ADVANCED IONOSPHERIC CHIRPSOUNDING

Dissertation

Submitted in fulfilment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

of Rhodes University

by

ALLON WILLIAM VICTOR POOLE

November 1983

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Firstly, I would like to thank my colleagues Geoff Evans and John Fisher, whose skilful collaboration on this project has been indispensible. Both have accomplished their contributions in the most satisfactory way. John Fisher was responsible for the design and construction of the FFT analyser which still outperforms most commercially available devices. Geoff Evans has not only designed and constructed the controller for the new sounder, but has developed a comprehensive operating system which facilitates the input and editing of experimental programme schedules. I could not have wished for two more conscientious colleagues.

Secondly, I would like to thank Bill Wright, Klaus Bibl, and Dick Grubb, who are amongst the most informed people on this subject in the international sphere, and who took time off at international conferences to enlighten me about various aspects of digital sounding.

I should also like to thank my brother, Graham, Gerhard de Jager and Justin Jonas for fun-filled discussions and help with the complexities of Fourier transforms. Justin has also been of invaluable help with various programming problems on the mainframe computer. I also much appreciate the early help given to me by Clive Way-Jones, largely in introducing me to the behaviour of modern integrated circuitry.

Much of the final product of this research has been through the vehicle of the Rhodes University Computing Centre, and I would like to thank Mike Lawrie and his staff for willing help from time to time with the mainframe computer.

I should like to thank the Department of Transport and the Co-operative Scientific Programmes division of the Council for Scientific and Industrial Research for providing the finance which sustained this research. Furthermore, my thanks are due to Rhodes University for providing the opportunity.

Finally, and most importantly, I would like to thank my supervisor, Professor J.A. Gledhill, for the indulgent fashion in which he allowed this research to continue in his department when at times the possibility of completion must have seemed remote. I very much appreciate his patience in this regard.

Allon Poole

October, 1983.

ABSTRACT

This dissertation reports research into the theory and practical application of linear frequency modulated ionospheric sounding, as an alternative to the more usual technique of pulse modulation. A comparison of this technique with that of conventional pulse sounders is given, based on the concepts of matched filters and ambiguity functions for both modulations. A theory is developed to relate the group range and phase velocity of the ionospheric target to the phase and frequency of the difference signal at the receiver output. A method is then described whereby the group range and phase velocity of the reflection point as well as the amplitude, arrival angle and polarisation mode of the reflected energy can be measured. A description of the implementation of the technique is given together with some initial results. Finally, some suggestions for improvements are given.

CHAPTER 1

PROGRESS AND HISTORY OF IONOSONDES AND IONOSPHERE RESEARCH

From the turn of the century, when the ionosphere was merely a "conducting layer" postulated to explain the daily variation in the earth's magnetic field, until the present, ionospheric research has been interwoven with the development and evolution of the ionospheric sounder or "ionosonde". As this monograph concerns an advancement in the field of ionosondes, and the application of these new techniques to modern ionospheric problems, the tradition is maintained, and it is thus appropriate to review the history of this interdependence.

It was Balfour Stewart [1] who, in 1878, suggested that a conducting layer in the upper atmosphere could explain daily variations in the direction and intensity of the earth's magnetic field. At that time, long distance radio frequency propagation was as yet unknown, and it was only after Marconi [2,3] succeeded in communicating with radio signals across the Atlantic in 1901 that the idea was revived to explain the phenomenon. Independently, and at the same time, in 1902 Kennelly [4] in the U.S.A. and Heaviside [5] in England re-postulated the existence of a conducting layer. Heaviside further suggested that the conductivity of the region might be due to ions, possibly produced by solar radiation.

Although Marconi was not attempting to measure anything, he was using a a primitive radio frequency transmitter and receiver, and as his signals were propagating by means of ionospheric reflection, this event could be considered a first embryonic step in the evolution of the modern ionosonde.

Eccles [6], in 1912, provided the first theoretical background for propagation of radio waves through a medium containing charged particles, which was expanded in 1924 by Larmour [7]. The Eccles-Larmour theory was

later refined by Appleton, Hartree, Goldstein and others (Mitra [8]). However, up to 1925, although evidence was mounting in support of the conducting layer theory, no experimental proof was available. Much of the evidence centered around the observation of frequency-selective fading at receiving sites not co-located with the transmitter. In 1910, G.W. Pierce [9] predicted this phenomenon and correctly attributed it to destructive interference of the "direct" wave from the transmitter with the "indirect" wave reflected from the hitherto unproven reflecting layer. Lee de Forest [10] described experimental evidence for the fading and offered it in support of Pierce's prediction. In a private communication (now lost), Pierce ventured a calculation which evaluated the height of the reflector at the midpoint as 196 miles [16]. Although de Forest disagreed with this evaluation, and it remained unpublished, it was the first attempt to "measure" the height of the ionosphere.

In 1923, Smith-Rose and Barfield [11] drew attention to the two outstanding problems involved in the propagation of radio waves over the earth's surface:

(i) why was long distance propagation possible, and

(ii) what was the explanation for the large and rapid variation of the signal intensities observed?

