulator, produced the impulse sequence (+++++++++-+++-+-+-+-+---+---------) and thus specified the zero pattern of $M(s)$. A value of X of 1.24 was used. This value of X yields a good energy ratio (17.0) and very low autocorrelation sidelobes (0.00068 of the central lobe). Also, the resulting ratio of σ_k to ω_k results in a $\vartheta(\omega)$ function which has sufficient width to give a smooth $\tau_g(\omega)$ function but is not so wide that $\tau_g(\omega)$ is unduly smoothed so that its features are obscured.

e(t,f), for an impulse sequence signal, consists of lines in the t-direction of the (t,f) plane and is periodic in the f-direction*. Fig. 7 shows a plot of the real part of e(t,f) evaluated for the signal having the zero pattern and the X value specified above; two periods in the f-direction and shown. This form of display was used by Singletom and Poulter[9]. Areas of low energy concentration appear as rapid alternations of Re $\left[e(t,f)\right]$ for the reason explained in reference 7. It is evident from this plot that the energy concentration approximates the required form. Fig. 8 shows one period of $\tau_g(\omega)$ for this signal, computed by evaluation of the expression (14). The real and imaginary parts of the sequence are presented in table 1 and the envelope of the pulse train is shown in fig. 9.

5. Remarks on Sequence Design.

The foregoing presents a method by which the distribution of the energy of a Huffman sequence in the time-frequency plane can be

---

*From equations (3) and (12), the t-f energy density distribution of the complex envelope, n(t), of the transmitted signal is related to that of m(t) by convolution with rect (t/T) in the t-direction (interpolation by 'zero order hold') and by multiplication with Tsinc(T.f) in the f-direction (which effects substantial attenuation outside the fundamental period).

controlled. However there are other properties of the sequence which one wishes to control such as the form of the sequence envelope. Unfortunately, the design of Huffman sequences remains a matter in which a degree of trial-and-error is required. The following remarks may assist when a sequence is to be designed for a particular application.

When half of the z-plane zeros lie within the unit circle and half lie outside, the envelope of the pulse train tends to be relatively symmetrical about the mid pulse. If more than half of the zeros lie inside the unit circle the energy of the pulse train is concentrated towards its front; the more zeros that lie within the unit circle, the greater is the frontal concentration. Conversely, if more than half of the zeros lie outside the unit circle, the energy is concentrated to the rear*. For most applications, where

---

*These statements can be expressed quantitatively. The normalised first moment of the pulse energies, $\overline{t_1} = (\sum_{k=o}^{N} kT\, |C_k|^2)/\sum_{k=o}^{N} |C_k|^2$, may be taken as a measure of the centre of the temporal energy distribution of the signal. This quantity may be expressed in terms of the group delay of the signal by using the expression given by Ville[10] for the firstmoment of the squared envelope of a signal (adapted for discrete-time, periodic-spectra signals). The expression is

$$\overline{t_1} = \frac{1}{2\pi} \lim_{F\to\infty} \frac{\int_{-F}^{F} |Mj2\pi f)|^2\, \tau_g\,(2\pi f)\, df}{\int_{-F}^{F} |M(j2\pi f)|^2\, df},$$

which is linear in $\tau_g(\omega)$. The spectrum magnitude is the same for all sequences having a particular autocorrelation function. Thus by inspection of equation (16), it can be seen that the more zeros lie outside the z-plane unit circle, the more positive is $\overline{t_1}$. When half the zeros lie inside the unit circle $\overline{t_1}$ has the value $NT/2$.

it is preferable that the energy should be distributed uniformly along the pulse train with consequent large energy ratio it seems correct, as Huffman suggested, to choose half of the zeros to lie on each circle.

When half of the zeros lie inside the unit circle, then if X is made very large the energy of the pulse train becomes concentrated in the central pulses. As X approaches unity, the energy becomes concentrated at the extremes of the train. As a general rule, recomputing the pulse sequence with a smaller value of X reduces the amplitudes of the central pulses relative to the amplitudes of the extreme pulses. The opposite is true when the sequence is recomputed with a larger value of X.

The energy ratio plotted as a function of X for a particular zero pattern may possess many quite sharp peaks. To choose the sequence having the maximum energy ratio for a given zero pattern it is therefore necessary to compute the sequences corresponding to a succession of closely spaced values of X.

6. Conclusion.

A method has been presented which enables the roots of the polynomial which represents a Huffman sequence to be chosen so as to exert a degree of control on the form of t-f energy density distribution. This can be done without affecting the impulse-equivalent character of its autocorrelation function.

There exists a duality between Huffman sequences and the "angle coded" signals devised by Voelcker[11], which are specified by the locations of their zeros in a complex time (instead of frequency) plane. The results of this paper can be applied to the design of angle coded signals with little more than a change in notation and terminology. (They were in fact first worked out for signals of this type).

