Modem Design Employing Real-Time Channel Evaluation

by

Martyn Joseph Shaw

submitted for the degree of

Doctor of Philosophy

to

The University of Warwick

Department of Engineering

April, 1989.

Contents

List of Figures and Tables

Acknowledgements

I would like to take this opportunity to thank my supervisor, Dr. B. Honary, without whose assistance, encouragement and persistence this work would not have been completed. Also I would like to thank Professor M. Darnell of Hull University, for his valuable input.

Thanks are due to Mr. K. Foster of the EES department, Coventry Polytechnic, for his invaluable assistance with the mathematics of Appendices D and G, Dr. C. Obrey and Mr. J. Tabor of the mathematics department of Coventry Polytechnic for their help and Mr. K. Hulstead of Warwick University for his help in the work of Appendix H.

Of course, my colleagues in the Hull-Warwick Communications Research Group must not be forgotten. Without the countless conversations over cups of tea, I should never have made it.

I would also like to take this opportunity to acknowledge the financial support of the SERC, who have sponsored this project and myself as a student.

I must also thank my parents; without them, I would not be.

Last but not least, thank you Gill for being there.

Declaration

The following is a list of the publications that have been made or submitted of the work contained in this thesis, along with the sections of the thesis to which they relate. None of the original work in this thesis was started before 1st April, 1986 when the author began his SERC studentship.

The author was first registered with the Councils for National Academic Awards (CNAA) through Coventry Lanchaster Polytechnic for an M.Phil. from April 1986 until April 1987. The registration was then transferred to be for a Ph.D. at the same establishment. From 1st April 1988, the author was then accepted at Warwick University to register for 1 year full time before submission of this thesis.

Publications and relevant thesis sections

1. Shaw M. "Channel Evaluation Technique by Zero Crossing Measurement", Coventry Lanchester Polytechnic research report number CEES 1/87.

Relevant thesis sections:- 2.1, 2.2, 2.3, 2.4, 2.6, 5.1 and Appendices A to F.

2. Darnell M., Honary B.K., Shaw M.J. "An Adaptive Automatic Repeat Request Coding System Employing Real Time Channel Evaluation", 3rd Joint USSR-Swedish International Workshop on Information Theory, Sochi, USSR, May 1987.

Relevant thesis sections:- 2.1, 2.2, 2.4 and 5.1.

3. Honary B., Shaw M.J. & Darnell M. "A New ARQ Transmission Scheme Involving Zerocrossing Channel Evaluation", Electronics Letters, No. 12, pp592-3, 12 May 1988.

Relevant thesis sections:- 2.2 and 5.1.

4. Shaw M.J., Honary B. & Darnell M. "An RTCE-Assisted ARQ Transmission Scheme: Design & Performance" Fourth International Conference on HF Radio Systems and Techniques, IEE Conference Publication No. 284, April 1988.

Relevant thesis sections:- 2.1, 2.2, 2.3, 3.1, 5.1, 5.2 and 5.3.

5. Shaw M.J., Honary B. & Darnell M. "Optimisation of Parameters in an MFSK Transmission System", Electronics Letters, Vol. 24, No. 12, June 1988.

Relevant thesis sections:- 4.1, 4.2, 4.3 and 4.4.

6. Shaw M.J., Honary B. & Darnell M. "Higher Order Crossing Analysis Applied to Signal Detection and Evaluation of Radio Channels", to be published in the proceedings of the IEE ICAP 89 conference, Warwick University, April 1989.

Relevant thesis sections:- 3.1, 3.2, 3.3, 4.7, 5.4 and 5.5.

7. Shaw M.J. & Honary B. "A new synchronisation technique for MFSK demodulation", submitted to IEE Electronics letters, March 1989.

Relevant thesis sections:- 4.7.

Abstract.

The thesis deals with the area of digital communications over the high frequency (HF) radio channel. It is divided into three main areas of research, the first two dealing with new methods of real time channel evaluation (RTCE) and the third with multi-frequency shift keying (MFSK) modulator and demodulators for use on HF circuits. The HF channel is simulated for use in the investigations.

The first of the RTCE methods employs the technique of counting the zerocrossings of the received signal. The method is investigated mathematically and by simulation. A hybrid error correction/repeat request system is simulated which employs the RTCE method. This is tested over the simulated radio channel and is shown to give a gain in throughput compared to conventional systems.

The second RTCE technique is an extension of the first, whereby higher order crossings (HOX) are counted. These HOX are the zerocrossings of a new series generated by differentiating or integrating the original signal. They are also investigated by mathematical analysis and by simulation.

The third area of work is concerned with MFSK signalling. The method is investigated with respect to the various parameters of the basic format and conclusions are reached as to the best parameters to be used in a number of situation. A demodulation scheme is devised whereby demodulation, synchronisation and channel evaluation are combined into one set of operations. The method is simulated by computer and implemented on a digital signal processor (DSP). Results are presented from the method working over both a simulated HF channel and in the practical situation.

Chapter 1. Introduction.

1.0 Introduction.

In this thesis, methods of reliable transmission of digital data over the high frequency (HF) radio channel are investigated. This is the frequency band from 2 to 30 MHz. Since the HF radio channel is a particularly variable transmission medium, the approach adopted has been to attempt to find a transmission system that is inherently reliable over HF, is easily adapted to changing conditions and lends itself to the simultaneous analysis of the channel. The channel analysis provided by the system can then be used to adapt the transmission system itself, in order that the fullest possible use may be made of the available resources.

The work has been approached in light of, and to a certain extent inspired by, the availability of cheap and plentiful signal processing power [Pearce, Baker & Carter 1988]. This has been made possible by, and will be increased by, advancing microchip technology which is enabling systems of previously unheard-of complexity to be developed at reasonable cost.

1.1 Organisation of the thesis.

The organisation of the thesis is as follows. In the remainder of this chapter we first attempt to identify the milestones in the advancement of HF communications in terms of error correction and detection, modulation techniques and so on. Secondly we develop a model of the channel in terms of the basic mechanisms that may cause errors, indicating previous methods and techniques to overcome these. In this way, an overview of the problems involved in HF communications will be built up.

Thirdly, we briefly review different digital modulation methods, indicating their suitability or otherwise for reliable communications over

the HF channel. We also review previous methods of real-time channel evaluation (RTCE).

The second chapter deals with a channel evaluation system developed to estimate the channel signal to noise ratio (SNR) and provide some degree of error detection which is independent from any coding or signal embedded into the basic modulation. This works by counting the number of zerocrossings, in a fixed period, of the voice-band (roughly 0-3kHz) signal coming out of the receiver. It is found that this count is related to the modulation format and to the SNR.

The third chapter deals with an extension of the zerocrossing counting technique developed in chapter 2. The method essentially counts the number of zerocrossings of the signal, as before, along with the number of maxima and minima, turning points and so on. This is done by first differencing or summing the sampled signal a number of times before counting the number of zerocrossings. These counts are termed higher order crossings (HOX) and contain more information than the simple zerocrossing counts.

Chapter four is concerned with the optimisation of a particular transmission and reception system in light of the knowledge gained above, along with considerations of the HF channel characteristics. The particular system chosen is multi-frequency shift keying (MFSK).

In the fifth chapter we present results of the use of the systems devised in previous chapters. These results are derived from both computer simulations and from practical on-air tests using an HF link on a relatively short UK link, under both day and night-time winter conditions.

Chapter six reviews and summarises the main results and conclusions drawn from the body of the work. It also outlines some ideas for further work, developing some of the ideas put forward in the thesis and some ideas that were thought of as a result of background research.

1.2 A Brief History of HF Radio Communications and Some Related Facts.

In the following section we present a brief historical view of the advancements of HF radio communications since its inception. Included in this are a few of the more important results in information theory and coding since these have had a considerable impact on the development of HF systems.

There is a saying "there is nothing new under the sun" and in the field of reliable communication, this is as true as everywhere else. For instance, we note below the use of repetition codes for error protection. This is really not very different from repeating yourself when giving instruction to somebody so that they will not misunderstand you. We also note that automatic repeat request (ARQ) systems have a direct analogy in speech. In conversation, we say 'yes' when we have understood something and 'pardon' if we are not clear. This is almost identical to the protocol used by an ARQ system. Many of the other concepts discussed here have parallels in everyday life also and these often help in understanding the basis of what is happening. Speech and writing particularly can be seen to have a lot of redundancy inherent in them, which is why we can understand somebody even on a noise telephone line or read a letter written by somebody with bad handwriting.

The following is a rather incomplete list of events and discoveries in the history of HF communications which will perhaps help the reader to gain some historical overview of the subject area discussed in this thesis. The information has been gleaned from Betts [1967, 1970], Berlekamp [1974], Stremler [1982], Thompson [1983], Ralphs [1985], Ratcliffe [1974] and other sources.

Date	Event
1894	Lodge demonstrates propagation of short-wave signals.
1896	Marconi supports Lodges' work with his own experiments.
1901	Marconi receives signals in Newfoundland from Cornwall.
1902	MacDonald 'proves' that transatlantic radio waves travel by bending around the earths surface, 'stuck' to the surface of the sea.
	Heaviside suggests that the upper atmosphere reflects radio waves since it should be conductive due to rarefication.
1903	Solar radiation suggested as the source of the ionosphere.
1916	Marconi develops a spark-gap transmitter on the 3m wave-length, using a parabolic directional antenna.
1919	Valve transmitters introduced on 15m band.
early 1920s	Marconi transmits from a station in Cornwall to his yacht at sea in order to test directional antennae.
	Good reception established from England to Australia.
1925	Appleton establishes the existence of the ionosphere.
	First ionospheric sounding measurements are made.
	Nyquists theorem published.
1927	The Empire Radio Link is set up throughout the commonwealth.
1931	The International Telegraph Union (ITU) arranges frequency allocations for fixed, mobile and broadcast transmissions.
	Weekly measurements of the noon E-layer critical frequency were started. These have been continued to the present time.
1932	Telex system was started by cable.
	Frequency multiplexing of AM signals on voice-band cables first started.
mid 1930s	Radio teleprinter links established.
1940	Detailed records of F-layer started.
early 1940s	Error detection codes using single parity checks were used for electromechanical computers by Bell Systems.

World war 2	Frequency multiplexing of frequency-shift keyed (FSK) signals on radio circuits started.
post war	FSK and multi-tone modem systems developed.
1947	Hamming describes the first perfect single error correcting, 4 data bit code.
	New frequency allocations made and a governing body set up.
	Multi-frequency-shift keying (MFSK) first proposed and analysed by Kotel'nikov.
1948	Shannons capacity theorem first published.
1953	First automatic repeat request (ARQ) system set up.
	MFSK systems first proved to be useful practically.
1955	Elias constructs first binary convolutional code.
1957	Frequency diversity introduced to combat fading.
	Sputnik launched as the first communications satellite.
1960	Reed Solomon (RS) codes first described.
1962	Telstar communications satellite launched.
	'Topside' ionospheric sounding from satellites started.
1963	MFSK first used as a long distance point-to-point communications system from London to Delhi.
	Comité Consultatif de International Radio (CCIR) recommended frequency stability for radio equipment was 30 parts in 10^6.
mid 1960s	Satellites first used for long distance communications.
1966	CCIR recommended frequency stability for radio equipment was 15 parts in 10^6.
1967	Viterbi algorithm first published.
1969	MFSK first used to transmit newspapers to the QEII.
1970	Automatic frequency selection using sounding of available channels proposed.
1978	The term real-time channel evaluation (RTCE) was coined to describe the measurement and use of channel parameters in real time.

1.4 Physical properties of the HF channel and their effects on a signal.

In order to understand the effects of the ionosphere, it is useful to have a model of the way in which HF signals are propagated in the medium. Different effects can be explained by models of greater or lesser complexity and so here we start with a very simple model to explain simple effects and build up more complexity as and when required. The basic effects described here are

i) Additive noise.
ii) Time delay.
iii) Skywave propagation.
iv) Doppler shifts.
v) Differential time delay (multipath propagation).
vi) Rayleigh fading of a single channel, time dispersal.
vii) Gross long-term effects of ionospheric variations.
iix) Critical frequency, maximum working frequency, optimum working frequency
ix) Ionospheric sounding.

1.4.1 Additive noise.

Additive noise is, as its name implies, the random or unknown signal that is added to the required (transmitted) signal before we receive it. The way in which we can specify the characteristics of any particular type of noise are by its probability density function (pdf) and its spectral characteristics. The most common form of noise that is considered in communications systems is additive white Gaussian noise (AWGN). This has a flat power spectral density (PSD), at least in the section of the spectrum under consideration, and a Gaussian pdf, known elsewhere as a Normal pdf [Massey 1971 pp75-80]. This form of pdf occurs naturally in

mathematics as a limit of the binomial distribution, under certain conditions [Spencer et al 1980 pp478]. It is easily described mathematically and is representative of the noise found in at least some communications links. The Gaussian pdf has the form

$$p(x) = \frac{1}{\sigma\sqrt{(2\pi)}}.\exp(-x^2/2.\sigma^2) \qquad (1.1)$$

where the mean has been set to zero since it is generally of little concern to us.

Additive noise may cause random errors by disturbing the received signal in some unknown way, possibly causing a wrong decision to be taken as to what signal was transmitted. The pattern of errors produced by AWGN is such that the probability of one symbol being received in error is independent of all of the other symbols and each symbol has an identical probability of error.

Other types of noise may not have a flat PSD and may have some pdf other than Gaussian. We may include interference in this description, whether deliberate or not, since it may be considered as noise of a particular type. An example of this would be an interfering tone where the PSD would be almost a spectral line and the pdf would be that of a sinusoid. Another example of non-Gaussian noise would be impulsive noise caused, perhaps, by lightning or by ignition coils. This would have a flat PSD but a non-Gaussian pdf. Much work has been done monitoring interference levels [Gott, Wong & Dutta 1982] and attempts made to model and predict them [Laycock et al 1988]. Generally it is better to avoid interference than correct for it in the received signal [Maslin 1987 pp92-102]. This may be done by many methods including selecting interference free channels [Tesla 1986] and using directional antennae to 'null out' interfering signals [Cvetkovic et al 1988].

The ratio of signal power to noise power (SNR) is normally expressed

in decibels (dB) and this is given by

$$SNR\ dB = 10*\log10(\text{signal power} / \text{noise power}) \qquad (1.2)$$

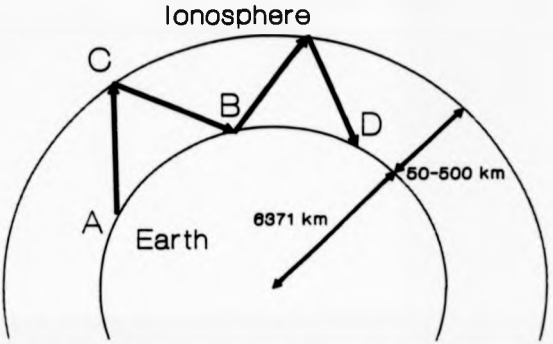
1.4.2 Time delay

Time delay is due to the fact that the radio signals travel at a finite velocity approximately equal to the speed of light ($3.0\times10^{8}\ ms^{-1}$). This of course means that the signal arrives at the receiver slightly after it left the transmitter. This alone causes few, if any, problems and indeed can be used to find out how far apart the transmitter and receiver are if suitably accurate clocks are available. We shall see later however that when there are two or more paths by which the signal is travelling to the receiver, the difference in the time that it takes for the two signals is quite significant with regard to errors. The difference in propagation time we term differential time delay (DTD).

1.4.3 Skywave propagation.

Skywave propagation is the phenomena whereby HF radio signals can travel around the globe. They do this by refraction from an ionised layer in the upper atmosphere. This ionised layer is generally reckoned to be at a height above ground of between 50 and 500 km [Rishbeth & Garriott 1969 pp2] and is caused by radiation emanating from the sun. It is very variable but for now we will model it as a single, perfectly reflecting layer. This model is sufficient to explain how HF signals may be received 'over the horizon' and indeed, all around the globe. If we consider Figure 1.1, we can see that the signal leaving the transmitter at A will not be able to travel directly to the receiver at B, due to the curvature of the earth. However, signals leaving the antenna at A will go up to the reflecting layer at C and be reflected back to the receiver at B. This is

Figure 1.1 The Ionosphere as a Radiowave Reflector



the basic principle of skywave communication. In addition to this, signals may be reflected from the earth at B and travel back to the reflecting layer and down for a second time to D. This is termed two-hop propagation. In practice it is possible for many hops to occur between, say, England and Australia. Note that in Figure 1.1, the reflecting layer and the radius of the earth have not been drawn to scale, since the virtual reflection height is only 0.8 to 8% of the earths mean radius.

1.4.4 Doppler shifts.

Doppler shifts are explained here since only a minor modification of the propagation model described above is sufficient to explain them. They are generally not a major source of error on HF links since they are normally associated with other phenomena which have far greater effects on distorting a signal [Ralphs 1985 pp56].

The way in which we can explain Doppler shifts is to consider the

reflecting layer previously described to be moving up (down). This means that as time goes on, the time that it takes one part of a signal to travel from A to B (Fig 1.1) will be longer (shorter) than the time taken by a signal leaving the source earlier. This scaling of time is may be interpreted as a shift in frequency but we note that there is an approximation involved when signals of non-zero bandwidth are being considered. The effect may be illustrated by considering two signals, separated in radian frequency by Ω rad/s around a center radian frequency of w_o, undergoing a scaling in time. We write the two signals as

$$x1 = \sin((w_o + \Omega/2).t) \tag{1.3}$$

$$x2 = \sin((w_o - \Omega/2).t) \tag{1.4}$$

and then consider them scaled in time

$$x1 = \sin((w_o + \Omega/2).t.(1+K)) \tag{1.5}$$

$$x2 = \sin((w_o - \Omega/2).t.(1+K)) \tag{1.6}$$

where K is related to the speed of motion of the layer. Rewriting these equations we get

$$x1 = \sin(w_o + w_o.K + \Omega/2.(1+K)).t) \tag{1.7}$$

$$x2 = \sin(w_o + w_o.K - \Omega/2.(1+K)).t) \tag{1.8}$$

and the Doppler shift of $w_o.K$ can be clearly seen. The difference between the two frequencies is now $\Omega.(1+K)$, an increase by a factor of $1+K$.

If we try and estimate the value of K in the above equations, we find that for a typical Doppler shift of 2Hz and a center frequency of, say, 10MHz, $K = 3.2 \times 10^{-8}$. Hence the tones are moved apart by a factor of $(1 + 3.2 \times 10^{-8})$ which we may consider to be undetectable and of theoretical interest only.

Typical values of Doppler shift are in the range of a few cycles to tens of cycles per second which equates to a change in path length in the order of a few metres per second. Since the signal is reflecting off the layer at an oblique angle, this means that the layer is moving somewhat slower than this. This effect means that any system employed over a sky-wave link must be insensitive to frequency translations of the order of the expected Doppler shifts. This is not normally a problem in most non-coherent systems and the whole effect may well be masked in the instability of transmitter and receiver oscillators anyway.

In the case of mobile communications, and in particular aircraft communications, Doppler shift is also caused by the actual motion of the transmitter or receiver [Ralphs 1985 sec. 12.8]. A suggested value for the Doppler shift that could be present under these conditions is 70Hz or so and if these conditions are present, more attention must be paid to compensating for the effects.

1.4.5 Differential time delay (Multipath propagation).

As mentioned above, DTD can be a cause of errors in an HF communications system and it can do this in several ways. The first problem which we discuss is 'frequency-selective fading' [Maslin 1987 pp78]. This can be illustrated by examining what happens when we transmit a sinusoid of angular frequency w and receive two versions of it, perhaps by a single hop and a double hop path. We assume that the signals are of identical amplitude and have a DTD of δt s. We write the signal received as

$$\sin(w.t) + \sin(w.(t + \delta t)) \qquad (1.9)$$

or

$$2.\sin(w.t + w.\delta t/2)\cos(w.\delta t/2) \qquad (1.10)$$

whereby it can be seen that if

$$w.\delta t/2 = n.\pi/2, \qquad n \text{ an odd integer} \tag{1.11}$$

then we receive nothing at all. If, on the other hand, n is an even integer, we receive a much stronger signal than either of the two originals.

From (1.10) it is clear why the effect is known as 'frequency-selective fading' since, for a given DTD, certain frequencies are affected more than others. This is in contrast to flat fading which will be discussed below. The effect on the signal transmitted is to change its PSD and it is obvious that the error rate of a given system in noise will be modified by this, particularly if data is being conveyed in the frequency components of the signal, such as in FSK systems. Since equation (1.10) is in the form of a linear filter, it is theoretically possible to recover the original signal from that which has been received, assuming that the DTD, and hence inverse filter, can be found or estimated. This technique is sometimes used in practice but obviously in the example given above, the frequency which is completely lost cannot be recovered. Also it is clear that if the signal was being received with AWGN, the PSD of the noise would be modified by the inverse channel filter, even though it had not been affected by the forward path filter.

Typical values of DTD range from 0 to 2 or 3 ms although they may be as long as 25ms. The possible values of w, when transposed to baseband, must be in the basic signalling bandwidth of 0-3kHz and so, using (1.11), we can take an example of a 2ms delay, n=1 for the lowest null in the spectrum, and we find

$$\begin{aligned} w &= \pi/\delta t \text{ rad/s} \qquad (1.12)\\ &= \pi/2*10^{-3} \text{ rad/s}\\ &= 2.\pi.250 \text{ rad/s} \end{aligned}$$

and so we get nulls at 250 Hz, 750 Hz, 1250 Hz and so on. These may obviously be a problem, particularly in an MFSK system where the null could easily fall on one particular tone frequency.

The problems of frequency selective fading are not as simple as the above analysis would indicate since the layers that are reflecting the signals are normally moving, as described in section 1.4.4. This means that the DTD may well be changing and so we get the effect whereby the frequency that is faded out will scan across the signalling bandwidth in some unknown fashion. This may be observed directly in the error patterns of some particularly types of systems where the signals are frequency division multiplexed into the same channel bandwidth.

The second problem involved with DTD on a communications link is that of intersymbol interferance (ISI) [Stremler 1982 pp367-370]. This is where the later arriving of two signals is carrying information from one symbol while the earlier arriving signal has gone on to the next symbol. This starts to be a problem when the DTD is equivalent to a few percent of the basic symbol period being used and is often a limiting factor for HF communications. This is particularly true of binary signalling where we can see that if the DTD is 2ms and we require this to be less than 2% of the bit time, we can have a maximum bit rate of 1/(2ms*50) = 10 bits per second (bps).

1.4.6 Rayleigh fading.

Rayleigh fading is perhaps one of the biggest problems of HF communications and is caused by the reflecting surface considered above not being a reflecting surface at all but rather a refracting layer. The refractive index of the layer changes with height causing the signal to be gradually bent around and back towards earth [Maslin 1987 pp60-62]. The way in which the refractive index changes is not smooth but subject to random variations and this may be modelled as a large number of scatter-

ing particles [Wozencraft & Jacobs 1965 pp529-530]. These scatterers can all be thought of as individual, moving, reflectors whose cumulative effect is to 'bend' an incident signal back towards the earth. Since each scatterer is at a slightly different position in space, each part of the transmitted signal travels a slightly different path to the receiver. This results in the signal being spread, or dispersed, in time by a small amount. Also, each of the received signals will interfere constructively or destructively with the others, resulting in amplitude and phase modulation of the signal as the scatterers move. It turns out [Wozencraft & Jacobs 1965 pp530-532] that the envelope amplitude pdf of the received signal will be Rayleigh, that is of the form

$$p(x) = \frac{2.x}{\sigma}.\exp(-x^2/\sigma) \qquad (1.13)$$

where x is the relative amplitude

σ is the standard deviation of the distribution

and the phase will be uniformly distributed on 0 to $2.\pi$ [Wozencraft & Jacobs 1965 pp529-32]. The amplitude modulation results in a frequency spreading of the signal such that if we transmitted a single tone, the received signal would have a Gaussian PSD. The standard deviation of the Gaussian PSD is termed the rms fade rate (f_{rms}) and is equal to σ in equation 1.13. It and can also be directly related to the average number of times that the envelope of the signal crosses through its median value in one second [Ralpha 1985 pp49].

Values of f_{rms} quoted by CCIR for testing HF communications systems are between 0.1 and 10 Hz but higher and lower values may be experienced in practice.

When a signal exhibits slow (with respect to the symbol rate) Rayleigh fading in a constant noise, the effective SNR will be changing.

This means that the probability of error will be changing and in certain periods, will be higher than at others. This produces bursts of errors in the received data stream which are normally combated by coding schemes specifically designed for this type of effect [Honary 1981, Honary & Farrell 1982].

1.4.7 Effects of gross ionospheric variations.

The 'scattering particles' refracting HF signals discussed in the previous section are in fact electrons and ions of various gasses such as hydrogen and oxygen which have been produced by the action of the sun on the neutral molecules of the upper atmosphere [Rishbeth 1969 chapter 3]. As the radiation from the sun arriving at a particular part of the ionosphere changes, so does the ion density and it is this which is critical in reflecting a signal. If the ion density is not high enough, a signal arriving at the ionised layer will pass through it, without a significant portion being reflected. Higher frequencies require a higher ion density to be reflected.

Short-term changes in the ion density result in fading of the signal being refracted. These fades are either known as fast or slow fades, depending on their mean period with respect to the basic symbol period of the system under consideration. They are also known as 'flat' fades because they are not, at least in the signalling bandwidth, frequency dependent. This is in contrast to the frequency selective fades mentioned above.

Variations in the amount of solar radiation being received are due firstly to whether it is day or night-time at the point in question [Rishbeth 1969 pp106]. As the sun sets on the ionosphere, the ionosphere cools and contracts and some of the ions recombine to form neutral, non-reflecting molecules again. The cooling and contracting effect is greater however and so the ion density goes up, enabling more signals to

be refracted. This creates the situation whereby medium-wave signals from the continent interfere strongly with local signals in Britain, making it almost unusable in the evening. At dawn the opposite happens; the ionosphere expands faster than ions are produced and the ion density plunges, inhibiting signal propagation. This is known as the 'dawn dip', since the maximum frequency which can be propagated drops significantly.

As well as these daily variations of the ionosphere, longer term effects have been noted, particularly those associated with sun spots. Activity on the surface of the sun becomes more and less energetic on a cycle lasting roughly 11 years [Maslin 1987 pp149] and it is this activity that sends out the ultraviolet and X-ray radiation that ionises the earths atmosphere [Rishbeth 1969 pp106]. This means that there is a significant 11 year cycle in the ion density of the ionosphere.

Prediction of the maximum frequency that will propagate is something that has been done for many years, with varying degrees of reliability [Maslin 1987 pp149-155]. The long-term levels can be predicted but short term effects cannot, resulting in the need for ionospheric measurements.

1.4.8 Critical frequency, Maximum working Frequency and Optimum working frequency.

The critical frequency, f_{max_0}, was alluded to in the above section and is the minimum frequency for which a signal will penetrate the ionosphere when it is travelling vertically from the source [Rishbeth & Garriott 1969 pp51-54]. Alternatively, it may be defined as the maximum signal frequency which will be reflected when travelling vertically. Either way, it gives a measure of the ion density overhead. It is a useful measurement, particularly if the figure is available for the midpoint of the path over which propagation is required. Signals of a frequency higher than f_{max_0} will be refracted if they meet the ionosphere at an angle since they travel further through the refracting layer and hence

have more time to be bent.

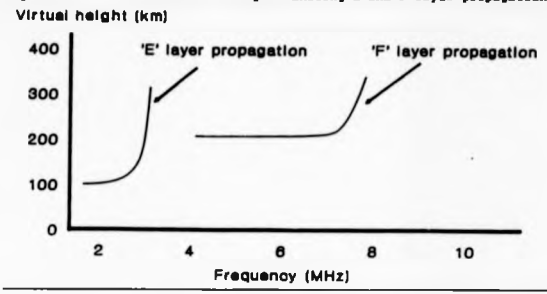
The maximum usable frequency (MUF) [Maslin 1987 pp64] is the maximum frequency that will be refracted on a particular path and hence is related to f_{max} at the midpoint of the path (single hop propagation) as described above. At this frequency, the absorption of the signal is least and so the received signal is strongest, a good reason for using frequencies close to it.

The median MUF is the MUF that is predicted for a given month and, by definition of 'median', this means that for 50% of the time the actual MUF will be below this level. In order that the predicted values can be used with 90% availability, the optimum working frequency (OWF) is defined 15% lower than the median MUF [Watts 1967 pp10]. When scheduling fixed-frequency communications, attention is paid to the OWF with the result that that part of the spectrum will become crowded. As noted below, if an adaptive system is implemented it can locate and use propagating frequencies close to the MUF that will not be allocated for fixed services.

1.4.9 Ionospheric sounding

In the model of the ionosphere given so far, we have not mentioned that the ionosphere is in fact split into several layers. The most important of these for HF communications are the E and F layers which extend from roughly 50-140km and 150-500km respectively [Watts 1967 pp9]. The effects of these layers can be found in real time by sounding the ionosphere, using either vertical or, perhaps more commonly now, oblique sounders. Oblique sounders work by transmitting a tone, swept up in frequency at a known rate and starting at a known time. The signals can then be picked up by a suitable receiver, sweeping up in frequency at the same rate. The received signal strength and relative time delay, or virtual reflection height, can be plotted so that reflections from the various

Figure 1.2 An idealised ionogram showing E and F layer propagation.



layers can be observed. A typical trace from one of these pieces of equipment is rather complex and difficult to interpret but some typical types of reflections can be seen in a somewhat idealised form in Figure 1.2 [Rishbeth & Garriott 1969]. The reflections from E and F layers have been shown. More complicated effects which will not be described here can also be seen on real ionograms, such as the ordinary and extraordinary rays of particular signals. Readers are referred to Rishbeth & Garriott [1969] and Betts [1967] for further details.

In modern HF communications systems, sounding of allocated frequencies is becoming more and more common as a method of finding 'good' frequencies to use from a pool of available (licensed) frequencies [Hague 1988]. There is concern however that in a few years time there will be so much sounding, ionospheric probing and frequency selection going on that the actual bandwidth left for data will be unacceptably reduced. This will only be overcome by responsible use of the HF channel.

1.5 Current modem types and attempts to overcome the problems.

There are several basic methods of modulating a radio frequency (RF) carrier, that is, changing it in some recognisable way according to some

data that it is required to transmit. Each method has it's own peculiar advantages and disadvantages for use in HF communications and it is the purpose of this section is to review the various schemes with particular reference to the particular problems noted in section 1.4 above. For basic reference to each of the schemes the reader is referred to Stremler, [1982 chapter 10] where each is described and the error rates under AWGN are given.

The particular effects of the HF channel that we wish to concentrate on are

a) AWGN.
b) Other noise forms, impulsive, spectral lines, etc.
c) Doppler shifts.
d) ISI.
e) Frequency selective fading.
f) Rayleigh fading.
g) Flat fading.

The basic modulation methods that we identify are

i) Amplitude modulation
ii) Phase modulation
iii) Frequency modulation
iv) Improved and combined systems
v) Broad-band (spread spectrum) systems

and these will be considered one at a time. It has been assumed throughout that the rate at which symbols are transmitted is constant and each has a fixed period.

1.5.1 Amplitude modulation (AM).

In amplitude modulation, we change the amplitude of a fixed frequency RF carrier according to the data that we wish to transmit. The binary form of this is simply to turn the carrier on and off for 1s and 0s respectively. If we wish to transmit more bits per symbol, we must use more levels, normally using a power of two so that there will be a whole number of bits per symbol. This naturally means that the constellation points become closer together and so immunity to noise is reduced.

One particular problem associated with AM is that of fading, since it is the amplitude of the signal that is being used to convey information. In the binary case, this means that the decision boundary, which should be half way between the '1' and '0' level, will be changing and if this cannot be tracked, a worse error performance will result. In the multi-level case the problem is worsened since, for a given transmitter power, the signal constellation points have to be quite close together and so the decision boundaries need to be tracked with even greater accuracy.

A further practical problem with AM is that the transmitter is being operated with a low peak-to-mean power ratio, resulting in a higher cost transmitter for the same transmitted signal power.

1.5.2 Phase Modulation.

The principle of phase modulation is to take a carrier and change its phase according to the data that we wish to transmit during a given period of time. The most simple example of phase modulation is that of phase reversal keying (PRK) [Stremler 1987 pp582] whereby we simply transmit the carrier or the inverse of the carrier, depending on the data. More complex systems switch the phase to one of a larger number of possible values, usually evenly spaced between 0 and 2.π, although not

exclusively so [Divsalar 1984, Honary, Shayan-Arani & Darnell 1989]. As the number of bits per symbol is increased, the distance between the constellation points decreases and so the immunity to noise is reduced.

A particular problem with using phase modulation on HF channels is that the HF channel is not stable in phase. This means that you cannot use a fixed phase reference in the demodulator but must alter the reference according to previously demodulated symbols. This is fairly complex and becomes more so if coding is also employed. Another problem is that the symbols tend to be quite short and it can be shown that ISI can be a limiting factor on the signalling speed.

1.5.3 Frequency modulation.

The principle of frequency modulation is that we take a carrier and change its frequency according to the data that we wish to transmit. The most simple example of frequency modulation is that of binary frequency shift keying (FSK) whereby we simply transmit one tone or another, depending on the data. More complex systems switch the frequency to one of a larger number of possible values, usually evenly spaced through the signalling bandwidth. As the number of bits per symbol, and hence the number of tones, is increased, the distance between the constellation points can be kept constant by making the frequencies orthogonal. This may be achieved by making the symbols longer, since the frequency spacing for orthogonality is equal to the reciprocal of the symbol period.

Frequency modulated systems in the HF environment generally suffer from less problems than phase modulated systems because there are less frequency disturbing effects to distort the waveform. Doppler shifts are normally sufficiently small that they will not cause one tone to be mistaken for another and the frequency dispersing effects of fading do not spread the energy to a very large extent. A more major problem is that of frequency selective fading, although this can be counteracted by use of

in-band frequency diversity techniques [Proakis 1983 pp470]. Due to the long (compared to the other systems mentioned) symbol periods, effects of ISI are reduced. The rapid changing of phase of signals that occurs under Rayleigh fading do however cause errors in multi-tone frequency modulated schemes due to the long integration times required by the longer symbols. These errors are not related to any additive noise [Ralphs 1985 pp59].

Frequency modulated schemes are easier to synchronize than frequency shift systems because there is no ambiguity in the digits.

1.5.4 Improved and combined systems.

In recent years there has been some work done on improving various aspects of AM, PM and FM signalling [see for example de Buda 1972, Chase 1976, Amoroso 1976 & 1979, Eaves & Wheatley 1979, Aulin & Sundberg 1981a and b, Hambley & Tanaka 1984, Sundberg 1986, Padovani 1986, Stremler 1987]. These have concentrated to a large part on altering the modulating waveforms in order to improve the PSD of the modulated signal, that is to say, to reduce the power in the sidelobes [Amoroso 1980]. This is generally done by making sure there are no sudden changes in phase of the transmitted waveform [Sundberg 1986], a technique which can result in a better behaved spectrum. This can be achieved without loss of error immunity but at the expense of complexity.

A second area of improvement has been to combine together coding with modulation in order to increase the effective distance between constellation points without increasing the bandwidth. This technique was pioneered by Ungerboeck [1982] and has generally been advanced for the AWGN channel. Great care must be exercised when attempting to apply schemes advanced for the AWGN channel to the HF environment.

One way in which the basic methods can be varied to provide some specific advantage in the HF environment is to frequency multiplex several signals together and use them to transmit the same information

[Chase 1973]. This provides in-band frequency diversity and can help to combat frequency selective fading and narrowband interference. Another way is to use coding systems specifically designed or modified to combat the burst error patterns that are typical of the fading channel [Monary 1981]. These will not be discussed here.

1.5.5 Broad-band (spread spectrum) systems.

In order to overcome the varying nature of the HF propagation medium, several different broad-band techniques have been proposed and used. The basic idea behind all of these systems is that if the transmitted signal has a very wide PSD, that is to say the symbol energy spread over a large bandwidth, then any frequency selective effects such as fading or narrow band noise will only affect a small part of the total signal [Ince & Schemel 1986]. The statistics of the broad-band channel are nearer to being stationary than those of a narrowband channel by consequence of the central limit theorem.

We can find several examples of broad-band modulation methods. The first is chirp modulation [Gott & Newsome 1971, El-Khamy & Shaaban 1988] where a signal is swept up or down a broad-band, depending on the data. A derivative of chirp modulation is differential phase-shift keying of a swept signal [Gott & Karis 1974], whereby the signal is always swept in one direction through the broad band and information is conveyed in its phase.

A second method of spreading the spectrum of a signal employs transmitting a pseudorandom sequence or its inverse for each data bit, using a conventional binary modulation format [Dixon 1976 pp13]. The spectrum is broadened due to the faster rate of switching the carrier, by a factor related to the length of the sequence. A further method of spectral widening is to use a conventional FSK modulation but to allow a much larger choice of possible frequencies to transmit [Dixon 1973 pp26]. This is

known as frequency hopping.

The practical problem for potential users of spread spectrum or broad-band systems, is that of transmitter and receiver procurement and design. Since a radio user is generally licensed for a limited number of narrowband frequencies, hardware is generally designed for that purpose. Due to these reasons, these systems are not considered further here.

1.6 A review of previous HFCE work.

Since the inception of HF communications, there has been the necessity to provide some form of channel evaluation due to the basic time-varying nature of the channel. In previous years this was achieved both by human operators listening to the received signals and estimating their suitability in real time and by the use of ionospheric prediction algorithms for long-term frequency allocation. However, as the supply of trained radio operators from the war years has dried up and the ability of prediction algorithms to predict short-term changes of the HF channel has not increased sufficiently, the need for automated systems has come to the fore.

Much work has been done in recent years on automated channel analysis for both channel selection and for estimating in-band parameters of specific channels. The reader is referred to a paper by Darnell [1983] for a clear review of work up to that time and to a later paper by the same author [1986] for more up-to-date work. In the next few sections we attempt to outline some of the various methods that have been used or proposed for channel analysis, including sounding, limited channel monitoring and in-band channel evaluation.

1.6.1 Sounding

Sounding has already been mentioned in section 1.4.9 above as a method of measuring the ionosphere in general. We note here that sounding, either on a number of fixed frequencies or across a wide range of frequencies, can be used as a source of channel analysis data [Maslin 1987 pp155]. Its uses are mainly in assessing which channels have a low time dispersal, ie have only a single propagating mode [Maslin 1987 pp158]. This is useful in frequency selection since a channel that is actually propagating must obviously be used and one with reasonably low distortion is preferred. The method does not, however, tell us much about the interference or noise on a particular channel and this may well be the limiting factor in reliable communications [Maslin 1987 pp164].

1.6.2 Limited channel monitoring.

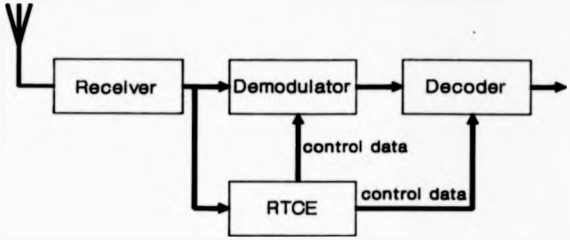
Monitoring of the available (allocated) channels is a useful technique in assessing which channel should be used at any one time since it can give an estimate of the likely error rates due to noise and interference. The monitoring can be either passive [Laycock, Morrell, Gott & Ray 1988], where the average signal strength can be measured, its spectrum analysed and so on, or it can be active, where the channel is tried out or 'probed' periodically and the error rate assessed from the received signal [for example Hague, Jowett & Darnell 1988]. The later of these methods will give an indication of the condition of the channel in terms of both unwanted signals and propagation conditions whilst the former will not tell whether the channel will actually propagate. Both can, however, provide useful information.