Their initial experiments to demonstrate the downward passage of radio energy from a reflecting layer, based on a form of direction finding, were unsuccessful. Their technique was correct, but the radio frequencies that they worked at (33-115kHz) were too low in the presence of the highly conducting ground plane. Later experiments [12] in 1926, with frequencies in the range 600-1000 kHz, were successful, but not before Appleton and Barnett [13], in 1925, published their famous paper in which they not only demonstrated the downward passage of reflected energy, but also gave an

estimate of the angle of arrival of the energy (60 to 70 degrees from the horizontal), evidence that the reflected energy was elliptically polarised in confirmation of the Eccles-Larmour theory, an estimate of the reflection coefficient (0.2% to 6.0 %), and a minimum value for the electron density (10^5 cm^{-3}). They are thus justifiably credited with the first experimental demonstration of the existence of a reflecting layer. From the ionospheric sounding viewpoint, it is interesting to note that the experimental technique that they employed was not the one which in later years was to find favour with routine ionospheric investigators. Their technique was to vary continuously the transmitter frequency by a known amount (from 780 to 760 kHz) and observe maxima in the interference pattern between the ground and reflected waves. Their technique was thus essentially a phase comparison of the reflected energy with a local standard reference (the direct ray). As will be seen, this idea has regained popularity in recent times, but in the intervening half century most height estimation has been based on the method of Breit and Tuve [14].

In the year following the publication by Appleton and Barnett, Breit and Tuve [14], using a different method, published a paper confirming the existence of a relecting layer. Their method involved amplitude modulation of the carrier frequency, achieved by supplying the plates of the transmitter output tubes with an alternating voltage (425 Hz) and varying the grid voltage so that the transmitter only conducted for 1/3 to 1/5 of a cycle. This gave "humps" of radio energy of approximate length .5 to .8 ms, repeated every 2.3 ms. It was an improvement on the method of Appleton and Barnett, because it allowed not only the resolution of the echo from the ground wave, but also of individual echoes from the same parent pulse. This enabled them to make the observation that some fading occurred on echoes independently of the direct ray, and also to observe multiple reflections.

Tuve and Dahl [15] soon improved the technique by using a multivibrator to generate shorter, squarer pulses of length 150 μs at a pulse repetition frequency of 80 Hz. Tuve was later to express disappointment at the rate with which investigators took up the pulse sounding technique; nevertheless pulse sounding was to become the accepted method for generations of sounders, indeed, the most up-to-date sounders at the present time are still fundamentally pulse oriented devices.

It is not difficult to see why the pulse technique became so successful. Referring for the moment to the concept of aperture and resolving power, it has been noted by O.G. Villard [16] that Breit and Tuve, possibly unwittingly, were the first to apply increased aperture in the frequency domain in order to improve resolution in the time domain to resolve their multiple reflections. The reason for the success of the pulse technique is thus twofold: the pulse is a very easily synthesized broadband signal, given the technological limitations of the era until quite recently, and the interpretation of the returned echo bears most directly to our fundamental experience of space and time.

Regular pulse soundings of the ionosphere began at the National Bureau of Standards in the U.S.A. in 1929, where T.R. Gilliland [17] was responsible for the design of the first swept frequency ionosonde in May 1933. The frequency span of this ionosonde was small, from 2.5 to 4.4 MHz. It is of interest that his son, C.R. Gilliland, was involved in the construction of the first Barry Research "chirp" sounders, on which the subject of this monograph is based. Meanwhile, in England, manual sweep frequency soundings were begun just prior to 1932 and were reported by Appleton and Naismith [18].

Ionograms have been recorded on a regular basis at the Radio Research Station at Slough (now moved to Didcott, Oxfordshire, and called the

Rutherford-Appleton Laboratory) from 1935 until the present day.

The next step in the evolution of the pulse sounder was in 1946 when Sulzer [19] introduced a system of automatic tracking of the transmitter and receiver. By the advent of the International Geophysical Year in 1957, there were approximately 160 pulse sounders in operation worldwide. Villard [16] has produced a graph that plots the number of published articles concerning the ionosphere, against time, from 1920 to 1970. This graph shows a exponential growth of interest, starting with around 5 articles in 1925 to over 1000 in 1970, with the expected discontinuity during World War II.

Although pulse sounders have been the mainstay of ionospheric sounding equipment over the years, other methods have also been implemented. Of these one of the most interesting is the so-called "phase ionosonde" that is associated with the name of J.D. Whitehead [20,21]. This device compromises between the phase comparison technique of Appleton and Barnett which is more accurate for echo range determinations, and the pulse method of Breit and Tuve which supplies the necessary echo resolution. The main drawback seems to be that the extraction of the group range information from the now more complex recording format reguires 5 times the effort of normal scaling, so that it was really only of use in specialised experiments requiring the increased accuracy.

As will be shown in Chapter 2, increased range resolution is obtained by increasing the effective bandwidth of the sounding signal. In the case of simple pulses, this is achieved by shortening the pulse length and decreasing the rise time. Because of the need to maintain constant energy output and thus signal-to-noise ratio, this led to attempts to "compress" pulses by transmitting for a longer period, but phase coding the longer pulse in such a way that it would appear shorter after analysis. Suitable "codes" are those for which the autocorrelation function has a high value of

peak to sidelobe ratio and include the various Barker codes (binary) and Huffman codes (complex). A limitation on the length of the coded pulse, when the transmitter and receiver are co-located, is set by the necessity to cease transmitting before the return of the echo with minimum delay (approximately 667 μs for reflection from the E-region) to avoid desensitizing the receiver. This limits the ratio of the uncompressed to compressed pulse (compression ratio) to about 13. A list of references concerning coded pulse ionosondes is given by M.L. Phillips [22].

In the 1970's, a major development in the evolution of ionospheric sounders took place with the construction of what have become to be known as "digital" or "advanced" sounders. Two of the chief proponents of these devices are J.W. Wright [24] and K. Bibl [25] who have had a major hand in the development of the "Dynasonde" and "Digisonde" respectively. The chief characteristics of the digital sounder can be summarised as follows:

(i) control of the entire sounder function including frequency synthesis, frequency selection, antenna configuration and sounding schedule by computer,

(ii) numerical rather than analogue evaluation of the data.