7. Acknowledgement.

The duality between Huffman sequences and angle coded signals was remarked by E. Titlebaum. The author is grateful to Professor H.B. Voelcker for pointing out this fact in a private letter.

- 13 -

8. References.

1. P.M. Woodward, "Probability and Information Theory with Applications to Radar", London: Pergamon, 1953.

2. C.E. Cook, and M. Bernfeld, "Radar Signals", New York: Academic Press, 1967.

3. D.A. Huffman, "The generation of impulse equivalent pulse trains", IRE Trans. on Information Theory, vol. IT-8, pp S10-S16, September 1962.

4. J. Injeyan, "Generation of Huffman Codes", Technical Memorandum RQ65EE2, Canadian General Electric Company, Toronto, 1965.

5. D.C. Coll and J.R. Storey, "Ionospheric Sounding using coded-pulse signals", Radio Sci. J. Res., NBS, vol. 68D, pp 1155 - 1159, October, 1964.

6. E.C. Titchmarsh, "The zeros of certain integral functions", Proc. London, Math. Soc., vol. 25, pp 283-302, 1926.

7. A.W. Rihaczek, "Signal energy distribution in time and frequency", IEEE Trans. on Information Theory, vol. IT-14, pp 369-274, May, 1968.

8. H. Inose and Y. Yasuda, "A unity bit coding method by negative feedback", Proc. IEEE, Vol. 51, pp 1524-1535, November 1963.

9. R.C. Singleton and T.C. Poulter, "Spectral analysis of the call of the male killer whale", IEEE Trans. on Audio and Electroacoustics, vol. AU-15, pp 104-113, June 1967.

10. J. Ville, "Theory and application of the notion of complex signal", Cables et Transmission, vol. 2, pp 61-74, 1948. (English translation by I. Selin, Rand Corporation report T-92, August 1958.)

11. H.B. Voelcker, "Toward a unified theory of modulation", (Parts I and II) Proc. IEEE, vol. 54, pp 340-353 and pp 735-755, March and May 1966.

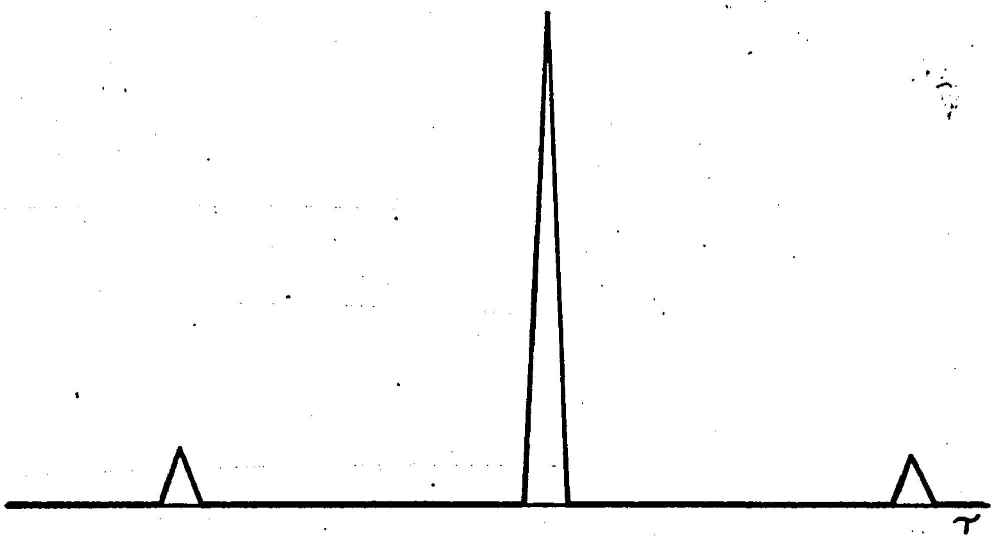


Fig.1. Autocorrelation of Huffman pulse sequence.