1.6.3 In-band channel monitoring.

In-band channel monitoring is a term for the analysis of a channel while it is use. This is necessary since we wish to know how the current channel is performing in order that we may decide whether or not to change to another frequency that may have been found by one of the above techniques. It could also be useful to know if there are narrow band interferers that we could adapt our signal to avoid, removing the necessity to change channels. This is perhaps one of the techniques' most potentially useful attributes since it may obviate the need for such of the sounding and probing that is currently polluting the spectrum.

In-band channel monitoring has been attempted by several different methods including spectral analysis, pilot-tone phase perturbation analysis [Watts 1970], pseudo-error counting, soft-decision analysis [Zolghadr, Honary & Darnell 1988] and ARQ repeat request counting. Three further methods developed by the author are presented in chapters 2, 3 and 4 with simulated and practical results from them in chapter 5. The first of these methods employs zerocrossing analysis [Darnell, Honary & Shaw 1987, Honary, Shaw & Darnell 1988, Shaw, Honary & Darnell 1988], the second higher order zerocrossing analysis [Shaw, Honary & Darnell 1989] and the third, a modulation derived technique for MFSK signals [Shaw, Honary & Darnell 1989]. All of these methods are based on the 'Parallel RTCE' approach, that is to say there is an extra module at the receiver side of the communications link that does the RTCE procedure at the same time as the demodulator is working. This concept is illustrated in Figure 1.3.

Figure 1.3 The parallel RTCE concept.



This page intentionally left blank.

Chapter 2. RTCE Employing Zerocrossings.

2.1 Introduction.

In the following chapter, the general techniques for RTCE based on zerocrossings is described. In section 2.2 the basic RTCE method will be discussed and the basis of the theoretical approach will be laid down. Details of various approaches to solution of the equations derived in the theory are presented in appendices A to D. Section 2.3 will describe the application of the method to several different modulation techniques and present some simulation results of these systems in additive white Gaussian noise (AWGN). Section 2.4 will describe the simulation of the HF radio channel and section 2.5 will present some results obtained regarding the RTCE method over the simulated channel.

The method investigated here for channel evaluation is to look at the zero crossings of the input to the demodulator. The block diagram of the general arrangement of this type of RTCE system is shown in Figure 1.3. Since, for a given modulation method, the bit error probability (P_{be}) can be calculated from E_b/N_o [Edwards 1973], a relationship is sought between the average number of zero crossings per second (AZCPS) and E_b/N_o. The distribution of times between zero crossings is also investigated.

In order to show the effect that we are looking at in a very straightforward way, some typical PSK waveforms at various SNRs have been plotted. These are shown in Figure 2.1a-d. Only a few cycles are shown and the noise is of the same form in each case, just scaled up or down to suit. The zero crossings can be seen and it is simple to see how their number increases as the SNR goes down. The signal shown is modulated by alternate 1's and 0's although most of the simulations have been done with a continuous stream of 0's to make programming a little easier and faster.

Figure 2.1a
BPSK waveform. All signal.

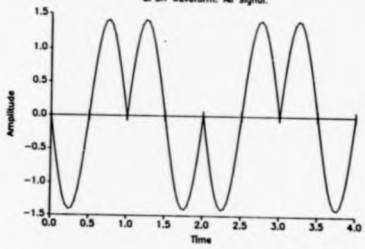


Figure 2.1b
BPSK waveform. 5dB SNR.

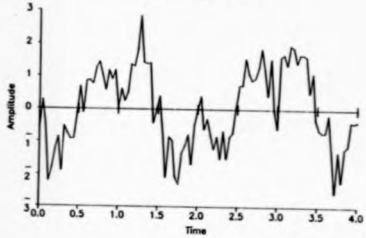


Figure 2.1c
BPSK waveform. 0dB SNR.

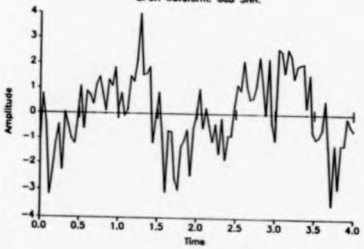
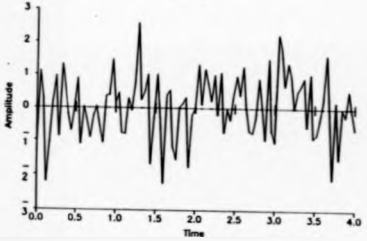


Figure 2.1d
All noise.



2.2 Theoretical Treatment of Zerocrossings

We consider sampled values of a signal which comprises of a single frequency of unit amplitude sinewave with AWGN. This can be written in the form

$$f(n.DT) = \sin(w.n.DT+\theta) + N(n.DT) \qquad (2.1)$$

where n is an integer

DT is the sampling period

θ is an arbitrary phase

N(n.DT) is a sample of Gaussianly distributed noise of zero mean, variance σ^2, each sample independant

The signal has a bandwidth (B), limited by the Nyquist rate, of 1/(2.DT).

We wish to predict the probability of a zero crossing from this description of the signal. That is to say, the probability that two consecutive samples will differ in sign, averaged over a period of time much greater than one period of the sinusoid. In order to do this we first define P_{∞} to be the probability of a zerocrossing in the time interval (n-1).DT < t < n.DT

Due to the (probabilistic) symmetry of the waveform f(t) about zero, the probability of a positive going zero crossing, P_{+-}, is the same as the probability of a negative going zerocrossing P_{-+}. Therefore

$$P_{\infty}(n.DT) = P_{+-}(n.DT) + P_{-+}(n.DT)$$
$$= 2 \; . \; P_{+-}(n.DT) \qquad (2.2)$$

and, writing the waveform in terms of some arbitrary time t, it is easy to see that

$$P_{+-}(t) = P(\; f(t) > 0 \; \&\& \; f(t-DT) < 0 \;) \qquad (2.3)$$

and since the noise samples are independent [Spencer 1977 pp486]

$$P_{+-}(t) = P(\ f(t) > 0\ ).P(\ f(t-DT) < 0\ ) \qquad (2.4)$$

and we know

$$P(\ f(t) > 0\ ) = P(\ \sin(w.t) > -N(t)\ ) \qquad (2.5)$$
$$= P(\ -\sin(w.t) < N(t)\ )$$

and this is given by [Stremler 1982 pp675-7]

$$= Erfc(\sin(w.t)/\sigma) \qquad (2.6)$$

where

$$Erfc(x) = \int_x^{\infty} \frac{\exp(-x^2/2)}{\sqrt{(2.\pi)}}\, dx \qquad (2.7)$$

so from (2.3) and (2.6) we get

$$P_{+-}(n.DT) = Erfc(\sin(t.w)/\sigma).Erfc(-\sin((t-DT).w)/\sigma]$$
$$= Erfc(\sin(t.w)/\sigma).Erf(\sin((t-DT).w)/\sigma) \qquad (2.8)$$

and from (2.2)

$$P_n(t) = 2.Erfc(\sin(n.DT.w)/\sigma].Erf(\sin((n-1).DT.w)/\sigma] \qquad (2.9)$$

This expression (2.9) now needs to be averaged over a whole period of the sinusoid to give the result

$$P_{\infty} = w/\pi . \int_{0}^{t=2.\pi/w} Erfc[\sin(w.t)/\sigma] . Erf[\sin(w.(t-DT))/\sigma] . dt \qquad (2.10)$$

This gives the probability of a positive or negative going zero-crossing in the time interval between two samples. We also define a 'system dependant' variation of this called 'average zerocrossings per second' (AZCPS), which is equal to P_{∞}/DT. This has been used in some stages of the study.

Unfortunately, equation (2.10) is not very tractable mathematically but has been attacked by several different methods. These attempts at finding an easily calculated solution to (2.10) are given in appendices A to D. Previous work on the continuous version of this signal done by Rice [1946], which is given in appendix E.

In order to compare the results of (2.10) with the case considered by Rice, the noise spectrum is assumed to be flat in the range 0 to 1/(2.DT) and to have a value of N_{∞} watts/Hz. We can now calculate the expected number of zerocrossings per second with respect to E_b/N_{∞} by use of equations (E.1) and (2.10), using the relationship

$$\begin{aligned} E_b/N_{\infty}\ dB &= 10.\log(E_b/N_{\infty}) \\ &= 10.\log((S/\text{bit rate}).(B/N)) \\ &= 10.\log((S.B)/(\text{bit rate}.\sigma^2)) \\ &= 10.\log(S/\sigma^2) + 10.\log(B/\text{bit rate}) \end{aligned} \qquad (2.11)$$

where S signal power

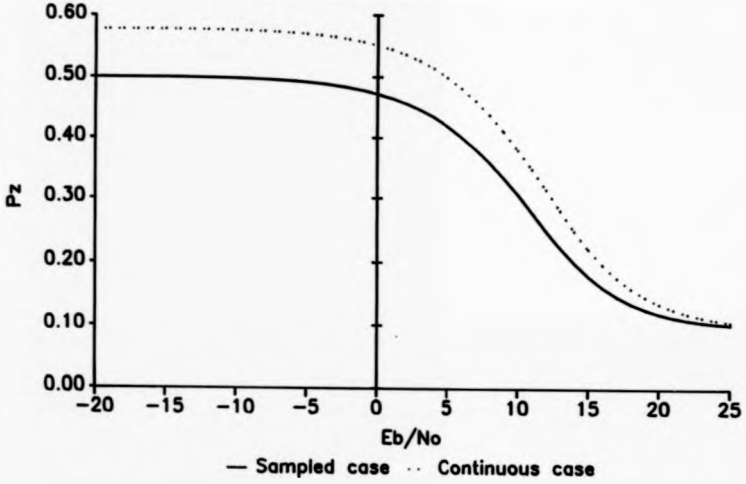
N noise power

σ^2 variance of the noise samples

and the bit rate is in bits per second (bps)

The calculations of (2.10) and (E.1) have been made and the results plotted in Figure 2.2, extended over a larger range of E_b/N_{∞} than would normally be encountered so that the extremities of the curves can be seen. Here 10.log(B/bit rate) = 11.5dB. It is clear that at the low

Figure 2.2
Zerocrossing probability (Pz) vs Eb/No
Frequency = fs/20



signal and, i.e. the waveform is nearly all noise, the curves, whilst being the same shape, are of different magnitudes. From consideration of Rices earlier work [Rice 1945], we find that there is a factor of $\sqrt{(4/3)}$ difference.

In the case of the all noise waveform, sampled at the Nyquist rate, we would expect the number of zerocrossings per second to be equal to the highest frequency. The reason being that each sample is independent and so has equal probability of being above or below the zero line. Therefore the probability of two successive samples being on opposite sides of the zero line is a half and, since we have two samples per cycle of the highest frequency present, the average number of zerocrossings per second is equal to the highest frequency. This is indeed what our analysis predicts. Rice's formula (E.1) predicts a higher figure and a demonstration of this difference is given below.

We have considered the waveform to be bandlimited to the Nyquist bandwidth and hence the original signal to be reconstructible by use of the formula [Cappallini 1979]

$$wf(t) = \sum_{n=-\infty}^{\infty} x(n.DT).sinc((t-n.DT).\pi/DT) \qquad (2.12)$$

where

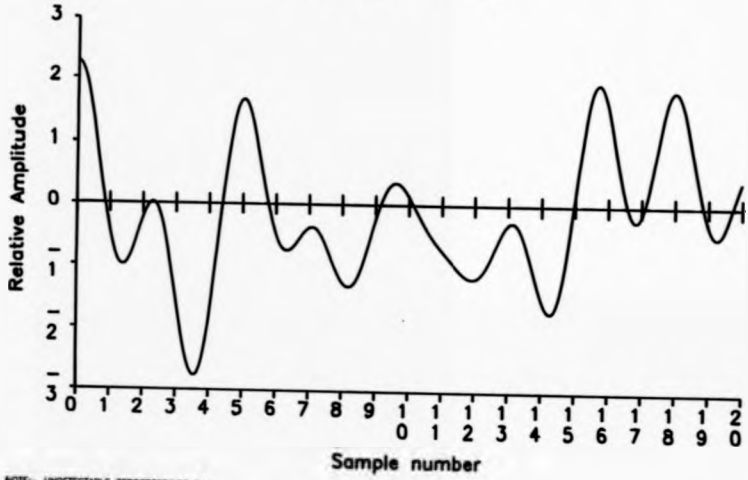
$$sinc(x) = \frac{\sin(x)}{x}$$

t = time in seconds

x(n.DT) = value of sample t = n.DT

which has solutions at t = n.DT of x(n.DT) and is bandlimited to 1/(2.DT) as required. A waveform reconstructed in this manner using 21 random samples is shown in Figure 2.3. An example of the 'extra' zerocrossings

Figure 2.3
Noise waveform showing an 'undetectable' pair of zerocrossings



NOTE:- UNDETECTABLE ZEROCROSSINGS IN INTERVAL 3

predicted by Rice can be seen in the 3rd interval and it is simple to see how they will be missed in a sampled system.

2.3 Applications to Channel Evaluation in AWGN.

In order to investigate the use of zero crossing counting as a valid method of channel evaluation, several different modulation methods have been simulated in software. White noise has been added and the signals then demodulated. In parallel, the zerocrossings have been counted and various measurements and conclusions have been made.

The general method used for all of these modulation schemes has been to generate signals over many bits at different SNR's. Equal steps of E_m/N_m in dB's have been taken and the standard deviation, σ, of the noise samples has been calculated by use of (2.11). The noise itself is generated by a subroutine which gives a Gaussian distribution, generated from two uncorrelated random numbers with flat distribution. Appendix F has the details of how this is done.

Zero crossings are detected by a change in sign of the demodulator input. The 'actual' time is then calculated by linear interpolation thus

From similar triangles

$$pd/(tz-(t-DT)) = d/(t-tz) \qquad (2.13)$$

in a few lines we get

$$tz = t-DT.d/(d+pd). \qquad (2.14)$$

The time of the previous zero crossing is stored and so the difference can be calculated and a density function of differential times between crossings built up.

The simulated demodulator is a correlation (matched filter) receiver which in this case is easy to implement as the timing and phase of the received signal are known exactly. The integrator is replaced by a sum and scaling is ignored as the decision level is set to zero.

2.3.1 Phase Shift Keying (PSK).

Binary PSK, or phase reversal keying (PRK) [Stremler 1982 pp303], is perhaps one of the best known form of data modulation and so we have taken it as the first example.

In Figure 2.4, the zerocrossings of a large number of bits have been taken and averaged out. The results are shown in terms of AZCPS for the system that we used. The signal shown previously (Figure 2.1a-e) is similar to the system simulated here but is modulated by alternate 1's and 0's. The experiments were done with a continuous stream of 0's to make the simulation a little easier and faster. This obviously has some effect on the number of zerocrossings since the signal will generally miss a zerocrossing if the data changes from one bit to the next. Random data should be assumed in the practical case.

The density function generated by the zero crossing detector has been plotted and is shown in Figure 2.5. The data has been taken over 10,000 bits with a sampling period of 0.07 seconds, bit rate of 1 bit per second and a carrier frequency of 1Hz. The bandwidth is limited only by the Nyquist bandwidth due to the sampling; the maximum frequency at the input to the demodulator being given by 1/(2.sample time).

In the limit when SNR -> ∞ and the number of samples -> ∞, we would expect the zero crossings to be distributed around 1/(2.carrier frequency) as there are 2 zero crossings per cycle of the input signal. In

Figure 2.4
AZCPS vs Eb/No
BPSK simulated case.

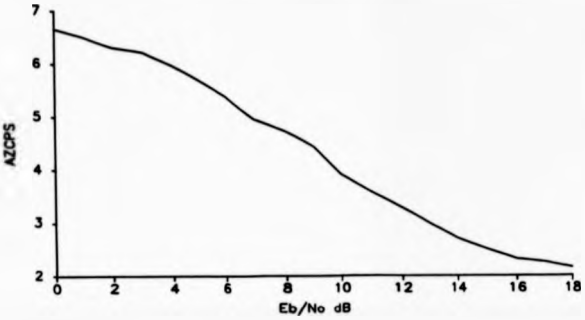
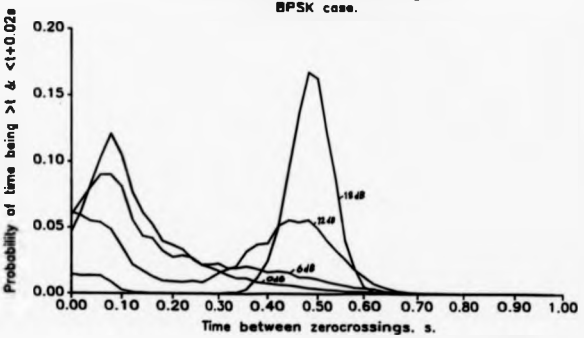


Figure 2.5 Probability density function of times between zerocrossings - BPSK case.



the other limit as SNR -› 0, we would expect the zero crossings to be distributed around the sampling rate or, in other words, around 1/(2.maximum frequency) for the reasons discussed above. This can indeed be seen to be happening in Figure 2.5.

When we plot the AECPS against E_b/N_o dB (see Figure 2.6) we can see that it is monotonically decreasing with increasing E_b/N_o dB. This implies that the number of zero crossings per second can be directly related to the BER and this can be seen in Figure 2.6. This curve is derived from the probability of error $P_e=Erfc\sqrt{(2E_b/N_o)}$ [Stremler 1982 pp583] for PSK signalling and from the simulation results. The curve can be seen to converge towards 2 crossings per second as E_b/N_o -› ∞ and towards 7 crossings per second as E_b/N_o -› 0. This is due to the reasons discussed above.

It is clear from these results that in a sampled system such as this we can, as expected, estimate the SNR of a waveform by simply counting the number of zero crossings.

2.3.2 Frequency Shift Keying (FSK).

FSK modulation involves transmitting one of two frequencies in each bit time, depending on the input signal [Stremler 1982 pp574]. The data presented here was derived in a similar way to that for the PSK case. We have used a centre frequency of 1.5Hz with a deviation ratio of 1. The simulation was done over 10,000 bits as before and the results are very similar. This time the average number of zerocrossings per second, Figure 2.7, has it's lower limit at 3 due to the fact that we are transmitting alternate 1's and 0's. For the PSK case the limit was 2. In the probability density plot, Figure 2.8, there are obviously 2 peaks at low noise levels and these are due to the 2 different frequencies being transmitted. The zero crossings per second against log BER graph, Figure 2.9, can be seen to be very similar to the PSK case, as we would expect.

Figure 2.6 AZCPS vs log10(BER)
BPSK simulated case.

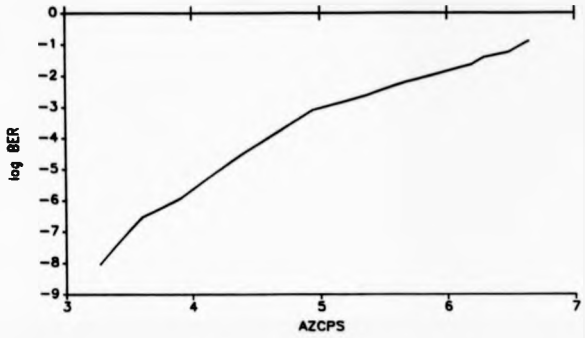


Figure 2.7 AZCPS vs Eb/No
FSK simulated case

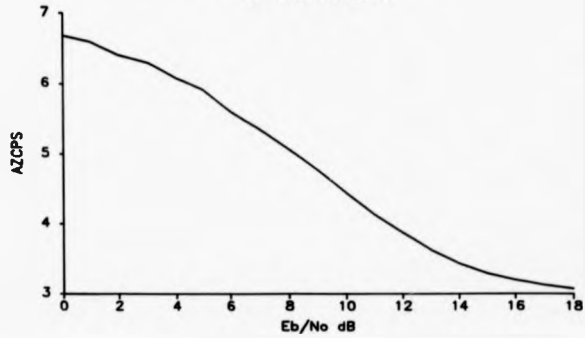
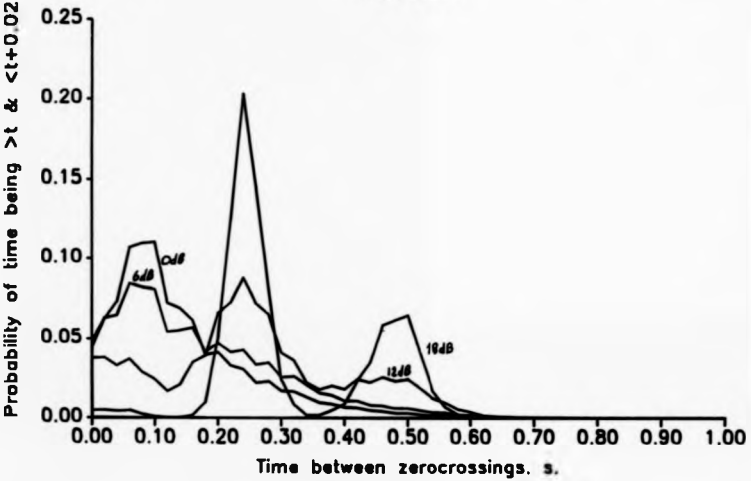


Figure 2.8 Probability density function
of times between zerocrossings
FSK case



2.3.3 Minimum Shift Keying (MSK).

Minimum shift keying is similar to FSK except that the two frequencies are not orthogonal and are spaced closer together on the frequency axis. Each bit has an effect over two symbol intervals [Stremler 1982 pp596]. The data presented here was derived in a similar way to that for the previous cases. The centre frequency used was 0.75Hz with a deviation ratio of 0.5. The data for E_b/N_o up to 8dB was taken over 10,000 bits and thereafter using one million bits. This method takes a long time to simulate as each channel (in phase and quadrature components) must be considered separately [Stremler 1982 pp599, de Buda 1972]. The data transmitted was alternating 1's and 0's on each channel. The theoretical BER is related to E_b/N_o by exactly the same equations as for the PSK case [Stremler 1982 pp601].

It can be seen from Figure 2.10 that the simulated results for AZCPS is monotonic with E_b/N_o and it is clear from Figure 2.11 that, once again, the BER is easily predicted from the average number of zerocrossings per second. These results further validate this work as being useful for real time channel evaluation.

2.3.4 Generalisation to other modulation schemes.

It is clear that for any constant-amplitude, continuous-phase type modulation waveform [Sundberg 1986], the AZCPS is not dependant on the modulation method but rather upon the carrier frequency, assuming that an equal number of 1's and 0's are transmitted. This makes the curve in Figure 2.2 valid for any of this class of modulation schemes with little further work. The relationship between E_b/N_o and bit-error-rate (BER) can be found by standard methods [Edwards 1973] for the modem being considered, and hence a relationship between AZCPS and BER can be found. This is the relationship that is required for RTCE systems and several ex-

Figure 2.9 AZCPS vs log10(BER)
FSK simulated case

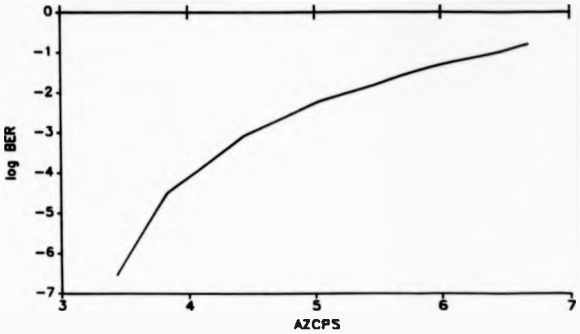


Figure 2.10 AZCPS vs Eb/No
MSK simulated case

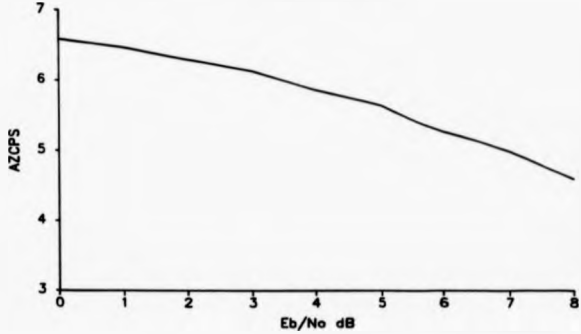
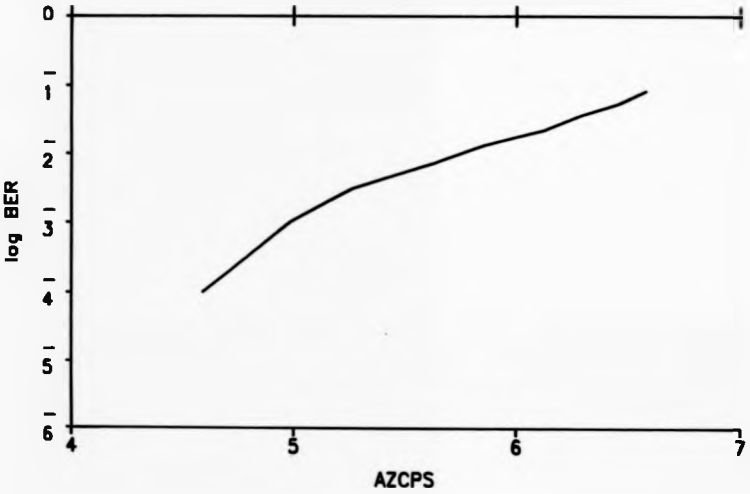


Figure 2.11 AZCPS vs log10(BER)
MSK simulated case



amples have been presented above.

In the case of amplitude shift keying (ASK or OOK), we can see that when the signal is a 1, i.e. keyed on, the average number of zerocrossings can be found from (2.10) with 3dB added to E_b/N_o due to the peak-to-mean ratio of the signal. When the signal is keyed off, the AZCPS is simply $f_m/2$ due to the noise. Hence the result

$$AZCPS_{am} = AZCPS_{+3dB}/2 + f_m/4 \qquad (2.15)$$

where $AZCPS_{am}$ = average zerocrossings per second for amplitude modulation plus noise

$AZCPS_{+3dB}$ = average zerocrossings per second for a sinusoid plus noise at 3dB above E_b/N_o of AM waveform.

2.4 Simulation of the Rayleigh Fading HF Channel.

The problem of simulating a Rayleigh fading channel has been tackled in many different ways in the past [Porter 1968, Watterson 1969a, 1969b, 1969c, 1970, Chase 1973, CCIR 1974, Ralphs 1976, Matley 1977, Ream 1978, Dawson 1984 etc.] but since we are interested in simulating modems and channels on a mainframe computer, it was decided to simulate the modulator and channel at the same time. Hence the signal which would emerge from a digitiser on the output of an idealized HF receiver has been simulated directly.

Our program can simulate a groundwave component and up to 4 skywave components. Each resolvable component of the received signal is generated independently and can have the following parameters

differential time delay

Doppler shift

Rayleigh fading rate

relative amplitude.

We can also have an additive Gaussian noise component of any specified level. The sampling rate can be fixed to any appropriate figure but it also fixes the bandwidth due to the Nyquist criterion. No other filtering has been applied as this is very system dependent and so this 'ideal' case only is considered. Filters could be added at a later date if and when this proves necessary.

The model used is based on CCIR recommendations [Watterson 1970, CCIR 1974] which is the generally accepted multipurpose model for multipath propagation simulators [Dawson 1984]. The computer implementation has a few marked differences to the normal method, but has the same overall effect. The differences, implementation and verification are outlined below. The actual computer program, written in 'C', can be found in appendix I where the modulator section is an 8-tone MFSK system.

Using the parameters mentioned above, the value of the signal from each path at the sampling instant is calculated. The signals from each path are then summed, along with the specified-amplitude Gaussian white noise. This signal is then sent to the demodulator. Differential time delays greater than 2 symbol times cannot be handled by the program. This is not a problem, however, as real channels with more than about 1 symbol period differential delay would be virtually unworkable due to intersymbol interference.

The Rayleigh fading is generated using two independent Gaussian white noise sources, digitally filtered to have a Gaussian power spectral density [CCIR 1974 sec.3]. The two noise sources are then used to generate the Rayleigh amplitude and phase information [Wozencraft & Jacobs 1965]. In the case of there being two magnetoionic components of a par-

ticular ray which are not resolvable in time, we consider each component to be resolvable in frequency. This gets over the problem of having to generate a complicated spectrum and reduces all of the spectra to Gaussian shape. This in turn enables one set of filter constants to be used for all the components. Since it is a digital filter, the rms frequency response, or Rayleigh fading rate, can be fixed by the rate at which we re-calculate the gain of each pair of filters (in-phase and quadrature components). Each pair of filters for each resolvable component has the same bandwidth and updating times. The filters are normalized to give unity power gain.

The modulation method simulated can be selected by simply changing a few lines of the program, as can the bit rate, sampling rate, data structure and so on. In the initial tests of the simulator, PSK modulation was used at a bit rate of 200 bps and a center frequency of 200Hz. The sample period was set to 1ms giving a bandwidth of 500Hz. The fractional out-of-band power was therefore 20dB down [Amoroso 1980]. This was used as a 'typical' system for some initial tests. The bit rate is around the maximum useable on an HF link due to intersymbol interference

In order to verify the error characteristics of the program, it has been run using the CCIR recommended parameters for a moderate channel [CCIR 1974 sec 4.1]. The parameters used are two equal amplitude skywave components with a 0.5ms delay between them and a fading rate of 0.1Hz. Gaussian white noise is added at 3dB below the individual signal levels.

The output of a demodulator following the channel simulator, in the form of hard decisions, has been analysed for burst error lengths using another program. A burst is defined as a sequence of demodulated data with more than 40% of the bits in error [Honary 1981]. The results are given in table 2.1 and are typical of an HF link.

In order to provide some verification of the Rayleigh fading simulation, a run has been done over 2.5 seconds (500 bits) at an rms fade rate of 4Hz. The output is shown in Figure 2.12. The upper half of the enve-

Table 2.1. Bursts of errors in a simulated HF link.

Burst length	Occurrences
1	54
2	2
3	39
4	1
5	60
7	5
9	3
10	1
11	4
13	4
14	1
15	6
19	2
21	4
23	3
25	80
31	1
33	10
34	1
44	1
48	1
51	7
59	1
100	2
101	21

lope has had superimposed upon it the Rayleigh attenuation and the median value of the envelope . The steps that can be seen in the envelope are due to the finite period (ie a period greater than the system sampling period) between the Rayleigh fading filters being updated.

The rms fading frequency is defined as the rms of the Gaussian power spectrum of the fading signal. It can be shown [Ralphs 1985 pp49] that the fading rate, defined as the number of positive going crossings of the signal envelope through its median value per unit time, is related to the rms fading rate by

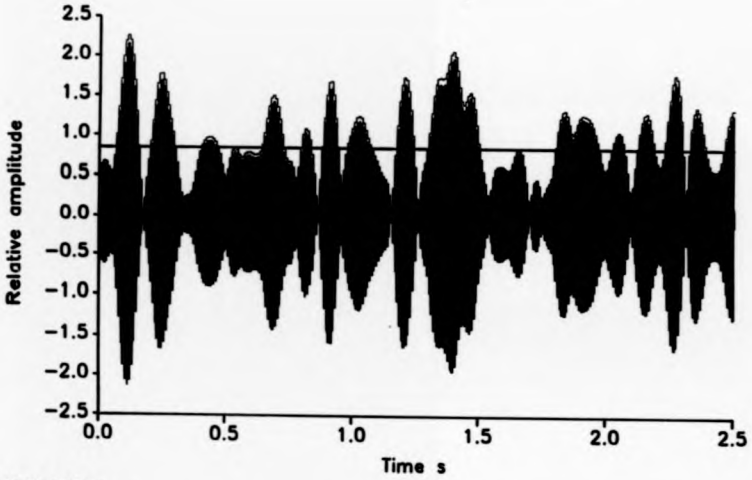
$$\text{fading rate} = 1.47 * (\text{rms fading frequency}) \qquad (2.16)$$

It can be seen in Figure 2.12 that there are 15 positive going crossings of the signal envelope through the median value. The theory says there should be 4Hz*2.5s*1.47 = 14.7, and so the simulation is assumed to be verified.

The Rayleigh amplitude and phase for a much longer run have been plotted on a polar plot (Figure 2.13). This is a representation of the 2 dimensional random walk that can be thought of as producing the Rayleigh fading [Ralphs 1976, Excell 1985]. The actual 'walk' from the simulation in Figure 2.12 is shown in Figure 2.14 [Ralphs 1985 pp51]. The number of cycles of fading can be seen by the number of positive going crossings of the circle representing the median value of the envelope.

Interference has not been simulated here, although it would be possible to add this to the program using similar methods to Dawson [1984]. The problem with this is that there are no standard interference models available although some work has been done on this [Dawson 1984, Perry & Abraham 1986].

Figure 2.12
Rayleigh fading signal from HF simulator
4 Hz rms fading rate.



500 bits @ 200 bps.
Median value of envelope shown.

Figure 2.14 Rayleigh amplitude and phase diagram for the simulation in Fig 2.12. Median circle shown.

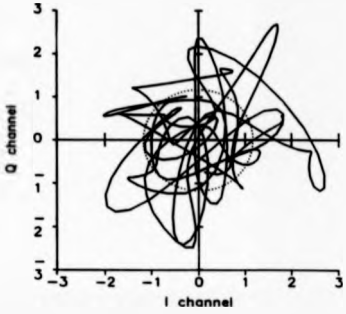
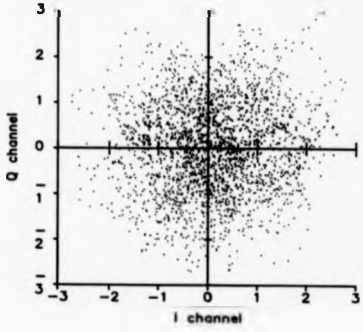


Figure 2.13 Rayleigh amplitude and phase diagram for 20 seconds



2.5 Treatment of Zerocrossings in a Rayleigh Fading Channel.

In the case of a Rayleigh fading channel, we can still use the zerocrossing counting technique to evaluate the channel but now we must consider the amplitude of the signal as remaining constant over the length of time that we take to make our assessment. This means that the assessment must be made in a short length of time with respect to the fading period. This has to be traded off against the period needed for an accurate assessment and so a compromise must be made.

It is not clear from the theory what the 'best' time interval will be and so a simulation has been done to try some different intervals. The simulation has been done using the simulator described in section 2.4 above. We have simulated one skywave path with an rms fading frequency of 0.1Hz and a bit rate of 200 bps for a 5 minute period. Noise was introduced at a level of -3dB relative to the mean signal level. The magnitude variation over this period is shown in Figure 2.15.

The error rates and zerocrossing rates have been averaged over 100 bits (0.5s), 200 bits (1s) and 2000 bits (10s) and the results shown in Figures 2.16a-b, 2.17a-b and 2.18a-b. For the 0.5s and 1s averages, only half of the data is shown for clarity. In order to see how well the zerocrossing data is estimating the bit-error-rate (BER), AZCPS against $\log_{10}$BER for the 3 cases have been plotted and the results are shown in Figure 2.19a-c. The theoretical result for fixed E_b/N_o's is also shown for reference. It can be seen that the difference between the 0.5 and 1 second averages is not that great but when we go up to 10 seconds, the similarity to the theoretical curve gets worse. It is believed that this is due to the averaging process used and the non-linear relationship between the AZCPS and the BER. Clearly the 10 second average case would be of little use to us in the RTCE environment anyway because it cannot give us an approximation to the real signal quality since it's estimation time is too long with respect to the rms fading frequency.

Figure 2.15
Magnitude variation over a 5 minute
Rayleigh fading simulation.

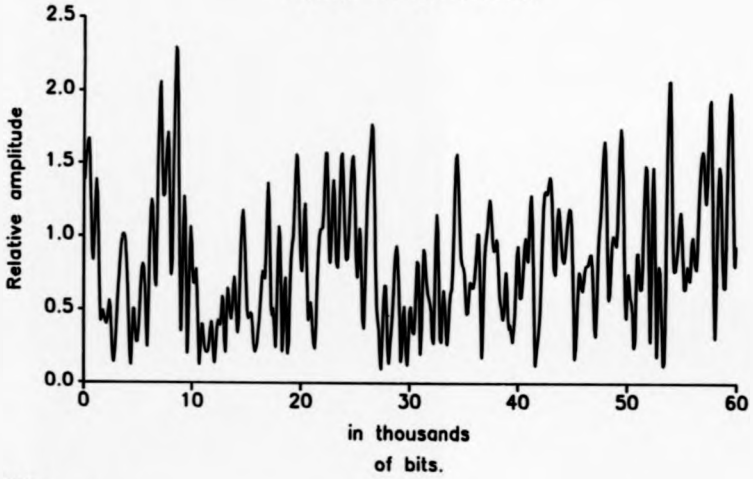


Figure 2.16a
Errors averaged over 100 bits (0.5 s).

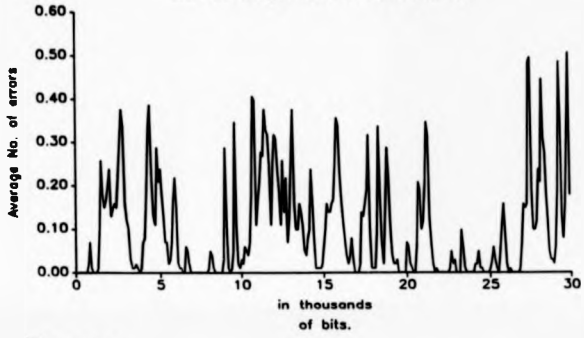


Figure 2.16b
AZCPS over 100 bits (0.5 s)

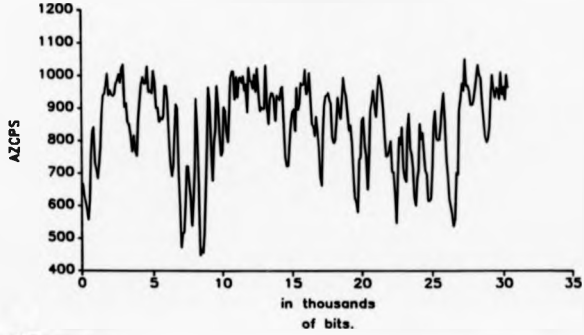


Figure 2.17a
Errors averaged over 200 bits (1 s)

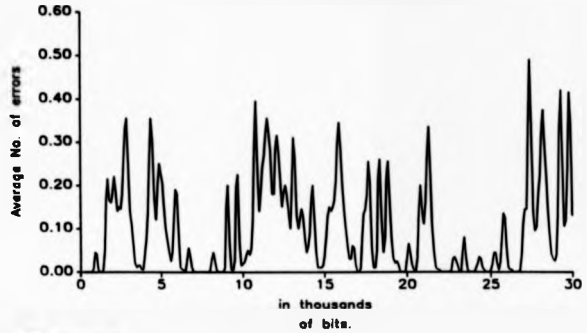


Figure 2.17b
AZCPS over 200 bits (1 s)

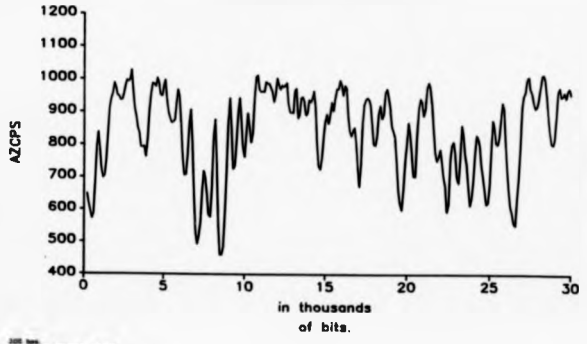


Figure 2.18a
Errors averaged over 2000 bits (10 s)

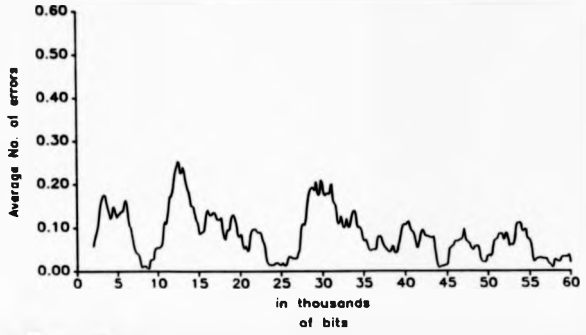


Figure 2.18b
AZCPS over 2000 bits (10 s)

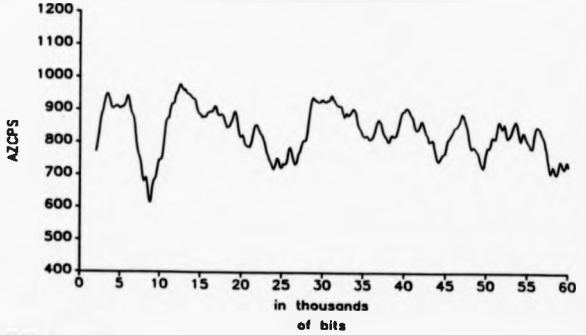


Figure 2.19a
AZCPS vs log BER. 100 bit average.