The digital sounder is much more flexible in its application than any of its analogue predecessors. This is due to the increased flexibility of the numerical approach to the analysis over the restricted analogue equivalent, or in modern jargon, the flexibility of the "software" compared with the "hardware" approach. As an example, one can cite the problem of establishing the angle-of-arrival of an ionospheric echo. An analogue approach is described by R.H. Clarke [26]. In this paper, two spatially diverse antennas are connected through cables of unequal length to a summing circuit so that destructive interference at certain frequencies produce nulls in the ionogram trace. The frequency at which these nulls occur will

alter if the phase relationship of the two signals is altered by off-vertical reflection, allowing the evaluation of the arrival angle. The digital approach would be to sample the two antenna signals, evaluate the absolute phases, calculate the arrival angle and direct the result to the peripheral of choice. The main point is that this in no way prejudices the evaluation of other parameters such as Doppler velocity, signal amplitude etc., and the sounder is not dedicated to the single task of arrival angle measurement, as in the analogue case.

The digital sounder is thus the current state-of-the-art equipment used for ionosphere sounding. However, because the apparatus is expensive, digital sounders are at present only implemented at relatively few selected stations around the world, often as support for larger facilities such as incoherent scatter radars. The general capabilities have been summarised by J.W. Wright and A.K. Paul [27]. The parameters which are known or measured as functions of space and time are the radio frequency, the group range (virtual height), the echo phase and amplitude. From these can be calculated, in real time if necessary, the angle of arrival, the polarisation, and the Doppler velocity (strictly, the rate of change of phase range with time). In the Dynasonde, real time electron density profiles and electron collision frequencies can be calculated. From this information many things can be deduced, including neutral temperatures, electron plus ion temperatures, winds, tides and turbulence, currents, electric fields and particle precipitation (Wright and Paul [27]).

Finally, a development that had to await the necessary technological advance that would enable the synthesis of a linear frequency modulated sweep, was the "chirp" or FMCW ionosonde first described by G.H. Barry [23] in 1971. This is really another form of pulse compression technique, but with a compression ration of the order of 50,000. It is indeed the purpose

of this monograph to expand on the principle of the analogue FMCW ionosonde to allow a digital capability such as that which has been evolved for the "Digisonde" and "Dynasonde". Both these devices are primarily pulse ionosondes. One could also say that for these digital devices, no pulse compression technique is employed, although some form of integration of individual pulses is used to improve signal to noise ratios (for instance, an echo is "identified" if it satisfies certain criteria m out of n times). No attempt has been made to implement digitisation on ionosondes that employ any form of compression technique, on the grounds that the analysis becomes more complicated.

If this monograph is to be summarised in terms that will allow it to fit into the foregoing history and development of ionospheric sounding equipment, it is that this work sets out to describe the theory and application of the FMCW technique to digital sounding, and thereby add another advancement to knowledge in this field.

REFERENCES: CHAPTER 1.

(1) Balfour Stewart, Encyclopaedia Brittannica (Ninth Edition), 16, 181 (1878).

(2) Marconi G., Electrician, 49, 362 (1902).

(3) Marconi G., Proceedings of the Royal Society, 17, 208 (1902).

(4) Kennelly A.E., Elec. World and Eng., 15, 473 (1902).

(5) Heaviside O., Encyclopaedia Brittannica (Ninth Edition), 33, 215 (1902).

(6) Eccles W.H., Proceedings of the Royal Society A, 87, 79 (1912).

(7) Larmour J., Philosophical Magazine, 48, 1025 (1924).

(8) Mitra S.K., "The Upper Atmosphere", 2nd Edition, The Asiatic Society

Monograph Series, Calcutta (1952).

(9) Pierce G.W., "Principles of Wireless Telegraphy", page 139 (1910).

(10) de Forest L., Electrician, 69, 369 (1912).

(11) Smith-Rose R.L. and Barfield R.H., Proceedings of the Royal Society A, 107, 587 (1925).

(12) Smith-Rose R.L. and Barfield R.H., Proceedings of the Royal Society A, 110, 580 (1926).

(13) Appleton E.V. and Barnett M.A.F., Proceedings of the Royal Society A, 109, 621 (1925).

(14) Breit G. and Tuve M.A., Physical Review, 28, 554 (1926).

(15) Tuve M.A. and Dahl O., Proceedings of the Institute of Radio Engineers, 16, 794 (1928).

(16) Villard O.G., Radio Science, 11, 847 (1976).

(17) Gilliland T.R., Proceedings of the Institute of Electrical Engineers, 23, 1076 (1935).

(18) Appleton E.V. and Naismith R.,Proceedings of the Royal Society A, 137, 36 (1932).

(19) Sulzer P.G., Electronics, 19, 137 (1946).

(20) Whitehead J.D. and Malek A., Journal of Atmospheric and Terrestial Physics, 25, 599 (1963).

(21) Whitehead J.D. and Kantarizis E., Journal of Atmospheric and Terrestial Physics, 29, 1483 (1967).

(22) Phillips M.L., IEEE Transactions on Antennas and Propagation, AP-22, 785 (1974).

(23) Barry G.H., IEEE Transactions on Geoscience Electronics, GE-9, 86 (1971).

(24) Wright J.W., Proceedings of the IEEE, 57, 481 (1969).

(25) Bibl K., Patenaude J.A., Reinisch B.W. and Vargas-Vila R., Report

AFCRL-71-0001, USAF Cambridge Research Labs. (1970).

(26) Clarke R.H., Planetary Space Science, 19, 1387 (1971).