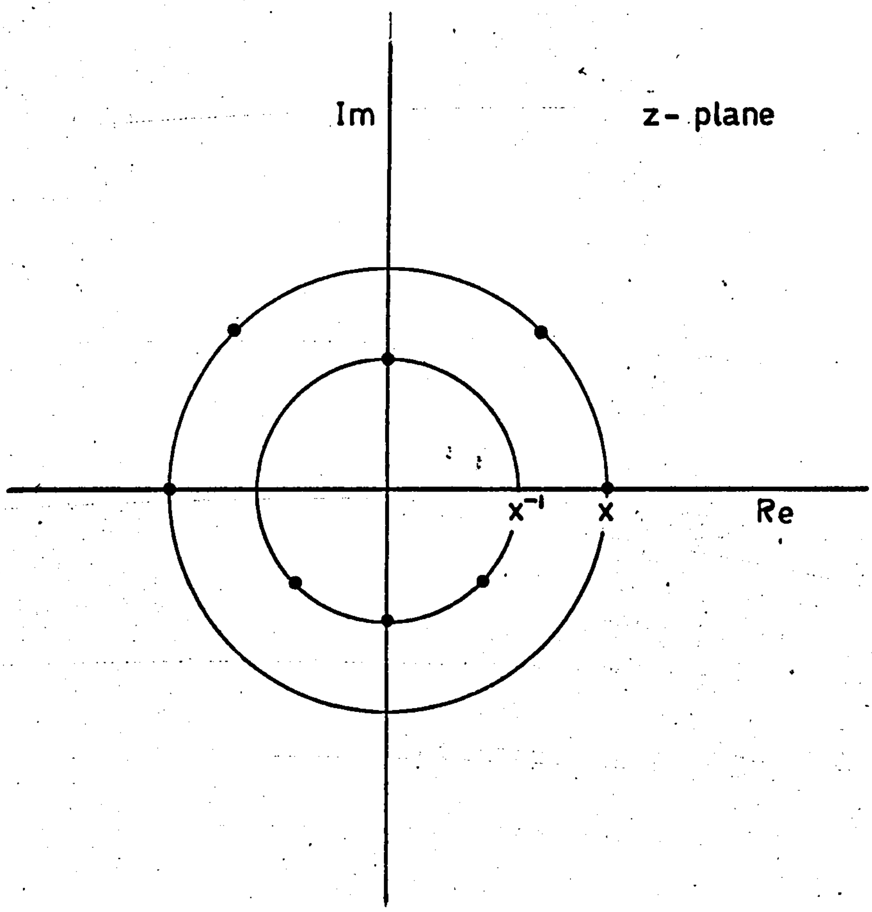


Fig.2. z-plane zero pattern of Huffman sequence.

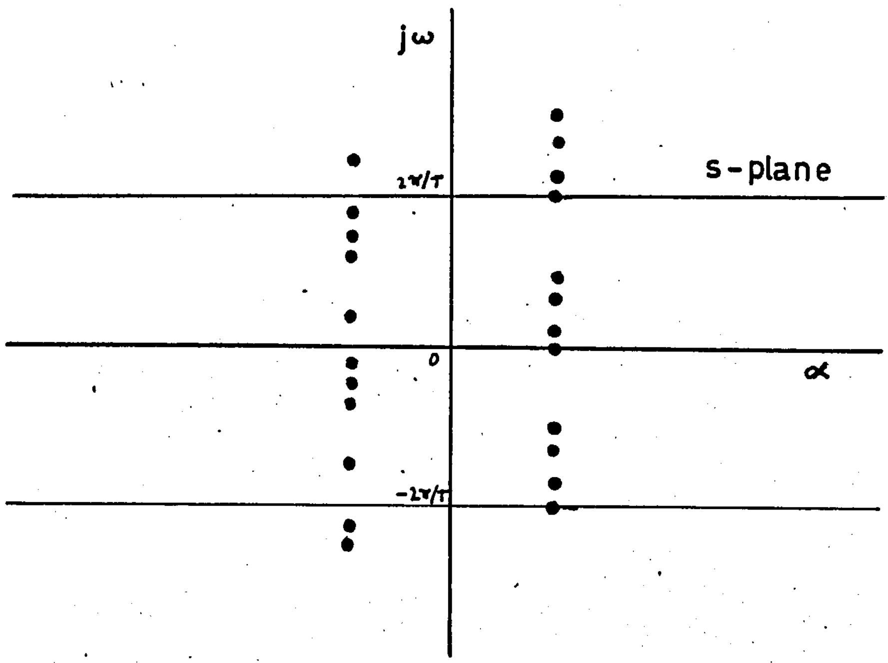


Fig. 3. s - plane zero pattern.