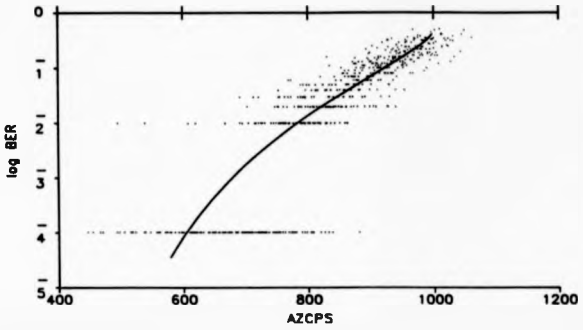


Figure 2.19b
AZCPS vs log BER. 200 bit average

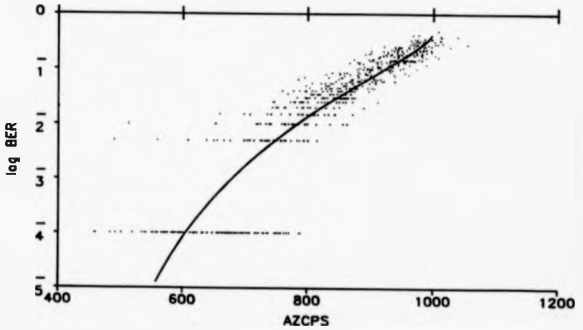
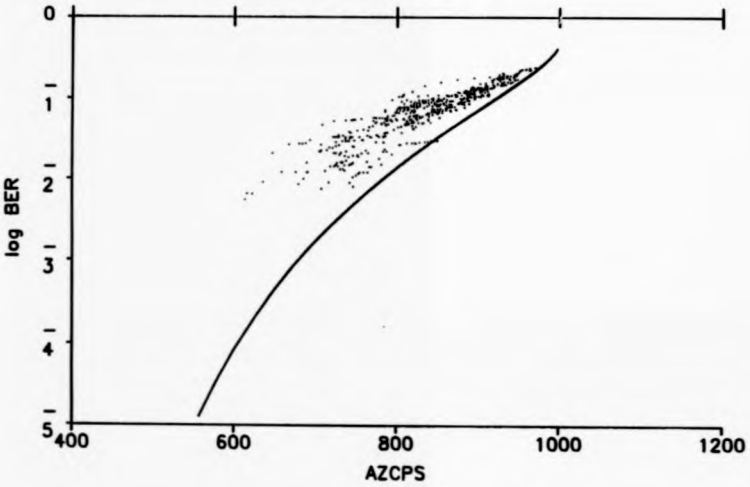


Figure 2.19c
AZCPS vs log BER. 2000 bit average.



This page intentionally left blank.

Chapter 3. Higher Order Crossings.

3.1 Introduction to higher-order crossings.

In chapter 2, a method of channel analysis was presented which relied on the zerocrossings of a sampled waveform. In this chapter, the work will be extended to cover the zerocrossings of new series which are generated by taking the differences or sums of the elements in the original series. The differencing or summing may be done many times in order to produce many series. This work has been developed as a result of some papers by Kedem [1984, 1986a-c] in which he discusses 'higher order zerocrossings' (HOX). The extra information that can be gained from looking at these different order zerocrossings will be used to provide a higher degree of accuracy to the RTCE method previously described. An MFSK demodulator is also described which relies on the HOX as a method of signal detection. Its performance is compared to the noncoherent matched filter detection scheme described in chapter 4.

The difference between the method described by Kedem [1984, 1986a-c] and the one that we consider is that in his theoretical work, he works with a time series with a Gaussian probability density function (pdf) and (possibly) sinusoidal components [Kedem 1986a]. In our method, we consider a sinusoid with additive Gaussian noise. The pdf of our waveform is therefore not Gaussian but rather the convolution of a Gaussian pdf with that of a sinusoid [Rice 1946 Fig 1]. Hence the analysis of Kedem is of no use to us.

Kedem does however use a useful normalization, that of scaling the time series such that the sampling interval becomes 1s. Under this transformation, AZCPS becomes the probability of a zerocrossing per sample which has previously been denoted $P_{=}$. This gives a maximum normalized frequency (by the Nyquist sampling theorem) of π rad/s or 0.5Hz. This normalization is quite useful in that it makes the calculations independ-

ent of the system under consideration and simplifies notation and so has been adopted throughout.

3.2 Theoretical treatment of higher order zerocrossings.

Further to the theoretical analysis presented in chapter 2 on the zerocrossing probability of a time series, the work can now be extended to higher order crossings. The probability of a zerocrossing, P_0, of a sampled waveform has been defined as the probability of two successive samples having opposite sign. HOX are defined in a similar way except that the original series is first differenced or summed, possibly many times, before the crossings are counted. We designate the zerocrossing probability of the original series as 0^{th} order zerocrossings. Differencing gives a positive order and summing gives a negative order. The set of HOX for the orders being considered will be known as a HOX vector. This is a different notation to Kedem [1986a] but serves our purpose better. Here we consider the theoretical expression of P_0 for the once differenced series as the other orders are given by an analogous procedure.

The effect that the differencing or summing procedure has on the series is that the noise samples can no longer be considered independent, ie they have some correlation. Also the sinusoidal component changes in magnitude and phase. This can most easily be seen by an example.

Suppose that we have a sinusoid of unity amplitude in AWGN and describe each sample thus

$$f(n.\delta t) = \sin(w.n.\delta t+\theta) + N(n.\delta t) \qquad (3.1)$$

where the $N(n.\delta t)$ are independant noise samples of standard deviation σ and θ is an arbitrary phase which, without loss of generality, can be ignored.

The series is normalised by setting δt to 1s and then the differ-

ences between successive samples are taken. We get the series

$$f'(n.\delta t) = [\sin(w.n) + N(n)] - [\sin(w.(n-1)) + N((n-1))] \quad (3.2)$$

$$= A.\sin(w.n+\theta) + N'(n) \quad (3.2a)$$

where

$$A = 2.\sin(w/2)$$

$$\theta = (w-\pi)/2$$

$$N'(n) = N(n) - N(n-1)$$

θ is constant and (as will be seen later) can be dispensed with.

If we consider a sample of f'() we can see that the variance of the new noise samples, N'(n), is $\sigma'^2=2.\sigma^2$, since the noise samples are independant. In addition, we can calculate the correlation coefficient, β, of the noise by use of the formula [Spencer et. al. 1980, page 486]

$$\beta = \mu_{11}/(\sigma'.\sigma')$$

$$= \mu_{11}/(2.\sigma^2) \quad (3.3)$$

where μ_{11} is the covariance of the 2 samples under consideration and is given by

$$\mu_{11} = E[N'(n).N'(n-1)] - E[N'(n)].E[N'(n-1)] \quad (3.4)$$

and E[x] denotes the expected value of x. The mean value of the noise samples is zero and so the second term of (3.4) is zero and we have

$$\mu_{11} = E[[N(n)-N(n-1)].[N(n-1)-N(n-2)]] \quad (3.5)$$

but since E[x+y] = E[x] + E[y] and the expected value of the product of two indepandent random variables of zero mean is zero,

$$\mu_{11} = E\{ -N((n-1))^2 \}$$
$$= -\sigma^2$$

and therefore, from (3.3)

$$B = -1/2 \qquad (3.6)$$

The values of σ'', A and B have been derived for the once differenced case but in general can be derived for any of the summing and differencing cases. The values for all of the cases are given in table 3.1.

A general formula for the number of zerocrossings of any order n can now be found in terms of the parameters σ, w and B. The general notation D(n,σ,w) has been used here for this value, noting that σ now refers to the standard deviation of the noise in the series under consideration. The probability of a zerocrossing in g(n), the general series, is found in a similar way to (2.5). That is, we find the probability of one of a

---

Table 3.1. Time series variance, amplitude and correlation parameters for different orders of differencing and summing.

Series	Order n	Variance σ''^2	Correlation coefficient B	Amplitude A
original	0	σ^2	0	1
once differenced	1	$2.\sigma^2$	-1/2	$2.\sin(w/2)$
twice differenced	2	$6.\sigma^2$	-2/3	$4.\sin^2(w/2)$
n times differenced	n	$\sum_{r=0}^{n} \left[\frac{n!}{r!(n-r)!}\right]^2 \sigma^2$	-n/(n+1)	$(2.\sin(w/2))^n$
once summed	-1	$2.\sigma^2$	1/2	$2.\cos(w/2)$
twice summed	-2	$6.\sigma^2$	2/3	$4.\cos^2(w/2)$
n times summed	-n	$\sum_{r=0}^{n} \left[\frac{(-n)!}{r!(-n-r)!}\right]^2 \sigma^2$	n/(n-1)	$(2.\cos(w/2))^n$

pair of successive samples being positive and the other negative. This expression is then averaged over all possible phases of the sinusoid to give the result. The distribution of the pair of noise samples is bivariate normal with correlation coefficient (ß) given in table 3.1. Hence we have the two samples

$$g(n-1) = A.\sin(w.(n-1)+\theta) + N'(n-1) \qquad (3.7a)$$

$$g(n) = A.\sin(w.n+\theta) + N'(n) \qquad (3.7b)$$

and the probability distribution of the noise [Johnson & Kotz 1972]

$$P(N'(n-1)=x,\ N'(n)=y) = \frac{\exp\{\ -(x^2-2ßxy+y^2)/(2.(1-ß^2).\sigma^2)\ \}}{2.\pi.\sigma^2\sqrt{(1-ß^2)}} \qquad (3.8)$$

using x and y as dummy variables.

From this we can obtain the probability of a positive going zerocrossing, $Pz+$, ie when $g(n-1) < 0$ and $g(n) > 0$, as

$$Pz+ = P(\ N'(n-1) < -A.\sin(w(n-1)+\theta),\ N'(n) > -A\sin(wn+\theta)\ )$$

$$= \frac{1}{2.\pi.\sigma^2\sqrt{(1-ß^2)}} \int_{y=-A\sin(wn+\theta)}^{\infty} \int_{x=-\infty}^{-A.\sin(w(n-1)+\theta)} \exp\{\ -(x^2-2ßxy+y^2)/(2.(1-ß^2).\sigma^2)\ \}\ dx.dy \qquad (3.9)$$

Now, setting n=1, integrating over all phases of θ, making a change of variable and changing the limits slightly we get

$$Pz+ = \frac{1}{4(\pi.\sigma)^2\sqrt{(1-ß^2)}} \int_{-\pi}^{\pi} \int_{A.\sin(\theta)}^{\infty} \int_{-\infty}^{A.\sin(\theta-w)} \exp\left[\frac{-(x^2-2ßxy+y^2)}{2\sigma^2(1-ß^2)}\right].dx.dy.d\theta \qquad (3.10)$$

and we note that, due to symmetry, the probability of positive and negative going zerocrossings are the same, giving us

$$D(n,\sigma,w) = \frac{1}{2(\pi.\sigma)^2\sqrt{(1-\beta^2)}}\int_{-\pi}^{\pi}\int_{A.\sin(\theta)}^{A.\sin(\theta-w)} \exp\left[\frac{-(x^2-2\beta xy+y^2)}{2\sigma^2(1-\beta^2)}\right].dx.dy.d\theta \quad (3.11)$$

where A and β are given in table 3.1.

Equation (3.11) is of a similar form to equation (2.5) and reduces to it in the case when n=0 (and so β=0). It is also mathematically intractable and so the values of D(n,σ,w) have been sought by several different methods.

The first method has been to simulate a signal of the required frequency, add random noise and count the crossings and higher order crossings. The problem with this is that it takes a large number of samples to get an accurate result.

The second method, detailed in appendix G, has been an analysis employing a Fourier transform technique similar to that of appendix D. This method works well for moderate values of SNR but it has proved difficult to decide when to terminate the resulting nested summations. This is a particular problem at higher SNRs as the waveform which is being analysed becomes almost square and so has large high frequency components.

The third method is detailed in appendix F and employs 2 dimensional modified Gaussian integration of a function derived from (3.11). The integrals converge to an accuracy of 1 in 10^6 for SNRs up to 5dB. The results have been checked against results from a simulation done over 10^8 samples and are reliable. A further modification is required for convergence at higher SNRs.

Some plots of HOX versus SNR and frequency have been prepared from the simulation results and can be seen in Figure 3.1a-h. There is one graph for each frequency in an 8-tone MFSK system and each graph covers HOX orders from -4 to 4. They show the probability of a zerocrossing between a pair of samples at a particular SNR. The upper line on each graph represents the highest HOX order due to the fact that the differencing

Figure 3.1. Higher order crossing probabilities vs SNR for each tone of an 8 tone MFSK system.

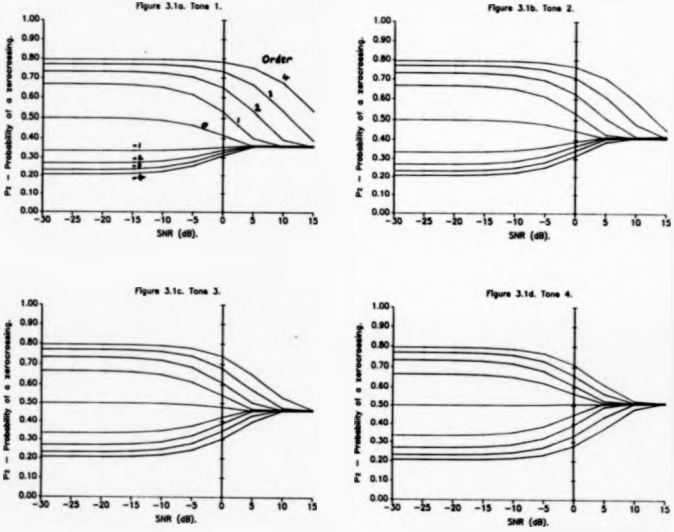


Figure 3.1e. Tone 5.

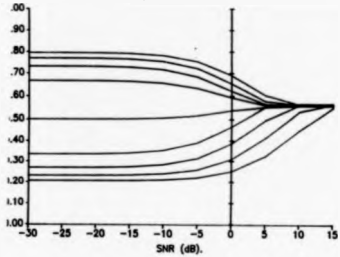


Figure 3.1f. Tone 6.

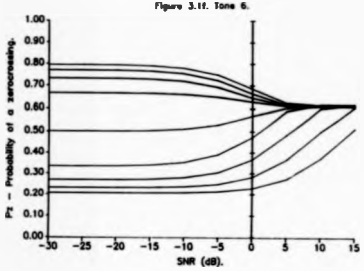


Figure 3.1g. Tone 7.

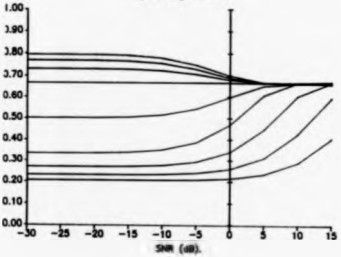
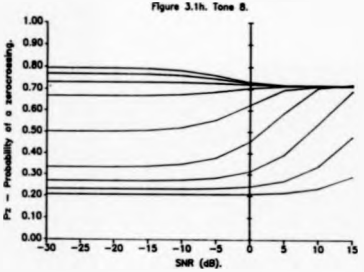


Figure 3.1h. Tone 8.



procedure is essentially differentiating or high-pass filtering, causing the number of crossings to increase. Summing is like integrating or low-pass filtering and so causes the number of crossings to decrease. The difference in the plots at high SNRs is due to the different frequencies present in each case and the similarity at low SNRs is due to the fact that the noise is swamping the signal in each case.

From the plots it is clear that the HOX behave in a similar way to the zero order crossings with respect to frequency and SNR.

3.3 Application of HOX

3.3.1 MFSK Demodulation using HOX

In order to use the counts of HOX over each symbol for demodulation purposes, the hardware system to be used must first be characterised. This is due to the fact that the analysis presented above is valid only when the noise being added to the system is Gaussian and there is no correlation between consecutive noise samples. In practice, systems can only be approximated to this due to the fact that the noise will not generally be either white or Gaussian. Systems may be characterised by finding the HOX vector values for each of the symbols in an MFSK system under various noise conditions and storing the results in a look up table. In order that the mean and standard deviations may be found to a reasonable degree of accuracy, the data must be taken over a sufficiently large number of symbols. Typically, orders from 4 to -4 may be considered for perhaps 10 different SNRs.

To demodulate a signal, the HOX vector must be found for each symbol. This vector, which hopefully characterises the received symbol, may then be compared with each of the stored reference vectors in turn and a measure of the 'distance' between them calculated. The distance is taken by finding the differences between the vector elements. Each difference is then divided by the standard deviation appropriate to that element, in

order to weight them according to the relative amount of information in each. The sum of the squares of the weighted differences is then taken as the distance measure from the received MOX vector to the reference vector. The reference vector with the smallest distance from the received MOX vector is then selected to be the one demodulated. Both symbol and SNR information are thus available and can be used for later analysis of reliability. The smallest distance found could also be used as soft decision information for an appropriate type of decoder. Symbol timing has been assumed to be known throughout this procedure.

Some results employing the above procedure for an 8-tone MFSK modem are presented in section 5.4 and they show that the demodulation technique using MOX is not as good as results from more conventional techniques. The reason for this can be seen in Figure 3.2a-c where we plot the pdf of zerocrossings occurring in a symbol of a typical system. The system chosen for this example is a 4-tone MFSK modulator (see Table 4.4, chapter 4) operating over an AWGN channel at various SNRs. Tone 0 has 3 cycles per symbol whilst tone 3 has 6. The numbers of zerocrossings are twice these figures. Figure 3.2a shows the pdfs due to different orders of zerocrossings, taken at one particular SNR and tone. It shows how the shape of the pdf is affected by the order of MOX taken and the particular tone present. It also gives some idea of how large the deviation of counts over individual symbols is. Figure 3.2b shows the 0^{th} order zerocrossings pdf for different tones. It can be seen that even at 5dB SNR, there is a large amount of overlap between the curves, resulting in the poor error rate performance found for the MOX MFSK demodulator. Figure 3.2c shows how the pdf of the zerocrossings changes with SNR for a particular tone. It can be seen that below about -5 dB the curves are very similar.

In order use MOX for reliable demodulation, we need to either increase the difference between the means of the pdfs for different tones, or reduce their standard deviations in order to increase their distin-

Figure 3.2a
Probability density function of HOX.
Tone 0, 5dB SNR.

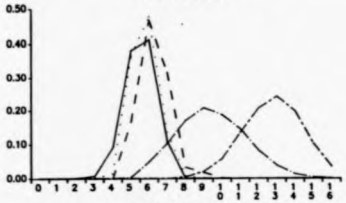


Figure 3.2b
Probability density function of HOX.
Order 0, 5dB SNR.

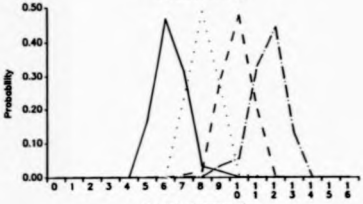
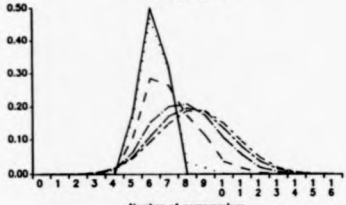


Figure 3.2c
Probability density function of HOX.
Tone 0, Order 0.



guishability. The first of these ideas implies using a wide-shift FSK system where the tones are not at their minimal spacing for orthogonality. This has not been investigated further since these types of system are not very bandwidth efficient [Schwartz 1980 pp217]. The second suggestion implies that we need a larger numbers of samples of the waveforms over which to count the zerocrossings. This also cannot be done without slowing down the signalling rate, unless it can be done over several symbols. The system is then of no use for demodulation and only its channel evaluation properties are preserved. This is investigated further in the following section.

3.3.2 RTCE using HOX

In order to use HOX for channel evaluation we must first use some conventional demodulator and synchronisation system to find which symbol was transmitted and the relative timing. We can then use the HOX vector derived from the signal to estimate the SNR in a similar way to that described in the previous section. The difference is that we now already know the frequency of the received signal and the symbol timing information, except in the case when an error is made by the conventional demodulator. This means that the search through the look up table of reference vectors is reduced to only searching the vectors corresponding to the already demodulated symbol. Also, since the SNR is assumed to be changing relatively slowly compared to the symbol rate of the system (a definition of the 'slow fading channel' [Ralphs 1985 pp52]), it is possible to make the channel assessment from several successive symbols. As mentioned in the previous section, this increases the time over which HOX are being counted and hence helps to reduce the standard deviation of the results. Thus more reliable results can be expected than those for the demodulator previously described.

Results from the above procedure, using the same parameters as in

the previous section, are given in section 5.5. The tests have shown that the method does estimate the channel SNR quite well. The period over which the SNR assessment can be made is a function of the fade period, whilst the use of the assessment is a function of the coding system employed. This means that the assessment period needs to be adjusted with reference to the overall communications system and the channel over which it is to be used.

Overall, the MOX RTCE method has a high complexity compared to the system to be described in chapter 4, since a conventional demodulator is required as well as the MOX counting circuitry/software. However, it may prove to be useful as an auxiliary RTCE method in some circumstances.

This page intentionally left blank.

Chapter 4. Multi-frequency shift keying (MFSK).

4.1 Introduction

The history and development of multi-frequency shift keying (MFSK) has been exposed by Ralphs [1985] and we shall be referring to this text on many occasions. The basic assumption that we shall make here is that MFSK can be a 'good' method of modulation for the HF channel (compared to more conventional binary FSK and PSK systems) due to its long element period. This long element period has two advantages. The first of these can be seen from consideration of some basic information theory which is normally derived for the AWGN channel. This work [Shannon & Weaver 1949] states that for a given bandwidth and SNR there is a maximum rate of information at which one can transmit information with negligible errors. This can be illustrated by consideration of a system transmitting one of M possible orthogonal signals in AWGN at a fixed data rate and using 'matched filter' [Stremler 1982 pp406-412] detection. We can see that as M increases and the data rate is held constant, each symbol carries more information and hence the symbols must be longer. Therefore we will be integrating the noise over a longer period and so its standard deviation will be reduced [Bozic 1979 pp88]. Hence the amount of noise at the output of the matched filter will be reduced, enabling a more reliable estimate of the transmitted symbol to be made. This argument is similar to the 'sphere hardening' argument which can be used to prove Shannons' result [Wozencraft & Jacobs 1965 pp323-341]. In this case the 'dimensions' are represented by the orthogonal matched filters. This is reasonable since the concept of Cartesian co-ordinates is that they are at right angles to each other or, put another way, they are orthogonal. Therefore the concepts of 'orthogonal signals' and 'dimensions' are equivalent. The relationship between the symbol period and M will be investigated further in section 4.2 with specific reference to MFSK systems.

The second and perhaps main advantage of the long MFSK element periods is the increased immunity from multipath effects. This immunity is due to the fact that a signalling element can be long compared to the multipath delay of the channel. Hence any time 'overlap' between successive received symbols, as a percentage of the element period, is reduced as M increases. As a result, intersymbol interference (ISI) problems are reduced. A similar thing can be said of symbol timing recovery errors.

In contrast to the advantages of a long symbol period, there are reasons why the element period cannot be extended indefinitely. The most obvious of these is that the decoding delay will be very long with a very long symbol. Secondly, and more importantly in most cases, a long symbol requires very good stability of the transmitter and receiver clocks and, more particularly, of the channel. The HF channel in general exhibits phase and frequency perturbations and so limits the accuracy to which the timing may be held.

Another ionospheric effect which is not overcome simply by making the symbols longer is the of frequency selective fading due to multipath effects. This is the effect, described in section 1.4.5, whereby if we have two equal amplitude skywave components with differential time delay δt, the two signals will cancel if their frequency is an odd integer multiple of $1/2.\delta t$ since they will arrive in exactly opposite phases.

From that which has been mentioned above, it is clear that the symbol period used has to be a compromise between many different factors. In section 4.2 we shall investigate how many tones should be used with respect to the noise level of the channel and whether there is slow Rayleigh fading or not. It is assumed that a fixed bandwidth is available and that 99% of the power of the signal should fit into this bandwidth.

In section 4.3, the PSD of MFSK signals is investigated so that the optimisation procedure of section 4.2 can be applied. Results from the optimisation procedure are described in section 4.4 and methods of implementing the resulting signal format are described in section 4.5. Sec-

tion 4.6 discusses the means of demodulating the signals designed and in section 4.7, the implementation algorithm is described.

4.2 Optimisation of the number of tones.

Here we shall try and consider the parameters of an MFSK system in the environment of HF radio communications [Shaw, Honary & Darnell 1988b]. It is assumed throughout that the bandwidth of the system is fixed and cannot be changed. This is a reasonably practical assumption since many transmitters and receivers will not have variable intermediate frequency (IF) filters. Also, frequency allocations in the crowded HF bands do not normally allow for a very wide-band signal, even if interference levels made such things practical. Since the actual bandwidth is not of concern to us here, we will work in terms of bits per second per hertz of bandwidth (bps/Hz).

In order to consider the practical parameters which we can use, we shall consider the AWGN channel case and see if we can gain any insight into the way in which the various parameters of the MFSK system interact. We shall try and maximize the information throughput [Abdel-Ghaffar & McEliece 1986], or system data rate, with respect to several parameters. These are

N, the number of tones. N integer ≥ 2

R, the code rate $0 < R \leq 1$

Te, the element period (symbol period)

DF = 1/Te, the tone spacing for orthogonality

etc..

As stated above, we shall assume that the bandwidth, B, is fixed. In the case of the HF channel which we are considering, this is typically 3kHz and is limited by transmitter licence. An approximate expression for

the bandwidth is given by [Ralphs 1985 p36]

$$B = (M-1+G)/Te \text{ Hz} \qquad (4.1)$$

where

G is the total number of tone element spaces (on the frequency axis) beyond the highest and lowest tones giving a 99% power containment bandwidth.

This formula is clear if realise that DF = 1/Te is the minimum tone spacing required for orthogonality and there are M-1 spaces between the tones.

It is suggested [Ralphs 1985 p36] that a typical value of G is 4 throughout a large range of values of M. This will be investigated further in section 4.3.

Obviously, if we have M tones, then each symbol carries $\log_2(M)$ bits of channel information. The symbols last for Te seconds and so the channel bit rate is given by

$$\log_2(M)/Te \text{ channel bps} \qquad (4.2a)$$

or in terms of normalised bit rate as

$$Ne = \log_2(M)/(Te.B) \text{ channel bps/Hz} \qquad (4.2b)$$

which from (4.1) is given by

$$Ne = \log_2(M)/(M-1+G) \text{ channel bps/Hz} \qquad (4.2c)$$

Taking coding into account, the system data throughput N, ie the rate at which data is delivered to the user, is therefore given by

$$N = R.Ne \text{ user bps/Hz} \qquad (4.3)$$

and from (4.2c) and (4.3) we get

$$H = R.\log_2(M)/(M-1+G) \text{ user bps/Hz} \qquad (4.4)$$

This is our expression for the system data throughput.

The maximum amount of error-free data that can be transmitted over the channel is presumably going to be a function of M and SNR. We call this quantity the system capacity. It can be obtained by calculation or approximation for the channel under investigation. For now we can write

$$\text{capacity} = C(M) \text{ user bits per channel symbol.}$$

From the coding theorem, we know that we must transmit at a rate less than this to get reliable communication. Hence we can say

$$R.\log_2(M) \leq C(M) \text{ user bits per channel symbol} \qquad (4.5)$$

and substituting this in (4.4) gives

$$H \leq C(M)/(M-1+G) \text{ user bps/Hz} \qquad (4.6)$$

Hence in order to maximize the system data rate, H, we need to maximize (4.6), at equality, over the possible values of M. In order to do this we need to know C(M). This can be found by use of [Gallager 1968 p74]

$$C(M) = \max_{Q()} \sum_{k,j} Q(k).P(j|k).\log_2\left[\frac{P(j|k)}{\sum_i Q(i).P(j|i)}\right] \text{ bits/symbol} \qquad (4.7)$$

where

Q(k) are the input symbol probabilities

P(j|k) are the transition probabilities

$\sum Q(i).P(j|i)$ is the probability of receiving symbol j.

$\sum Q(k) = 1$

$Q(k) \geq 0$

the 'max' is taken over all possible input probability assignments. In our case the k and j variables run from 1 to M and obviously the P(j|k)s are dependant on M. This means that, in order to maximize M, we need to solve (4.7) for the Q(k) over all values of M and then insert these values in (4.6) to find Hmax. This procedure will need to be repeated for all values of SNR so that when the SNR is known, an appropriate value of M can be selected, along with an appropriate value of R.

In order to make some approximation of (4.6) we can make some assumptions on the parameters in (4.7). Since we are assuming a Gaussian channel with equal input probability assignments we can see that

$$Q(k) = 1/M \quad \text{for all } k \tag{4.8a}$$

$$\sum Q(i).P(j|i) = 1/M \quad \text{for all } j \tag{4.8b}$$

$$P(j|k) = Ps/(M-1) \quad j \neq k \tag{4.8c}$$

$$= 1-Ps \quad j = k \tag{4.8d}$$

where

Ps = symbol error probability.

Hence

$$C(M) = Ps.\log_2(Ps.M/(M-1)) + (1-Ps).\log_2((1-Ps).M) \text{ bits/symbol} \tag{4.8e}$$

An expression for Ps can be found in many sources [Ralphs 1985 p31, Prokis 1985 p212, Lindsey 1964] for the AWGN case and it is given by

$$Ps = \sum_{r=1}^{M-1} \frac{(-1)^{(r+1)}(M-1)!}{(r+1)!(M-r-1)!} \cdot \exp\left[\frac{-r}{(r+1)} \cdot \frac{Es}{No}\right] \qquad (4.9)$$

where

No is the noise density in watts/Hz and so Es/No is the signal to noise energy ratio in the signal filter at the end of a symbol.

Since Es is given by S.Ts, Ts=(N-1+G)/B and No is given by N/B, Es/No becomes (N-1+G).S/N.

Using (4.6), (4.8a) and (4.9) we can now find an upper bound on the error-free data throughput of an MFSK system in AWGN for various values of M.

In order to find the upper bound on error free data throughput in the case of slow (with respect to the symbol rate) Rayleigh fading, we need to replace (4.9) with an expression for Ps for this case. This is given by [Lindsey 1964]

$$Ps = \sum_{r=1}^{M-1} \frac{(-1)^{(r+1)}(M-1)!}{r!(M-r-1)!} \cdot \frac{1}{1+r+r.Es/No} \qquad (4.10)$$

(We must note here that this expression is also given in Ralphs, pp52, but is typographically incorrect in several respects.)

In order to make use of the equations given above to investigate the possible advantages of using different numbers of tones, we must now look at the power spectral density of the various systems. This is so that we can set the parameter 'G' to an appropriate figure.

4.3 Power Spectral Density Calculations.

In this section, the power spectral density (PSD) of MFSK signals for various values of M is investigated. This will enable us to find the

value of 'G' to use in the analysis of the previous section.

The PSD of MFSK signals can be found by use of formulae given by Anderson and Salz [1965] as

$$\frac{G(\alpha)}{A^2T} = \frac{1}{M} \cdot \sum_{m=1}^{M} \left[\frac{1}{2} \operatorname{sinc}^2(\theta n) + \frac{1}{M} \cdot \sum_{n=1}^{M} J.\operatorname{sinc}(\theta m).\operatorname{sinc}(\theta n) \right] \qquad (4.11a)$$

where

α is the frequency, normalised to the tone spacing.

$$\theta n = (\alpha - an/2).\pi \qquad (4.11b)$$

$$an = 2.n - M + 1 \qquad n = 1,2...M \qquad (4.11c)$$

$$J = \frac{\cos(\theta n + \theta m) + \cos(\theta n + \theta m - 2.\pi.\alpha)}{2.(1 + \cos(2.\pi.\alpha))} \qquad (4.11d)$$

$$\operatorname{sinc}(x) = \sin(x)/x$$

It can be seen from 4.11d that J tends to ∞ when $\alpha = 1/2,\ 3/2$ etc. and this means that there are spectral lines in the spectrum at the tone frequencies. Each of these contain $1/M^2$ of the total energy in the spectrum [Anderson & Salz 1965].

The PSDs have been plotted for several different values of M and are given in Figure 4.1a-e . The frequency axis is in terms of α, that is normalised so that the units are the number of tone intervals away from the center frequency. Only the upper half is shown for clarity since the spectra are symmetrical.

It is now possible to find the 99% power containment bandwidth of MFSK signals with respect to M. This enables evaluation of the value of G to use in the equations of section 4.2. In order to do this, we have plotted the signal power rejected by an ideal bandpass filter with cutoff at π/Te Hz from the center frequency. This has been done for those values of M that we have been considering and has been found by numerical

Figure 4.1a
PSD of 2 tone MFSK or binary FSK

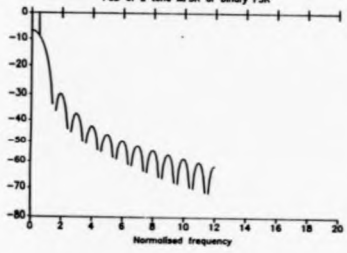


Figure 4.1b
PSD of 4 tone MFSK

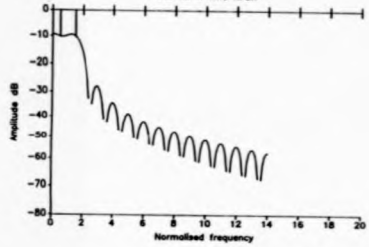


Figure 4.1c
PSD of 8 tone MFSK

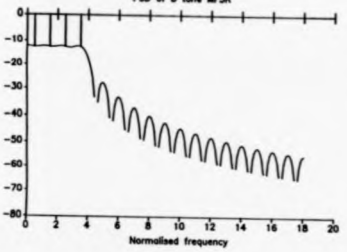


Figure 4.1d
PSD of 16 tone MFSK.

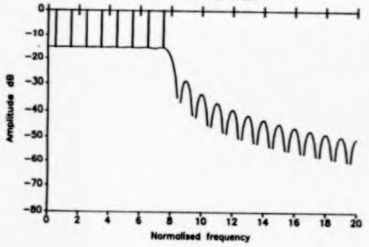


Figure 4.1e
PSD of 32 tone MFSK

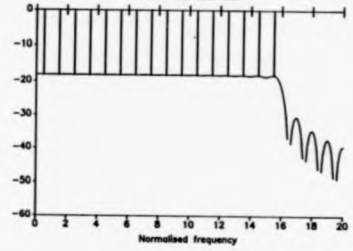


Figure 4.2a
CPSD of 2 tone MFSK or binary FSK

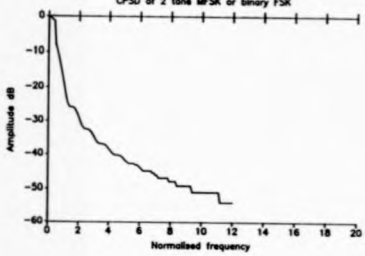


Figure 4.2b
PSD of 4 tone MFSK

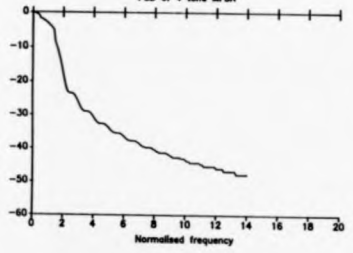


Figure 4.2c
CPSD of 8 tone MFSK

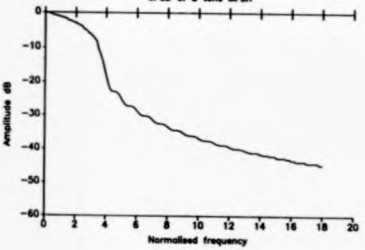


Figure 4.2d
CPSD of 16 tone MFSK.

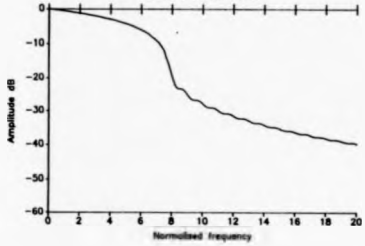


Figure 4.2e
CPSD of 32 tone MFSK.

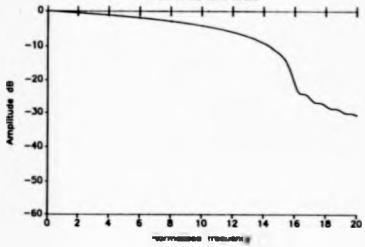


Figure 4.3
Fractional out of band power for MFSK
signals, scaled for constant bit rate

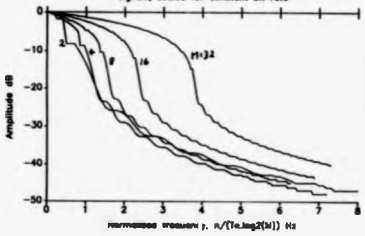
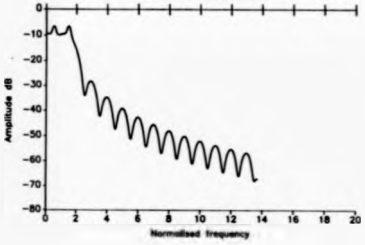


Figure 4.4. Fast Fading PSD.
4 tone MFSK, Frms=0.1*DF.



integration of the PSD formulae (4.11a-d). Special care was needed around the spectral lines at each tone frequency. These plots are shown in Figure 4.2a-e and have been re-plotted in Figure 4.3 with the frequency axis scaled so that the rate (in bps) is the same for each case. This shows the bandwidth expansion of approximately M/log(M) discussed in the following section. A selection of relevant points from these plots are given in table 4.1. Note that 99% power containment occurs when -20 dB is rejected.

It can be seen from table 4.1 that since the B is given by M=1+G, G works out to be around 1.2 for the cases that we have taken. This is within the main lobe of the PSD and is considerably narrower than the estimate given by Ralphs [1985 pp34-36]. The value G=4 specified by Ralphs always gives the bandwidth containing the main lobe and the first sidelobe each side of the center frequency. The difference in our calculations and Ralphs' is due to the fact that he uses a somewhat inaccurate approximation for the PSD calculations.

These results lead us to the conclusion that setting G at 2 is a reasonable thing to do for all cases of interest. This means taking in the main lobe of the PSD only and gives a normalised bandwidth of M=1.

In the case of fast fading signal conditions, the received signal spectrum is broadened [Ralphs 1985 pp57]. An example of the PSD of a

Table 4.1. Power spectral characteristics of various MFSK systems.