(27) Wright J.W. and Paul A.K., NOAA Special Report: "Toward Global Monitoring of the Ionosphere in Real Time by a Modern Ionosonde Network: The Geophysical Requirements and Technological Opportunity" (1981).

CHAPTER 2

REVIEW OF RADAR PRINCIPLES AND DISCUSSION OF PULSE AND CHIRP TECHNIQUES

Contents:

CHAPTER 2

REVIEW OF RADAR PRINCIPLES AND COMPARISON OF "PULSE" AND "CHIRP" TECHNIQUES

2.1 Introduction

To the best of the author's knowledge, all work to date on the theory and implementation of digital sounders and descriptions of their various "modes" or measuring formats have centered around the traditional pulse ionosonde. As this monograph deals with an alternative technique of Frequency Modulated Continuous Wave (FMCW) or "chirp" modulation, it is appropriate to investigate the two approaches theoretically and assess their relative merits.

Such a comparison must necessarily involve the principles and theory of radar, so this chapter will begin with a brief review of these fundamentals, based on the concepts of the matched filter and the ambiguity function.

Using these two mathematical tools, the relationship between the matched filter and its realisation in practice is discussed for pulse and chirp modulations. The chapter then proceeds with a review of the theory behind the assessment of measurment accuracy and echo resolution for these two modulations. It should be emphasised that this analysis is in no way intended to demonstrate superiority of one modulation over the other, but rather to provide some mathematical insight into the differences between the two methods as it relates to the description of chirp modulation as an alternative.

Review of Radar Principles: Matched filter

All signals can be represented in the exponential form (Appendix A):

$$s(t) = u(t)e^{j2\pi f_o t} \qquad (2.1)$$

where $u(t) = a(t)e^{j\Phi(t)}$

is called the complex modulation.

$a(t)$ = real amplitude

f_o = carrier frequency

The matched filter [2] for a particular signal s(t) is the optimum filter when s(t) is corrupted by white noise. If S(f) is the frequency spectrum of the signal s(t) to which the filter is to be matched, then the matched filter will be optimum if it has frequency spectrum ([2], section 3-2):

$$H(f) = K.S^*(f)e^{-j2\pi fT_o}$$

where K = gain constant

$e^{-j2\pi fT_o}$ = linear phase term necessary for real filters in order to make them physically realisable (i.e. the filter may not respond before it has been excited). The spectrum of the response of the filter to the signal s(t) will be:

$$Z(f) = (1/2)H(f).S(f)$$

$$= (K/2)S(f).S^*(f)e^{-j2\pi fT_o} \qquad (2.2)$$

Taking the Fourier transform of both sides gives the response of the filter to s(t) as a function of time:

$$z(t) = (K/2) \int_{-\infty}^{\infty} s(\eta).s^{*}(\eta + T_o - t)d\eta \qquad (2.3)$$

The output is seen to be a maximum when $t = T_o$. Since this filter is known to be optimum under a wide variety of conditions [2], it is the task of the engineer to try to implement it as closely as possible. An illustration of the relationship between the theoretical matched filter and the practical application for the cases of pulse and chirp modulations is given below.

Pulse modulation

The transmitted waveform can be represented in exponential form according to equation (2.1):

$$s_T(t) = u(t)e^{j2\pi f_o t}$$

where f_o is the carrier frequency. As a first approximation, a pulse can be represented by a burst of R.F. energy of constant amplitude A and finite duration δ, for which the complex modulation is given by ([2], equ. 3.10):

$$u(t) = A\ \mathrm{rect}(t/\delta) \qquad (2.4a)$$

$$\mathrm{rect}(t/\delta) = \begin{cases} 1, & |t| <= \delta/2 \\ 0, & |t| > \delta/2 \end{cases} \qquad (2.4b)$$

This serves not only as the definition of the simplest pulse, but also as a definition of the function rect(a/b) ascribed to P.M. Woodward ([1], pg. 287).

For simplicity, we will assume that the target velocity is zero so that there is no Doppler shift on the carrier frequency f_o. The received signal

can then be represented by:

$$s_R(t) = K_1 u_T(t - \tau_R) e^{j2\pi f_o (t - \tau_R)}$$

$$= K_1 s_T(t - \tau_R)$$

where K_1 = attenuation factor

τ_R = signal delay

In this case, the matched filter can be implemented directly. The receiver performs the function of amplifying the received signal, and for optimum rejection of noise, the receiver bandpass or filtering characteristics should match that of the received signal as closely as possible. Since we do not know $s_R(t)$ exactly because τ_R is unknown, we can match the filter to the known value of $s_T(t)$. The response to the signal $s_R(t)$ of a filter matched to $s_T(t)$ as a function of time, will be, from equation (2.3):

$$z(t) = (K/2) \int_{-\infty}^{\infty} s_R(\eta) . s_T^*(\eta + T_o - t) d\eta$$

$$= (K.K_1/2) \int_{-\infty}^{\infty} s_T(\eta - \tau_R) . s_T^*(\eta + T_o - t) d\eta$$

This will have a maximum value when $T_o - t = -\tau_R$ i.e. when $t = T_o + \tau_R$. Thus, τ_R can be evaluated by observing the time at which the receiver output peaks.

It should be noted that in practice the receiver will normally remove all or part of the carrier frequency f_o and the filter will operate not on $s_R(t)$ but on the "demodulated" signal given by:

$$s_c(t) = s_R(t) e^{-j2\pi f_o t}$$

$$s_c(t) = u_R(t)e^{-j2\pi f_o \tau_R}$$

$$= K_3 u_T(t - \tau_R)$$

$$K_3 = e^{-j2\pi f_o \tau_R}$$

The filter would then be matched to the known modulation $u_T(t)$ for which the time domain response would be given by:

$$z(t) = (K.K_3/2) \int_{-\infty}^{\infty} u_T(\eta - \tau_R).u_T^*(\eta + T_o - t)d\eta$$

This process is illustrated in figure (2.1).