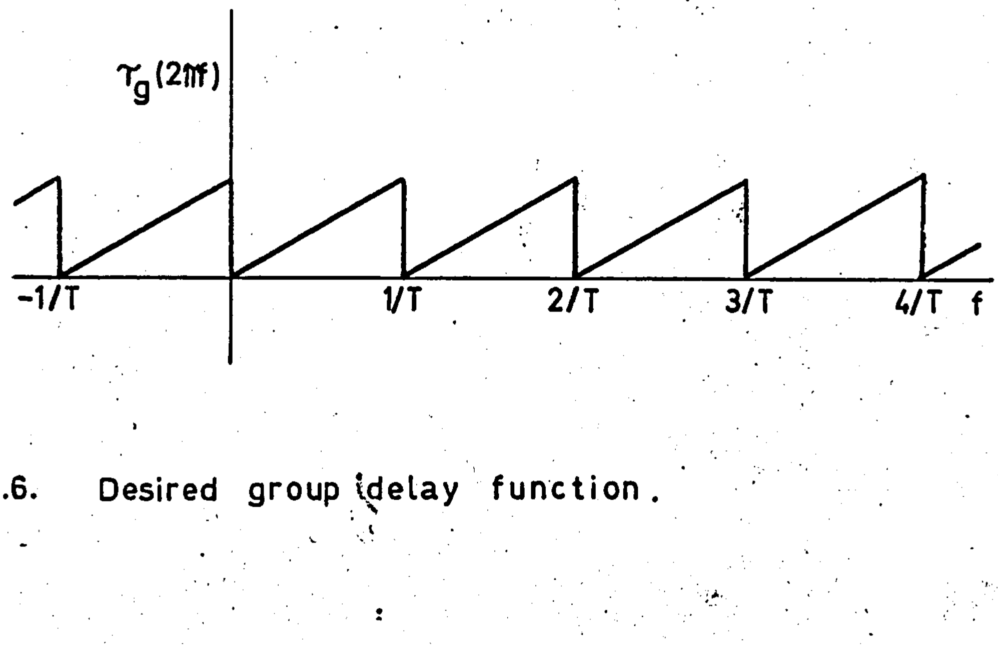


Fig.6. Desired group delay function.

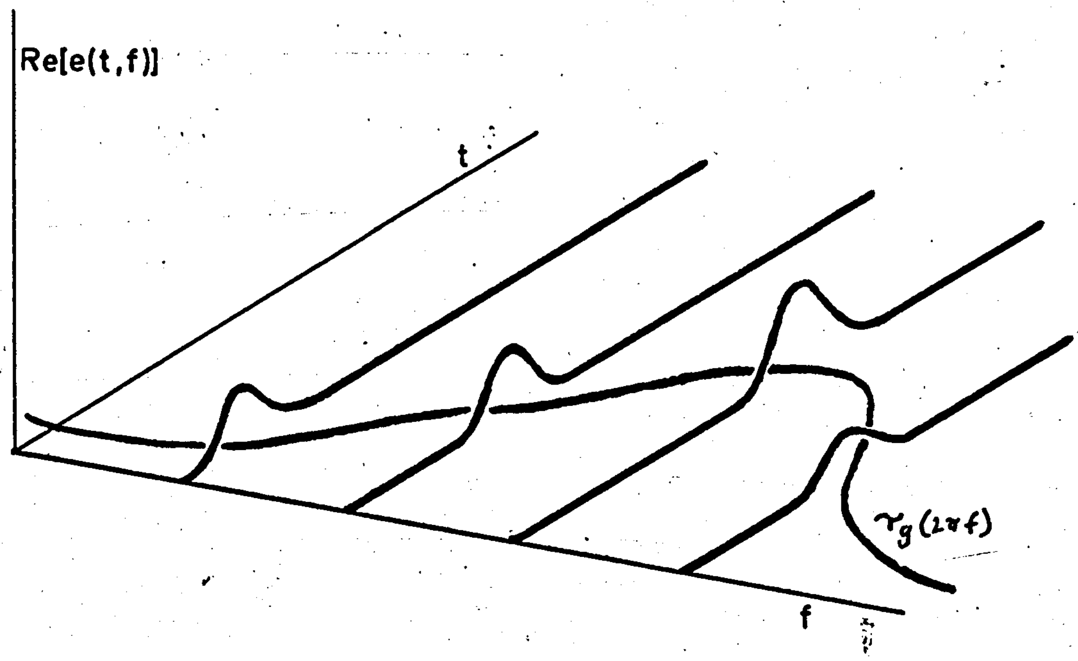


Fig. 4. Group delay as normalised first moment of the time-frequency energy distribution.