M	99% power B, α	Rejected power, dB 1st node	2nd node	3rd node	4th node	5th node
2	2.13	-25.97	-32.68	-37.08	-40.36	-42.84
4	4.23	-23.57	-29.07	-32.74	-35.59	-37.85
8	8.21	-22.96	-27.35	-30.33	-32.61	-34.53
16	16.08	-23.32	-26.79	-29.12	-30.95	-32.49
32	31.8	-24.30	-27.05	-28.89	-30.31	-31.56

4-tone MFSK signal under these conditions has been calculated to illustrate this. This has been done by convolving the spectrum given by (4.11) with the Gaussian power spectrum which would result from the fading process 'modulating' a constant carrier [Ralphs 1985 pp57]. The result is shown in Figure 4.4 and it can be seen that although the spectral lines have been 'smeared' in frequency, the overall spectrum has not been broadened much at all. The case that has been plotted is a quite extreme example where the rms fade rate, fr, is 10% of the tone spacing. In the systems that we have considered the tone spacing is typically around 250Hz and so this amount of fade would be around 25Hz, which is higher than that normally encountered [Ralphs 1985 pp60]. In some cases however where the number of tones is very high, ie they are closely spaced, this could be a significant quantity. The 99% power containment bandwidth for this example has been found to be 4.26 times the tone spacing, which is still within the main lobe of the PSD and so does not alter the conclusion that the normalised bandwidth may be taken as M+1 for all cases of interest.

4.4 Results of optimisation procedures.

From equations (4.6), (4.8a), (4.9) and (4.10) given above, we can now find the optimum value of M required for an MFSK system operating over either a flat channel or a Rayleigh fading channel, with additive Gaussian noise.

Some results of these calculations can be seen in Figures 4.5a-c to 4.8a-c and table 4.2 and are explained below. Results are presented for G=2 and G=4 for non-fading signals and G=2 for slow Rayleigh fading signals.

It can be seen from Figures 4.5a-c that for a fixed value of SNR, the number of tones used affects the maximum error-free data throughput considerably. The curves rise with M to a maximum throughput and then

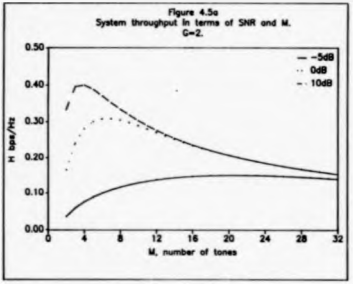


Figure 4.5a
System throughput in terms of SNR and M.
G=2.
H bps/Hz
M, number of tones
−5dB
0dB
10dB
0.00
0.10
0.20
0.30
0.40
0.50
0
4
8
12
16
20
24
28
32

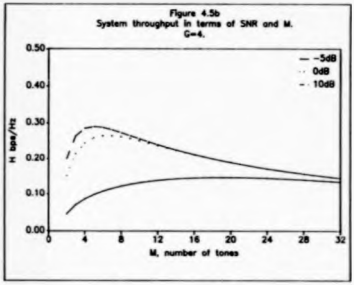


Figure 4.5b
System throughput in terms of SNR and M.
G=4.
H bps/Hz
M, number of tones
−5dB
0dB
10dB
0.00
0.10
0.20
0.30
0.40
0.50
0
4
8
12
16
20
24
28
32

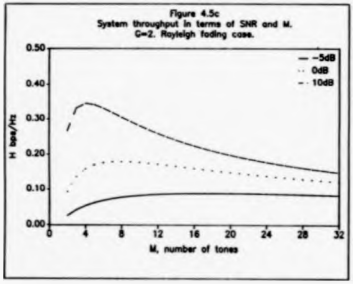


Figure 4.5c
System throughput in terms of SNR and M.
G=2. Rayleigh fading case.
H bps/Hz
M, number of tones
−5dB
0dB
10dB
0.00
0.10
0.20
0.30
0.40
0.50
0
4
8
12
16
20
24
28
32

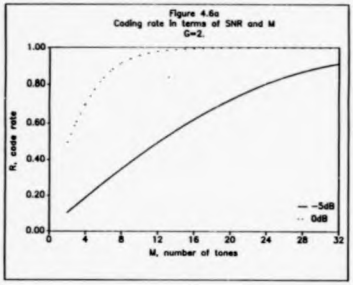


Figure 4.6a
Coding rate in terms of SNR and M
G=2.
R, code rate
1.00
0.80
0.60
0.40
0.20
0.00
0
4
8
12
16
20
24
28
32
M, number of tones
-5dB
0dB

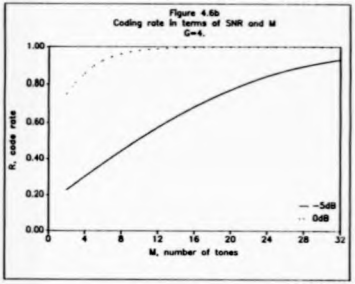


Figure 4.6b
Coding rate in terms of SNR and M
G=4.
R, code rate
1.00
0.80
0.60
0.40
0.20
0.00
0
4
8
12
16
20
24
28
32
M, number of tones
-5dB
0dB

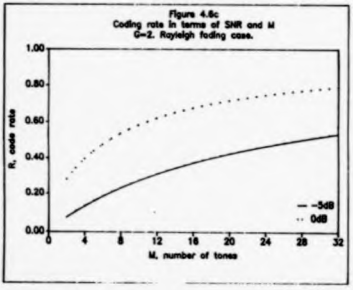


Figure 4.6c
Coding rate in terms of SNR and M
G=2. Rayleigh fading case.
R, code rate
1.00
0.80
0.60
0.40
0.20
0.00
0
4
8
12
16
20
24
28
32
M, number of tones
-5dB
0dB

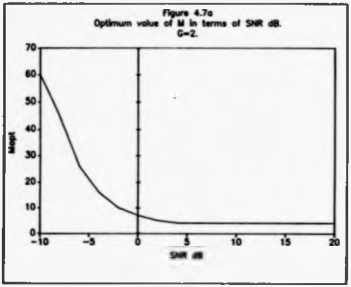


Figure 4.7a
Optimum value of M in terms of SNR dB.
G=2.
Mopt
SNR dB
70
60
50
40
30
20
10
0
−10
−5
0
5
10
15
20

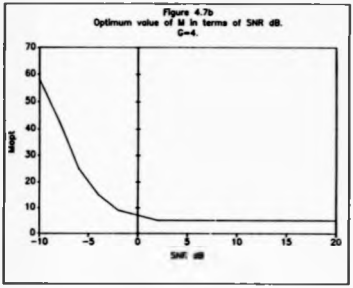


Figure 4.7b
Optimum value of M in terms of SNR dB.
G=4.
Mopt
SNR dB
70
60
50
40
30
20
10
0
−10
−5
0
5
10
15
20

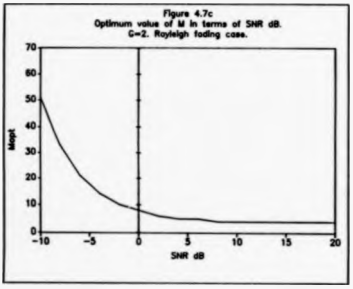


Figure 4.7c
Optimum value of M in terms of SNR dB.
G=2. Rayleigh fading case.
Mopt
SNR dB
70
60
50
40
30
20
10
0
−10
−5
0
5
10
15
20

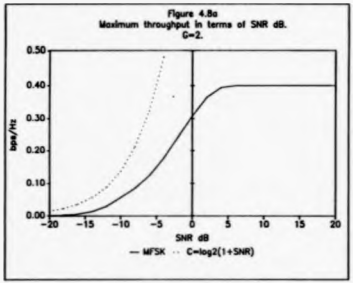


Figure 4.8a
Maximum throughput in terms of SNR dB.
G=2.

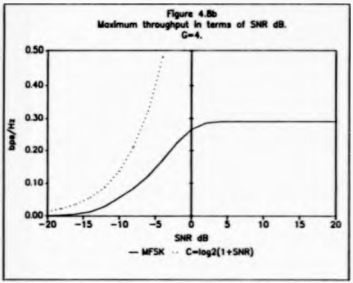


Figure 4.8b
Maximum throughput in terms of SNR dB.
G=4.

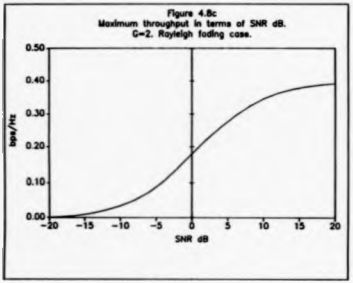


Figure 4.8c
Maximum throughput in terms of SNR dB.
G=2. Rayleigh fading case.

drop off again. This basic shape is due to the fact that, for a constant bandwidth, the symbol rate is roughly proportional to 1/M and the amount of data per symbol is proportional to $\log_2(M)$. Hence the data rate is roughly proportional to $\log_2(M)/M$. This has a maximum at M=3 and decays away as M increases. The actual shape seen is modified from this due to the required coding and the fact that the symbol rate is proportional to 1/(M-1+G) rather than 1/M. It can be seen that there is an increase in throughput when we use G=2 compared to G=4 due to the fact that more of the bandwidth is available to put the tones in. Hence they can be further apart for a fixed value of M and have a shorter period.

In Figs 4.6a-c, the coding rate required to achieve error free communications is shown and in Figs 4.7a-c we show the optimum values of M for a range of SNR dB. The throughput at these optimum values is shown in Figures 4.8a-c in terms of bps/Hz of bandwidth available. In Figures 4.8a-b the theoretical maximum value for error-free throughput [Shannon & Weaver 1949] is also shown for comparison.

With these results we can see that with the normal MFSK format, operating over the channels considered, one cannot get a very high throughput compared to that theoretically possible. This is particularly true at SNRs above about 0dB. It is to be expected however since we are only using one signal level in each of the dimensions available to us and so even when there is no noise, the throughput is quite limited. In order to approach Shannons limit, we would have to allocated more data bits to

---

Table 4.2. Optimum values of M, data throughput and coding rate required for various MFSK systems at different SNRs.

	G=2			G=4			Rayleigh fading, G=2		
mean SNR dB	Mopt	Rmax	R	Mopt	Rmax	R	Mopt	Rmax	R
-10	60	0.056	0.58	58	0.056	0.58	51	0.035	0.32
0	7	0.31	0.82	7	0.26	0.88	8	0.18	0.51
10	4	0.4	0.86	5	0.29	0.9	4	0.35	0.74
20	4	0.4	0.86	5	0.29	0.9	4	0.39	0.84

each dimension, implying possibly that we should amplitude modulate as well as frequency modulate the carrier. For HF communications, the limitations may have to be accepted since the information throughput of a practical system is determined both by the theoretical throughput of the channel and by other, more practical, constraints. These include the desire to have a constant amplitude signal format in order that the transmitter may be operated most efficiently ie at a high peak-to-mean ratio.

Further considerations affecting the choice of modulation for HF communications include the frequency dispersion of the channel, oscillator stability, characteristics of the practical coding scheme employed, and so on [Betts 1967, Stremler 1982, Ralphs 1985, Maslin 1987 etc.]. These problems and means of combating them are discussed further in chapter 6.

4.5 Modulator Design.

In this section, we discuss the practical design and implementation of the MFSK signalling format. The environment which we are considering is, as stated previously, that of an HF transceiver feeding directly into an analogue to digital (A-to-D) convertor and a digital signal processor (DSP). This is seen as a modern alternative to conventional hardware since it is possible to change and adapt the modem very rapidly with any detected change in the operating conditions. The basic block diagram of this system is shown in Figure 4.9.

A good point to start our design considerations is with the spectral characteristics of our transceiver system. In the practical case this means that we need to know the power spectral density at the output of the receiver, assuming that the transmitter is being fed with white noise and the channel is clear. In general the frequency response can be approximated to a shape having very little output from 0Hz to some value and then rise to a central 'flat' portion of the band at, say, LB. Then

Figure 4.9 Block diagram of a modern HF communications terminal.

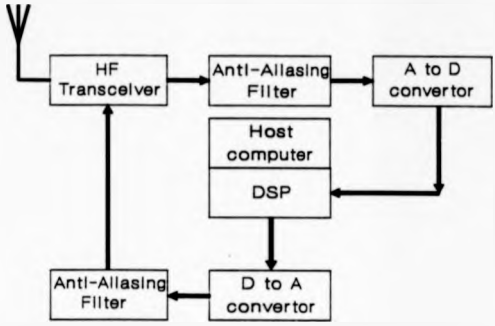
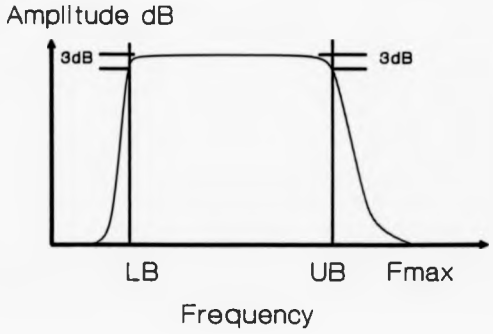


Figure 4.10 Idealised spectral characteristics of an HF communications link.



it will be 'flat' to some frequency, say UB. It will then die away again to the system noise floor at, say, Fmax. This idealised PSD is illustrated in Figure 4.10.

If we say that we are interested in putting 99% of the power of our signal in the flat part of our band, and we also require M tones, then we can say

$$B = UB - LB \qquad (4.12a)$$

and since

$$B = (M-1*G)/Te \qquad (4.1)$$

and

$$DF = 1.0/Te \qquad (4.12b)$$

where

DF is the frequency spacing for signal orthogonality

we find

$$DF = (UB-LB)/(M-1*G) \qquad (4.12c)$$

and the required tones are

$$Fn = LB + n*DF \qquad (4.12d)$$

where $0 \leq n < M$

Fn are the tone frequencies

This is fine in theory or in the analogue case but, since we are implementing a sampled system, we must expect that the sampling frequency, fs, will have some effect also. The value of fs should be set at the minimum value possible in order to have the minimum amount of processing to do (avoiding redundancy). Also, this gives the maximum amount of time between samples to do the processing operations in. These considerations imply setting fs = 2*Fmax and hence the sampling period, DT, to 1/(2*Fmax).

Since we have already calculated the symbol or element period, Te, above, we now see that, due to signal processing considerations, we wish Te to be an integer multiple of DT. Also we require that each symbol con-

tains an integer number of cycles of the chosen frequency. This avoids phase discontinuity, a condition that has been assumed in the spectral analysis above and which helps to keep spectral sidelobes low.

As a result of the above considerations, the value of Te that we attempt to use is set at the next multiple of DT up from that calculated. This results in a slight narrowing of the spectrum and so a small amount of the available bandwidth not being used. The actual amount will depend on the parameters being used and an example will be given below.

From the above considerations we now form an algorithm to calculate the frequencies, the number of samples per symbol and the number of cycles per symbol for each tone. This algorithm is shown in table 4.3. It has been used for designing the signal formats for all of the MFSK modems used in this thesis and, as an example, table 4.4 shows the calculations for the M=4, G=2 modem used in chapter 5. The required practical spectral characteristics of the back to back transmitter/receiver system are shown in Figure 4.11.

It can be seen in table 4.4 that the symbol period has gone up from 2.63ms to 2.72ms and the bandwidth has been reduced to 100 * (2573.53-735.29)/(2600-700) % = 97.75% of its original when we compare the values allowable by the original specification to the implementable solution. This is due to the implementation practicalities.

4.5.1 Modulator implementation I

The way in which the modulator has been implemented on a DSP is to first read data from a data source in a suitable binary or M-ary form corresponding to the symbol to be transmitted. This number is then used by the DSP as a reference to look up the address of a stored sine table for that tone. N waveform samples are then read sequentially out of memory and passed to a D-to-A convertor. The resulting analogue signal is lightly filtered before being fed to the audio input of the transmitter

Figure 4.11. Spectral characteristics of a practical HF communications link.

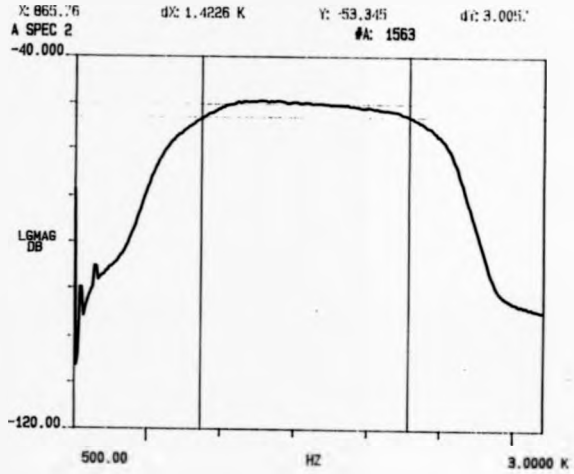


Table 4.3 An algorithm for tone arrangement in DSP MFSK modem

Given LB, the lower 3dB point of the passband

UB, the upper 3dB point of the passband

fs, the sampling rate

DT = 1/fs the sample period

N, the required number of tones

G, the frequency guard space

Perform the following calculations, using lb and ub as variables

1. B = UB - LB, the usable bandwidth.
2. DF = B/(N-1+G) the maximum tone spacing.
3. Ts = 1/DF the minimum symbol period.
4. Since the symbol period must be a multiple of DT and be ≥ Ts, the minimum number of samples is given by

$$N = \{Ts/DT\}$$

where {x} denotes the smallest integer greater than or equal to x.

5. The new symbol period is given by Ts = N*DT
6. The new frequency spacing is given by DF = 1/Ts
7. The new lower 99% power containment frequency is given by

$$lb = DF * \{LB/DF + G/2\} - DF*G/2$$

since the lowest tone frequency is an integer multiple of DF in order that there are an integer number of cycles of each tone in a symbol period.

8. The new upper 99% power containment frequency is given by

$$ub = lb + (N-1+G) * DF$$

and has changed due to steps 4 to 7.

9. Check that bandwidth constraints are not exceeded.

If ub > UB

Increase the symbol period by the minimum amount by increment-

ing N. Return to step 5.

10. The frequencies are now given by

$$F(n) = lb + (n+G/2)*DF \qquad 0 \le n < M$$

The number of samples per tone is N

The number of cycles per tone is $lb/DF + G/2 + n$

Table 4.4 Application of to the design of a M=4, G=2 MFSK modem

From Figure 4.11 we find

UB = 2600 Hz LB = 700 Hz Fmax = 3125 Hz

and we specify

M = 4 G = 2

Therefore $fs = 2*Fmax = 2*3125 = 6250$ Hz

and $DT = 1/6250\ s = 160\ \mu s$

Following the algorithm in table 4.3

1. $B = UB - LB = 2600 - 700 = 1900$ Hz
2. $DF = B/(M-1+G) = 1900/(4-1+2) = 380.0$ Hz
3. $Te = 1/DF = 1/380.0 = 2.63$ ms
4. $N = (Te/DT) = (2.63*10^{-3}/160*10^{-6}) = (16.45) = 17$
5. $Te = N*DT = 17*160\mu s = 2.72$ ms
6. $DF = 1/Te = 1/2.72ms = 367.65$ Hz
7. $lb = DF * (LB/DF) = 367.65 * (700/367.65) = 735.29$ Hz
8. $ub = lb + (M-1+G) * DF = 735.29 + (4-1+2) * 367.65 = 2573.53$ Hz
9. $ub > UB$? No so terminate procedure.
10. $F(n) = lb + (n+G/2) * DF \qquad 0 \le n < M$

$= 735.29 + (n+1) * 367.65$Hz

The number of samples per tone is N = 17

The number of cycles of tone n is $lb/DF + G/2 + n = n+3$

---

or recorded on standard audio tape for later trials. An example of the ideal and un-filtered waveform are shown in Figure 4.12.

Two systems have been implemented and the power spectral densities measured whilst transmitting random data. The results are shown in Figure 4.13a-b. The systems used are M=8, G=4 (Fig 4.13a) and M=4, G=2 (Fig 4.13b) and the modulator is transmitting random data.

4.5.2 Modulator implementation II

The above method of modulation is very effective but does not utilise a DSP very efficiently. It has been chosen for some stages of the work however due to the nature of equipment available at the time and for the ease of modifying parameters.

An alternative, and more memory efficient, method of generating the modulator output waveform is to look at the mathematical function which we are trying to generate. The waveform samples are given by

$$\sin(2\pi s(s+Q)i/N) \qquad (4.13a)$$

where

$0 \le s \le M$ is the symbol to be transmitted

$0 \le i \le N$ is the sample number within the symbol

$Q = 1b/DF + G/2$ is the integer number of tone intervals between 0Hz and the lowest tone.

Since the sin function takes its argument mod(2π), this is equivalent to

$$\sin(((s+Q)i) \bmod (N) 2\pi/N) \qquad (4.13b)$$

where

y mod(x) means the remainder after dividing y by x.

This may be implemented by use of a sine wave look-up table, is a

Figure 4.12
Ideal and unfiltered output of modulator
Tones 1 and 4.

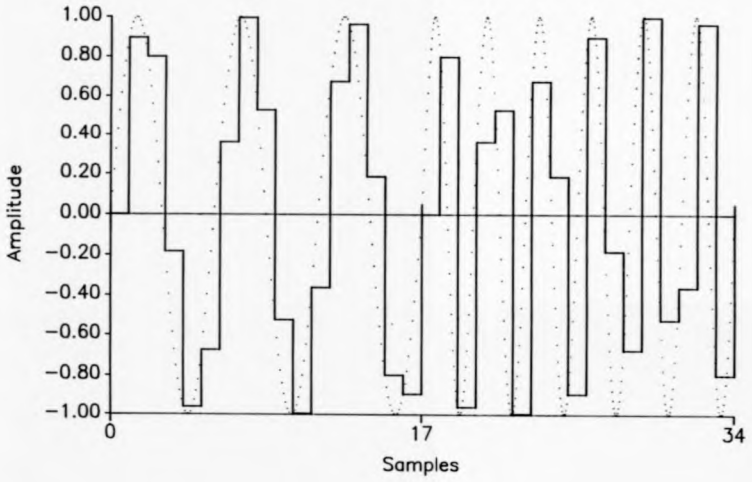


Figure 4.13. Spectral characteristics of MFSK signals. Random data.

Figure 4.13a. M=8, G=4.

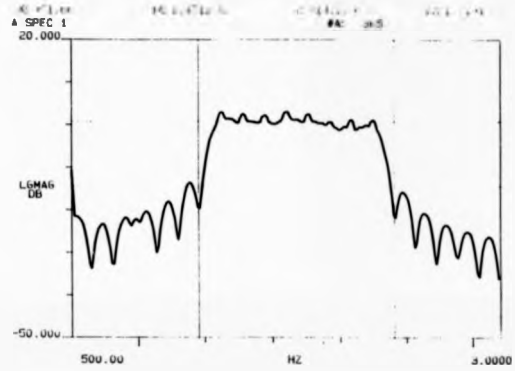


Figure 4.13b. M=4, G=2.

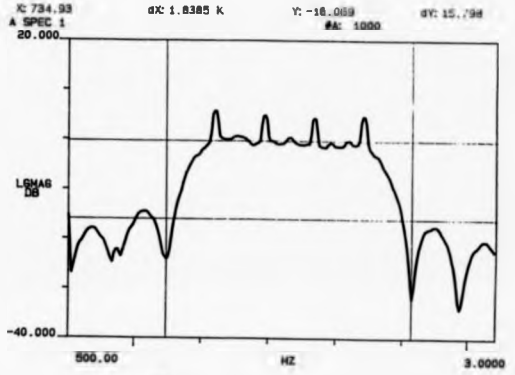


table of N samples of a sinusoid, from sin(0) to sin(2*π*(N-1)/N). The element of the table that is taken is

$$((s*Q)*i) \bmod (N) \qquad (4.13c)$$

The values given by (4.13c) for the example given above are shown in table 4.5. It can be seen that this method of finding the samples to be output is more efficient in terms of memory requirements since only N sine elements need to be stored.

4.6 Demodulator Design Considerations.

In order to find which of the many methods of demodulation are particularly suited to DSP implementation, it is informative to consider a vector representation of an MFSK signal. When dealing with an MFSK demodulator, we are concerned with deciding which of M orthogonal signals was generated at the modulator, and when. As we have seen, the way in which the signals are realised in practice is to use M tones that are or-

Table 4.5. Elements to be found in sinewave lookup table.

	Symbol			
Sample number	0	1	2	3
0	0	0	0	0
1	3	4	5	6
2	6	8	10	12
3	9	12	15	1
4	12	16	3	7
5	15	3	8	13
6	1	7	13	2
7	4	11	1	8
8	7	15	6	14
9	10	2	11	3
10	13	6	16	9
11	16	10	4	15
12	2	14	9	4
13	5	1	14	10
14	8	5	2	16
15	11	9	7	5
16	14	13	12	11

thogonal over a symbol period Ts is spaced in frequency at some integer multiple of 1/Ts [Wozencraft & Jacobs 1965 pp223-227]. Since each frequency is orthogonal to each other and each frequency has an in-phase and quadrature component, we are actually dealing with a vector in 2.N dimensions.

In practice, we have to regenerate all the possible transmitted signals at the receiver in order to compare them with the received one. This poses problems with stability of both the source reference oscillators, receiver local oscillators and, particularly in the case of the HF channel, the transmission medium itself. The result is that it is normally possible recreate the frequencies sufficiently well but not the phase. Therefore the phase of the tones is presumed to carry no information and the magnitude of the signal at each frequency only is considered. This has been implicitly assumed in the generation of the signals at the modulator since only the 'sine' component of each frequency is modulated by the data. The demodulator will be noncoherent [Prokis 1985 pp141] and our received vector need only be considered in the positive sector of N dimensional space.

It is well known [Wozencraft & Jacobs 1965 pp239] that using a bank of filters matched to the set of possible transmitted signals is the optimum method of signal recovery. The outputs of the filters are sampled at the end of a symbol period and the largest selected as being the best estimate of the transmitted signal. This method has been used extensively in the past for orthogonal signals as it is theoretically optimum [Wozencraft & Jacobs 1965 pp240] and is reasonably easily implemented. An exactly equivalent method, in terms of performance, is to cross correlate the received signal with all of the possible transmitted signals and then compare these values at the end of a symbol in a similar way [Wozencraft & Jacobs 1965 pp234]. This method has not been so popular in the past due, it is suspected, to the necessity for a multiplying element [Wozencraft & Jacobs 1965 pp235], not readily available in analogue cir-

cuitry. In the DSP environment however, this is available to us and it will be seen in section 4.7 that a very efficient algorithm can be built up around this method.

A further problem in the design of a demodulator, and probably the biggest one, is synchronisation of the demodulator symbol timing clock to the symbol transitions. Imperfect synchronisation is one of the major causes of error in digital communication systems and this is particularly true of HF channels due to the fading nature of the multipath propagation. This can be seen quite easily if we consider receiving a signal from two paths with a differential path delay of, say, 1ms. We wish to be synchronised on the strongest of them but their relative strengths will be changing, resulting in a 1ms change in the timing as they fade up and down. It can be seen from the example in table 4.4 that this represents a large portion of a symbol interval in a typical system.

Traditionally, most radio communications synchronisation requirements have been met by a process of 'initial' synchronisation, via say a preamble code, followed by some sort of 'flywheel' synchronisation procedure [Darnell & Honary 1988]. This normally involves either the use of timing information inserted into the data stream, timing information being superimposed onto the signal or simply relying on the accuracy of the clocks at either end [Ralphs 1985 chapter 8]. The technique for synchronisation developed here has the advantages of not requiring these 'overheads' but functions entirely on the unmodified, information-bearing signal. This type of synchronisation is termed 'modulation derived synchronisation' (MDS) [Shaw, Honary & Darnell 1989, Shaw & Honary 1989] and is particularly suited to DSP implementations of demodulators. If reception is interrupted for any reason, the procedure will recover synchronisation automatically after the discontinuity. This type of 'intrinsic' synchronisation has been attempted before for MFSK signals, but has not been successful [Ralphs 1985 pp122].

Going back to thinking of MFSK as a vector moving in the positive

sector of N dimensional space, we see that at a symbol transition the vector is lying exactly on one of the axes. Away from the transition the vector has components in more than one of the directions, although its overall length will remain fairly constant. It is this characteristic which we can use to derive the synchronisation information. If we take the ratio of the largest component of the vector to its magnitude then this function has a strong component at the symbol period and maxima at the symbol transitions.

4.7 Implementation of Correlation Demodulator and Synchronisation system.

We now consider how a demodulator can be efficiently implemented on a DSP. We shall assume the signal format is the same as that designed in section 4.5. At sample time n we correlate the signal over the last N samples with both the in-phase and quadrature versions of the possible transmitted signals. We then square the two components and add them to give a value proportional to the square of the amount of energy at each frequency (in each dimension). This gives the N dimensional vector which we denote $\underline{I^2(n)}$, with elements

$$I^2(n,m) = \left[\left[\sum_{i=n-N}^{n} f(i).\sin(w_m.i.DT)\right]^2 + \left[\sum_{i=n-N}^{n} f(i).\cos(w_m.i.DT)\right]^2\right] \quad (4.14a)$$

where $0 \le m < N$ is the tone number

n is the sample time

f(i) are the time samples

DT is the sample period

w_m is the frequency in question.

These summations can be calculated efficiently on a DSP by realising

that the value of the $I^2(n,m)$ is independant of the absolute phase of the reference vector. Hence we can write

$$I^2(n,m) = \left[\left[\sum_{i=n-N}^{n} f(i).\sin(w_m.i.DT + \theta(n,m))\right]^2 + \left[\sum_{i=n-N}^{n} f(i).\cos(w_m.i.DT + \theta(n,m))\right]^2\right] \qquad (4.14b)$$

where the $\theta(n,m)$ are arbitrary. If we fix $\theta(n+1,m) = \theta(n,m) + w_m$, we realise that in the expression for $I^2(n+1,m)$ and $I^2(n,m)$ the two summations differ only in the addition of a new point at one end and the deletion of the oldest point at the other. This is the method of calculation employed and, coupled with the fact that only two new terms need be calculated for each sample point, it leads to an efficient computational method, independant of N. The units of the $I^2(n,m)$ elements are energy^2, that is they are proportional to the square of the symbol energy.

The vector describes a stepped course through the signal space as n varies, the course being determined by the data being transmitted, along with any coding present. In practice it will also be modified by the transmitter and the channel due to the finite bandwidth, noise, Doppler shift, pure time delay and so on. The amount by which it is disturbed will determine the symbol error rate in the demodulation process.

For each value of n then, we have the vector $\underline{I^2(n)}$ from which we wish to recover the original signal. First we must recover the clock and in order to do this we find the ratio of the largest vector element to the sum of all the vector elements thus

$$D^2(n) = \frac{I^2(n,r)}{\sum_i I^2(n,i)} \qquad (4.15)$$

where $D^2(n)$ is the confidance level at time slot n.

$I^2(n,r)$ is the largest of the $I^2(n,m)$

$\sum_i I^2(n,i)$ gives the squared magnitude of the vector.

and the i variable goes from 0 to M-1.

Now, if n falls on a symbol transition and there is no noise and symbol p was transmitted we have

$$I^2(n,p) = Es^2 \qquad 0 <= p < M \qquad (4.16)$$

$$I^2(n,q) = 0 \qquad q <> p$$

and $D^2(n) = 1$

where Es is proportional to the symbol energy.

If, on the other hand, n is not coincident with a symbol transition then the vector will represent some other point in M dimensional space and $D^2(n) \leq 1$. The $D^2()$ series will therefore have a strong component at the clock rate and can be used for clock recovery.

In order to recover the phase of the data clock from the $D^2()$ series, we must assume that we know, within certain limits, the symbol period. The $D^2()$ series can then be filtered by a bandpass filter and the phase recovered. The form of the filter required will be considered further in section 4.7.1 below but it is worthwhile noting here that when a series of the same symbol occurs, the value of $D^2(n)$ will remain constant. The filtering applied to the series must therefore have sufficient 'flywheel' effect so as not to allow these events to inhibit the maintenance of correct synchronisation.

Formulae (4.14a) to (4.16) have been discussed with respect to the perfect, noise-free communication channel. For practical channels with noise, the $D^2(n)$ values will still contain the synchronisation information but will also be determined by channel and noise characteristics. It

can be seen that, at a symbol transition, the value of $D^s(n)$ is in fact a measure of the ratio of signal-plus-noise in the selected sub-channel to total noise. As such it can be used as a 'detection confidance', or source of RTCK data and has been arrived at at little or no extra effort. Such RTCK data can be used to improve the performance of both demodulation and decoding procedures and this has been investigated further in section 5.1. It could also be used to assist in modifying the transmission format if required, for instance by optimising of the number of MFSK tones, M, in response to channel state [Shaw, Honary & Darnell 1988].

Figure 4.14 shows the $D^s()$ series and $I^s(n,m)$ values for a sequence of symbols, using a 4 tone system.

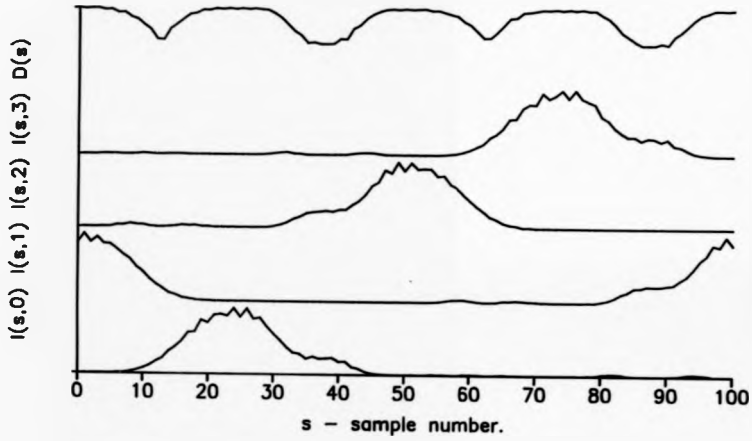
4.7.1 Filter design for synchronisation recovery from the $D^s()$ series.

We shall now consider what sort of filtering should be applied to the $D^s()$ series in order to most accurately track the clock phase.

Since we are considering the implementation of this system on a DSP, we are confronted with the problems of both what the filters' response should be and how to implement it. As far as implementation is concerned, we have the basic choices of finite impulse response (FIR) and infinite impulse response (IIR) filters [Ahmed & Natarajan 1983]. We are not restricted to simple linear operations for the filter, although it may well be easier to use a well known filter type and introduce any non-linear functions afterwards.

The choice between IIR and FIR filters seems to be quite straight forward if we consider a signal being received over an HF link. The signal may be good quality for some period of time and then fade out or be swamped by interference. It may then return to being good quality. In this instance, if we had synchronised on the original signal and been able to hold the timing over the period of bad SNR, we should be able to resume demodulation again immediately. If we try and do this using an FIR

Figure 4.14. Matched filter outputs and synchronisation signal from 4 tone MFSK demodulator.



filter, the output could have died away completely during the fade and we would have no timing information. If, however, we had been using an IIR filter, there may still be enough energy in the filter for us to regain synchronisation again immediately, or at least provide some initial guidance since an IIR filters response carries on for a long time, as its name implies. It is difficult to envisage a scenario where we would be better off using an FIR filter than an IIR one and so this decision has been made.

The question of the amplitude and phase response of the filter is a more difficult one to make. We may start by listing some of the more obvious characteristics which we wish to have.

i) The center frequency should be at the reciprocal of the symbol period.

ii) The bandwidth characteristic should be wide enough to cope with any effects which may be interpreted as time offsets in the system. This includes offsets in the transmit and receive oscillators and sampling clocks, as well as time-distorting effects of the propagation medium such as that brought about by the mechanism producing Doppler shift (section 1.4.4). It should only be a few percent of the centre frequency which implies a high Q filter.

iii) The phase response should be well defined and known over at least the frequency region considered in (ii) above. This is because the information that we are trying to recover is in the phase of the signal and when we detect a peak (or zerocrossing) in the output of the filter, we wish this to be at a fixed time offset from the symbol transition that we are looking for. If the phase is not linear then this time will be frequency dependant.

iv) In the case of the multipath signals which are likely to be encountered on the HF channel, and in the absence of any equalisation, we wish to synchronise on the strongest of the multipath components. The problem here is that if one path gets stronger while the one being syn-

chronised on fades out, there may be a rapid change of synchronisation required. The filter needs to track this change and this would require a larger bandwidth than the above conditions would suggest, but only for a short period. If the strongest can always be found, we have a type of space diversity [Ralphs 1985 chapter 5] working in our favour. That is to say, we pick the signal which has travelled by the path with least attenuation.

v) The amplitude response in the passband of the filter need not be particularly flat, since it is the phase of the signal which we are interested in. The amplitude of the $D^2()$ series is independant of the signal amplitude, but may well be disturbed by non-Gaussian noise.

Since no theoretical relationship is known between the channel conditions and the PSD of the $D^2()$ series, the series has been derived by simulation on a mainframe computer and measured. This has been done for the CCIR [1974] recommended good, moderate and poor channels. The modem used was a 4-tone MFSK modem with LB=700Hz, UB=2600Hz and G=2 (see table 4.5). The sampling rate was 6.25kHz and the resulting symbol rate was 367.65 symbols per second, with 17 samples per symbol. The $D^2()$ series for 4000 randomly selected symbols was output via an D-to-A convertor and the spectrum measured on a spectrum analyser.

It can be seen in Figures 4.15a-c that there is a spectral peak with its centre at the symbol repetition rate, as expected. There are also spectral peaks at half that frequency, probably due to the fact that the series remains at a constant level when there is a repetition of a symbol, an event that occurs roughly 25% of the time. This second peak can be seen to be between 4.5 and 6.5 dB down from the main peak, depending on the channel parameters chosen.

If Figure 4.16a-c, we show the section of the spectrum that we are particularly interested in. It can be seen from the cursors on the plots that the 6dB points are at a spacing of 51.6Hz for the good, 52.0Hz for the moderate and 53.6Hz for the poor channel conditions. This gives us

Figure 4.15. Spectral characteristics of synchronisation signal.
Figure 4.15a. CCIR good conditions.

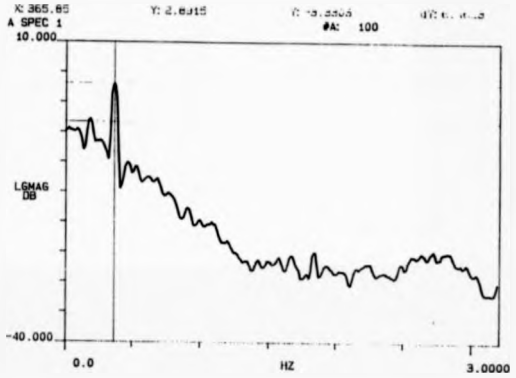


Figure 4.15b. CCIR moderate conditions.

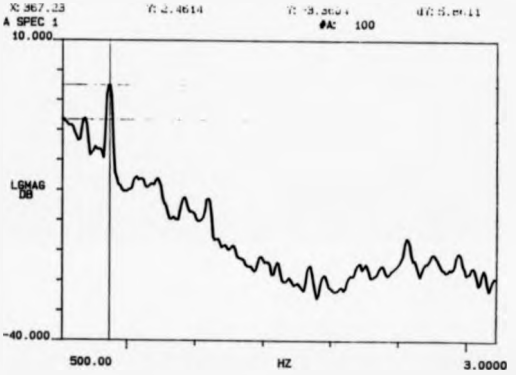


Figure 4.15c. CCIR poor conditions.

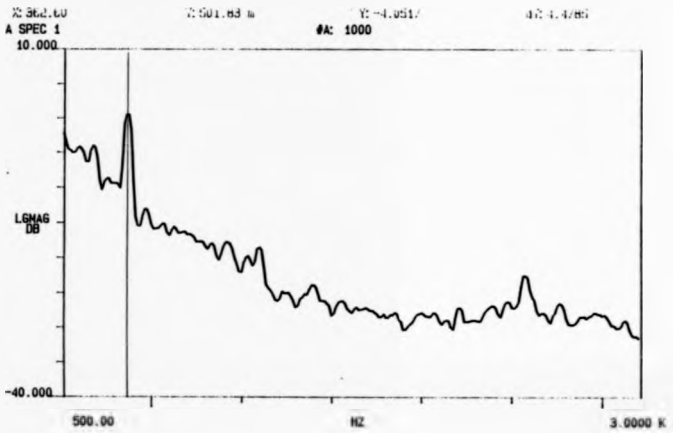


Figure 4.16. Spectral characteristics of synchronisation signals—expanded scales.