Chirp modulation

The transmitted signal is once more represented by:

$$s_T(t) = u(t)e^{j2\pi f_o t}$$

where in this case the complex modulation is given by ([2], equ. 3-112):

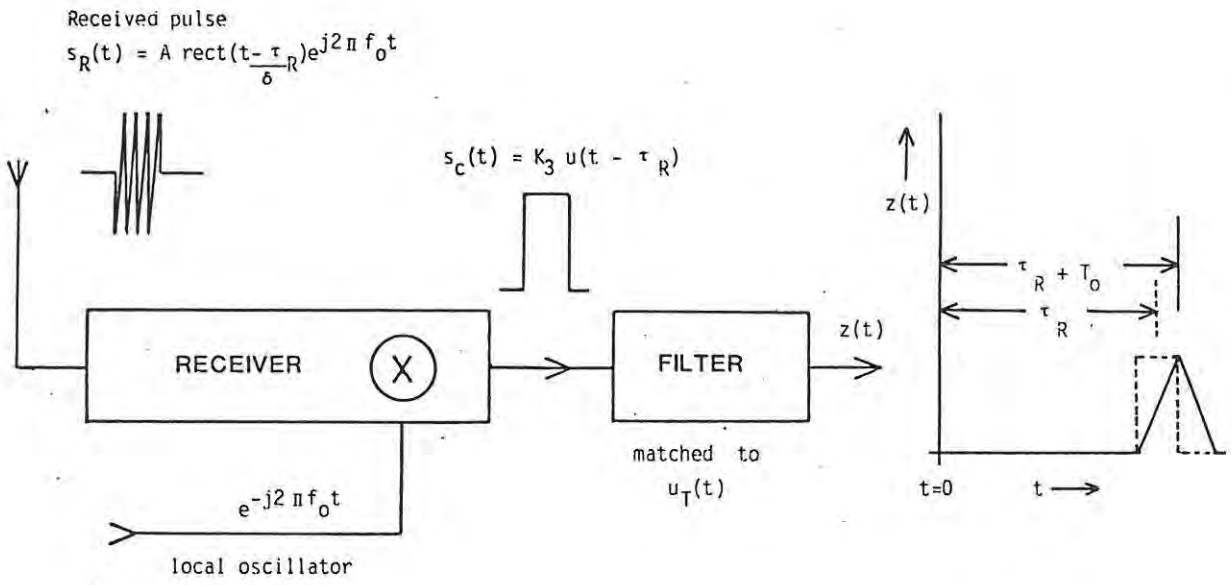
$$u(t) = Ae^{j\pi k t^2} rect(t/\delta)$$

which represents a burst of R.F. energy of amplitude A, length δ, but with a frequency that is increasing linearly at rate k. If we assume that the target is stationary as before, the received signal can be represented by:

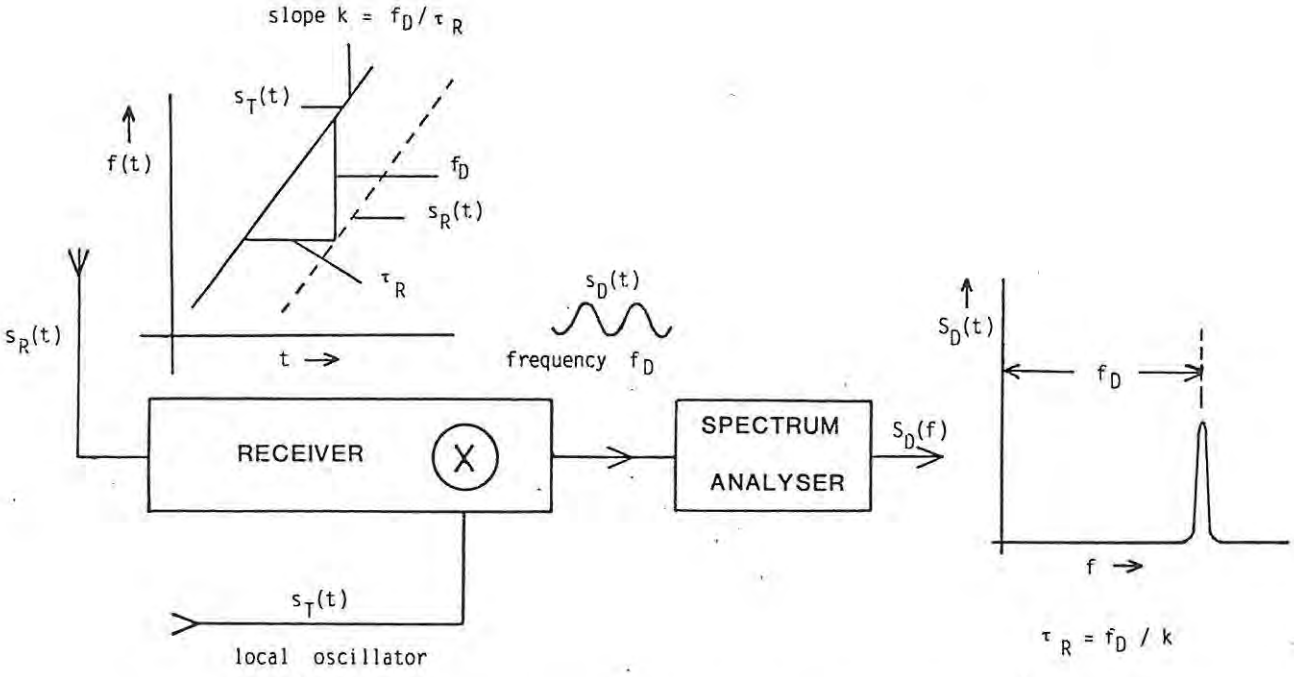
$$s_R(t) = Ks_T(t - \tau_R) \qquad (2.5)$$