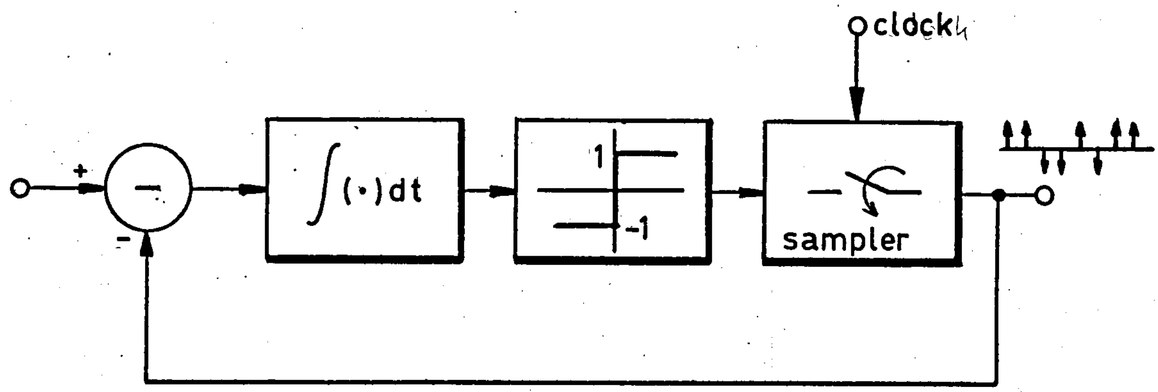


Fig. 5. Delta-sigma modulator.

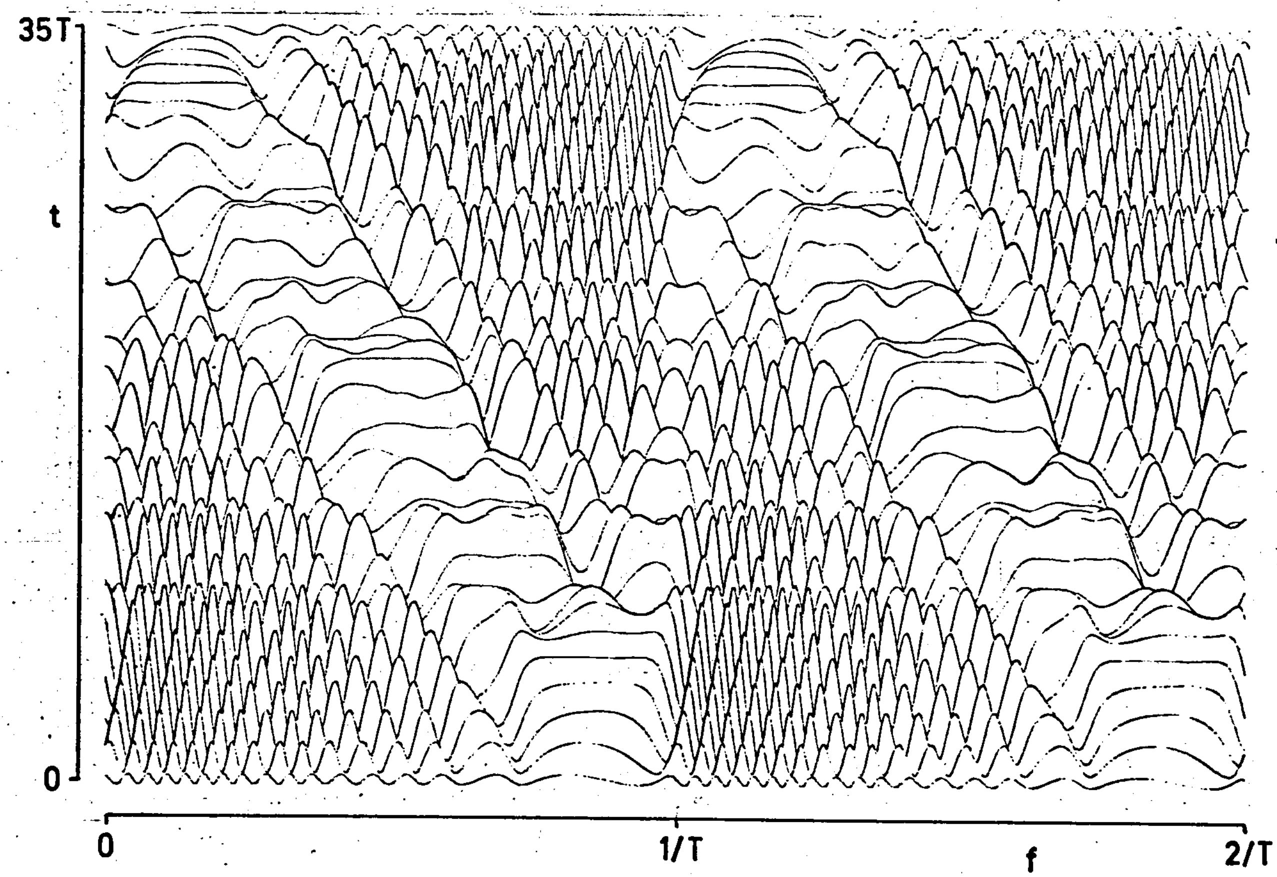


Fig.7. t - f energy distribution of synthesised signal .