Figure 4.16a. CCIR good conditions.

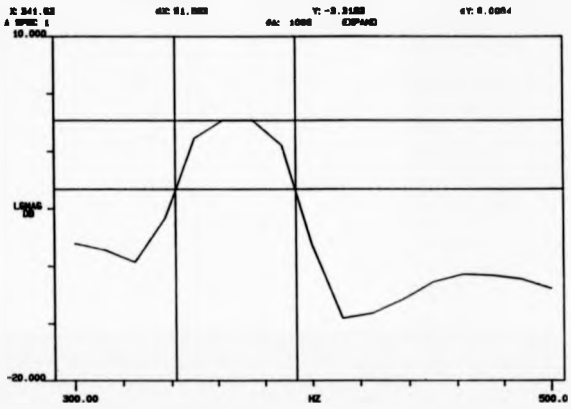


Figure 4.16b. CCIR moderate conditions.

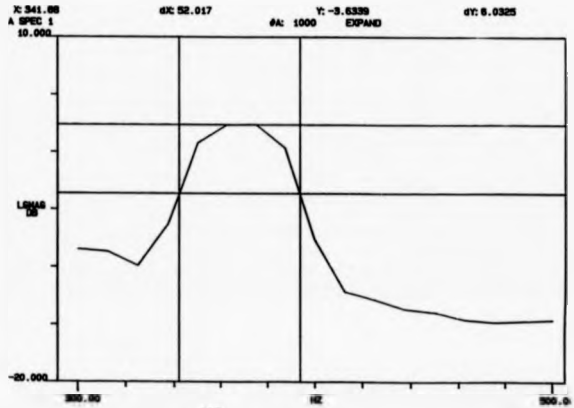
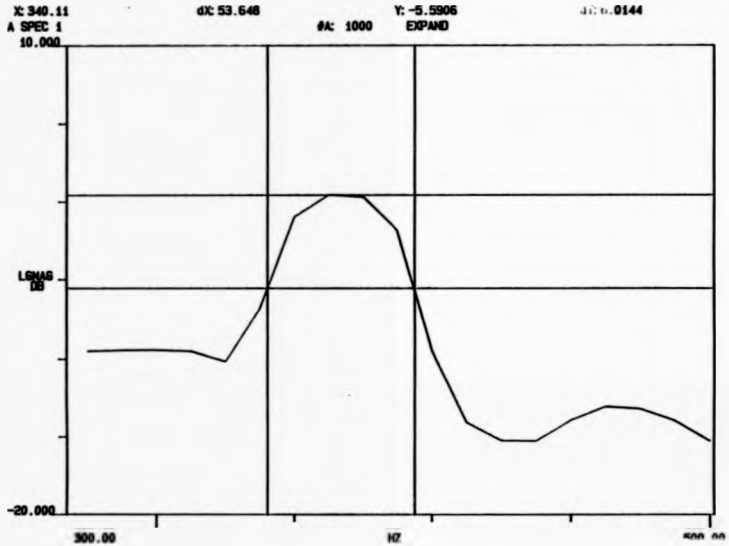


Figure 4.16c. CCIR poor conditions.



the cutoff frequencies of the required second-order filter that we need to filter the series and recover the clock phase. From consideration of these results, it was decided to implement a second-order Butterworth bandpass filter with a normalised center frequency of 1/17 and a normalised upper 3dB point at 1/18. These values equate to frequencies of 367.65Hz and 390.625Hz respectively and set the lower 3dB point at 346.02Hz.

4.7.2 Filter implementation.

There are several problems involved in implementing IIR digital filters on a DSP. The first of these is that of stability and arises particularly when we require feedback constants close to unity. In some situations the round-off errors in the iterative calculations may cause the effective feedback to become, or exceed one and hence cause instability. This is a particular problem with fixed-point processors such as the one which the authors are using (a Texas instruments TMS32010). The second problem is that of overflow and its associated non-linearities and again this is a particular problem on fixed-point processors.

In order to circumvent these problems and also check the design of the filters, the authors have used a spreadsheet package on a personal computer (PC) to calculate the filter constants from standard formulae and then to simulate the filter with its maximum possible input signal. This makes it very easy to check for overflow problems and also to check the impulse and step responses. The step response and maximum input response of the filter designed for the $D^{2}()$ series detailed in the last section can be seen in Figure 4.17a-b as an example of the usefulness of this technique.

Figure 4.17a. Step response of bandpass filter for $\dot{D}()$ series. Center frequency 1/17.DT Hz. 100 unit step.

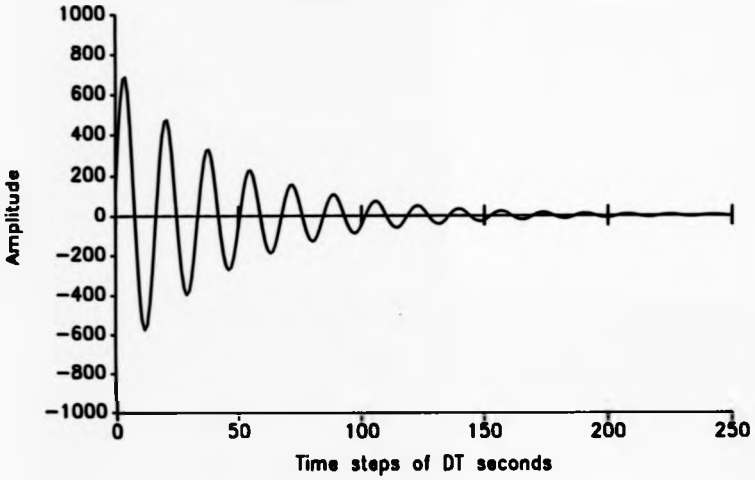
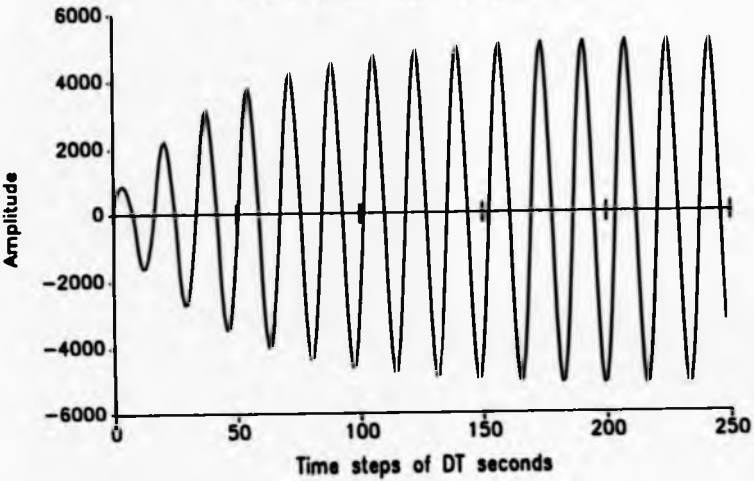


Figure 4.17b
Response of bandpass filter to maximum input. Center frequency 1/17.DT Hz



4.7.3 Synchronisation verification.

In order to verify how well the demodulation and synchronisation procedure detailed above works in practice, the algorithm has been implemented on both a DSP and a mainframe computer. The demodulated data and $D^*()$ series have been stored on disk for off-line analysis. The data has then been analysed for both symbol and synchronisation errors. A synchronisation 'slip' is defined as when an extra symbol is inserted into the data stream (a slip up) or a symbol is missing from the data stream (a slip down).

In order to detect errors, we must compare a copy of the transmitted data with the received one and count an error for every difference that we find. Therefore we must know when the copy is synchronised with the received sequence. The way in which this has been done is to arrange that the transmitted data is a short sequence, of length P, with good autocorrelation properties, repeated over and over again. There are now only P unique possible offsets of the transmitted symbol sequence where synchronisation can occur. In order to find the errors and synchronisation slips, we therefore implement P simple 1st order IIR filters, one for each of the offsets. The input to the p^{th} filter is taken as '1' or '0' depending on whether or not the received symbol corresponds to the equivalent symbol in the p^{th} offset original sequence. After several symbols have been processed in this manner, the filter with the largest output corresponds to the correct offset. Any variation of the received sequence from the one detected to be in synchronism with the original sequence, must be a symbol error. This enables us to detect both random errors and bursts of errors where the received error symbols may or may not be independent from the transmitted ones.

In the case of a synchronisation slip, the filter output corresponding to the previously correct offset will start to fall, whilst the filter output corresponding to the new correct offset will rise. The time

constants of the filters are arranged such that the two outputs will cross over after Q symbols. In the interval when one filter output takes over from another as having the largest amplitude, a synchronisation slip must have occurred Q^{th} previous interval. In order to not count Q errors in this instance, we must arrange the synchronisation finding filters to be acausal with respect to the error checking procedure, that is, we must 'look forward' Q symbols for the outputs of the filters with respect to where we are currently checking for errors.

The procedure is illustrated in table 4.6 where P=4 and Q=3. In this table, we show the received symbol sequence in the first column, along with its errors and synchronisation errors. In the next four columns, we show the four possible transmitted sequences which are the time-shifted variants of the original signal. The following four columns show the logical (true=1, false=0) results of the comparison of the possible transmitted sequences with the received one. It can be seen that for the first 10 symbols, the 4^{th} offset sequence is the correct one. These binary sequences are the inputs for the synchronisation tracking filters. The filter outputs can be seen in the following four columns, followed by the recovered synchronisation offset. This is derived by finding which filter has the largest output Q symbols further on. Errors are detected by comparing the received data with the offset transmitted data version indicated by the synchronisation tracking system. If they correspond then there is no error and this is indicated by a 0 in the final column.

This system works well for checking the performance of modems and, if required, can be made to produce tables relating which symbol was transmitted and which was received, in order to see if the error symbols are random or related to the transmitted symbol. This function has been used in the testing, described in section 5.4, of the HOX demodulator described in section 3.3.1. It has been found that the error symbols are close, in the frequency domain, to the transmitted ones.

Table 4.6 Synchronisation verification and error checking routine

Rx data	Notes	Possible original data	Result of compare	Filter outputs 1	2	3	4	Recovered sync data	errors
1		3 2 4 1	0 0 0 1	0.00	0.00	0.00	0.16	4	0
4		1 3 2 4	0 0 0 1	0.00	0.00	0.00	0.29	4	0
2		4 1 3 2	0 0 0 1	0.00	0.00	0.00	0.41	4	0
3		2 4 1 3	0 0 0 1	0.00	0.00	0.00	0.50	4	0
1		3 2 4 1	0 0 0 1	0.00	0.00	0.00	0.58	4	0
4		1 3 2 4	0 0 0 1	0.00	0.00	0.00	0.65	4	0
2		4 1 3 2	0 0 0 1	0.00	0.00	0.00	0.70	4	0
3		2 4 1 3	0 0 0 1	0.00	0.00	0.00	0.75	4	0
1		3 2 4 1	0 0 0 1	0.00	0.00	0.00	0.79	4	0
4		1 3 2 4	0 0 0 1	0.00	0.00	0.00	0.82	4	0
3	slip up	4 1 3 2	0 0 1 0	0.00	0.00	0.16	0.69	3	0
1		2 4 1 3	0 0 1 0	0.00	0.00	0.29	0.58	3	0
4		3 2 4 1	0 0 1 0	0.00	0.00	0.41	0.49	3	0
2		1 3 2 4	0 0 1 0	0.00	0.00	0.50	0.41	3	0
3		4 1 3 2	0 0 1 0	0.00	0.00	0.58	0.35	3	0
3	error	2 4 1 3	0 0 0 1	0.00	0.00	0.49	0.45	3	1
4		3 2 4 1	0 0 1 0	0.00	0.00	0.57	0.38	3	0
2		1 3 2 4	0 0 1 0	0.00	0.00	0.64	0.32	3	0
3		4 1 3 2	0 0 1 0	0.00	0.00	0.70	0.27	3	0
4	error	2 4 1 3	0 1 0 0	0.00	0.16	0.58	0.23	3	1
2	error	3 2 4 1	0 1 0 0	0.00	0.29	0.49	0.19	3	1
2		1 3 2 4	0 0 1 0	0.00	0.25	0.57	0.16	3	0
3		4 1 3 2	0 0 1 0	0.00	0.21	0.64	0.13	3	0
1		2 4 1 3	0 0 1 0	0.00	0.17	0.70	0.11	3	0
4		3 2 4 1	0 0 1 0	0.00	0.15	0.75	0.09	3	0
2		1 3 2 4	0 0 1 0	0.00	0.12	0.79	0.08	3	0
3		4 1 3 2	0 0 1 0	0.00	0.10	0.82	0.07	3	0
1		2 4 1 3	0 0 1 0	0.00	0.09	0.85	0.06	3	0
4		3 2 4 1	0 0 1 0	0.00	0.07	0.87	0.05	3	0
2		1 3 2 4	0 0 1 0	0.00	0.06	0.89	0.04	3	0
4	error	4 1 3 2	1 0 0 0	0.16	0.05	0.75	0.03	3	1
1		2 4 1 3	0 0 1 0	0.13	0.04	0.79	0.03	3	0
2	error	3 2 4 1	0 1 0 0	0.11	0.20	0.66	0.02	3	1
2		1 3 2 4	0 0 1 0	0.09	0.16	0.72	0.02	3	0
1	error	4 1 3 2	0 1 0 0	0.08	0.30	0.60	0.02	3	1
1		2 4 1 3	0 0 1 0	0.07	0.25	0.67	0.01	3	0
1	slip down	3 2 4 1	0 0 0 1	0.06	0.21	0.56	0.17	3	1
3	error	1 3 2 4	0 1 0 0	0.05	0.34	0.47	0.14	4	1
2		4 1 3 2	0 0 0 1	0.04	0.28	0.40	0.28	4	0
2	error	2 4 1 3	1 0 0 0	0.19	0.24	0.33	0.24	2	1
1	not error	3 2 4 1	0 0 0 1	0.16	0.20	0.28	0.36	3	1
1	error	1 3 2 4	1 0 0 0	0.30	0.17	0.24	0.30	4	1
1	error	4 1 3 2	0 1 0 0	0.25	0.30	0.20	0.25	1	1
1	error	2 4 1 3	0 0 1 0	0.21	0.25	0.33	0.21	2	1
1	not error	3 2 4 1	0 0 0 1	0.18	0.21	0.27	0.34	3	1
1	error	1 3 2 4	1 0 0 0	0.31	0.18	0.23	0.28	4	1
1	error	4 1 3 2	0 1 0 0	0.26	0.31	0.19	0.24	4	1
1	error	2 4 1 3	0 0 1 0	0.22	0.26	0.32	0.20	4	1
1		3 2 4 1	0 0 0 1	0.18	0.22	0.27	0.33	4	0
4		1 3 2 4	0 0 0 1	0.15	0.18	0.23	0.43	4	0
2		4 1 3 2	0 0 0 1	0.13	0.15	0.19	0.52	4	0
3		2 4 1 3	0 0 0 1	0.11	0.13	0.16	0.60	4	0
1		3 2 4 1	0 0 0 1	0.09	0.11	0.14	0.66	4	0
4		1 3 2 4	0 0 0 1	0.08	0.09	0.11	0.72	4	0

This page intentionally left blank.

Chapter 5. Practical Evaluation of Techniques.

5.1 An Example of an ARQ System Employing Zerocrossing RTCE

In this section, we describe the use of the RTCE system explained in chapter 2 in a simulated system. This system is an adaptive data transmission scheme employing error correcting Reed-Solomon (RS) codes combined with automatic repeat request (ARQ) [Sklar 1988] for retransmission and code rate adaption, based on the zerocrossing RTCE [Darnell, Honary, Shaw 1987]. There is a demonstrable improvement in performance and reliability over a simulated HF channel using the new technique, in comparison with conventional constant rate codes and ARQ systems.

Error detection incorporated with ARQ is widely used for error control in data communications systems [Schwartz 1980 pp546-550] since as a method of error control it is simple and provides high system reliability. In most ARQ error control systems, a high rate error detecting code, say an (n,k) linear block code, incorporated with a certain protocol, is used. At the receiver, the decoder computes the syndrome [Peterson 1961] from the received codeword. If the syndrome is zero, the received codeword is assumed to be error-free and is delivered to the user. If the syndrome of the received word is not zero, the presence of errors is detected, and the receiver requests a retransmission, via a feedback channel, of the same codeword. Retransmission continues until the codeword is successfully received. With this error control system, erroneous data is delivered to the user only if the receiver fails to detect the presence of errors. Thus the reliability of an ARQ system normally depends on the choice of the code used for error detection and is generally very high.

The main feature of ARQ techniques is that the throughput of the system is not constant and falls rapidly with increasing error rates [Honary & Farrell 1986]. This is due to the system adapting to the changing noise level of the channel. Normally the system will have a maximum

throughput determined by the rate of the error detection code being used. The advantage of the method described here is that the error detection algorithm is based upon the use of the zerocrossing RTCE technique. This enables the system to be matched to the channel capacity without the use of an error detecting code. Hence any code used is used only for forward error correction (FEC) [MacWilliams & Sloane 1978] and this opens up the possibility of using no coding when the channel error rate is very low. This results in an increased throughput and/or a lower system BER than has previously been possible.

Four coding levels have been used for the system

a) no coding;

b) Reed-Solomon (RS) (15,9) error correction code;

c) shortened by 3 RS error correction code (RS-3);

d) shortened by 6 RS error correction code (RS-6).

Where the shortened codes are obtained from the Reed-Solomon (15,9) [MacWilliams & Sloane 1978] code by setting the first 3 or 6 symbols to zero and not transmitting them [Honary 1981].

The system operates as follows. Initially the data is transmitted without coding. At the receiver, before demodulation, the RTCE provides an estimation of the BER. A decision is now made as to whether

i) the information block can be accepted. If so, an acknowledgement (ACK) is sent to the transmitter along with the next required code rate. This allows the code rate to be adapted to the current channel capacity. or, ii) the information block cannot be accepted. In this case, a re-transmission must be requested and the system adapted to another, more powerful, coding level. Thus a non-acknowledgment (NAK) is sent to the transmitter along with the next coding rate to be used.

The process is repeated for the subsequent blocks. Thus the code rate for a block is determined from the zerocrossing density of the pre-

vious time interval.

In the case under consideration, the zerocrossing counting technique is used to evaluate the channel. In order to do this the SNR of the signal is considered to be constant over the period of time taken to make the assessment and then the assessment is assumed to be valid for the following data block. This means that the assessment interval must be short with respect to the fading rate. This consideration must be set against the relatively long period ideally required for an accurate assessment of the channel and thus a compromise has to be made. It is not clear from theoretical considerations what the optimum period for channel evaluation is; it is presumably dependant on the actual channel conditions. Since the maximum block length of the code used here is 60 bits, this is the period that was chosen.

The data presented here was derived from a software simulation of the proposed system. In the simulation, the values of the parameters in the HF simulator which have been used are 1 skywave with an rms fading frequency of 0.1Hz. Various noise levels have been used to give the required range of average channel error rates. The bit rate is 200 bps.

Results have been obtained for sets of data with different overall error rates. These sets of data have been used with

a) RS (15,9) error correction code

b) RS-6 error correction code

c) RS (15,9) normal error detection ARQ system

d) RS-6 normal error detection ARQ system

e) RS-6 as a 1 symbol error correcting and 4 symbol error detecting ARQ system

f) the adaptive system described above.

The throughput and reliability [Lin & Costello 1983] results are given in Figures 5.1 and 5.2. The term 'throughput' is taken to mean (the

Figure 5.1 Throughput characteristics for various ARQ systems. For systems, see text.

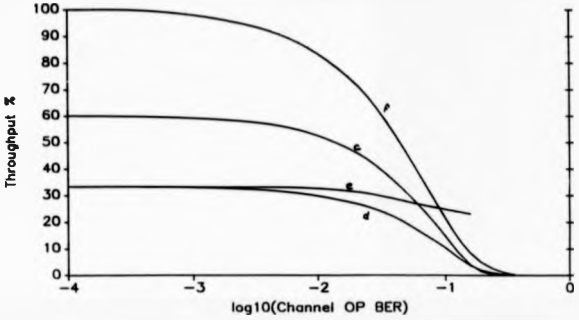
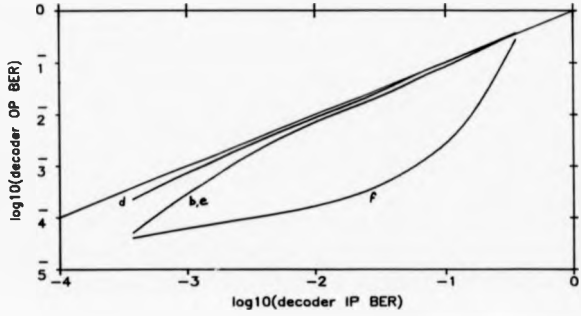


Figure 5.2 Decoder input to output BER reliability characteristics. For systems see text.



number of bits delivered to the user)/(the number of bits transmitted over the channel) and the term 'reliability' is taken to mean the user bit error rate compared to the channel bit error rate ie the effectiveness of the system in isolating the user from channel errors. For cases a) and b), the gain of the codes is not as great as would be expected for a random error distribution. This is because the codes have a fixed power and so can correct only a fixed number of errors in each block. The errors are, however, distributed in bursts of high density, separated by periods of very low density. Therefore the codes fail in the high error density periods and successfully correct errors only in the periods of low error density. This gives a smaller overall gain than the random error case with similar average BER.

In cases c) to e) where the codes are being used for error detection, the error rates obtained were very small as nearly all the errors are detected. However, the power of the code is not matched to the channel most of the time and so its efficiency is low. This is apparent from the throughput curves.

In the case of the new adaptive system, the transmission rate is being matched to the channel capacity by altering the code rate according to the RTCE. This results in the throughput being very much better than for all of the other systems considered due to the fact that the no-coding condition is being used during the frequent periods of low noise. The only conditions where some of the other systems have higher throughputs is in exceptionally high noise conditions where most of the systems would be unusable anyway due to output BER considerations. The output BER of the adaptive system is also very much lower than that of the other systems employing FEC. This is due to the fact that a much larger number of bits are being received with only a comparable number of errors being made. The error rate of the new system seems to have a residual lower limit, as can be seen in Figure 5.2. This is thought to be due to i) to the occasional very short bursts of noise, not detected by

the RTCE method due to its necessarily finite evaluation period, and ii) to the fact that the zerocrossing counting technique detects the probability of error and not symbols in error. This means that a detected low probability of error does not mean that any particular block will have no errors. These effects may be compensated for in a more complex system by combining the RTCE method with some form of very high rate (close to 1) error detection method to ensure that these residual errors are picked up.

5.1.1 Problems with the zerocrossing counting method of RTCE

In the case of the simulations described in section 5.1 above, the sampling rate of the system was set to 4kHz, giving a bandwidth of 2kHz. The signal was centred on 200Hz and the bit rate was 200bps. This results in the signals' PSD being centered on 200Hz. The assumption is then made that the relationship between zerocrossings and SNR will follow a similar relationship to that shown in Figure 2.2, ie the curve going asymptotically to 400 crossings per second at high SNRs due to the sinusoidal signal, and to 2000 crossings per second in the case of all noise, due to the reasons discussed in section 2.2.

In practice, we find that in order to use the available spectrum efficiently, we should arrange that the signals' PSD be approximately centrally distributed in the available 'flat' section of the spectrum. This is at 1kHz in this case and results in the signal giving 2000 zerocrossings per second at high SNRs, exactly the same as that for the noise. This means that no information is available from simply counting the zerocrossings in a fixed period of time.

The problem just described does not present itself in the same way for the case of MFSK signals, due to the fact that at different times there are different frequencies being transmitted. The centre frequency of the PSD is still arranged in the same manner of course and so the

problem still exists with the long-term average number of zerocrossings per second. However, the number of zerocrossings per symbol changes according to the frequency present in the symbol being transmitted. This can be seen in Figures 3.1a-h and means that the technique can still be used for these types of signals. This is described in section 3.3 with results being given in section 5.5 below.

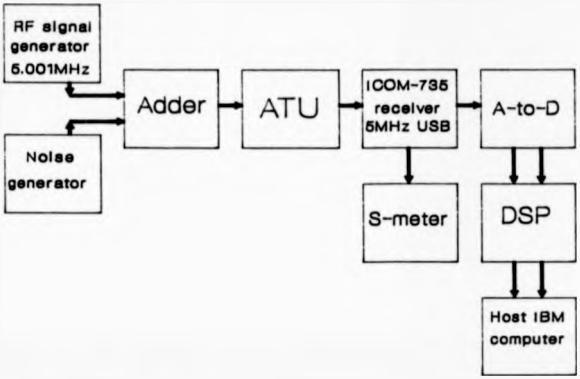
5.2 Radio Test for Zerocrossing Verification.

In order to evaluate the HF channel by the zerocrossing method, we need to know the zerocrossing characteristics of the receiver that we wish to use. In the theoretical analysis of chapter 2, we have considered the PSD of our baseband noise to be flat and the sampling to be at the Nyquist rate. These assumptions will only be approximately true in the practical case due to the fact that the PSD at the audio output of the receiver will be shaped by the receiver filters and will not, in general, be flat. Also, as the bandwidth will extend beyond the nominal 3kHz, we need to sample at a higher rate than may be thought necessary by the Nyquist rate for the channel. This situation is also described in section 4.5.

In order to measure the zerocrossing characteristics of a typical arrangement, a system was set up to sample the output of a typical receiver at 10kHz, digitize it, read the samples into a DSP and then count the zerocrossings. The DSP was a Texas instruments TMS32010 on a SKY-320 board, interfaced to the receiver via A-to-D and D-to-A convertors of local construction. This board is plugged into a bus slot on an IBM PC. The receiver used is an ICOM-735.

Signal and noise were fed into the receiver aerial tuning unit (ATU) via a resistive adder. The noise level was fixed at a convenient level of 7 on the receiver 'S' meter and then removed. The signal was then applied and adjusted to give the same reading. The noise was then re-connected to

Figure 5.3 Receiver test setup



give a reference level. This level has been taken as 0dB in the following discussion although it must be noted that the automatic gain control (AGC), which provides the signal for the 'S' meter, senses the average signal level rather than rms or power level. This will lead to an offset of around +1dB in the measurements due to the different mean to rms ratios of the sinusoid and the noise. The RF signal generators' level was switched in 1dB steps to give the different relative SNR levels required. The system setup is shown in Figure 5.3.

Results were obtained of zerocrossings in a fixed block length. The block size was chosen to be the same as that used in the simulated system of section 5.1 above, that is the equivalent of 60 bits at 200bps. This means that each sample block is 0.3s long and consists of 3000 samples. Samples were read into the DSP board and zerocrossings counted in one block whilst the next block was being read. The data was stored and when the information from sufficient blocks had been accumulated, the arithme-

tic mean and standard deviation of the zerocrossing counts was calculated. In order to determine how many blocks of data were required to give a reliable estimate of the mean number of zerocrossings in a block at the particular SNR, data was first taken over 100 blocks (30s) and then the standard deviations obtained were used to determine how many blocks would be required to give accurate averages at the various SNRs. This is calculated by use of the central limit theorem [Hodge & Seed 1972] and the formula

$$d = x.\sigma/\sqrt{n} \qquad (5.1)$$

where d = required tolerance

σ = standard deviation of samples

n = number of samples required

x = number of standard deviations on the normal distribution needed for the required accuracy.

Since the number of zerocrossings is integer, we set $d=0.5$ and for 98% certainty of being correct, we set $x=2$. Hence we obtain

$$n = 4.\sigma^2/0.25$$
$$= 16.\sigma^2 \qquad (5.2)$$

The following values of n were taken for ranges of SNR

SNR	sd, σ	n
-∞ to -4	12 to 18	5000
-3 to 0	7 to 12	2500
1 to ∞	0 to 6	800

The results are shown in Figure 5.4 and table 5.1. They can be seen to be similar in form to theoretical results obtained earlier and shown in Figure 2.2. There were a few problems in obtaining these results due to the long-term stability of the amplitude of the noise generator not being what it should. Also the frequency stability of the receiver and

Figure 5.4 Results of reciever test.

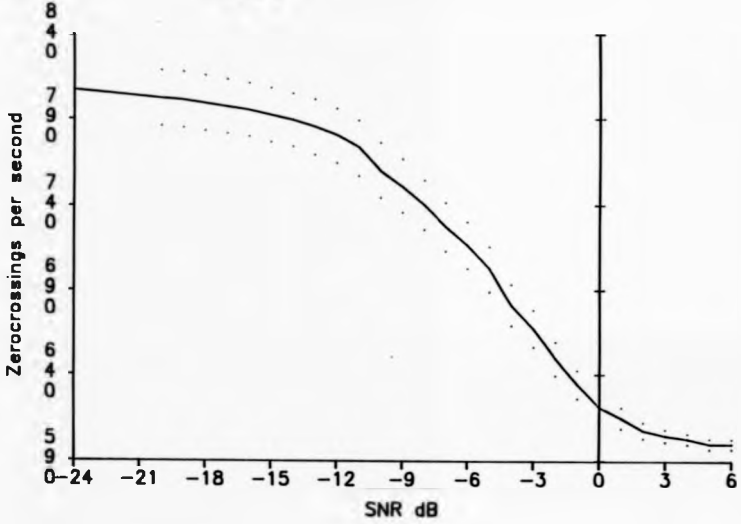


Table 5.1 Receiver zerocrossing test data.

SNR dB	zcs	sd	zcs+sd	zcs-sd	number of blocks	16*sd²
∞	807	16.25	823.25	790.75	5 000	4 223
-20	802	16.32	818.32	785.68	5 000	4 261
-19	801	16.43	817.43	784.57	5 000	4 319
-18	799	16.29	815.29	782.71	5 000	4 246
-17	797	16.04	813.04	780.96	5 000	4 117
-16	795	15.81	810.81	779.19	5 000	3 999
-15	792	16.02	808.02	775.98	5 000	4 107
-14	789	15.70	804.70	773.30	5 000	3 943
-13	785	16.33	801.33	768.67	5 000	4 266
-12	780	15.81	795.81	764.19	5 000	4 000
-11	773	15.75	788.75	757.25	5 000	3 970
-10	760	15.50	775.50	744.50	5 000	3 846
-9	751	15.44	766.44	735.56	5 000	3 814
-8	740	14.75	754.75	725.25	5 000	3 483
-7	727	14.38	741.38	712.62	5 000	3 309
-6	716	13.77	729.77	702.23	5 000	3 035
-5	702	13.07	715.07	688.93	5 000	2 734
-4	681	12.03	693.03	668.97	5 000	2 315
-3	667	11.08	678.08	655.92	2 500	1 964
-2	649	9.86	658.86	639.15	2 500	1 554
-1	634	8.19	642.19	625.81	2 500	1 073
0	621	7.14	628.14	613.86	2 500	815
1	615	5.82	620.82	609.18	800	542
2	608	4.59	612.59	603.41	800	337
3	605	3.66	608.66	601.34	800	214
4	603	3.19	606.19	599.81	800	163
5	600	3.15	603.15	596.85	800	158
6	600	2.97	602.97	597.04	800	141

signal generator cannot always be relied on, except after very long periods at a constant temperature. The signal generator was used to generate a 1MHz signal which was then demodulated using USB to give a 1kHz tone. This tone was found to change over a period of 25 minutes, the test period at lower SNRs. Therefore, both the amplitude and frequency had to be re-calibrated at the start of each measurement.

5.3 Practical Test of Channel Analysis System.

In order to test the channal analysis system, analogue signals are derived by using the receiver on single sideband (SSB) mode, tuned 1kHz off a known carrier to give a 1kHz beat frequency. A system was set up to

read the number of zerocrossings in a block of 3000 samples. A look-up table, derived for zerocrossings versus SNR as detailed in section 5.2 above, was then used to determine the SNR. This was output as a voltage to a chart recorder. The voltage drive to the 'S'-meter of the receiver was also recorded as an indication total received signal power.

If the noise level is constant and there is no interference, the 'S'-meter reading recorded is roughly proportional to the signal strength. These are the conditions that were present on one of our allocated frequencies (5.75MHz) on 14-7-87. The chart recording is shown in Figure 5.5. It can be seen that from 13:48 to 13:56 of the test, the signal strength was very high and for most of the time the analyser was giving its maximum output of 15 dB. At some points however, the signal strength drops below S2 or S3 and this is detected as a more noisy signal by the RTCK system. After 14:00, the signal fades out completely and this can be seen on both traces.

A further test was done using 'Radio 1' as a source. This is shown in Figure 5.6 and shows that for this "noise" data, the zerocrossing figure is very variable, as expected.

Several hours of similar tests have also been done using the receiver on LSB at 5.001MHz to give a 1kHz beat from the 5MHz standard frequency transmissions. These signals are transmitted from, it is believed, Rugby, England and Neuchatel, Switzerland [Giles & Jessop 1977 pp135-137], alternating at 5 minute intervals and so produce a range of different fading and non-fading conditions. The system was set up to give a voltage output proportional to the number of zerocrossings in each block.

These tests produced amplitude traces of typical Rayleigh fading signals, along with a trace proportional to the number of zerocrossings detected. Once again, the SNR could be tracked by observation of the zerocrossing count. The results are shown in Figure 5.7a and b. The following effects can be seen:-

Figure 5.5 Chart recording at 5.75MHz. S-meter and zerocrossing RTCE.

Block length 0.3s. 14-7-87, 13:48 - 14:02 BST.

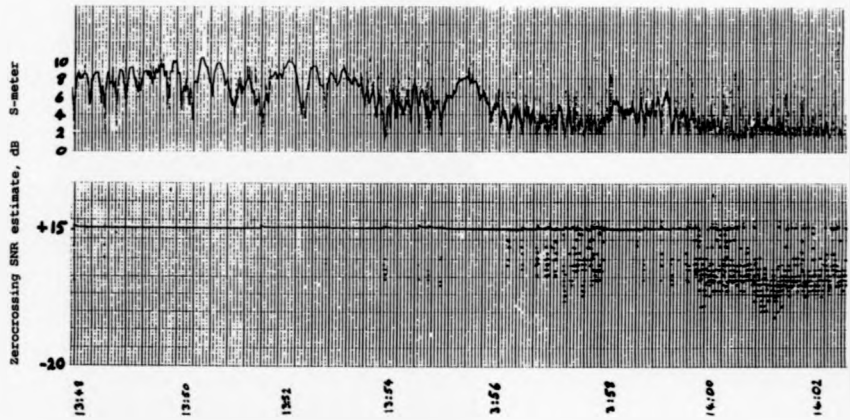


Figure 5.6 Chart recording at 1089kHz. S-meter and zerocrossing counts.

Block length 0.3s.

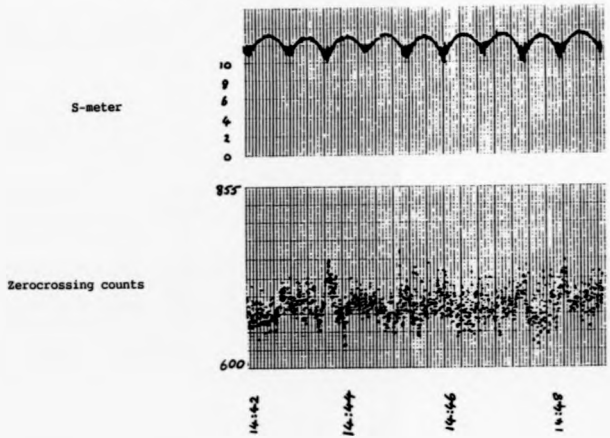


Figure 5.7a Chart recording at 5MHz. S-meter and zerocrossing counts.

Block length 0.3s. July 1987, 11:45 - 11:57 BST.

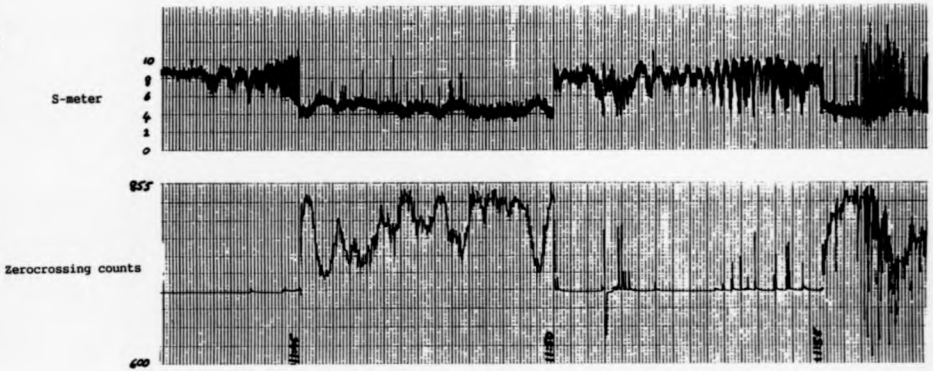
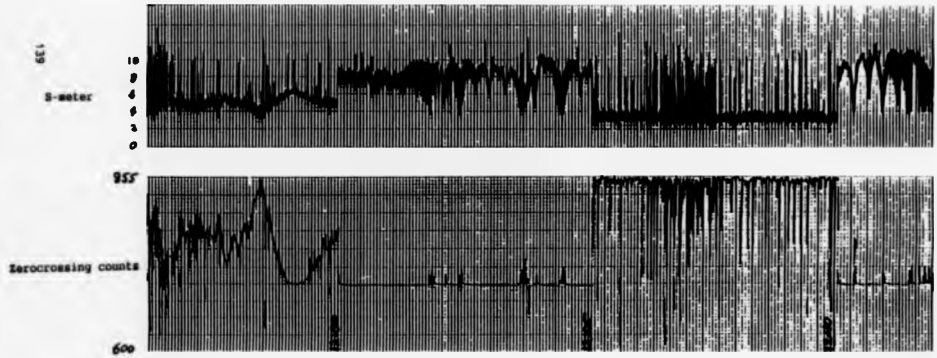


Figure 5.7b Chart recording at 5MHz. S-meter and zerocrossing counts.

Block length 0.3s. July 1987, 11:57 - 12:11 BST.



Time	Effect
11:44-11:45	Fast fading on S-meter signal
11:45-11:50	Fading between S4-6, detected in detail by the zero-crossing count
11:51, 11:53-11:50	Deep fades detected by the zerocrossing count
11:56-11:57	Bursts of high SNR detected by the zerocrossing count
11:59-12:00	Slow fade detected by zerocrossings

These results clearly show that the zerocrossing detector can analyse the channel very well between signal strengths of S2-4 but for signals above this level an almost constant level of zerocrossings is found. In an attempt to counteract this problem of limited dynamic range, the length of the blocks used to count the zerocrossings over was increased from 3000 samples (0.3s) to 10,000 samples (1s). The results from this are shown in Figure 5.8. It can be seen that they are similar to the results in Figure 5.7 except that the range of S-meter readings over which the analyser works has changed to some extent. Resolution has not been improved very much due to the fact that we now need to scale the output by a factor of 4 to keep within the range of our D-to-A convertor.