where K is the attenuation constant. The receiver performs the task of multiplying the two waveforms $s_T(t)$ and $s_R(t)$ together and discarding the sum frequency, thus the receiver output can be represented by:

$$s_D(t) = (1/2)s_T(t).s_R^*(t)$$



FIGURE(2.1) An illustration of the application of a matched filter for the case of a pulse of length δ delayed by τ_R. The delay is obtained directly by observing the maximum in the receiver output as a function of time.



FIGURE(2.2) An illustration of the application of a matched filter for the case of a chirp signal with linear FM rate k, delayed by τ_R. The delay is obtained by observing the maximum in the receiver output as a function of frequency, through the relation $\tau_R = f_D / k$.

If this signal is spectrum analysed by means of the Fourier transform, the spectrum of the output will be given by (Appendix A, equ. A4):

$$S_D(f) = (1/2) \int_{-\infty}^{\infty} s_T . s_R^* e^{-j2\pi ft} dt \qquad (2.6)$$

Combining equations (2.5) and (2.6) gives:

$$S_D(f) = (K.K_2/2) \int_{-\infty}^{\infty} s_T(t) . s_T^*[t + (f/k - \tau_R)]dt$$

where $K_2 = \exp[j\pi k\{(f/k)^2 - 2\tau_R f/k\}]$ is a constant phase term.

Changing the variable of integration gives:

$$S_D(f) = (K.K_2/2) \int_{-\infty}^{\infty} s_T(\eta) . s_T^*[\eta + (f/k - \tau_R)]d\eta$$

Comparing with equation (2.3) for the matched filter shows that the amplitude of the filter output z(t) will equal $S_D(f)$ (the amplitude of the component of the difference signal of frequency f) when

$$K = K_1 K_2$$

and $$T_o - t = (f/k - \tau_R)$$

and in particular, will be a maximum when

$$T_o - t = (f_D/k - \tau_R) = 0$$

i.e. $$\tau_R = f_D/k$$

Thus, determination of the frequency f_D at which the maximum occurs in the spectrum affords a way to measure τ_R. The implementation of the matched filter for the chirp case is illustrated in figure (2.2).

2.3 Review of Radar Principles: Accuracy and resolution

It is important that a clear distinction be drawn between the concepts of accuracy and resolution since both are important but separate properties of any radar signal.

Accuracy: By this is meant a measure of the error that can be expected in the measurement of the range and/or velocity of a single point target and can be expressed as a function of the standard deviations of those errors, as follows [1],[2],[5]:

$$\text{range accuracy} = c.\sigma_\tau/2 \tag{2.7}$$

$$\text{velocity accuracy} = c.\sigma_\nu/2f_o \tag{2.8}$$

c = velocity of light

In ionospheric sounding, both the range and the velocity of the point target are unknown, and the standard deviations of the measurement errors are given by: ([5],Section 5.1, equs. (63) and (64))

$$\sigma_\tau^2 = 1/\beta_o^2(2E/N_o)(1-\alpha^2/\beta_o^2 t_o^2) \tag{2.9}$$

$$\sigma_\nu^2 = 1/t_o^2(2E/N_o)(1-\alpha^2/\beta_o^2 t_o^2) \tag{2.10}$$

β_o = "rms angular bandwidth" (rad s^{-1})

t_o = "rms time duration" (s)

α = "time-phase constant", a measure of the cross-coupling between the range and velocity measurements (radians).

E = energy in real received signal

$N_o/2$ = noise power per Hz for the real noise waveform

Definitions of β_o, t_o and α are given in Appendix A. These formulas have been derived for the simultaneous measurement of range and velocity, however, if either parameter is known from an independent information source, or if the range-velocity coupling α is zero, these formulas reduce to:

$$\sigma_\tau = 1/\beta_o(2E/N_o)^{1/2} \tag{2.11}$$

$$\sigma_\nu = 1/t_o(2E/N_o)^{1/2} \tag{2.12}$$

For both sets of formulae, it is seen that as the signal-to-noise ratio $(2E/N_o)$ increases, the standard deviations of the expected errors tends to zero. However, the standard deviations will only decrease as the inverse square root of the S/N ratio.

The accuracy of other measurements such as angle of arrival, can likewise be shown ([3],[2] Section 4-2) to be inversely proportional to the antenna aperture (or separation in the case of two antennas) and is independent of the choice of signal modulation.

Resolution: While the term "accuracy" is used to describe the measurement of some parameter in an echo from a single point target, the concept of resolution is used to describe the ability to identify two targets closely spaced in terms of the measured parameter. As before, the term "parameter" is restricted here to time and frequency, which are related to the range and Doppler velocity by:

$$\text{Range} = R = c\tau_R/2 \tag{2.13}$$

$$\text{Velocity} = V = c\nu_R/2f_o \tag{2.14}$$

$$\tau_R = \text{signal delay}$$

ν_R = Doppler frequency shift

(i) Range resolution: If one assumes equal velocities and that the amplitudes of the returned echoes are equal, it is possible to define the range resolution constant T_r for the two targets closely spaced in range as ([5], Section 5.2.1):

$$T_r = 1/\beta_e$$

where β_e is called the "effective bandwidth" (for definition, see Appendix A)

$$\text{Thus } T_r = \frac{\int_{-\infty}^{\infty} |U(f)|^4 \, df}{\left[\int_{-\infty}^{\infty} |U(f)|^2 \, df\right]^2} = \frac{\int_{-\infty}^{\infty} |R(\tau)|^2 \, d\tau}{4E^2} \qquad (2.15)$$

E = total energy in signal

$R(\tau)$ = autocorrelation of u(t)

The numerator of equation (2.15) is the area under the square of the autocorrelation function of u(t) and should be minimised to achieve good range resolution (small T_r).