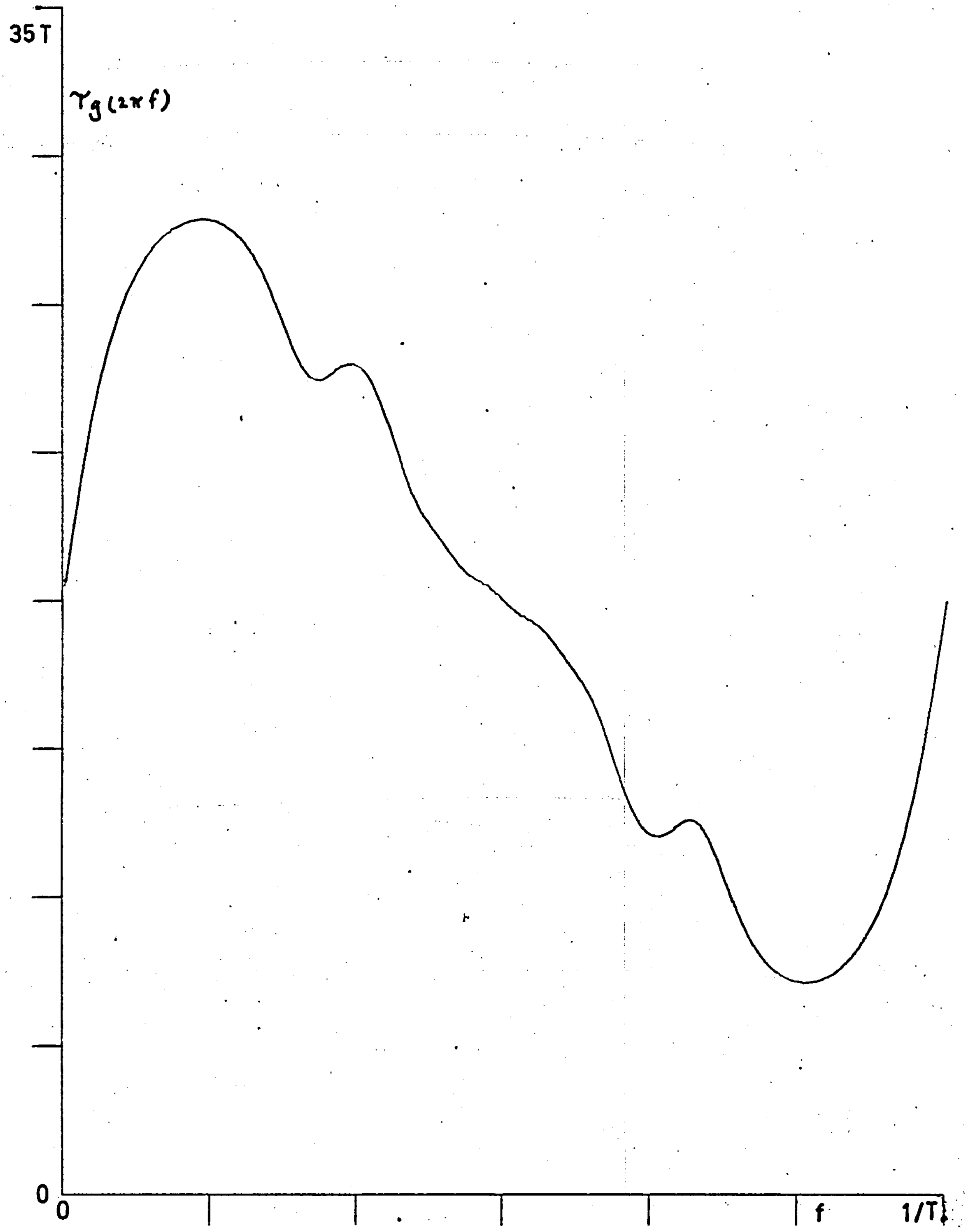


Fig.8. Resulting group delay function.