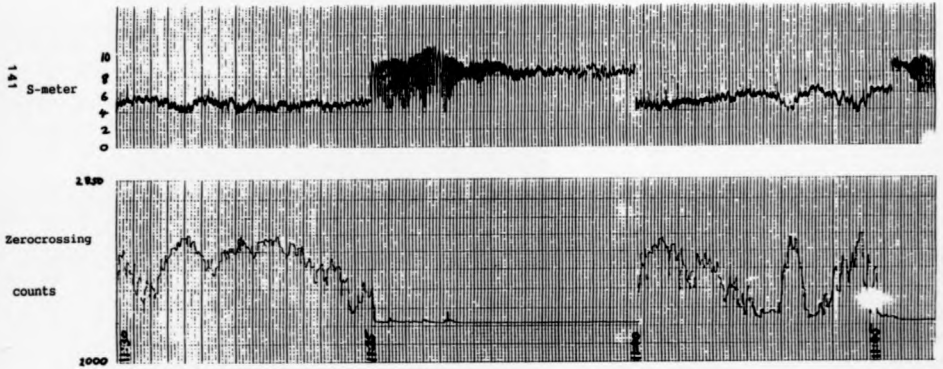
In fast fades, the system cannot respond fast enough to keep up with the signal strength. It is clear that the block size should be decreased to, perhaps, 100-1000 samples and then some sort of filtering applied before passing the output out to the recorder. The problem with this is that the filter would possibly need to adapt to the rate of the fading.

5.4 MFSK demodulation using HOX

Using the procedure detailed in section 3.3.1, some tests have been done on MFSK demodulation using HOX [Shaw, Honary & Darnell 1989]. These tests have been done on a mainframe computer, using the HF channel simulator described in section 2.4 and Appendix I. Results have been

Figure 5.8 Chart recording at 5MHz. S-meter and zerocrossing counts.

Block length 1s. July 1987, 11:30 - 11:45 BST.

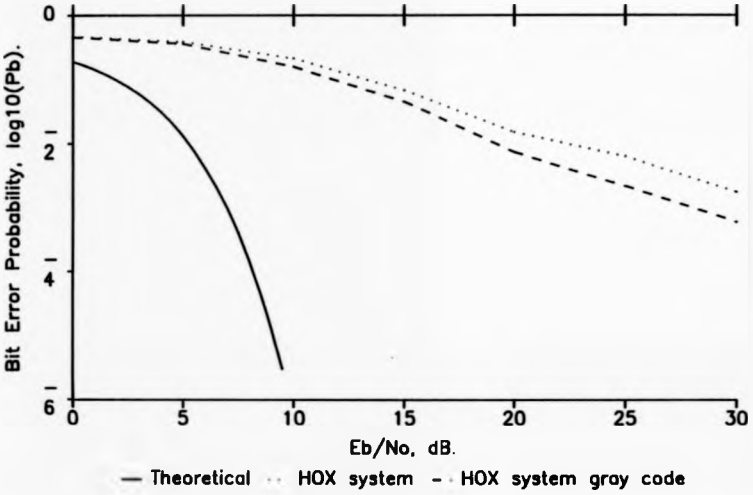


plotted in terms of the bit error probability, P_b, versus E_b/N_o and these are shown in Figure 5.9. The theoretical result for a noncoherent matched filter demodulator working on the same signal have been shown for comparison.

Since the system described has the characteristics that when an error is made, it tends to demodulate an adjacent symbol, the P_b values shown are somewhat pessimistic. By this we mean that the error rate improvement that could be achieved with a suitable coding system will be higher, for a given P_b, than that of a system where the errors are not correlated with the symbol transmitted. The degree of gain may be calculated by investigation of the mutual information in the signal. A simple scheme that has been applied to reduce P_b has been simply to use Gray coding [Stremler 1982 pp555] in the binary to octal mapping. This reduces P_b somewhat for a given symbol error rate since only a single bit error is made when an adjacent symbol is demodulated in error, instead of an average of 1.5 bits in error. The results are also shown in Figure 5.9 for comparison.

It can be seen that this system is many dB worse than the theoretical performance for noncoherent matched filter detection; the reasons for this are discussed in section 3.3.1. There are two basic reasons. The first is that there are very few samples of the waveform in each symbol, in the order of 25 (see the example in section 4.5 for more details of this). This results in the standard deviation of the zerocrossing count being large compared to the number of crossings being counted. Subsequently the amount of information that can be gained from the counts is low. Secondly, the difference in the number of zerocrossings in adjacent symbols of an orthogonal MFSK system is only 2 due to the fact that there need only be 1 cycle difference in their waveforms [Ralphs 1985 pp20]. This means that any deviation from the theoretical value can cause an error. The system could gain, however, in some implementations since it only requires simply implemented functions for its operation.

Figure 5.9. 8-tone HOX MFSK demodulator-
error rates.



5.5 RTCE using HOX

An example of the system described in section 3.3.2 has been implemented in software. It employed an 8-tone MFSK demodulator of the form described in section 4.7 to perform the demodulation and synchronisation functions necessary. The results from two, 40 second sections of the simulation are shown in Figures 5.10a-b. In these Figures, the data has been taken in blocks of 30 symbols and the number of errors counted. We have also averaged the estimated SNR, found by the technique described in section 3.3.2, over the 30 symbols and displayed the results in dB. Two different channels have been simulated. The first is a single fading skywave with a fade rate of 0.1Hz and 0dB mean SNR. The results from this are shown in Figure 5.10a. Figure 5.10b shows a similar set of results but in this case there is a second independant skywave with the same characteristics as the first. There is a differential time delay of 1ms.

5.6 Channel evaluation using $D^2()$ series data

5.6.1 Results from simulated systems

As noted in section 4.7, the values of $D^2(n)$ derived in the MFSK demodulator synchronisation procedure are, at the symbol transition point, a straight forward function of Es/No. This fact has been exploited as a means of RTCE by taking the average value of the $D^2(n)$ values at the detected symbol transition point, averaged over several symbols in order to reduce the variance of the result. This gives a reliable estimate of the SNR in the AWGN channel and also on slowly fading channels. In this instance, by slow we mean channels that are reasonably constant over the period taken to make the assessment.

The example presented here is that of an 8-tone MFSK system, operating firstly over a single skywave channel with 0dB mean SNR and a 0.1Hz rms fade rate and secondly over 2 skywave paths with 1ms path delay

Figure 5.10a. Error rates and HOX channel evaluation. 1 skywave. 0dB mean SNR. 0.1Hz fade rate.

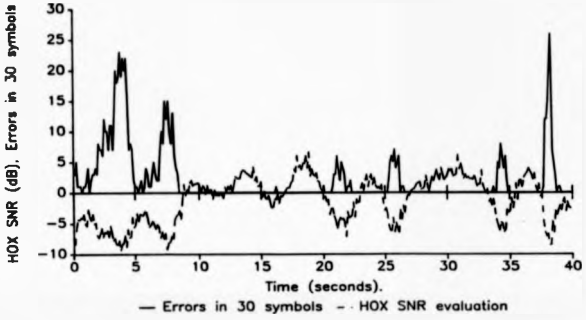
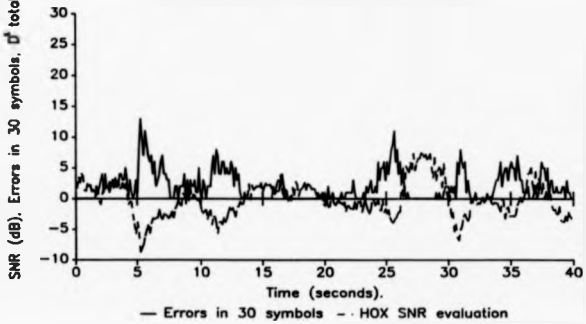


Figure 5.10b. Error rates & HOX SNR channel evaluation results. 2 skywaves 0dB mean SNR, 0.1Hz fade rate, 1ms DTD.



and otherwise similar conditions. Results can be seen in Figures 5.11a and b, where the results of the MOX channel evaluation procedure detailed in section 3.3.2 are repeated. The demodulator used here was the same one used to synchronise the MOX RTCE system. The results have been taken over 30 symbols since this is a convenient compromise between the symbol period and the fade period. It can be seen that there is close correlation between the error rates and the channel evaluation results.

5.6.2 Results from practical system

Using a similar system to that described in the section above, some practical tests have been made on the $D^s()$ method of demodulation, synchronisation and RTCE. The modem used was the 4-tone MPSK modem detailed in Table 4.4. The results from these tests can be seen in Figures 5.12a-b where the average $D^s()$ values and error rates are shown for blocks of 30 symbols which is equivalent to 81.6ms. During some periods, the error rates can be seen to be very bad and this is due to loss of synchronisation in the demodulator which results in roughly 75% errors.

All of the results shown have been taken from tests made on an HF link from Hull University to Warwick University, a distance of around 170 kilometres (106 miles) at a bearing of 242°. The data has been taken under night-time and daytime conditions on a frequency of 5.75MHz. The data transmitted was a pseudorandom (pn) sequence of 63 symbols over the Galois field GF(4) [MacWilliams & Sloane 1976]. Times referred to are in Greenwich mean time (GMT) and the HF conditions were generally thought to be favourable, a large peak in the sunspot number being due in late 1989.

Figure 5.12a is taken from a test starting at 15:27 on 14-2-89. The channel is obviously fading with a period of around 175 blocks or 14.3s (175 blocks . 30 symbols/block . 17 samples/symbol . 160µs/sample). Around block number 610, the signal can be seen to rise to a good SNR and then fall again in a period of around 14 blocks or 1.1s. This is accompa-

Figure 5.11a. D'() RTCE and HOX
channel evaluation. 1 skywave.
0dB mean SNR, 0.1Hz fade rate.

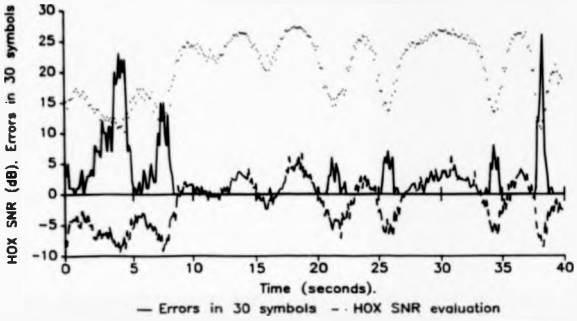


Figure 5.11b. D'() RTCE and HOX
channel evaluation results. 2 skywaves
0dB mean SNR, 0.1Hz fade rate, 1ms DTD.

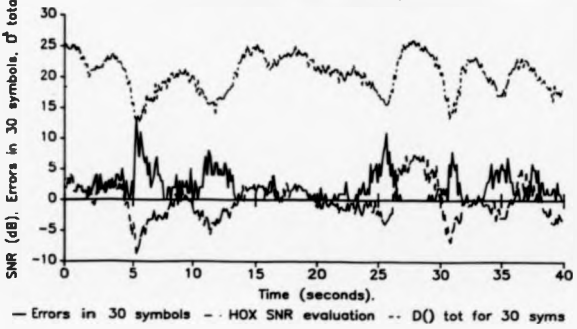


Figure 5.12a. Average $\hat{D}(n)$ values and error rates in 30 symbol blocks. Taken at 15:27, 14-2-89.

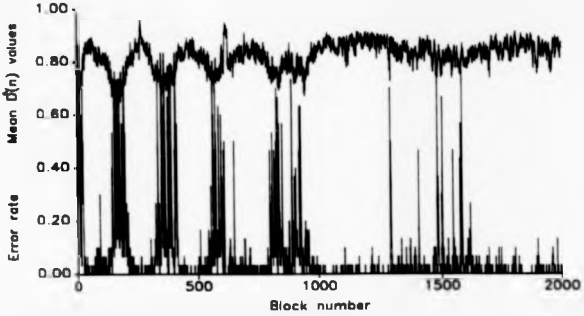
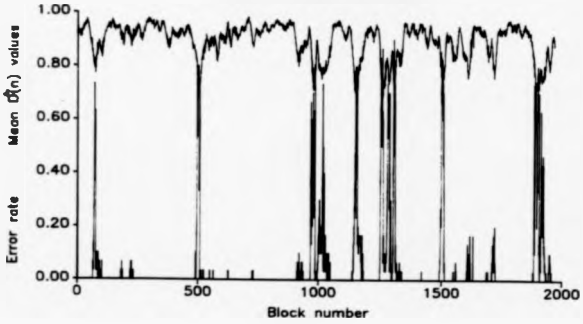


Figure 5.12b. Average $\hat{D}(n)$ values and error rates in 30 symbol blocks. Taken at 18:50, 16-3-89.



nied by a period of no errors and is thought to be due to a different mode of propagation fading in and out. Throughout this run, there are periods where the error rate is catastrophically bad and this is due to synchronisation errors. It is thought that this could be due to the fact that the tones were being received 17Hz low from where they should have been (4.6% of the tone spacing) due to receiver mis-tuning which was only discovered after the event.

Figure 5.12b is taken from a test done at 18:50 on the 16-3-89. It can be seen in this case that the modem was synchronised much better although there are periods where the synchronisation may be presumed to be lost. The channel can be seen to be fading in a typical sort of manner, with a fade periods of around 450 blocks (37s) at the beginning of the run and 140 blocks (11s) from the middle onwards. During the first half of this test, the signal is seen to have deep fades at 70 and 500 blocks, and to be fading in a more minor way in between.

This page intentionally left blank.

Chapter 6. Conclusions and further work

6.1 Conclusions

6.1.1 Zerocrossing RTCE.

In chapter 2 it was shown that the method of counting zerocrossings is a viable method of channel evaluation for use both in the AWGN channel and the Rayleigh fading channel, with the rider given in section 5.1.1. The analysis of the method has been approached from both theoretical and simulation points of view and some insight has been gained into the problems and advantages of applying the method to the HF channel.

The RTCE adaptive ARQ system detailed in section 5.1 was shown to have a distinct advantage over both constant rate coding systems and conventional ARQ systems. This was made possible by use of an independent channel evaluation method, the gain in throughput being attributable to the ability to use the 'no coding' condition when the channel is very good. It is noted that similar gains could be made by use of other independent channel evaluation systems.

The channel simulation method detailed in section 2.4 and appendix I is believed to be a useful contribution to the software tools available to the researcher working in the field of HF communications. It also provides some insight into the form of the baseband signal that can be expected to arrive over an HF channel.

6.1.2 Higher order crossings.

The uses of higher order crossings for channel evaluation and demodulation of MFSK signals has been investigated in chapter 3 from both theoretical and applications viewpoints. From the results given in section 5.4 and shown in Figure 5.9, it is clear that demodulation of MFSK signals by higher order crossing counting is possible, but very in-

efficient. The reasons for this are expounded in section 3.3.1. From section 5.5 and Figure 5.10 it is clear that the use of MOX for RTCE is a very real possibility, although the need for a separate demodulator may make the technique obsolete before it is used.

6.1.3 MFSK.

MFSK has been investigated as a candidate for reliable communications over HF radio channels. Some previously held assumptions on the spectral characteristics have been dispelled leading to calculation of the best number of tones to use for different channel conditions. Efficiency of the format at various SNRs has been shown, leading to some suggestions on improving the basic modulation method which are outlined in section 6.4 below.

The implementation of MFSK modulators and demodulators on DSPs has been considered and efficient algorithms produced for both. As part of the demodulator design, a new method of synchronisation has been developed which depends for its operation on the availability of DSP technology at the communications system receiver. It illustrates how the DSP architecture at the receiver can be used in a manner that is uniquely digital, rather than simply simulating classical analogue processing procedures.

The synchronisation procedure developed has the major practical advantage of requiring no specific synchronisation overheads in the form of preambles, sub-modulation, periodically inserted timing signals or extra bandwidth. It also has the advantages of being able to cope with variations in the absolute transit time of the signal, not requiring highly stable timing clocks and of being able to regain synchronisation after periods of synchronisation loss.

In addition to providing synchronisation, the demodulation algorithm developed in chapter 4 also provides real-time channel information. It

has been shown that this may be used to provide an indication of channel error rates in both simulated and live situations.

It is hoped that the work performed here could lead to a resurgence of interest in MFSK as a signalling format on the HF channel.

6.2 Further work

The following sections of work contain ideas that have come about during the course of this research. There is only passing connection between some of these ideas whilst others are inextricably linked. They have not been developed by the author although since they were first put on paper, some of them have been developed and improved by other members of the Hull-Warwick research group. In these cases, reference has been made to the work completed and in progress. The ideas are presented here as possible further areas of research for the future, without any claim as to their individual merit.

6.3 Technological improvements applied to digital communications.

Shannons capacity theorem gives us a theoretical value for the error free capacity of a channel, given that the time taken to evaluate the data is infinite. In most circumstances however, this is of little practical use as any practical demodulation/decoding scheme must give a result in a finite period of time and can be expected to give a non-zero probability of error.

The modem designer has the job of achieving a certain performance from the system that is being designed. In doing this, the theoretical channel capacity must be kept in mind but, probably more importantly, certain practical restrictions must also be satisfied. These restrictions may be on bandwidth, acceptable error rate, permissible error distribution, decoding delay, transmitter power and so on. The list may be virtu-

ally endless. The designer may find that the channel conditions are not all that is desired and so the question is posed, 'How can modern technological improvements help in achieving the aims of the modem designer?'.

6.3.1 Quantisation and DSP Usage.

As has already been mentioned, a major impetus of the work in this thesis has been the availability of cheap and plentiful signal processing power in the form of general purpose DSP integrated circuitry. In order to make use of this circuitry, an analogue signal coming over a channel must first be bandlimited (to prevent aliasing), sampled and digitized (by means of an A-to-D convertor) before any information can be extracted. The sampled signal must then be processed by the DSP in real time in order to recover the transmitted data. There are a number of parameters involved in this process which influence the possible throughput and reliability of the data delivered to the final user. These may be divided between factors within the control of the designer and factors beyond his/her control. They include:-

i) The number of quantization steps in the A-to-D convertor.

ii) The degree of 'clipping' at the A-to-D input.

iii) The observation time, or time delay, of the processing.

iv) The signal to noise ratio.

v) The required or acceptable throughput.

vi) The required or acceptable probability of error.

vii) The allowable bandwidth of the signal.

If theoretical relationships can be found between the parameters listed above, then they would be very useful in optimizing a communications system with regard to the practical constraints imposed on the designer. An example of this would be that of a fixed bandwidth AWGN chan-

nal, an A-to-D convertor of fixed resolution, a fixed amount of processing power limiting the observation time and a required maximum mean probability of error; it is required to calculate the maximum data rate through a system that would be possible for a given signal to noise ratio. This would enable both the system design to be as near to optimum as possible and the evaluation of how far from optimum any particular system may be.

The type of relationships mentioned above could also enable the accurate calculation of where extra effort in the design and costing procedure would offer improvements in performance. This involves estimating the relative 'cost' of each of the factors involved. An example of this would be to find if getting a higher resolution A-to-D convertor would be more cost effective than boosting the speed of the DSP in a given environment.

6.3.2 Improved Signal Formats

Current design techniques for new signalling formats seem to rely on modifying known techniques to provide some specific advantage. For example, the bandwidth requirements of binary and M-ary PSK/PSK systems have been reduced by using MSK [Stremler 1982 pp596] style modulation schemes [de Buda 1972, Amoroso 1976, Eaves & Wheatley 1979, Amoroso 1979, Aulin & Sundberg 1981a-b, Hambley & Tanaka 1984, Sundberg 1986, Padovani & Wolf 1986 etc.]. In these schemes the information from each bit is spread over several channel symbols and the phase changes are smoothed out to reduce the magnitude of spectral sidelobes.

It is assumed that at least part of the reason for basing these new modulation methods on older conventional binary types of modulation schemes is both conceptual and the difficulty of constructing hardware for a modulator giving anything but a sinusoidal - derived waveform. With DSP technology, most of the constraints of actually generating signals

once they have been designed have been overcome, with the result that we are free to design signals to convey information in an optimum way, in whatever way we choose to define 'optimum'. Signals generated by calculation and generation in a DSP in this way are bandlimited due to the sampling theorem and should be filtered accordingly. However, their detailed spectrum could be manipulated within the Nyquist constraint to match the available signalling spectrum. It is thought that signals could be adapted to have the appropriate spectrum for a channel with, for example, frequency selective fades or narrow band interference.

Additionally, signals may be designed specifically for the decoding method that it is intended to use, which should in turn be a maximum likelihood (Viterbi) decoder [Viterbi 1971].

6.4 Advantages, disadvantages and possible improvements to MFSK.

As has been mentioned in chapter 4, MFSK has particular advantages for use over HF radio, compared to binary modulation schemes, due to its long symbol period. This gives it relative immunity to intersymbol interference. It does however suffer from a very low data rate, compared to that theoretically possible, at high signal to noise ratios. It would therefore seems advantageous to investigate the basic modulation scheme further with the idea of preserving the symbol period but allowing for adaptation of the data rate to suit channel SNR conditions. Several ideas for doing this are presented below.

The PSD of MFSK, particularly when M becomes large, is already quite good in terms of sidelobe reduction with respect to main lobe width (see section 4.3), although it is thought that improvements could also be made in this area. These are dealt with when considering possible coding schemes applicable for MFSK modulation.

6.4.1 Increased in-band monitoring of MFSK.

As described in chapter 4, when dealing with MFSK we are concerned with M orthogonal signals. The way in which these signals are realised in practice is to use M tones that are orthogonal over a symbol period Ts. In theory we are using a system with 2.M dimensions, that is the real and imaginary parts of the M tones. In practice however we have to regenerate the M tones at the receiver in order to get the original signals back. This poses problems due to drift in the oscillators at both the transmitter and receiver sides and the phase coherence of the channel. As a result, normally the phase of each tone is presumed to carry no information and the magnitude of the signal at each frequency only is considered. This is what has been done in the demodulator detailed in chapter 4 and has previously been implemented in hardware [Ralphs 1985 chapter 9].

It is proposed that a system could be developed to track both the frequency and phase of each of the signalling frequencies. The information content of the signal could still be contained in the real component of each tone while the combination of real and imaginary components could be used to monitor the phase of the individual tones. This would enable the channel to be assessed in terms of sub-channel absolute signal and noise levels, phase stability and frequency shifts such as Doppler shifts. If the system were to be adaptive, and conditions suitable, the extra dimensions that become available could be used for the transmission of data.

How then may such a system be implemented? Since there are M tones then, on average, each tone will only be transmitted every 1/M symbols and hence if phase locked loops (PLL) were to be used to track the individual tone frequencies, they would require a narrow bandwidth and high stability compared to a normal binary system such as PSK or FSK. This is of course relatively easy to arrange on a DSP. The overall system would require a modified PLL arrangement that had one PLL per tone to track in-

dividual tone drifts and a master control to track changes common to all the tones, such as those due to RF carrier drift.

6.4.2 PSD control for MFSK.

Considering the current MFSK system and looking at the M-dimensional vector representing the signal being transmitted, we see that the M symbols are represented by the vectors with one '1' and M-1 '0's. The collection of these points is known as the constellation diagram. In between the symbol transitions, the vector traces out a locus in M-dimensional space from one of the constellation points to another. The course of the vector is dictated by the way in which the signal is constructed. In the case of MFSK it is the vector equivalent of switching from one tone to another at the symbol transition time. The PSD of the signal may be varied by switching between the tones in a different manner [Anderson & Salz 1965], in a similar way to the MSK type methods mentioned in section 6.1.2 above; in particular, the spectral lines may be abolished by altering the modulation index.

6.4.3 Coding for MFSK.

An idea for upgrading of the MFSK signalling format in terms of the data throughput comes from some work where the optimum input pdfs', or constellation points, have been calculated for certain forms of signalling. This work has been done for the continuous, scalar channel [Smith 1971], the sampled and quantized scalar channel [Honary, Ali & Darnell 1989] and the 2 dimensional continuous channel [Einarsson 1979]. It is not known whether the work has been done for larger numbers of dimensions although it obviously could be. In the absence of these results, it seems safe to assume that the basic feature of the previous results, that of the optimum constellation points not having equal probabilities

of occurrence, will carry over to higher numbers of dimensions. This means that currently we could find the optimum constellation points to use for our upgraded signalling format but unfortunately, as far as the author is aware, a coding scheme that can make full use of the uneven probabilities is not available. This is obviously an area in the development of codes that requires considerable research: the design of 'good' codes with specified rates and optimum symbol probabilities.

6.4.4 Conclusions on upgrading MFSK.

In conclusion then, we can say that MFSK modulation could be upgraded by use of a well chosen set of constellation points and a complementary coding scheme. In addition, the locus traced out between the constellation points may be chosen for the most desirable timing recovery and/or bandwidth control characteristics. As mentioned in section 6.3 above, these opportunities have been made available by the proliferation of DSP equipment, relieving us of the natural constraints of 'old fashioned' analogue hardware.

6.5 Synchronisation systems.

In section 4.6, a synchronisation system termed modulation derived synchronisation (MDS) was described. In this system, the relative timing of a received signal was assumed to be unknown and so the demodulation was performed at all of the possible timing offsets, ie once for every sample. A decision is then made as to which offset is most likely to be correct, based on the N dimensional vector positions at the various times.

An extension of this technique has been made [Honary & Zolghadr 1989] whereby the time period over which the parallel operations are performed, before the timing decision is made, is extended. This is made

possible by applying convolutional coding to the data before transmission. Matched filter detection is then implemented at the receiver in the same way as in the MDS system described previously.

In order to understand how this new method works, assume there are N samples per symbol, 'n' is an arbitrary timing instant and 'm' refers to one of the M possible received tones. The matched filter outputs $I^2(n,m)$ are assigned as metrics on the appropriate branches of the first trellis of a Viterbi decoder. Outputs $I^2(n+N,m)$ are assigned to the second trellis, $I^2(n+2*N,m)$ to the third and so on for one complete decoding tree. The procedure is repeated with the values $I^2(n+1,m)$, $I^2(n+1+N,m)$, $I^2(n+1+2*N,m)$, etc. on a second, time delayed, Viterbi decoding tree and so on for the N possible time slots. Maximum likelihood decoding is then performed on each separate decoder. The decoder with the largest total path metric is found and the appropriate data is output to the user, timing information having been recovered. This procedure has been termed code assisted bit synchronisation (CABS).

A potential problem with this method is that the relative timing between each sample of the received signal is assumed to be fixed. It is therefore proposed that the form of the Viterbi decoder used in the CABS system could be a modified to allow for effects of the transmission medium and reception equipment that may be interpreted as timing variations. The basic idea is to make more paths available through the decoding trellises and then search for the most likely one as before. More paths may be made available by adding extra connections between the nodes of the trellises corresponding to different time shifts. These would have 'penalties' attached to them which must relate to the effects that are being compensate for.

The type of effect that could be overcome using this procedure would be varying group delays between different symbol transitions [Ralphs 1985 pp121], caused by filtering in the receiver stage. Compensation could be achieved by allowing the symbol periods to vary slightly from their

nominal values. To make this variation possible, extra branches may be added to the decoding trees, connecting the nodes of the tree from one timing offset to corresponding nodes on the immediately preceding and succeeding trees. For these extra branches a negative weight would be assigned, its value being dependant on the allowable symbol period variations. Decoding would follow its normal form, except for the addition of an extra stage between each trellis where a correction step may or may not be taken.

A second method of improving the CABS system would be to implement equalization, or diversity combining, of multipath signals. This could be effected by allowing the branch metrics to be some linear combination of $I^2(n,m)$ values normally assigned. The combination required would not be known in advance since it represents the impulse response of the channel which is of course unknown. However, the impulse response changes reasonably slowly on the HF channel so the optimum combination would only change slowly, reducing the amount of processing required.

6.6 Symbol Period Adaption.

Coding systems, in general, have a fixed 'power' or correction capability which should not be exceeded for the code to be effective. Also, if the code is not to represent an inefficiency, the system should be arranged such that all of its 'power' is used. This 'power' may be expressed in terms of a maximum permissible error rate at the input to the decoder which may in turn be related to a particular signal to noise ratio at the input to a demodulator driving it. The 'best' SNR for any particular demodulator/decoder combination is therefore a function both the code power and the demodulator characteristics.

The output error rate of a demodulator is generally a function of E_m/N_m, the symbol energy to noise energy ratio, or equivalently E_m/N_m, the bit energy to noise energy ratio. Since N_m is a function of the noise

power and the bandwidth of the demodulator only, it is constant for all symbol periods. E_s, or E_b, is function of both the signal power and the symbol period and so increases as the symbol period is increased. This means that if the period of the symbols being transmitted over a channel could be varied as a function of the noise power, only one coding scheme would be necessary; this could have great complexity advantages in terms of hardware and/or software. This scheme is to be compared to the systems generally put forward to adapt a system to the channel conditions [Darnell 1983, Darnell, Honary & Shaw 1987], such as the one in section 5.1, to vary the code rate depending on the SNR.

We can write the relevant equations for this scheme as

$$E_s = S.T_s$$

$$N_s = N/B$$

$$E_s/N_s = S.T_s.B/N$$

where

E_s is the symbol energy

S is the signal power

T_s is the symbol period

N_s is the noise energy

N is the noise power

B is the system bandwidth

therefore for fixed E_s/N_s, B, N and S, we have

$$T_s = K/N$$

where K is some system dependent constant.

One of the problems with varying the symbol length in a system is that of keeping the PSD characteristics of the system constant. For example, if the symbol period were allowed to vary in an MFSK system such

as those previously described, firstly the tones may no longer be orthogonal and secondly, if the symbols got longer, there would be 'dips' in the spectrum in between the tone frequencies, in a similar way to wide band FSK [Anderson & Salz 1965]. This problem does not seem insuperable however and may be tackled by suitable symbol design.

6.7 Further uses of RTCE information for channel simulation

It is thought that the data that may be obtained from channel evaluation systems, such as those described in this thesis, may in future be useful for channel simulation. This would enable more realistic channels to be simulated than the currently available CCIR [1974] recommended channels. As has been seen in section 5.6.2, changes in fade rate may be detected, along with unusual modes fading in and out, a feature not available on the CCIR model [Watterson & Coon 1969c].

This page intentionally left blank.

References

Abdel-Ghaffar K. & McEliece R.J. "The Ultimate Limits of Information Density", NATO ASI conference on 'Performance Limits in Communication Theory and Practice', Tuscany, Italy, 7-19 July 1986.

Abramowitz M. & Stegun I.A. "Handbook of Mathematical Functions", Dover, 1965.

Ahmad N. & Natarajan T. "Discrete-Time Signals and Systems", Reston publishing co. inc., Reston Virginia, 1983.

Amoroso F. "Pulse and Spectrum Manipulation in the Minimum (Frequency) Shift Keying (MSK) Format", IEEE COM 24, No. 3, pp381-384, March 1976.

Amoroso F. "The use of Quasi-Bandlimited Pulses in MSK Transmission", IEEE COM 27, No. 10, Oct 1979.

Amoroso F. "The Bandwidth of Digital Data Signals", IEEE Communications Magazine, pp13-24, Nov 1980.

Anderson R.R. & Salz J. "Spectra of Digital FM", BSTJ, pp1165-1189, July-Aug 1965.

Aulin T. & Sundberg C.E.W. "Continuous Phase Modulation-Part 1: Full Response Signalling", IEEE COM 29, No. 3, March 1981a.

Aulin T., Rydbeck N. & Sundberg C.E.W. "Continuous Phase Modulation -Part 2: Partial Response Signalling", IEEE COM 29, No. 3, March 1981b.

de Buda R. "Coherent Demodulation of Frequency-Shift Keying With Low Deviation Ratio", IEEE COM 20, pp429-435, June 1972.

Berlekamp E.R. (editor) "Key papers in the development of coding theory", IEEE press, 1974.

Betts J.A. "High Frequency Communications", English University Press Limited, 1967.

Betts J.A. & Broom R.S. "The use of pilot tones for real-time channel estimation of HF data circuits", University of Southampton,

Electronics Dept. internal report, 1974.

Batts J.A., Ellington R.P. & Jones D.E.L. "C.W. sounding and its use for control of h.f. (3-30MHz) adaptive systems for data transmission", Proc. IEEE, Vol. 117, No. 12, Dec 1970.

Bozic S.M. "Digital and Kalman Filtering", Edward Arnold (publishers) Ltd., 1979.

Calderbank A.R. & Sloane N.J.A. "An eight dimensional trellis code", Proc. IEEE, Vol. 74, no. 5, pp757-759, May 86.

Cappellini V. "Digital Filters and their Applications", Academic press, London, 1979.

CCIR report 549 "HF Ionospheric channel simulators", 1974.

Chase D. "A Combined Coding and Modulation Approach for Communication over Dispersive Channels", IEEE COM 21, No. 3, pp159-174, March 1973.

Cvetkovoc M., Davis D.E.N., Griffiths H.D. & Collins B.S. "An HF Direction Finding and Null Steering System Employing a Four Element Circular Array", IEE Conference publication number 284, 'HF Radio Systems and Techniques', pp221-5, London, 1988.

Darnell M. "Channel Evaluation Techniques for Dispersive Communications Paths", Communication Systems and Random Process Theory, Sijthoff & Noordhoff, 1978.

Darnell M. "Embedded Real-Time Channel Evaluation Techniques", AGARD lecture series No. 127 on Modern HF Communications, May-June 1983.

Darnell M. "The Design of Static and Mobile HF Communications Systems", AGARD-LS-145 'Propagation Impact on Modern HF Communications Systems Design', AGARD, March 1986.

Darnell M. & Honary B. "Multi-Functional Coding Schemes Applicable to Secure Communication", Proc. IEE Int. Conf. on 'Secure Communications', Conf. Proc. No. 269, October 1986.

Darnell M., Honary B.K. & Shaw N.J. "An Adaptive Automatic Repeat Request Coding System Employing Real Time Channel Evaluation", Proc. 3rd

Joint USSR-Swedish International Workshop on Information Theory, Sochi, USSR, pp52-63 (Russian), pp197-206 (English), May 1987.

Darnell M. & Honary B. "New Synchronisation Techniques Applicable to Multi-Tone Transmission Systems", Proc. AGARD EPP Specialist Conference on 'Propagation effects circuit performance of modern radio-systems with particular emphasis on those employing band-spreading', France, Oct 1988.

Davis P.J. & Rabinowitz P. "Methods of Numerical Integration", Academic Press, New York, 1975.

Dawson J. "An HF Simulator for use with Real Time Channel Evaluation Systems", AGARD Conference Proceedings No. 363, Athens, 1984.

Dingle R.B. "Asymptotic Expansions: Their Derivation and Interpretation", Academic press, London, 1973.

Divsalar D. & Yuen J.H. "Asymmetric MPSK for Trellis Codes", 'Globecom 84' conference, pp20.6.1-20.6.8, November, 1984.

Dixon R.C. "Spread Spectrum Systems", John Wiley & Sons, New York, 1976.

Eaves R.E. & Wheatley S.M. "Optimization of Quadrature-Carrier Modulation for Low Crosstalk and Close Packing of Users", IEEE COM 27, No. 1, pp176-185, Jan 1979.

Edwards J.D. "A comparison of modulation schemes for binary data transmission", The Radio and Electronic Engineer, Vol. 43, pp562-568, Sept 1973.

Einarsson G. "Signal Design for the Amplitude-Limited Gaussian Channel by Error Bound Optimization", IEEE COM 27, No. 1, pp152-158, Jan 1979.

Ellington R.P. "The use of sounding to control adaptive systems for high frequency radio data transmission", University of Southampton Electronics Department publication No. 201, Jan 1970.

El-Khamy S.E. & Shaaban S.E. "Matched Chirp Modulation: Detection and Performance in Dispersive Communication Channels", IEEE COM 36, No. 4, pp April 1988.

Erdelyi A., Magnus W., Oberhettinger F. & Tricomi F.G. "Higher Transcendental Functions - Vol. 2", Bateman Manuscript Project, McGraw Hill Book Co. Inc., 1954.

Excell P.S. "A fading channel simulator and interleaved burst-error control techniques", University of York final year project, April 1985.

Gallager R.G. "Information Theory and Reliable Communications", Wiley, 1968.

Gardner F.M. "A BPSK/QPSK Timing-Error Detector for Sampled Receivers", IEEE COM 34, No. 5, pp423-429, May 1986.

Giles T.G. & Jessop G.R. "The radio data reference book", published by RSGB, edition 4, 1977.

Gott G.F. & Newsome J.P. "H.F. data transmission using chirp signals", Proc. IEE, Vol. 118, No. 9, pp1162-1166, Sept 1971.

Gott G.F. & Karia A.J. "Differential phase-shift keying applied to chirp data signals", Proc. IEE, Vol. 121, No. 9, pp923-928, Sept 1974.

Gott G.F., Wong N.F. & Dutta S. "Occupancy measurements across the entire HF spectrum", AGARD Conference 'Propagation Aspects of Frequency Sharing, Interference and System Diversity', Paris, 1982.

Gradshteyn I.S. & Ryzhik I.M. "Table of Integrals, Series, and Products", Academic press, 1980.

Gray A., Mathews G.B. & MacRobert T.M. "Bessel Functions", MacMillan, 1931.

Hague J., Jowett A.P. & Darnell M. "Adaptive Control and Channel Encoding in an Automatic HF communication system", Fourth International Conference on HF Radio Systems and Techniques, IEE Conference Publication No. 284, London, pp17-23, 1988.

Hambley A.R. & Tanaka O. "Generalised Serial MSK Modulation", IEEE COM 32, No. 3, pp305-311, March 1984.

Hodge S.E. & Seed N.L. "Statistics and Probability", Blackie and Chambers, Glasgow, 1972.

Honary B.M. "Error correction techniques for bursty channels", Ph.D. thesis, University of Kent, 1981.

Honary B. & Farrell P.G. "Coding techniques for HF communications systems", Proc. fourth international conf. on systems engineering, Coventry, Sept 1986.

Honary B., Shaw M.J. & Darnell M. "A New ARQ Transmission Scheme Involving Zerocrossing Channel Evaluation", Electronics Letters, No. 12, pp592-3, 12 May 1988.

Honary B., Zolghadr F. & Darnell M. "A Code-Assisted Bit Synchronisation (CABS) Scheme", IEE Electronics Letters, Vol. 25, No. 23, 19 Jan 1989.

Honary B., Shayan-Arani F. & Darnell M. "Optimisation of Constellation Angles for Trellis Coded 4-PSK Signals", to be published in Electronics Letters.

Honary B., Ali F. & Darnell M. "Information Capacity of Additive White Gaussian Noise Channel with Practical Constraints", to be submitted to IEE Proc. I, 'Communications, Speech and Vision', 1989.

Ince A.N. & Schemel R.E. "Factors affecting the use and design of spread-spectrum modems for the HF band", IEE Proc. Vol. 133, Part F, No. 2, pp187-193, April 1986.

Johnson N.L. & Kotz S. "Distributions in Statistics: Continuous Multivariate Distributions", Wiley, 1972.

Kedem B. "On the sinusoidal limit of stationary time series", Ann. Stat., Vol. 12, pp665-674, June 1984.

Kedem B. "Spectral Analysis and Discrimination by Zero-Crossings", Proc. IEEE, Vol. 74, No. 11, pp1477-1493, Nov 1986a.

Kedem B. "A Stochastic characterisation of the sine function", Amer. Math. Monthly, Vol. 93, p430-440, 1986b.

Kedem B. "On the asymptotic variance of higher order crossings with special reference to a fast white noise test", Biometrica, Vol. 3.1, pp143-9, 1986c.

Kirkpatrick E.M. (ed.) "Chambers 20th Century Dictionary", W & R Chambers Ltd., Edinburgh, 1983.

Laycock P.J., Morrell M., Gott G.F. & Ray A.R. "A model for HF spectral occupancy", Proc. of 'Fourth International Conference on HF Radio Systems and Techniques', IEE publication No. 284, pp165-176, London 1988.

Lin S. & Costello D.J.Jr. "Error Control Coding: Fundamentals and Applications", Prentice-Hall, 1983.

Lindsey W.C. "Error Probabilities for Rician Fading Multichannel Reception of Binary and N-ary Signals", IEEE IT 10, pp339-350, Oct 1964.

MacWilliams F.J. & Sloane N.J.A. "Pseudorandom Sequences and Arrays", Proc. IEEE Vol. 64, No. 12, pp, Dec 1976.