(ii) Velocity resolution: Similarly, the velocity resolution constant F_r, for equal amplitudes and ranges, is given by:

$$F_r = 1/t_e$$

where t_e is known as the "effective duration" (see Appendix A):

$$F_r = \frac{\int_{-\infty}^{\infty} |u(t)|^4 \, dt}{\left[\int_{-\infty}^{\infty} |u(t)|^2 \, dt\right]^2} = \frac{\int_{-\infty}^{\infty} |K(\nu)|^2 d\nu}{4E^2} \qquad (2.16)$$

$K(\nu)$ = autocorrelation of the spectrum $U(f)$.

The numerator of equ.(2.16) is the area under the square of the autocorrelation of $U(f)$, which should be minimised for good velocity resolution.

(iii) Simultaneous resolution: The ambiguity function: In the ionospheric case, the range and the velocity are both unknown, and must be determined together. For this one must derive the "combined ambiguity constant", which is given by ([5], Section 5.4):

$$\text{ambiguity constant} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{|\chi(\tau,\nu)|^2 \, d\tau d\nu}{4E^2} \tag{2.17}$$

$$\text{where } \chi(\tau,\nu) = \int u(t)u^*(t+\tau)e^{-j2\pi\nu t}\, dt \tag{2.18}$$

$$= \int U(f+\nu)U^*(f)e^{-j2\pi f\tau}\, df$$

is the two dimensional autocorrelation function of the modulation $u(t)$ and is known as the ambiguity function.

Physical meaning can be assigned to the ambiguity function if one considers a filter matched to some delayed and Doppler shifted version of $s(t)$ with expected values τ_M and ν_M. If the signal arrives with delay and shift τ_R and ν_R, then the ambiguity function $\chi(\tau,\nu)$ is the amplitude of the complex modulation of the filter output, where $\tau = \tau_R - \tau_M$ and $\nu = \nu_R - \nu_M$ are the range and Doppler mismatches respectively ([1], Section 8.1; [2], Sections 3.2 and 3.3).

It would seem logical to attempt to minimise the numerator of equ.(2.17) as before. However, it can be shown that:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\chi(\tau,\nu)|^2 \, d\tau d\nu = |\chi(0,0)|^2 = 4E^2 \tag{2.19}$$

and thus the "ambiguity constant" = 1, and is independent of the choice of signal modulation. A plot of $|\chi(\tau,\nu)|$ as a function of τ and ν is known as an ambiguity diagram, which reveals graphically the resolution capabilities of the modulation u(t). Some of the more readily identifiable features of the ambiguity diagram are summarised below:

(i) $|\chi(\tau,\nu)|$ has its maximum value at (0,0), i.e. the filter will give maximum response when matched to the signal.

(ii) Values of $|\chi(\tau,\nu)|$ for τ,ν different from zero represent the response of the filter at mismatch. For good simultaneous resolution of range and velocity, one would hope for these values to be as small as possible. However, the volume under the surface $|\chi(\tau,\nu)|$, given by equ.(2.19), has been shown to be constant, thus one could only achieve small values of $|\chi(\tau,\nu)|$ in one area of the τ,ν plane at the expense of other areas. This phenomenon has been aptly expressed as follows: "Like slums, ambiguity has a way of appearing on one place as fast as it is made to disappear in another" (P.M. Woodward, [9]). Some modulations (notably pseudo-random coded) do succeed in achieving a sharp spike at (0,0) with the rest of the ambiguity volume almost equally distributed over the τ,ν plane. These are known as "thumbtack" ambiguity functions, but their prime disadvantage is the difficulty of implementing a matched filter. Values of $|\chi(\tau,\nu)|$ comparable with the central maximum $|\chi(0,0)|$ are known as "ambiguities" and are generally considered undesirable.

(iii) Setting $\nu = 0$ reduces equ.(2.17) to equ.(2.15), and similarly setting $\tau = 0$ reduces (2.17) to (2.16). Thus T_r and F_r are given by the areas under the function $|\chi(\tau,0)|^2$ and $|\chi(0,\nu)|^2$ respectively.

2.4 Discussion of the pulse and FMCW (chirp) modulations

Armed with the foregoing means of describing the properties of a radar signal, one is in a better position to discuss the relative merits of pulse and chirp modulations. In particular, one would like to answer the following questions:

(i) What exactly is meant by the terms "pulse" and "chirp" and how are they related mathematically, if at all?

(ii) How do the two modulations behave with regard to the accuracy of the range and Doppler measurements?

(iii) How do the two modulations behave with regard to the resolution of echoes closely spaced in range and/or velocity?