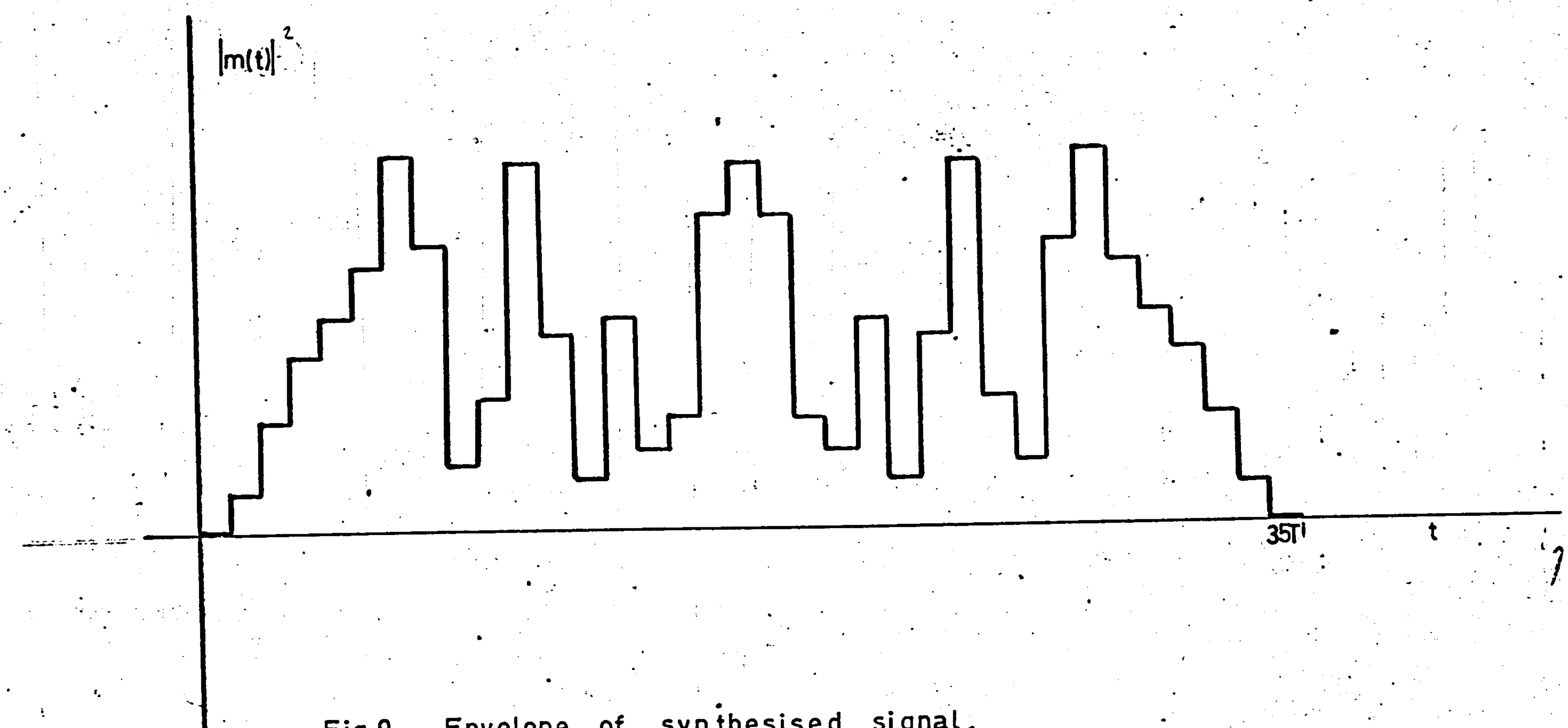


Fig.9. Envelope of synthesised signal.

TABLE 1

pulse number	Re	Im
1	1.00	0.00
2	0.00	3.07
3	-4.71	-1.63
4	5.00	-3.89
5	-0.46	6.96
6	-4.14	-6.53
7	9.11	1.35
8	-4.87	6.36
9	-3.47	-1.80
10	-3.83	-3.92
11	6.84	-5.98
12	4.25	5.01
13	0.94	3.26
14	-4.21	5.49
15	-3.62	-2.19
16	5.03	0.46
17	-7.14	-4.42
18	-9.16	-0.30
19	-7.14	-4.42
20	-5.03	0.46
21	-3.62	-2.19
22	-4.21	5.49
23	0.94	3.26
24	4.25	5.01
25	6.84	-5.98
26	-3.83	-3.92
27	-3.47	-1.80
28	-4.87	6.36
29	9.11	1.35
30	-4.14	-6.53
31	-0.46	6.96
32	5.00	-3.89
33	-4.71	-1.63
34	0.00	3.07
35	1.00	0.00

APPENDIX 'C'

Computing the Coefficients of Polynomials of High Order

The standard method of computing the coefficients of a polynomial from its roots was found to give rise to excessive error accumulation when used with polynomials of high order (greater than about 40). This appendix presents an alternative method which proved to be free from this difficulty.

The coefficients (C_0, C_1, ...,C_N) are related to the zeros of the polynomial ($Z_1, Z_2, \ldots, Z_N$) by

$$C_0 + C_1 Z + \ldots + C_N Z^N = (1 - Z/Z_1)(1-Z/Z_2) \ldots (1-Z/Z_N)$$

An Nth order polynomial consists of the product of N first order polynomials, the coefficients of the ith of which are $(1, -1/Z_i)$. Multiplication of two polynomials amounts to the discrete convolution of their coefficients. The conventional method of computing $(C_0, C_1, \ldots, C_N)$ is to perform the discrete convolution of $(1, -1/Z_1)$ with $(1, -1/Z_2)$ and then convolve the result with $(1, -1/Z_3)$, and so on. This method proves satisfactory for polynomials of low order when programmed in FORTRAN for an ICL 1905 computer. However, it is prone to error accumulation when used with polynomials of high order. This is shown by table C1(a) which shows the coefficients as computed for a set of 127 roots uniformly spaced around the unit circle.