MacWilliams F.J. & Sloane N.J.A. "The theory of error correcting codes", North Holland Publishing Co., Amsterdam, 1978.

Matley M. "A digital high-frequency multipath propagation simulator", The Radio and Electronic Engineer, Vol. 47, No. 7, pp305-314, July 1977.

Maslin N. "HF Communications - a systems approach", Pitman Publishing, London, 1987.

Massey L.D. "Probability and Statistics", McGraw-Hill Inc, New York, 1971.

Murakami H. "Multichannel Convolutional Coding Systems over a Direct Sum of Galois Fields", IEEE IT 24, No. 2, pp205-212, March 88.

Padovani R. & Wolf J.K. "Coded Phase/Frequency Modulation", IEEE COM 34, No. 5, pp446-453, May 1986.

Peterson W.W., "Error correcting codes", MIT Press, Cambridge Mass., 1961.

Pearce T.H., Baker A.C. & Carter C.G. "The Application of DSP Techniques in HF Receiver Design", IEE Conference publication number 284, 'HF Radio Systems and Techniques', pp205-9, London, 1988.

Perry B.D. & Abraham L.G. "A wideband HF interference and noise model based on measured data", IEE Conference publication number 284, 'HF Radio Systems and Techniques', pp172-6, London, 1988.

Porter G.C. "Error Distribution and Diversity Performance of a Frequency-Differential PSK HF Modem", IEEE COM 16, No. 4, pp567-575, Aug 1968.

Proakis J.G. "Digital Communications", International students edition, McGraw-Hill, Singapore, 1983

Ralphs J.D. "An HF channel simulator using a new Rayleigh fading method", The Radio and Electronic Engineer, Vol. 46, No. 12, pp579-587, Dec 1976.

Ralphs J.D. "Principles and practice of multi-frequency telegraphy", Peter Perigrinus Ltd. for IEE, 1985.

Ratcliffe J.A. "Scientists' reaction to Marconi's transatlantic radio experiment", IEE Proc., Vol. 121, No. 9, pp1033-1038, Sept 1974.

Ream N. "Simulation of Rayleigh fading by digital generation of amplitude and phase", The Radio and Electronic Engineer, Vol. 48, No. 11, pp567-572, Nov 1978.

Rice S.O. "Mathematical Analysis of Random Noise", Part 3. BSTJ, Vol. 24, pp46-108, 1945.

Rice S.O. "Statistical Properties of a Sine Wave Plus Random Noise", BSTJ, Vol. 26, p109-157, 1946.

Rishbeth H. & Garriott O.K. "Introduction to Ionospheric Physics", Academic press, 1969.

Schwartz M. "Information Transmission, Modulation and Noise", McGraw-Hill, 1980.

Shannon C.E., Weaver W. "The Mathematical Theory of Communication", University of Illinois press, 1949.

Shaw M.J., Honary B. & Darnell M. "An RTCE-Assisted ARQ Transmission Scheme: Design & Performance", Fourth International Conference on HF Radio Systems and Techniques, IEE Conference Publication

No. 284, pp43-50, April 1988a.

Shaw M.J., Honary B. & Darnell M. "Optimisation of Parameters in an MFSK Transmission System", Electronics Letters, Vol. 24, No. 12, pp737-739, June 1988b.

Shaw M.J., Honary B. & Darnell M. "Higher Order Crossing Analysis Applied to Signal Detection and Evaluation of Radio Channels", Proc. IEE ICAP 89 conference, Warwick University, pp2.68-2.74, April 1989a.

Shaw M.J. & Honary B. "A new synchronisation technique for MFSK demodulation", submitted to IEE Electronics letters, March 1989b.

Sklar B. "Digital Communications", Prentice Hall International Inc., London, 1988.

Smith J.G. "The Information Capacity of Amplitude and Variance Constrained Scalar Gaussian Channels", Information and Control, Vol. 18, pp203-219, Academic Press Ltd., 1971.

Spencer A.J.M. et. al. "Engineering Mathematics. Volume 1", Van Nostrand Reinhold, 1980.

Stremler F.G. "Introduction to Communication Systems", Addison-Wesley publishing co., 1982

Sundberg C.E. "Continuous Phase Modulation", IEEE Communications Magazine, Vol. 24, No. 4, pp25-28, April 1986.

Tesla D.D. "A method of channel occupancy monitoring in adaptive HF systems", IEE HF Conference, London, pp76-79, 1986.

Thompson T.M. "From Error-Correcting Codes Through Sphere Packing To Simple Groups", The Mathematical Association of America, 1983.

Ungerboeck G. "Channel Coding with Multilevel/Phase Signals", IEEE IT 28, No. 1, pp55-67, Jan 1982.

Viterbi A.J. "Convolutional codes and their performance in communications systems", IEEE COM 19, No. 5, Oct 1971.

Watterson C.C., Juroshek J.R. & Bensema W.D. "Experimental Verification of an Ionospheric Channel Model", ESSA Technical Report ERL 112-ITS 80, July 1969a.

Watterson C.C., Ax G.G., Demmer L.J. & Johnson C.N. "An Ionospheric Channel Simulator", ESSA Technical Memorandum ERLTM-ITS 198, Sept 1969b.

Watterson C.C. & Coon R.M. "Recommended Specifications for Ionospheric Channel and Atmospheric Noise Simulators", ESSA Technical Report ERL 127-ITS 89, Sept 1969c.

Watterson C.C. "Experimental Confirmation of an HF Channel Model", IEEE COM 18, No. 6, pp792-803, Dec 1970.

Wozencraft J.M. & Jacobs I.M. "Principals of Communications Engineering", Wiley, 1965.

Zolghadr F., Honary B. & Darnell M. "Embedded Convolutional Decoding", Fifth international conference on Digital Processing of Signals in Communications, IERE publication No. 82, pp247-255, Sept 1988.

This page intentionally left blank.

Symbols and abbreviations

AM	Amplitude modulation
ARQ	Automatic repeat request
ASK	Amplitude shift keying
A-to-D	Analogue to digital convertor
ATU	Aerial tuning unit
AWGN	Additive white Gaussian noise
AZCPS	Average zerocrossing per second
BER	Bit error rate
bit	Binary digit
bps	Bits per second
BPSK	Binary phase shift keying
B	Bandwidth
C(x)	Capacity as a function of x
CABS	Code assisted bit synchronisation
CCIR	Comité Consultatif International des Radiocommunications
cdf	Cumulative distribution function
CPSD	Cumulative power spectrum density
D-to-A	Digital to analogue convertor
$D^2(n)$	Detection confidence at time n
dB	Decibel
DF	Frequency spacing of MFSK tones
DSP	Digital signal processor
δt	Small time increment
DT	1/fs, the period between samples
DTD	Differential time delay
E_b	Bit energy
E_b/N_o	Bit energy/noise energy ratio
eg	for example
E_s	Symbol energy

f_{crit}	Critical frequency
FEC	Forward error correction
FM	Frequency modulation
Fmax	Maximum frequency of PSD
f_{rms}	Root-mean-square fading frequency
fs	Sampling frequency
FSK	Frequency shift keying
G	Number of tone intervals beyond upper and lower tones of MFSK system to give bandwidth containing 99% of power
H	User information rate in bps/Hz
He	Channel information rate in bps/Hz
HF	High frequency (2-30MHz)
HOX	Higher order crossings
Hz	Hertz, frequency in cycles per second
$I^2(n,m)$	Squared matched filter output for tone m at sample time n
ie	That is
IF	Intermediate frequency
IIR	Infinite impulse response
ISI	Intersymbol interference
ITU	International telegraph union
LB	Lower 3dB point
LHS	Left hand side
LSB	Lower sideband
M	Number of symbols in an MFSK system
MDS	Modulation derived synchronisation
MFSK	Multi-Frequency shift keying
ms	millisecond
MSK	Minimum shift keying
MUF	Maximum usable frequency
m,n	integer variables
N	Noise power, number of samples per symbol

NAK	Non-acknowledgement
N_o	Noise energy in watts/Hz
OWF	Optimum working frequency
OOK	On-off keying
P_e	Probability of bit error
PC	Personal computer
pdf	probability density function
P_e	Probability of error
PLL	Phase locked loop
PRK	Phase reversal keying
PSD	Power spectral density
PSK	Phase shift keying
Pz	Probability of a zerocrossing
Pz+	Probability of a positive going zerocrossing
Pz-	Probability of a negative going zerocrossing
R	Code rate
RF	Radio frequency
RHS	Right hand side
rms	Root mean square
RS	Reed-Solomon
RTCE	Real time channel evaluation
rv	Random variable
Rx	Receiver
s	Second
sd	standard deviation
S-meter	Signal strength meter
sinc(x)	sin(x)/x
SNR	Signal to noise power ratio
SSB	Single sideband
Te	Element period
Ts	Symbol period

Tx	Transmitter
UB	Upper 3dB point
USB	Upper sideband
w	Radian frequency
w_o	Centre radian frequency
zc	zerocrossing
θ	Phase angle
$\lceil x \rceil$	Smallest integer greater than or equal to x
β	Correlation coefficient
α	Normalised frequency

Appendix A

Probability of Zero Crossings for a Sinusoid with Additive Gaussian Noise - Look-up Table Approach.

The probability of a zerocrossing, averaged over a long period has been shown in chapter 2 to be given by

$$P_{ze} = 2. \int_{0}^{t=2.\pi/w} Erfc[\sin(w.t)/\sigma].Erf[\sin(w.(t-\delta t))/\sigma].dt \qquad (2.10)$$

This formula has been converted into a summation and programmed on a computer using a look up table for Erfc(x). Values of SNR dB have been taken over a large range and σ calculated from them. The values in the look-up table were previously calculated by numerical integration up to Erfc(3.0), and beyond that range by using an approximate formula [Stremler 1982 pp675].

The problem with this method is in assessing the errors in the calculations due to the finite step sizes taken for the integrations. The method is at best crude and time consuming but is quite useable if a single table of data is required. The results compare very favourably with simulated results and were used in the first part of this study.

Appendix B

Probability of Zero Crossings for a Sinusoid with Additive Gaussian Noise - Convergent series approach.

From the work in chapter 2 we know

$$P_{\infty} = w/\pi \int_{0}^{t=2.\pi/w} Erfc(\sin(w.t)/\sigma).Erf\{\sin(w.(t-\delta t))/\sigma\}.dt \qquad (2.10)$$

Putting $\theta=w.t$, $d\theta=w.dt$, $Erfc(x)=1/2(1-erf(x/\sqrt{2}))$, $Erf(x)=1/2(1+erf(x/\sqrt{2}))$ [Stremler 1982 pp676] and altering the limits of integration we get

$$P_{\infty}=1/(4.\pi)\int_{-\pi}^{\pi}\{1+erf(\sin\theta/(\sqrt{2}.\sigma))\}\{1-erf(\sin(\theta+w)/(\sqrt{2}.\sigma))\}.d\theta \qquad (B1)$$

Expanding out the brackets under the integral, we can integrate the constant term to leave three terms. We can use [Dingle 1973 pp1]

$$erf(x)=\frac{2.x}{\sqrt{\pi}}\sum_{s=0}^{\infty}\frac{(-x^2)^s}{s!\,(2s+1)} \qquad (B2)$$

and by changing the order of integration and summation, the $erf(\sin\theta/(\sqrt{2}.\sigma))$ term becomes

$$\frac{\sqrt{2}}{w.\sqrt{\pi}.\sigma}.\sum_{s=0}^{\infty}\frac{(-1)^s}{2s.\sigma^{2s}.s!.(2s+1)}.\int_{-\pi}^{\pi}\sin\theta.\sin^{2s}\theta.d\theta \qquad (B3)$$

and clearly this integral goes to zero for all s. A similar thing happens to the term involving $\theta+w$. This leaves

$$P_n=\frac{1}{2}-\frac{1}{2.\pi^2.\sigma^2}.\sum_{m=0}^{\infty}\sum_{n=0}^{m}\frac{(-1)^{m-n}}{2^{m-n}.\sigma^{2(m-n)}.m!.n!.(2.m+1).(2.n+1)}.\int_0^{2\pi}\sin^{2m+1}\theta.\sin^{2n+1}(\theta+w).d\theta \qquad (B4)$$

We also know [Gradshteyn & Ryzhik 1980 1.320 page 25]

$$\sin^{2n-1}(x)=\frac{1}{2^{2n-2}}.\sum_{k=0}^{n-1}(-1)^{n-k-1}.\left|\begin{matrix}2n-1\\k\end{matrix}\right|.\sin(2.n-2.k-1)x \qquad (B5)$$

where

$$\left|\begin{matrix}2a+1\\k\end{matrix}\right|=\frac{(2a+1)!}{(2a+1-k)!k!} \qquad (B6)$$

and putting 2a+1 for 2n-1 we get

$$\sin^{2a-1}x=\frac{1}{2^{2a}}.\sum_{k=0}^{a}(-1)^{a-k}.\left|\begin{matrix}2a+1\\k\end{matrix}\right|.\sin(2.a+1-2.k)x \qquad (B7)$$

so (B4) becomes

$$P_n=\frac{1}{2}-\frac{1}{2.\pi^2.\sigma^2}.$$

$$\sum_{m=0}^{\infty}\sum_{n=0}^{m}\frac{(-1)^{m-n}}{2^{2(m-n)}.\sigma^{2(m-n)}.m!n!(2m+1)(2n+1)}\sum_{k=0}^{m}\sum_{i=0}^{n}(-1)^{m-n-k-i}.$$

$$\left|\begin{matrix}2m+1\\k\end{matrix}\right|.\left|\begin{matrix}2n+1\\i\end{matrix}\right|.\int_{-\pi}^{\pi}\sin(2m+1-2k)\theta.\sin(2n+1-2i)(\theta+w)d\theta \qquad (B8)$$

The integral here is in the form

$$I=\int_{-\pi}^{\pi}\sin(A.\theta).\sin(B(w+\theta)).d\theta \qquad A,B \text{ integer} \geq 1. \qquad (B9)$$

Expanding $\sin(B(w+\theta))$ and multiplying out we get

$$I=\cos(B.w).\int_{-\pi}^{\pi}\sin(A.\theta).\sin(B.\theta).d\theta + \sin(B.w).\int_{-\pi}^{\pi}\sin(A.\theta).\cos(A.\theta).d\theta \quad (B10)$$

Clearly, the second integral goes to zero. Expanding the term under the first integral we get

$$I=\frac{\cos(B.w)}{2}.\int_{-\pi}^{\pi}\cos(A-B)\theta - \cos(A+B)\theta.d\theta \quad (B11a)$$

$$=\frac{\cos(B.w)}{2}.\left[\frac{\sin(A-B)\theta}{A-B}-\frac{\sin(A+B)\theta}{A+B}\right]_{-\pi}^{\pi} \quad A\neq B \quad (B11b)$$

$$=0 \text{ since } (A-B) \text{ and } (A+B) \text{ both integer } \neq 0 \quad (B11c)$$

$$=\frac{\pi.\cos(B.w)}{w} \quad A=B \quad (B11d)$$

and hence the integral in (B8) is non-zero only when A=B i.e. when $2m+1-2k=2n+1-2l$, $m-k=n-l$. This means that we can eliminate one of the summations and we choose to eliminate l. This gives $k=m-n+l$ and when $l=0$, $k=m-n$; when $l=n$, $k=m$; and so we get new limits for k. Also $k\geq 0$ and so $m-n\geq 0$ and hence we get a new upper limit for n of $n=m$. Putting all of this into (B8) we get

$$P_m = \frac{1}{2}-\frac{1}{2\sigma^2\pi}\sum_{m=0}^{\infty}\sum_{n=0}^{m}\sum_{k=m-n}^{m}\frac{(-1)^{m-n}(2m)!(2n)!\cos(2n+1-2(k-m+n))w}{\theta^{m-n}\sigma^{2m-2n}\,m!n!(2m+1-k)!(1-k+m+n)!k!(k-m+n)!} \quad (B12)$$

and putting $l=k-m+n$ to simplify the k summation gives

(B13)

$$P_e = \frac{1}{2} - \frac{1}{2\sigma^4\pi} \sum_{m=3}^{\infty} \sum_{n=0}^{m} \sum_{k=-n}^{n} \frac{(-1)^{m-n}(2m)!(2n)!\cos(1-2l+2n)w}{8^{m-n}\sigma^{2m-2n}\, m!n!(m+n-1+1)!(1-l+2n)!(l+m-n)!l!}$$

This formula should now be a convergent expansion for equation (B1). The expansion for the error function used is valid for all x but requires a sufficient number of terms to be calculated to give an accurate result. This number is approximately equal to $3.|x^2|$ [Dingle 1973 pp1]. In (B1) it can be seen that within the range of integration, the maximum value of x will be $1/(\sqrt{2}.\sigma)$. The signal to noise ratio in dB is given by

$$SNR\ dB = 10.\log(0.5/\sigma^2) \qquad (B14)$$

since the signal power is 1/2 and therefore we get the number of terms required as approximately

$$3*10^{[illegible]} \qquad \text{where SNR is in dB.} \qquad (B15)$$

We cannot easily calculate on our computer terms which involve factorials above 30. This leads us to discover that the maximum value of 'm' in (B13) above is 14 which gives a signal to noise ratio of approximately 7dB. Even at this signal to noise ratio, however, it has proved very difficult to calculate the points reliably and a termination point of -10dB seems to be the best that can be achieved.

Appendix C

Probability of Zero Crossings for a Sinusoid with Additive Gaussian Noise

-Rational Approximation and Gaussian Integration Approach

From the previous work we know

$$P_z = w/\pi \int_0^{t=2.\pi/w} Erfc(\sin(w.t)/\sigma).Erf(\sin(w.(t-\delta t))/\sigma).dt \quad (2.10)$$

We can put $\theta=w.t$, $d\theta=w.dt$ to make this integral more reasonable and we get

$$P_z=w/\pi. \int_0^{2\pi} Erfc(\sin(\theta)/\sigma).Erf(\sin(\theta+w)/\sigma).d\theta \quad (C1)$$

This integral cannot be solved in a closed form and so a numerical solution has been found, valid for a wide range of the parameters w and σ. To do this, an easily calculated expression for Erfc(x) and Erf(x) has been used. It is the rational approximation [Abramowitz & Stegun 1980 7.1.26 pp299]

$$erf(x)=1-(a_1t+a_2t^2+a_3t^3+a_4t^4+a_5t^5).\exp(-x^2)+\varepsilon(x) \quad (C2)$$

$$t=1/(1+p.x)$$

$$|\varepsilon(x)|\leq 1.5\times10^{-7} \quad \text{for all } x$$

$p=0.32759$ $a_1=0.254829592$ $a_2=-0.284496736$

$a_3=1.421413741$ $a_4=-1.453152027$ $a_5=1.061405429$

and also $Erfc(x)=1/2(1-erf(x/\sqrt{2}))$, $Erf(x)=1/2(1+erf(x/\sqrt{2}))$ [Stremler 1982 pp676]. In order to do the θ integration, a Gaussian-type numerical

method has been chosen and this has the general form [Abramowitz & Stegun 1980 pp887-8]

$$\int_a^b f(y).dy = \frac{b-a}{2}.\sum_{i=1}^{n} w_i.f(y_i)+R_n \qquad (C3)$$

where

$$R_n = \frac{(b-a)^{2n+1}.(n!)^4.2^{2n+1}}{(2n+1).[(2n)!]^3}.f^{2n}(E) \quad a<E<b \qquad (C4)$$

gives the error in calculation and

$$y_i=\frac{(b-a)}{2}.x_i + \frac{(b+a)}{2}.$$

x_i and w_i are found from the roots of Legendre polynomials [Davis & Rabinowitz 1975 pp27-28] and are tabulated to 21 decimal places [Abramowitz & Stegun 1980 pp887-8]. The order of integration, n, must be chosen and is related to the order of the polynomial expansion of f(y), should one exist. We do not have such an expansion but is was felt that the order should be quite low. 16th order was chosen and found to work well up to a signal to noise ratio of 10dB. The expression for the error in the integration, R_n, has not been calculated, due to the difficulty in finding the 32nd derivative of our function.

Appendix D

Probability of Zero Crossings for a Sinusoid with Additive Gaussian Noise

-Fourier Transform Approach

The average probability of a zerocrossing has been shown to be given by

$$Pz = 2.\int_{0}^{t=2.\pi/w} Erfc(sin(w.t)/\sigma).Erf(sin(w.(t-\delta t))/\sigma).dt \qquad (2.10)$$

Putting $\theta = w.t$, $d\theta = w.dt$, $Erfc(x) = 1/2(1-erf(x/\sqrt{2}))$, $Erf(x) = 1/2(1+erf(x/\sqrt{2}))$ [Stremler 1982 pp676] and altering the limits of integration we get

$$Pz = 1/(4.\pi)\int_{-\pi}^{\pi} \{1+erf(sin\theta/(\sqrt{2}.\sigma))\}\{1-erf(sin(\theta+w)/(\sqrt{2}.\sigma))\}.d\theta$$

It is easy to see that, since $erf(sin\theta/\sqrt{(2}.\sigma))$ is an odd periodic function of θ with period 2π, when the product is expanded two of the terms will vanish on integration leaving the expression

$$Pz = 1/2 - 1/(4.\pi)\int_{-\pi}^{\pi} erf(sin\theta/(\sqrt{2}.\sigma)).erf(sin(\theta+w)/(\sqrt{2}.\sigma)).d\theta \qquad (D1)$$

Now assume that $erf(K.sin\theta)$ is expanded in a Fourier series

$$erf(K.sin\theta) = \sum_{n=1}^{\infty} a_n(K).sin(n.\theta) \qquad (D2)$$

where

$$a_n = 1/\pi \int_{-\pi}^{\pi} erf(K.sin\theta).sin(n.\theta).d\theta \qquad (D3)$$

Introducing (D2) into (D1) yields

$$P_e = \frac{1}{2} - \frac{1}{4.\pi} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_m(1/(\sqrt{2}.\sigma)).a_n(1/(\sqrt{2}.\sigma)).\int_{-\pi}^{\pi} \sin(m.\theta).\sin(n\theta).d\theta$$

$$= \frac{1}{2} - \frac{1}{8.\pi} \sum_{m,n=1}^{\infty} a_m(1/(\sqrt{2}.\sigma)).a_n(1/(\sqrt{2}.\sigma)).\int_{-\pi}^{\pi} \cos(n.w)-\cos(n(w-2\theta)].d\theta$$

$$= \frac{1}{2} - \frac{1}{4} \sum_{m=1}^{\infty} a_m(1/(\sqrt{2}.\pi)).\cos(m.w) \qquad (D4)$$

It now remains to determine the coefficients. It readily shown from (D3) that $a_n=0$ for all even values of n. Thus it is left to evaluate

$$a_{2r+1}(K) = 1/\pi \int_{-\pi}^{\pi} erf(K.\sin\theta.\sin((2r+1).\theta).d\theta \qquad (D5)$$

Integrating once by parts

$$\pi.a_{2r+1}(K)=erf(K.\sin\theta).\left[\frac{-\cos((2r+1)\theta)}{2r+1}\right]_{-\pi}^{\pi} + \int_{-\pi}^{\pi} \frac{d(erf(K.\sin\theta)}{(2r+1).d\theta}.\cos((2r+1)\theta).d\theta$$

Since erf(0)=0 the first term on the RHS vanishes. Further

$$\frac{d\ erf(K.\sin\theta)}{d\theta} = \frac{2}{\pi}.\exp(-(K.\sin\theta)^2).K.\cos\theta$$

therefore

$$\pi.a_{2r+1}(K) = \frac{2K}{(2r+1)\sqrt{\pi}}.\int_{-\pi}^{\pi} \exp(-K^2.\sin^2\theta).\cos\theta.\cos((2r+1).\theta).d\theta$$

$$= \frac{K.\exp(-K^2/2)}{(2r+1)\sqrt{\pi}}.\left[H_{r+1}(K) + H_r(K) \right] \qquad (D6)$$

where

$$H_r(K) = \int_{-\pi}^{\pi} \exp(K^2.\cos(2.\theta)/2).\cos(2.r.\theta).d\theta$$

Expanding the exponential term [Gray 1931]

$$\exp(K^2.\cos(2.\theta)/2) = Io(K^2/2) + 2.\sum_{n=1}^{\infty} In(K^2/2).\cos(2.n.\theta)$$

and so

$$H_r(K)=Io(K^2/2).\int_{-\pi}^{\pi} \cos(2.r.\theta).d\theta + 2.\sum_{n=1}^{\infty} In(K^2/2).\int_{-\pi}^{\pi} \cos(2.n.\theta).\cos(2.r.\theta).d\theta$$

and clearly here the first integral is zero unless r=0, when its value is 2π. Also, the integral in the summation is zero unless n=r, when its value is π. Thus

$$H_r(K) = 2.\pi.I_r(K^2/2) \quad \text{for all } r \qquad (D7)$$

whence from (D6)

$$a_{r+1}(K) = \frac{2.K.\exp(-K^2/2)}{(2r+1).\sqrt{\pi}}.[\, I_{r+1}(K^2/2) + I_r(K^2/2) \,] \qquad (D8)$$

Combining these results together we get

$$P_E = \frac{1}{2} - \frac{\exp(-1/(2.\sigma^2))}{2.\pi.\sigma^2} \cdot \sum_{r=0}^{\infty} \frac{\cos((2r+1).w)}{(2r+1)^2} \cdot \{I_{\frac{r}{2}}(1/(4.\sigma^2)) + I_{\frac{r+1}{2}}(1/(4.\sigma^2))\}^2 \qquad (D9)$$

This has proved to be an easy method of calculating solutions to equation (3.5). The integer order modified Bessel functions are rapidly calculated from convergent series [Dingle 1973] and the summation in (D9) is rapidly convergent itself.

Appendix E

Probability of Zero Crossings for a Sinusoid with Additive Gaussian Noise - The Non-Sampled Case Considered by Rice.

Rice [1946] considers the current

$$I = Q.\cos(qt) + I_n$$

where

Q = amplitude constant

q = radian frequency constant

t = time

I_n = random noise current

and obtains the formula (equation 2.8 of [Rice 1946])

$$2.N(0) = M_n.[\exp(-\alpha).I_o(\beta) + b^2.Ie(\beta/\alpha,\alpha)/(2.\alpha)] \qquad (E1)$$

where

$2.N(0)$ = expected number of zeros per second

$M_n = \sqrt{(-v_n''/v_n)}/\pi$

v_n = mean square noise current

$= \sigma^2$

$$-v_n'' = 4.\pi^2.\int_0^\infty w(f).f^2.df$$

$w(f)$ = power spectrum of I_n

$\alpha = (a^2 + b^2)/4$

$\beta = (a^2 - b^2)/4$

$a = Q/\sqrt{(v_n)}$

$$b = Q.q/\sqrt{(-v_m'')} = 2.\pi.f_m/M_m$$

f_m = q/2.π frequency in Hz

$I_m(x)$ = Bessel function of order zero and imaginary argument

$$= \sum_{s=0}^{\infty} (x/2)^{2s}/s!s! \qquad \text{[Dingle 1973]}$$

$$Ie(k,x) = \int_0^x \exp(-u).I_m(k.u).du$$

$$= \sum_{n=0}^{\infty} (k/2)^{2n}.(2.n)!.A_m/(n!.n!)$$

$$A_m = 1 - [1 + x + x^2/2! \ldots + x^{2n}/(2.n)!].\exp(-x)$$

ie $A_0 = 1 - \exp(-x)$

$A_1 = 1 - [1 + x + x^2/2].\exp(-x)$ and so on.

This is derived from the integral for Ie(k,x), substituting in the power series for $I_m(k.u)$, changing the order of summation and integration and then repeatedly integrating by parts.

Equation (E1) can be evaluated by use of the power series representations of $I_m(x)$ and Ie(k,x) and when calculated, requires some re-scaling in order to be comparable with our previous results. This re-scaling is due to the fact that Rice calculates the average number of zerocrossings per second and we calculate the average number of zerocrossings per intersample space.

This formula has the advantage over (2.10) and its solutions in that the power spectral density of the noise can be specified to any arbitrary shape.

In order to compare Rices work with our own, the spectrum is assumed to be flat in the range 0 -> 1/(2.δt) and so w(f) becomes

$$w(f) = N_o \quad 0 < f < 1/(2.\delta t)$$
$$= 0 \quad \text{elsewhere}$$

where

N_o = power per Hz of the noise

δt = time between samples in the sampled case.

The total noise power is given by

$$v_o = N_o \text{ . bandwidth}$$
$$= N_o/(2.\delta t).$$

Now $-v_o''$ becomes

$$-v_o'' = 4.\pi^2.\int_0^{1/(2.\delta t)} N_o.f^2.df$$
$$= (4.\pi^2.N_o/3).[f^3]_0^{1/(2.\delta t)}$$
$$= 4.\pi^2.N_o/(3.2^3.\delta t^3)$$
$$= \pi^2.N_o/(6.\delta t^3).$$

And so we obtain N_o

$$N_o = \sqrt{(-v_o''/v_o)}/\pi$$
$$= \sqrt{(1/3.\delta t^2)}$$
$$= \sqrt{(1/3)}/\delta t$$

We now need to calculate a, b, α and β. We obtain a and b directly

$$a = Q/\sqrt{(v_o)}$$

$$b = Q.q/\sqrt{(-v_m'')}$$
$$= Q.q.\sqrt{(6.\delta t^2/(N_m.\pi^2))}$$
$$= Q.q.\sqrt{(6.\delta t^2/N_m)}/\pi$$

and hence α and β

$$\alpha = (a^2 + b^2)/4$$
$$= Q^2.\delta t.[1 + q^2.3.\delta t^2/\pi^2]/(2.N_m)$$

$$\beta = (a^2 + b^2)/4$$
$$= Q^2.\delta t.[1 - q^2.3.\delta t^2/\pi^2]/(2.N_m).$$

Now we can calculate the expected number of zerocrossings per second with respect to E_m/N_m by use of the formula (E1) given by Rice and the series expansions of $I_m(x)$ and $I_0(k,x)$.

The calculations have been made to 5 decimal places and plotted in Figure 2.2. Also displayed there is the results of the calculations done by the method of appendix A. It is clear that at the low signal end, ie the waveform is nearly all noise, the curves, whilst being the same shape, are of different magnitudes.

In the case of the all noise waveform, sampled at the Nyquist rate, we would expect the number of zerocrossings per second to be equal to the highest frequency. The reason being that each sample is independant and so has equal probability of being above or below the zero line. Therefore the probability of two successive samples being on opposite sides of the zero line is an half and, since we have two samples per cycle of the highest frequency present, the average number of zerocrossings per second is equal to that frequency.

Why then does Rice's formula predict a higher figure? In order to answer this question we look at some earlier work of his on the statistics of random noise [Rice 1945] in which he derives a formula for the

expected number of zeros per second of a random noise current. For the case that we are looking at of a flat noise power spectrum density in a band covering 0 to f Hz he gets

$$\text{expected number of zeros per second} = 2.f/\sqrt{3}$$
$$= 1.155\ f$$

and the formula we investigated previously gives the same answer as Q->0. The reason for this difference is that the system we are using to count the number of zerocrossings both in the simulations and in the methods of appendices A to D is not the same as that being considered in Rice's analysis of the continuous case.

Nyquist's sampling theorem says that we should have enough information from the samples to reconstruct exactly the original signal that we have sampled, but does not guarantee that the samples will always enable us to detect zerocrossings in the way in which we are trying to do. The 'missing' 15% of the zerocrossings must be due to situations such as that shown in Figure 2.3. It is very unlikely to get extra zerocrossings when the signal has changed sign because this would require two turning points of the signal between the samples and this is rare, if not impossible, due to bandwidth considerations. The occurrences of 'extra' zerocrossings must therefore happen mostly in pairs and so occurrences need only happen $(\sqrt{(4/3)}-1)*100/2$ = 7.7% of the time to cause the factor that we have found. This is surely the mechanism by which this factor is occurring but we have not been able to derive a mathematical expression for it.

We can get a bound for this probability by noting that, for the pure noise case, each sample has a 50% chance of being above or below the zero line and so there is a 50% chance of two successive samples being the same side of the zero line. Similarly, there is a 50% chance of the slope of the signal at each sample time being in the correct direction and so we get a bounding probability of 1/8 or 12.5%.

So, we need to look at the universe of signals that could possibly occur between the samples to cause an extra pair of crossings. The possible forms are shown in Figure E.1 and are symmetrical. The magnitudes of the samples are Gaussianly distributed, and the slopes have a symmetrical distribution on $-\infty \rightarrow +\infty$.

In order to reconstruct a bandlimited analogue signal from the samples, we can represent each sample $x(n.\delta t)$ by a bandlimited function, $y(t)$, having the same value at each sample time as the sample being considered ie for sample N

$$y(t) = x(N.\delta t) \qquad \text{at } t = N.\delta t$$
$$= 0 \qquad \text{at } t = n.\delta t \quad n \neq N$$

and for this we use the well known [Cappellini 1979] 'sinc' function given by

$$\frac{\sin(w_s.t/2)}{w_s.t/2} = \text{sinc}(w_s.t/2)$$

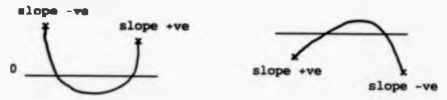
where

$w_s = 2.\pi.f_s = 2.\pi/\delta t$

f_s = sampling frequency

t = time

Figure E.1. Possible forms of signal between samples that could cause zerocrossings to go undetected.



and so we can use, for the sample at time N.δt

$$y(t) = x(N.\delta t).\mathrm{sinc}((t-N.\delta t).\pi/\delta t)$$

where y(t) is our bandlimited analogue function. Our total analogue signal wf(t) therefore becomes

$$wf(t) = \sum_{n=-\infty}^{\infty} \mathrm{sinc}((t-n.\delta t).\pi/\delta t).x(n.\delta t)$$

which has solutions @ t = n.δt of x(n.δt) and is bandlimited to 1/(2.δt) as required. A waveform reconstructed in this manner using 21 random samples is shown in Figure 2.3. The 'extra' zerocrossings can also be seen in the 3rd interval.

We now wish to find a description of the waveform between N.δt and (N+1).δt which is dependant upon x(N.δt) and x((N+1).δt). Without loss of generality, we set N = 0 and find

$$wf(t) = \sum_{n=-\infty}^{-1} \mathrm{sinc}((t-n.\delta t).\pi/\delta t).x(n.\delta t) + \mathrm{sinc}(\pi.t/\delta t).x(0)$$

$$+ \mathrm{sinc}((t-\delta t).\pi/\delta t).x(\delta t) + \sum_{n=2}^{\infty} \mathrm{sinc}((t-n.\delta t).\pi/\delta t).x(n.\delta t)$$

Now, all x(n.δt)'s are independent Gaussian distributed random variables (rvs) with variance var[x(n.δt)] = v_x. We therefore form the purely random component of wf(t) between n.δt and (n+1).δt as

$$noise(t) = \sum_{n=-\infty}^{-1} \mathrm{sinc}((t-n.\delta t).\pi/\delta t).x(n.\delta t) + \sum_{n=2}^{\infty} \mathrm{sinc}((t-n.\delta t).\pi/\delta t).x(n.\delta t)$$

and work out it's variance

$$\text{var}[\text{noise}(t)] = \text{var}[\ \Sigma + \Sigma\ ]$$

$$= \text{var}[\ \Sigma\ ] + \text{var}[\ \Sigma\ ]$$

$$= \Sigma\ \text{var}[\,] + \Sigma\ \text{var}[\,]$$

$$= v_{rv}.\{\ \sum_{n=-\infty}^{-1} (\text{sinc}((t-n.\delta t).n/\delta t))^2 + \sum_{n=2}^{\infty} (\text{sinc}((t-n.\delta t).n/\delta t))^2\ \}.$$

where v_{rv} is an rv.

So, wf(t) between t= M.δt and (M+1).δt is described by the sum of three Gaussianly distributed random variables (rv's), each multiplied by a time dependant function.

It is not known by the author how to take the analysis any further at this point since the equation for the waveform now has three independent rvs in it along with some quite intractable equations. Clearly in order to solve these equations, the thing to do is to integrate over all possible waveforms in this set and find the proportion that exhibit the characteristic under consideration. This has not been carried out although it is expected that the result would agree with Rice [1946].

Appendix F.

Generation of Gaussian Distributed Noise

In order to generate noise with a Gaussian distribution on a computer, we can make use of commonly available routines that generate random numbers with a flat distribution on (0,1). The way in which we can do this is as follows.

Consider two independent Gaussianly distributed random variables (rv's), x and y, and consider them to correspond to two-dimensional rectangular co-ordinates. Working out their distribution in terms of polar co-ordinates r,θ we put

$$x = r.\cos\theta \qquad \text{(F1a)}$$

$$y = r.\sin\theta \qquad \text{(F1b)}$$

Now, consider the elemental area on the xy plane, dx.dy, and the volume underneath it. Writing the probability distribution function (pdf) as f(x,y) we get

$$f(x,y).dx.dy = f(r,\theta).dr.d\theta$$

and from the independence and Gaussian statistics of x and y

$$f(x,y) = \frac{\exp(-(x^2+y^2)/(2.\sigma^2))}{2.\pi.\sigma^2} = \frac{\exp(-r^2/(2.\sigma^2))}{2.\pi.\sigma^2} \qquad \text{(F2)}$$

where

σ^2 variance of x and y

$r^2 = x^2 + y^2$

Transforming the differential area we have

$$dx.dy = r.dr.d\theta$$

and so from (F2) we get

$$f(r,\theta).dr.d\theta = \frac{r.\exp(-r^2/(2.\sigma^2)).dr.d\theta}{2.\pi.\sigma^2} \qquad (F3)$$

and

$$f(r,\theta) = \frac{r.\exp(-r^2/(2.\sigma^2))}{2.\pi.\sigma^2}$$

and clearly, this is independant of θ. In other words, θ is flatly distributed on 0 -> 2.π . We can therefore integrate out the angular variable

$$f(r) = \int_0^{2\pi} f(r,\theta).d\theta$$

$$= r.\exp(-r^2/(2.\sigma^2))/\sigma^2 \qquad (F4)$$

This expression (F4) now gives us the pdf of r. It is the well known Rayleigh amplitude distribution [Schwartz 1980 pp379]. We can work out the cumulative distribution by integrating the pdf from 0 to R thus

$$cdf(R) = \int_0^R f(r).dr$$

$$= \int_0^R r.\exp(-r^2/(2.\sigma^2))/\sigma^2.dr$$

$$= \left[-\exp(r^2/(2.\sigma^2)) \right]_0^R$$

$$= 1 - \exp(-R^2/(2.\sigma^2)) \qquad (F5)$$

This is a proper cdf since at R = 0 it has value 0 and as R -> ∞ it goes to 1. Hence the pdf is a proper one also. If we take a randomly distributed point on (0,1), say B, and let it represent a value of the cdf, we can solve (F5) in terms of R. R will then have a Rayleigh amplitude distribution. If we also generate a random phase using a rv on (0,1), say A, we can use (F1a) or (F1b) to give a Gaussianly distributed rv. Thus we put

$$B = 1-\exp(-r^2/(2.\sigma^2))$$

$$\ln(1-B) = -r^2/(2.\sigma^2)$$

and therefore

$$r = \sqrt{(-2.\sigma^2.\ln(1-B))}$$

$$= \sigma.\sqrt{(-2.\ln(1-B))}$$

and, using (F1a)

$$x = r.\cos\theta$$

$$= \sigma.\sqrt{(-2.\ln(1-B))}.\cos(2.\pi.A)$$

Since B is evenly distributed on (0,1), so is 1-B. Thus we have a formula to give numbers with a Gaussian distribution from two flat distributed ones. By use of (F1b) we can also get a second, independent rv.

This method of generating noise with a Gaussian distribution has been used extensively in the tests in this thesis.

Appendix G.