As an alternative to computing the discrete convolution of two sequences directly it is possible to :

(i) compute their discrete Fourier transforms

(ii) multiply the transformed sequences

(iii) compute the inverse discrete Fourier transform of the product sequence.

This provides an alternative method for computing the coefficients of a polynomial from its roots. The DFT $(A_0, \ldots, A_N)$ of the coefficient sequence is given by

TABLE C1

n	Re[C_n]	Im[C_n]	Re[C_n]	Im[Cn]
1	1.00E 00	-5.37E-09	1.00E 00	-6.59E-11
4	1.20E-07	-1.52E-06	-5.01E-11	3.87E-10
8	-1.21E-03	2.24E-02	-1.59E-10	-8.87E-13
12	-2.92E 02	1.19E 02	-4.13E-10	2.26E-10
16	-4.72E 05	-1.69E 05	-4.74E-10	-1.12E-10
20	-7.82E 07	-1.98E 07	-1.05E-11	2.90E-10
24	-5.57E 08	-5.79E 08	6.65E-11	3.74E-11
28	1.14E 12	-4.82E 11	-1.06E-10	-5.73E-11
32	8.83E 13	-7.15E 13	-5.20E-10	4.70E-10
36	2.00E 15	-2.26E 15	-2.13E-10	3.60E-11
40	2.77E 16	-2.88E 16	2.47E-10	8.29E-12
44	9.07E 16	-2.90E 17	-4.69E-10	7.15E-11
48	-1.30E 18	-1.93E 18	3.57E-11	4.66E-10
52	-5.57E 18	-7.78E 18	-4.46E-10	9.32E-11
56	-7.08E 18	-1.90E 19	1.59E-10	3.48E-10
60	-1.26E 19	-8.61E 18	-3.21E-10	-1.98E-10
64	-7.53E 18	2.07E 19	-6.18E-11	1.03E-09
68	6.01E 18	1.68E 19	5.63E-10	9.62E-11
72	8.10E 17	7.02E 18	2.55E-10	1.14E-10
76	-1.09E 18	4.70E 18	1.94E-10	4.26E-10
80	8.91E 17	1.56E 18	-9.03E-11	2.36E-10
84	4.34E 17	1.93E 17	-1.90E-10	2.04E-10
88	5.01E 16	1.74E 16	-1.51E-11	-1.84E-10
92	2.17E 15	2.52E 15	-6.65E-11	3.27E-10
96	9.35E 13	1.35E 14	-2.40E-10	3.00E-10
100	2.32E 12	8.62E 11	3.38E-11	-1.28E-10
104	1.01E 10	5.31E 09	3.59E-10	-1.34E-10
108	-1.19E 08	9.16E 07	3.23E-10	-8.28E-11
112	-5.31E 05	2.60E 05	2.63E-10	-9.04E-11
116	5.67E 01	-4.88E 02	-4.89E-10	2.70E-11
120	3.50E-01	-3.23E-01	1.65E-10	1.40E-10
124	4.37E-06	-4.88E-06	1.39E-10	5.86E-10
128	-1.00E 00	0.00E-01	-1.00E 00	-1.25E-09

(a) (b)

Coefficients, C_n, of the polynomial $1 - x^{127}$ as computed from its roots by (a) the direct method and (b) the DFT-based method. Compare the magnitudes of the intermediate coefficients, the correct values of which are all zero. (Only every fourth coefficient is displayed).

$$
\begin{aligned}
(A_0, \ldots, A_N) = \; & (1 - 1/Z_1,\ 1-W/Z_1,\ 1-W^2/Z_1,\ \ldots,\ 1-W^{N-1}/Z_1) \\
\times \; & (1 - 1/Z_2,\ 1-W/Z_2,\ 1-W^2/Z_2,\ \ldots,\ 1-W^{N-1}/Z_2) \\
& \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\
\times \; & (1 - 1/Z_N,\ 1-W/Z_N,\ 1-W^2Z_N,\ \ldots,\ 1-W^{N-1}/Z_N)
\end{aligned}
$$

where $W = e^{-j2\pi/N}$. This scheme is easily programmed and the coefficient sequence $(C_0, \ldots, C_N)$ is obtained from $(A_0, \ldots, A_N)$ by using a standard FFT subroutine to compute the inverse DFT. Table C1(b) shows the error in the case of roots uniformly spaced around the unit circle to be acceptably small even when the number of roots is large.