Higher Order Crossings - Theoretical Study

A general formula for the number of zerocrossings of any order can be found in terms of the parameters σ, the standard deviation of the noise, w, the frequency component present and ß, the correlation coefficient. The general notation D(σ,w,ß) has been used here for this value, noting that σ now refers to the standard deviation of the noise in the series under consideration. The probability of a zerocrossing in g(n), the general series, is found by using the bivariate normal distribution to give the probability that the first sample is positive and the second negative and then integrating this expression over all possible phases of g(n), thus

$$D(\sigma,w,\beta) = \int_{\theta=-\pi}^{\theta=\pi} \frac{1}{2(\pi.\sigma)^2\sqrt{(1-\beta^2)}} \int_{A.\sin(w.n+\theta)}^{\infty} \int_{-\infty}^{A.\sin(w.(n-1)+\theta)} \exp(-(x^2-2.\beta.x.y+y^2)/(2.\sigma^2(1-\beta^2)))].dx.dy.d\theta$$

or, setting n=0, we can write

$$D(\sigma,w,\beta) = \frac{1}{2(\pi.\sigma)^2\sqrt{(1-\beta^2)}} \int_{-\pi}^{\pi} \int_{A.\sin(\theta)}^{\infty} \int_{-\infty}^{A.\sin(\theta-w)} \exp\left[\frac{-(x^2-2\beta xy+y^2)}{2\sigma^2(1-\beta^2)}\right].dx.dy.d\theta \quad (G1)$$

In order to solve this equation, it is noted that, due to periodicity with θ, D(σ,w,ß) is the 'dc' value of the function

$$F(\sigma,w,\beta:\theta) = \frac{1}{\pi.\sigma^2\sqrt{(1-\beta^2)}} \int_{A.\sin(\theta)}^{\infty} \int_{-\infty}^{A.\sin(\theta-w)} \exp\left[\frac{-x^2-2\beta xy+y^2}{2\sigma^2(1-\beta^2)}\right].dx.dy \quad (G2)$$

By making the substitutions u=x/(√2.σ) and v=y/(√2.σ) we can write this as

$$P(\sigma,w,\beta:\theta) = \frac{2}{\pi.\sqrt{(1-\beta^2)}} \int_{B}^{\infty} \int_{-\infty}^{C} \exp\left[\frac{-(u^2-2\beta uv+v^2)}{(1-\beta^2)}\right] .du.dv \qquad (G3)$$

where

$$B = D.\sin(\theta),\ C = D.\sin(\theta-w),\quad D = A/(\sqrt{2}.\sigma)$$

This formula is in almost the same form as [Erdélyi, Magnus, Oberhettinger and Triconi 19XX, eqn.22., Mehlers formula]

$$\sum_{n=0}^{\infty} \frac{(w/2)^n.H_n(u).H_n(v)}{n!} = \exp\left[\frac{2uvw - (u^2+v^2).w^2}{(1-w^2)}\right]/\sqrt{(1-w^2)} \qquad (G4)$$

where $H_n(x)$ are the Hermite polynomials. Here we let w=β and put (G4) in the same form as the exponential in (G3)

$$\exp\left[\frac{-(u^2-2\beta uv+v^2)}{(1-\beta^2)}\right] = \sqrt{(1-\beta^2)}.\exp(-(u^2+v^2)).\sum_{n=0}^{\infty} \frac{(\beta/2)^n.H_n(u).H_n(v)}{n!} \qquad (G5)$$

Hence (G3) and (G5) give

$$P(\sigma,w,\beta:\theta) = \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(\beta/2)^n}{n!} \int_{B}^{\infty} \exp(-u^2).H_n(u).du \int_{-\infty}^{C} \exp(-v^2).H_n(v).dv \qquad (G6)$$

Now [Abramowitz & Stegun 1965 22.13.15]

$$\int_{0}^{x} H_n(t).\exp(-t^2).dt = H_{n-1}(0) - \exp(-x^2).H_{n-1}(x) \quad n>0 \qquad (G7a)$$

and

$$\int_0^x H_n(t).\exp(-t^2).dt = \int_0^x \exp(-t^2).dt = \sqrt{\pi}.\mathrm{erf}(x)/2 \qquad (G7b)$$

hence, noting that $\int_x^\infty = \int_0^\infty - \int_0^x$ and separating the n=0 term from (G6)

and readjusting the indices

$$F(\sigma,w,B:\theta) = \frac{[1-\mathrm{erf}(B)][1+\mathrm{erf}(C)]}{2}$$

$$- \frac{2}{\pi} \sum_{n=0}^{\infty} \frac{(\theta/2)^{n-1}}{(n+1)!}.[\exp(-B^2).H_n(B)].[\exp(-C^2).H_n(C)] \qquad (G8)$$

Fourier series for erf(D.sin(θ)):

Noting that erf(D.sin(θ)) is an odd function of θ and that sin(π-θ) = sin(θ), erf(D.sin(θ)) can be expressed as a Fourier series

$$\mathrm{erf}(D.\sin(\theta)) = \sum_{r=0}^{\infty} p_r(D).\sin((2.r+1).\theta) \qquad (G9)$$

The 'dc' component of the first term on the right hand side (RHS) of (G8) is therefore

$$\frac{1- dc\{\ \mathrm{erf}(B).\mathrm{erf}(C)\ \}}{2}$$

$$= [1-dc\{ \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} p_r(D).p_s(D).\sin((2.r+1).\theta).\sin((2.s+1).(\theta-u)]/2$$

$$= [1 - \sum_{r=0}^{\infty} p_r^2(D).\cos((2.r+1).w)/2]/2 \qquad (G10)$$

Previous calculation (Appendix D) has shown that

$$p_r(D) = \frac{2.D}{(2.r+1).\sqrt{\pi}}.\exp\left[-\frac{D^2}{2}\right].\left[I_{r-1}(D^2/2) + I_r(D^2/2)\right]$$

and therefore

$$p_r^2(D) = \frac{4.D^2}{(2.r+1)^2.\pi}.\exp(-D^2).[I_{r-1}(D^2/2) + I_r(D^2/2)]^2 \qquad (G11)$$

Where $I_n(x)$ is the modified Bessel function of the first kind. Putting (G11) in (G10) gives

$$= \frac{1}{2} - \frac{D^2.\exp(-D^2)}{\pi} \sum_{r=0}^{\infty} \frac{[I_{r-1}(D^2/2) + I_r(D^2/2)]^2.\cos((2.r+1).w)}{(2.r+1)^2} \qquad (G12)$$

The dc component of the first term on the RMS of (G8) is now known and it remains to evaluate the dc term of the series.

Fourier series for $\exp(-D^2.\sin^2(\theta)).H_n(D.\sin(\theta))$:

In order to evaluate the series in (G8), consider the function

$$G_n(x) = \exp(-x^2).H_n(x) \qquad (G13)$$

Now the Hermite polynomials have the recurrence relationship [Abramowitz & Stegun 1965 22.7]

$$H_{n+1}(x) - 2.x.H_n(x) + 2.n.H_{n-1}(x) = 0 \qquad n \geq 1 \qquad (G14)$$

hence, from (19) and (20)

$$G_{n-1}(x) - 2.x.G_n(x) + 2.n.G_{n-1}(x) = 0 \qquad n \geq 1 \qquad \text{(G15)}$$

Further, $H_n(x)$ and hence $G_n(x)$ is an odd function of x if n is odd and an even function of x if n is even. Therefore $G_n(D.\sin(\theta))$ is an odd periodic function of θ in the range $[-\pi,\pi]$ if n is odd and an even periodic function of θ in the range $[-\pi,\pi]$ if n is even. Thus if n is odd = 2.r+1, r ≥ 0

$$G_n(D.\sin(\theta)) = G_{2.r+1}(D.\sin(\theta)) = \sum_{s=0}^{\infty} q_{r,s}.\sin((2.s+1).\theta) \qquad r \geq 0 \quad \text{(G16a)}$$

and similarly if n is even = 2.r r ≥ 0

$$G_n(D.\sin(\theta)) = G_{2.r}(D.\sin(\theta)) = \sum_{s=0}^{\infty} p_{r,s}.\cos(2.s.\theta) \qquad r \geq 0 \quad \text{(G16b)}$$

Thus if n = 2.r+1

$$G_n(D.\sin(\theta)).G_n(D.\sin(\theta+w)) =$$

$$\sum_{s1=0}^{\infty}\sum_{s2=0}^{\infty} q_{r,s1}.q_{r,s2}.\sin((2.s1+1).\theta).\sin((2.s2+1).(\theta+w))$$

$$= 1/2 \sum_{s1=0}^{\infty}\sum_{s2=0}^{\infty} q_{r,s1}.q_{r,s2}.[\cos\{\,2.(s2-s1).\theta - (2.s2+1).w\,\}$$

$$- \cos\{\,2.(s2+s1+1).\theta + (2.s2+1).w\,\}\,]$$

and the dc terms in this only occur in the first cos term when s1=s2

$$dc = 1/2 \sum_{s=0}^{\infty} q_{r,s}^2.\cos((2.s+1).w) \qquad n = 2.r+1,\ r \geq 0 \qquad \text{(G17a)}$$

If n = 2.r, by a similar process

$$dc = 1/2 \sum_{s=0}^{\infty} (1+\delta_{r,s}).p_{r,s}^{2}.\cos(2.s.w) \qquad n=2.r,\ r \geq 0 \qquad (G17b)$$

where $\delta_{r,s}$ is the Kronecker delta function and is 1 when r=1, 0 otherwise.

Thus from (G12) and (G17) the dc value of (G8), is D(σ,w,s) can be found. It remains now to determine the $p_{r,s}$ and $q_{r,s}$.

Fourier series for $G_n(x)$:

$G_0(x)$.

Since it is known that $H_0(x) = 1$, then $G_0(x) = \exp(-x^2)$ and here we have $x = D.\sin(\theta)$ and we know [Gray 1931 pp36]

$$\exp(-D^2.\sin^2(\theta)) = \exp(-D^2/2).\exp(\ D^2.\cos(2.\theta)/2\ )$$

$$= \exp(-D^2/2).(\ I_0(D^2/2) + 2.\sum_{r=1}^{\infty} I_r(D^2/2).\cos(2.r.\theta)\ ) \qquad (G18)$$

and this gives us the Fourier series for $G_0(D.\sin(\theta))$ and from (G16b)

$$p_{0,s} = \exp(-D^2/2).I_0(D^2/2) \qquad (G19a)$$

$$p_{r,s} = 2.\exp(-D^2/2).I_r(D^2/2) \qquad s \geq 1 \qquad (G19b)$$

$G_1(x)$:

It is also known that $H_1(x) = 2.x$ and hence, using (G18)

$$G_1(D.\sin(\theta)) = 2.D.\sin(\theta).\exp(-D^2/2).\{I_0(D^2/2) + 2.\sum_{r=1}^{\infty} I_r(D^2/2).\cos(2.r.\theta)\}$$

which, after expanding the brackets, using a trigonometrical identity and collecting terms under the summation becomes

$$G_1(D.\sin(\theta)) = 2.D.\exp(-D^2/2).\sum_{s=0}^{\infty}\left[I_s(D^2/2) - I_{s+1}(D^2/2) \right].\sin((2.r+1).\theta)$$

Thus using (G18a)

$$q_{n,s} = 2.D.\exp(-D^2/2).[\ I_s(D^2/2) - I_{s+1}(D^2/2)\ ] \qquad s \geq 0 \qquad (G20)$$

$G_2(x)$:

It is also known that $H_2(x) = (4.x^2-2)$ and hence

$$G2(D.\sin(\theta)) = \exp(D^2/2).(4.D^2.\sin^2(\theta)-2).\{\ I_0(D^2/2) + 2.\sum_{r=1}^{\infty} I_r(D^2/2).\cos(2.r.\theta)\ \}$$

multiplying out the brackets and applying similar procedures as before, this becomes

$$G2(D.\sin(\theta)) = \sum_{s=0}^{\infty} p1,s.\cos(2.s.\theta)$$

$$+ 2.\exp(-D^2/2).\{ (D^2-1).I_0(D^2/2) - D^2.I_1(D^2/2)$$

$$+ \sum_{r=1}^{\infty} \cos(2.r.\theta)*[\ 2.(D^2-1).I_r(D^2/2) - D^2(I_{r-1}(D^2/2) + I_{r+1}(D^2/2))\ ]\ \}$$

giving

$$p_{1,0} = 2.\exp(-D^2/2).\{ (D^2-1).I_0(D^2/2) - D^2.I_1(D^2/2)\ \} \qquad (G21)$$

$$p_{1,s} = 2.\exp(-D^2/2).\{\ 2.(D^2-1).I_s(D^2/2)$$

$$- D^2.(I_{s-1}(D^2/2) + I_{s+1}(D^2/2))\ \} \qquad s \geq 1 \qquad (G22)$$

$G_m(x)$:

In order to find the rest of the $p_{r,s}$ and $q_{r,s}$ we can use the recurrence relationships (G15) and expansions (G16) for $G_m(x)$.

$G_m(x)$, n even:

Consider the cases when n is even ($n = 2.r$) and ≥ 2 ($r \geq 1$), from (G15)

$$G_{m-1}(D.\sin(\theta)) = 2.D.\sin(\theta).G_m(D.\sin(\theta)) - 2.m.G_{m-1}(D.\sin(\theta))$$

$$G_{2.r-1}(D.\sin(\theta)) = 2.D.\sin(\theta).G_{2.r}(D.\sin(\theta)) - 4.r.G_{2.r+1}(D.\sin(\theta))$$

$$= 2.D.\sin(\theta).\sum_{s=0}^{\infty} p_{r,s}.\cos(2.s.\theta) - 4.r.\sum_{s=0}^{\infty} q_{r-1,s}.\sin((2.s+1).\theta)$$

$$= D.\sum_{s=0}^{\infty} p_{r,s}.\sin((2.s+1).\theta) - D.\sum_{s=0}^{\infty} p_{r,s}.\sin((2.s-1).\theta)$$

$$- 4.r.\sum_{s=0}^{\infty} q_{r-1,s}.\sin((2.s+1).\theta)$$

Separating out the first term of the second sum and readjusting the indi-

cas we get

$$= D.\sum_{s=0}^{\infty} p_{r,s}.\sin((2.s+1).\theta) - D.\sum_{s=0}^{\infty} p_{r,s-1}.\sin((2.s+1).\theta)$$

$$- 4.r.\sum_{s=0}^{\infty} q_{r-1,s}.\sin((2.s+1).\theta) + D.p_{r,0}.\sin(\theta)$$

$$= \sum_{s=0}^{\infty} (D.(p_{r,s} - p_{r,s-1}) - 4.r.q_{r-1,s}) \ . \ \sin((2.s+1).\theta)$$

$$+ D.p_{r,0}.\sin(\theta)$$

expanding the left hand side (LHS) using (22a) and equating terms

$$G_{s,r-1}(D.\sin(\theta)) = \sum_{s=0}^{\infty} q_{r,s}.\sin((2.s+1).\theta)$$

$$q_{r,0} = D.(2.p_{r,0} - p_{r,1}) - 4.r.q_{r-1,0} \qquad r \geq 1 \qquad \text{(G23)}$$

$$q_{r,s} = D.(p_{r,s} - p_{r,s-1}) - 4.r.q_{r-1,s} \qquad r \geq 1, \quad s \geq 1 \quad \text{(G24)}$$

<u>$G_n(x)$, n odd:</u>

Now the case for n odd ($n = 2.r + 1$) and ≥ 3 ($r \geq 1$) must be considered.

$$G_{n-1}(D.\sin(\theta)) = 2.D.\sin(\theta).G_n(D.\sin(\theta)) - 2.n.G_{n-1}(D.\sin(\theta))$$

$$G_{2(r+1)}(D.\sin(\theta)) = 2.D.\sin(\theta).G_{2r+1}(D.\sin(\theta)) - 2(2.r+1).G_{2r}(D.\sin(\theta))$$

$$= 2.D.\sin(\theta).\sum_{s=0}^{\infty} q_{r,s}.\sin((2.s+1).\theta) - 2(2.r+1).\sum_{s=0}^{\infty} p_{r,s}.\cos(2.s.\theta)$$

$$= D.\sum_{s=0}^{\infty} q_{r,s}.\cos(2.s.\theta) - D.\sum_{s=0}^{\infty} q_{r,s}.\cos(2(s+1).\theta)$$

$$- 2(2r+1).\sum_{s=0}^{\infty} p_{r,s}.\cos(2.s.\theta)$$

$$= D.\sum_{s=0}^{\infty} q_{r,s}.\cos(2.s.\theta) - D.\sum_{s=1}^{\infty} q_{r,s-1}.\cos(2.s.\theta)$$

$$- 2(2r+1).\sum_{s=0}^{\infty} p_{r,s}.\cos(2.s.\theta)$$

$$= \sum_{s=1}^{\infty} \cos(2.s.\theta).\{ D.(q_{r,s} - q_{r,s-1}) - 2(2r+1).p_{r,s} \}$$

$$+ D.q_{r,0} - 2(2r+1).p_{r,0}$$

expanding the LHS using (G16b) and equating terms

$$G_{n,(s+1)}(D.\sin(\theta)) = \sum_{s=0}^{\infty} p_{r+1,s}.\cos(2.s.\theta)$$

$$p_{r+1,0} = D.q_{r,0} - 2(2r+1).p_{r,0} \qquad r \geq 1 \qquad \text{(G25)}$$

$$p_{r+1,s} = D.(q_{r,s} - q_{r,s-1}) - 2(2r+1).p_{r,s} \qquad r \geq 1,\ s \geq 1 \qquad \text{(G26)}$$

Equation (G8) can now be rewritten as $D(\sigma,w,\theta)$ by taking the dc parts of each term using equations (G12), (G17a) and (G17b). The summation over n must be split into the odd and even components. This gives

$$D(\sigma,w,\theta) = \frac{1}{2} - \frac{D^2.\exp(-D^2)}{\pi} \sum_{r=0}^{\infty} \frac{[I_{r+1}(D^2/2) + I_r(D^2/2)]^2.\cos((2.r+1).w)}{(2.r+1)^2}$$

$$- \frac{1}{\pi}.\left[\sum_{r=0}^{\infty} \frac{(\theta/2)^{2r-s}}{(2r+2)!}. \sum_{s=0}^{\infty} q^2_{r,s}.\cos((2.s+1).w) \right.$$

$$\left. + \sum_{r=0}^{\infty} \frac{(\theta/2)^{2r-1}}{(2r+1)!}.\left[\sum_{s=0}^{\infty} p^2_{r,s}.\cos(2.s.w) + p^2_{r,0} \right] \right] \quad (G27)$$

and the $p_{r,s}$ and $q_{r,s}$ can be determined by use of

$$p_{0,s} = \exp(-D^2/2).I_s(D^2/2) \quad (G19a)$$

$$p_{0,s} = 2.\exp(-D^2/2).I_s(D^2/2) \qquad s \geq 1 \quad (G19b)$$

$$q_{0,s} = 2.D.\exp(-D^2/2).[\ I_s(D^2/2) - I_{s-1}(D^2/2)\ ] \qquad s \geq 0 \quad (G20)$$

$$p_{1,s} = 2.\exp(-D^2/2).\{\ (D^2-1).I_s(D^2/2) - D^2.I_{s-1}(D^2/2)\ \} \quad (G21)$$

$$p_{1,s} = 2.\exp(-D^2/2).[\ 2.(D^2-1).I_s(D^2/2)$$

$$- D^2.(I_{s-1}(D^2/2) + I_{s+1}(D^2/2))\ \} \qquad s \geq 1 \quad (G22)$$

$$q_{r,s} = D.(2.p_{r,s} - p_{r,s-1}) - 4.r.q_{r-1,s} \qquad r \geq 1 \quad (G23)$$

$$q_{r,s} = D.(p_{r,s} - p_{r,s-1}) - 4.r.q_{r-1,s} \qquad r \geq 1,\ s \geq 0 \quad (G24)$$

$$p_{r-1,s} = D.q_{r,s} - 2(2r+1).p_{r,s} \qquad r \geq 1 \quad (G25)$$

$$p_{r-1,s} = D.(q_{r,s} - q_{r,s-1}) - 2(2r+1).p_{r,s} \qquad r \geq 1,\ s \geq 1 \quad (G26)$$

These formula must be calculated in the correct order. That is

factor	limits	formula
$P_{0,0}$		(G19a)
$P_{0,n}$	all $n \geq 0$	(G19b)
$Q_{0,n}$	all $n \geq 0$	(G20)
$P_{1,0}$		(G21)
$P_{1,n}$	all $n \geq 1$	(G22)
$Q_{1,0}$		(G23)
$Q_{1,n}$	all $n \geq 1$	(G24)
$P_{2,0}$		(G25)
$P_{2,n}$	all $n \geq 1$	(G26)
$Q_{2,0}$		(G23)
$Q_{2,n}$	all $n \geq 1$	(G24)

and so on.

Appendix H

Probability of a Higher Order Crossing - Numerical Approach

In chapter 3 a general expression for the probability of a higher order crossing was found and it is repeated here for reference.

$$D(n,\sigma,w) = \frac{1}{2(\pi.\sigma)^2\sqrt{(1-\delta^2)}} \int_{-\pi}^{\pi} \int_{A.\sin(\theta)}^{\infty} \exp\left[\frac{-(x^2-2\delta xy+y^2)}{2\sigma^2(1-\delta^2)}\right] .dx.dy.d\theta \quad (3.11)$$

The method of solution that is presented here is to use standard NAG routines on a mainframe computer. This is done by first applying a change of variable in order to get the equation in a form suitable for numerical integration. We put

$$z = \frac{x-\delta y}{\sigma\sqrt{(1-\delta^2)}}$$

which gives us

$$D(n,\sigma,w) = \frac{1}{\sqrt{2}.\pi^{3/2}\sigma} \int_{-\pi}^{\pi} \int_{A.\sin(\theta)}^{\infty} \exp(-z^2/\sigma^2).\mathrm{erf}(A.\sin(\theta-w)).dz.d\theta \quad (H1)$$

which is a double integral of standard functions and can be calculated using 2 dimensional modified Gaussian integration. The Fortran program to do this is presented below since it is thought that it is of fairly general application. The program is written in the form of a function that can be called from a main program. The parameters needed to call it are

N, the order

FREQ, the frequency as a fraction of the sampling rate

SNR, the signal to noise ratio in dB

and IFAIL, an indicator of whether the program has managed to converge to the answer within the requested precision.

The function returns the probability of a zerocrossing occurring between two samples.

```
      DOUBLE PRECISION FUNCTION ED(N,FREQ,SNR,IFAIL)
      DOUBLE PRECISION FREQ,SNR
      INTEGER N,IFAIL
      DOUBLE PRECISION W,A,SIGMA,BETA,ABSACC,ANS,THMIN,THMAX,PI,DUM
      DOUBLE PRECISION YMAX,C,DENOM,ENEG,X02AEF,SUM,FACT,M,FREQ1
      COMMON /PARMS/ W,A,SIGMA,BETA,YMAX,C,DENOM
      INTEGER NPTS,Q
      EXTERNAL PHI1,PHI2,F

      IF(N.LT.0) THEN
            FREQ1=1.0D0-FREQ
            Q=-N
            ED=1.0D0-ED(Q,FREQ1,SNR,IFAIL)
            RETURN
      END IF

      PI=4.0D0*DATAN(1.0D0)
      ENEG=X02AEF(DUM)
      W=PI*FREQ
      A=(2.0D0*DSIN(W/2.0D0))**N
      M=N
      BETA= -M/(M+1)

      SUM=0.0D0
      DO 99 I=0,M,1
            SUM=SUM+(FACT(M)/(FACT(I)*FACT(M-I)))**2
99    CONTINUE

      SIGMA=10.0D0**(SNR/10.0D0)
      SIGMA=DSQRT(0.5D0/SIGMA)
      SIGMA=SIGMA*DSQRT(SUM)
      ABSACC=0.000001
      THMIN=-PI
      THMAX=PI
      YMAX=SQRT(-ENEG*2.0D0)*SIGMA
      C=SQRT(2.0D0*PI)/(2.0D0*SIGMA*PI**2)
      DENOM=SIGMA*SQRT(1.0D0-BETA**2)
      IFAIL=1
      CALL D01DAF(THMIN,THMAX,PHI1,PHI2,F,ABSACC,ANS,NPTS,IFAIL)
      ED=ANS
      RETURN
      END
      DOUBLE PRECISION FUNCTION PHI1(THETA)
      DOUBLE PRECISION THETA,W,A,SIGMA,BETA,YMAX,C,DENOM
      COMMON /PARMS/ W,A,SIGMA,BETA,YMAX,C,DENOM
      PHI1=A*DSIN(THETA)
      IF(PHI1.GT.YMAX) PHI1=YMAX
      RETURN
      END
```

```
      DOUBLE PRECISION FUNCTION PHI2(THETA)
      DOUBLE PRECISION THETA,W,A,SIGMA,BETA,YMAX,C,DENOM
      COMMON /PARMS/ W,A,SIGMA,BETA,YMAX,C,DENOM
      PHI2=YMAX
      RETURN
      END
      DOUBLE PRECISION FUNCTION F(Y,THETA)
      DOUBLE PRECISION Y,THETA,W,A,SIGMA,BETA,YMAX,C,DENOM,V1,S15ABF
      COMMON /PARMS/ W,A,SIGMA,BETA,YMAX,C,DENOM
      V1=(A*DSIN(THETA-W)-BETA*Y)/DENOM
      F=C*DEXP(-(Y/SIGMA)**2*0.5D0)*S15ABF(V1,IFAIL)
      RETURN
      END
      DOUBLE PRECISION FUNCTION FACT(N)
      DOUBLE PRECISION FACT
      INTEGER N
      IF(N.GT.1) THEN
            FACT=N*FACT(N-1)
      ELSE
            FACT=1.0D0
      END IF
      RETURN
      END
```

This program has been found to converge to an accuracy of 1 in 10^6 for SNRs up to 5dB. The results have been checked against results from a simulation done over 10^6 samples and seem to be reliable. A further modification is required for convergence at higher SNRs.

Appendix I

A Program in C to Simulate An MFSK Signal Arriving Over A Rayleigh/Ricean Fading Channel

```
/* THIS PROGRAMME SIMULATES A RAYLEIGH/RICEAN FADING CHANNEL. */
/* IT SIMULATES A GROUND WAVE AND UP TO 4 INDEPENDENT SKYWAVES. */
/* FOR EACH SKYWAVE CHANNEL YOU CAN SPECIFY THE TIME DELAY WITH */
/* RESPECT TO THE GROUND WAVE, ITS RELATIVE MEAN AMPLITUDE, A FIXED */
/* DOPPLER SHIFT AND THE MEAN FADE RATE. */

/* THIS PARTICULAR VERSION SIMULATES AN MFSK SIGNAL BEING RECEIVED. */
/* THE SECTIONS THAT NEED TO BE MODIFIED FOR DIFFERENT MODULATION */
/* SCHEMES HAVE BEEN MARKED. */

/*************************************************************************/
/* Normal C include files                                                */
/*************************************************************************/

#include <math.h>
#include <stdio.h>

/*************************************************************************/
/* Definitions for Rayleigh fading generator                             */
/*************************************************************************/

#define N 20          /* No. of elements in filter for Rayleigh fade */
#define MM 6.0        /* No. of standard deviations covered by filter */

/*************************************************************************/
/* Definitions required for MFSK modulator                               */
/*************************************************************************/

#define M 8           /* No of MFSK tones */
#define G 4           /* Frequency guard space */
#define LB 700.0      /* Lower bandwidth edge */
#define UB 2600.0     /* Upper bandwidth edge */
#define BW (UB-LB)    /* Available bandwidth */
#define DT 1.6e-4     /* sampling period */

/*************************************************************************/
/* Start of main                                                         */
/*************************************************************************/

main()
{

/*************************************************************************/
/* Variables required by channel simulator                               */
/*************************************************************************/

     float pi = 4*atan(1.0);  /* pi */
     float twopi = 2.0*pi;    /* 2*pi */
     float noise();           /* noise subroutine */
     float dBtoR();           /* converts dB to a magnitude ratio */
     float t = 0.0;           /* time */
     float rxwf=0.0;          /* value of received waveform at t */
     float a;                 /* dummy */
```

```
    float mag=0.0;      /* sum of filter constant magnitudes squared */
    float sdns;         /* sd of additive Gaussian noise */
    float grndatten;    /* groundwave fixed attenuation */
    float skyt;         /* time with delay */
    float ns;           /* I chan noise component */
    float nc;           /* Q chan noise component */
    int skywave;        /* index of skywave under consideration */
    int nuskywave;      /* total number of skywaves */
    int data;           /* transmitted data */
    int dataS = 0;      /* transmitted data at t - PTD of skywave*/
    int dataL = 0;      /* previous tx data */
    int dataN = 0;      /* one before that tx data */
    int b,j;            /* dummy */
    int syms;           /* no. of syms so far */
    int totsyms;        /* total no. of syms */
    long dts=0;         /* number of samples so far */

/*************************************************************************/
/* The following arrays store the parameters for the various skywaves */
/* Their size may be changed if more skywaves are required            */
/*************************************************************************/

    float df[4];        /* Doppler shifts in skywave components */
    float ptd[4];       /* time delay in skywave wrt groundwave */
    float attn[4];      /* fixed attenuation in skywave components */
    float totatt[4];    /* total attenuation ie attn[]*raleigh fade */
    float phaze[4];     /* phase shift due to raleigh fading */
    float dt[4];        /* I & Q chan noise filter shift rates  */
    float dtnc[4];      /* time of next change to filter outputs */
    float filtconst[N+1];    /* constants for FIR digital filter */
    float randI[4][N+1];     /* FIR filter delay line, I channel */
    float randQ[4][N+1];     /* FIR filter delay line, Q channel */
    int point[4];       /* pointers to start of filter arrays */

/*************************************************************************/
/* These next variables are used by the modulator                        */
/*************************************************************************/

    float te;           /* temporary time variable for symbol period*/
    float freq[M];      /* symbol frequencies */
    float divs;         /* symbol spacing in frequency */
    float lb;           /* lower bandwidth cutoff point */
    int samples;        /* number of samples per symbol */

/*************************************************************************/
/*   Work out tone frequencies for MFSK format modulator                 */
/*************************************************************************/

    te=(M-1.0*G.0)/BW;          /* te would be this */
    te=te/DT;                   /* or this no of samples */
    samples=(int)(te+1.0);      /* a greater integer */
calc:
    te=(float)(samples)*DT;     /* so te is this now */
    divs=1.0/te;                /* so divs becomes this */
    lb=divs*((int)(1.0*LB/divs));       /* recalc new lb to even out No
of wfs per symbol */
    if((lb+divs*(M-1.0*G.0))>UB)  /* if signal does not fit in BW.... */
    {
        samples++;      /* make element a little longer... */
        goto calc;      /* and try again */
    }
```

```
/*************************************************************************/
/* Now calculate the tone frequencies                                    */
/*************************************************************************/

     for(j=0;j<M;j++)              /* we already know where to start */
     {
          freq[j]=divs*(j+G.0/2.0)+lb;  /* and increment size */
     }

/*************************************************************************/
/* Next section is general for channel simulator                         */
/*************************************************************************/

/*************************************************************************/
/* Read in parameters etc                                                */
/* All signal and noise levels are in dB rel. to 0.5watt                 */
/*************************************************************************/

     scanf("%d",&totsyms);     /* total no. of symbols to be simulated */
     scanf("%f",&a);           /* groundwave attenuation in dB */
     grndatten = dBtoR(a);     /* convert to v ratio */
     scanf("%f",&a);           /* read noise level in dB */
     sdoa = sqrt(0.5)*dBtoR(a);    /* calc sd of noise  samples */
     scanf("%d",&nuskywave);  /* read number of skywaves to sim */
     for(skywave=0;skywave<nuskywave;skywave++)
     {
          scanf("%f",&a);      /* this skywave atten dB */
          atten[skywave] = dBtoR(a);     /* scale to power of tx signal */
          scanf("%f",&df[skywave]);      /* Doppler shift */
          scanf("%f",&ptd[skywave]);     /* time delay wrt ground wave */
          scanf("%f",&a);                /* fade rate =frms*sqrt(2.ln2) */

/*************************************************************************/
/* For this eqn, see Ralphs, "An HF channel simulator using a new        */
/* Rayleigh fading method" The Radio and Electronic engineer. Vol 46     */
/* No 12 pp 579-587 Dec 1976 eqn 11. We calculate the noise filter       */
/* shift rate in the same way                                            */
/*************************************************************************/

          dt[skywave]=sqrt(pi*log(2.0))*MN/(N.0*2.0*sqrt(2.0)*pi*a);
     }

/*************************************************************************/
/* Init Gaussian PSD filters used for I and Q channel random walks       */
/* FIR filter has N element delay line, covers MN sds and has            */
/* Gaussian impulse response                                             */
/*************************************************************************/

     for(b=0;b<M;b++)
     {
          filtconst[b]=exp(-0.5*pow( (((float)b-M.0/2.0)*MN/M.0) , 2.0));

/*************************************************************************/
/* Sum squared constants for normalisation                               */
/*************************************************************************/

          msq+=filtconst[b]*filtconst[b];

/*************************************************************************/
/* Initialise delay line at the same time. Each element is a             */
```

```
/* Gaussianly distributed number and there are I and Q filters        */
/*********************************************************************/

          for(skywave=0;skywave<nuskywave;skywave++)
          {
               randI[skywave][b] = noise(1.0);
               randQ[skywave][b] = noise(1.0);
          }
     }

/*********************************************************************/
/* Want unity power gain and so take root of sum of mags**2 and scale */
/*********************************************************************/

     mag=sqrt(mag);
     for(b=0;b<N;b++)
     {
          filtconst[b]/=mag;
     }

/*********************************************************************/
/* Initialize filters by doing following:-                           */
/* Set pointer to start of delay line array                          */
/* Set up time for next update                                       */
/* Shift data along by moving pointer                                */
/* Input new random values to both channel filters                   */
/*********************************************************************/

     for(skywave=0;skywave<nuskywave;skywave++)
     {
          point[skywave]=0;
          dtnc[skywave]+=dt[skywave];
          point[skywave]=(point[skywave]+1)%N;
          randI[skywave][point[skywave]]=noise(1.0);
          randQ[skywave][point[skywave]]=noise(1.0);

/*********************************************************************/
/* Now calculate first filter output by FIR procedure                */
/*********************************************************************/

          ns=0.0;
          nc=0.0;
          for(b=0;b<N;b++)
          {
               ns+=filtconst[b]*randI[skywave][(point[skywave]+b)%N];
               nc+=filtconst[b]*randQ[skywave][(point[skywave]+b)%N];
          }

/*********************************************************************/
/* Calculate total attenuation as fixed atten times that due to fade  */
/* Also phase of fading signal                                       */
/*********************************************************************/

          totatt[skywave]=atten[skywave]*sqrt(0.5*(ns*ns+nc*nc));
          phaze[skywave]=atan2(ns,nc);
     }    /* next skywave */

/*********************************************************************/
/* The main body of the simulation starts here.                      */
/*********************************************************************/
```

```
    for(syms = 1;syms<= totsyms;syms++)     /* Loop for symbols */
    {

/*************************************************************************/
/* Read data. This is in octal here since we have an 8-PSK system        */
/*************************************************************************/

        scanf("%d",&data);

        do                  /* loop for samples of symbol */
        {

/*************************************************************************/
/* For each skywave we must first check whether fading vector needs      */
/* updating and if so do it. Then work out skywave mag and phase.        */
/* Then work out the arrival time, see if there is intersymbol           */
/* interference and work out waveform value                              */
/*************************************************************************/

            for(skywave=0;skywave<nuskywave;skywave++)
            {

/*************************************************************************/
/* Check if filter needs updating and if so, do it                       */
/*************************************************************************/

                if(t>dtnc[skywave])
                {
                    /* fixup update time */
                    dtnc[skywave]+=dt[skywave];
                    /* shift data along delay line */
                    point[skywave]=(point[skywave]+1)%N;
                    /* ip new value to filter */
                    randI[skywave][point[skywave]]=noise(1.0);
                    randQ[skywave][point[skywave]]=noise(1.0);
                    ns=0.0;   /* reset */
                    nc=0.0;   /* reset */
                    /* calculate new op from filter */
                    for(b=0;b<N;b++)
                    {
                        ns+=filtconst[b]*
                            randI[skywave][(point[skywave]+b)%N];
                        nc+=filtconst[b]*
                            randQ[skywave][(point[skywave]+b)%N];
                    }

/*************************************************************************/
/* Calculate total attenuation of skywave                                */
/*************************************************************************/

            totatt[skywave]=atten[skywave]*sqrt(0.5*(ns*ns+nc*nc));

/*************************************************************************/
/* Calculate phase angle of vector                                       */
/*************************************************************************/

                    phase[skywave]=atan2(ns,nc);
                }   /* end of updating procedure */

/*************************************************************************/
/* Now we check if the time delay is such that we are still receiving */
```

```
/* the last symbol or the one before that. We do this so that we use  */
/* the appropriate data for generating the waveform. This is the      */
/* intersymbol interference simulator                                 */
/**********************************************************************/

                    skyt = t-ptd[skywave];
                    if(skyt < (syms-1)*samples*DT)
                    {
                         if(skyt<(syms-2)*samples*DT)
                         {
                              dataS=dataM;
                         }
                         else
                              dataS=dataL;
                    }
                    else
                    {
                         dataS=data;
                    }

/**********************************************************************/
/* Finally we can calculate the signal that is arriving from this     */
/* skywave path. We add this to any received signals that we have     */
/* already calculated.                                                */
/**********************************************************************/
/* This is the modulator and so needs modifying if a different type   */
/* is required                                                        */
/**********************************************************************/

                    rxwf+=totatt[skywave]*
               sin(twopi*(freq[dataS]+df[skywave])*skyt+phaze[skywave]);
               }    /* loop back for next skywave */

/**********************************************************************/
/* Now we have done all of the skywaves and so we must do the         */
/* groundwave component                                               */
/**********************************************************************/
/* Again, this is the modulator and would require tailoring to        */
/* particular needs.                                                  */
/**********************************************************************/

               rxwf+=grndatten*sin(twopi*freq[data]*t);

/**********************************************************************/
/* And add the noise at the appropriate level                         */
/**********************************************************************/

               rxwf+=noise(sdom);

/**********************************************************************/
/* So now we have the received signal and can go ahead and output it  */
/**********************************************************************/

               printf("%f\n",rxwf);

/**********************************************************************/
/* Reset the time variable. It is based on an int to avoid rounding   */
/* errors as we increment the counter millions of times in a          */
/* reasonable sized simulation                                        */
/**********************************************************************/
```

```
            dts++;
            t = dts*DT;
            rswf = 0.0;
        } while( dts%samples != 0 );  /* Loop for samples in symbol */

/*************************************************************************/
/* That is the end of a symbol and so the "previous symbol" stores       */
/* are updated and we return to the top.                                 */
/*************************************************************************/

        dataM = dataL;
        dataL = data;
    }    /* next symbol */

/*************************************************************************/
/* End of main.                                                          */
/*************************************************************************/

}    /* end of main */

/*************************************************************************/
/* Gaussian noise generation. This is done via a Rayleigh distribution*/
/*************************************************************************/

float noise(sd)
float sd;
{
    float temp;
    float hash;
    float limit = 2147483647;
    float twopi = 8.0*atan(1.0);

    temp = cos(twopi*random()/limit);
    hash = sqrt(-2*log(random()/limit))*temp;
    hash = sd * hash;
    return(hash);
}

/*************************************************************************/
/* Function to convert input power in dB to voltage ratios               */
/*************************************************************************/

float dBtoR(a)
float a;
{
    return( exp(a/(20*log10(exp(1.0)))));
}
/*************************************************************************/
/* End of program                                                        */
/*************************************************************************/
```