(i)Mathematical relationship: If a "pulse" is considered to be a burst of R.F. energy of finite duration δ and constant amplitude A, as described in section 2.2, then from equations (2.4a) and (2.4b) the complex modulation is given by:

$$u(t)_{pulse} = A.rect(t/\delta) = \begin{cases} A, & |t| <= \delta/2 \\ 0, & |t| > \delta/2 \end{cases} \qquad (2.20)$$

The corresponding modulation for a "chirp" signal is given by ([2], equ. 3-112):

$$u(t)_{chirp} = A.rect(t/\delta)e^{j\pi\ kt^2} \qquad (2.21)$$

which corresponds to a burst of R.F. energy of length δ but with a frequency that is increasing linearly with time at a rate k. It must be remembered that (2.20) and (2.21) describe the complex modulations only; the signals themselves are evaluated through equation A.3, Appendix A.

A comparison of (2.20) and (2.21) reveals that what has been termed a pulse modulation can be thought of as a special case of a chirp modulation

when k = 0.

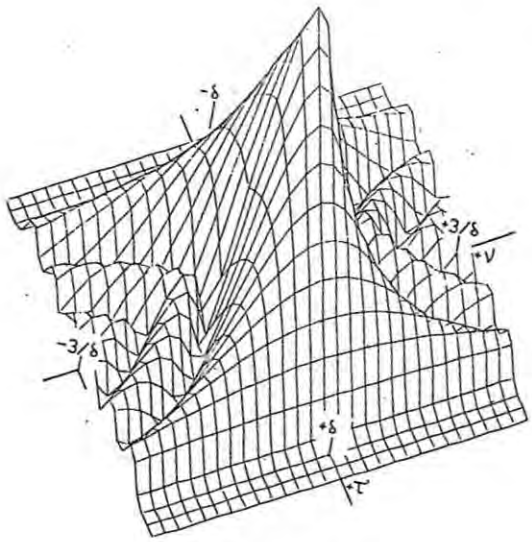
The ambiguity function for both modulations can be evaluated analytically from equ.(2.18) and can be expressed in the general form:

$$|\chi(\tau,\nu)|_{\text{chirp}} = \text{rect}(\tau/2\delta)e^{j\pi\nu\tau}\,(\delta - |\tau|)\frac{\sin[\pi(k\tau + \nu)(\delta - |\tau|)]}{\pi(k\tau + \nu)(\delta - |\tau|)} \quad \ldots (2.22)$$

which for the pulse case, k = 0, reduces to:

$$|\chi(\tau,\nu)|_{\text{pulse}} = \text{rect}(\tau/2\delta)e^{j\pi\nu\tau}\,(\delta - |\tau|)\frac{\sin[\pi\nu(\delta - |\tau|)]}{\pi\nu(\delta - |\tau|)} \quad \ldots (2.23)$$

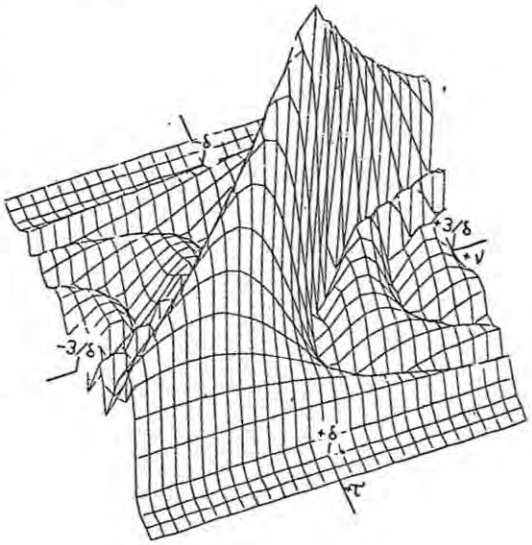
The ambiguity function for a simple pulse is plotted in fig.(2.3). Figures (2.4) and (2.5) show how the triangular intersection of the pulse ambiguity function with the range axis $(\tau,0)$ rotates clockwise as an increasing measure of chirp modulation k is applied. As it does so, the intersection of the ambiguity surface $|\chi(\tau,\nu)|$ with the τ axis becomes narrower, which illustrates that if ν is known, the resolution of the range parameter τ improves with k. This narrowing of the intersection with the τ axis is known as "pulse compression". However, if ν is unknown, the response of the filter can lie anywhere along the knife-edge and its rotation with respect to the (τ,ν) framework does not improve the combined resolution. The assumption in standard ionospheric chirpsounding is that ν lies near zero, limiting the area of the knife-edge surface over which τ may range. Thus, to the extent that this assumption is true, pulse compression is achieved.



$k = 0.$

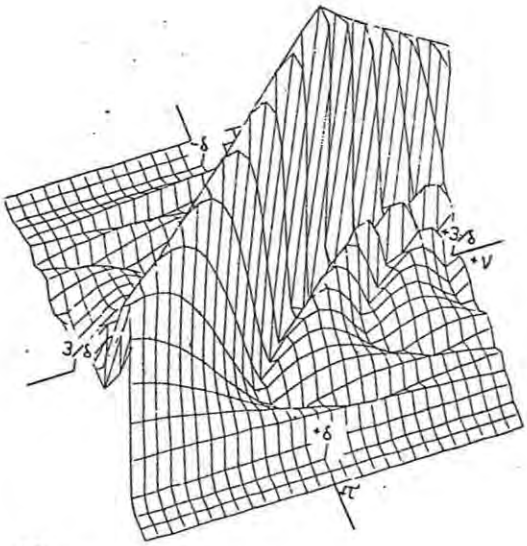
FIGURE(2.3)

$|\chi(\tau, \nu)|$



$k = 3/\delta^2$

FIGURE(2.4)



$k = 5/\delta^2$

FIGURE(2.5) Plots of the ambiguity function $|\chi(\tau,\nu)|$ for a rectangular pulse of duration δ for 3 different values of the linear frequency modulation constant k (arbitrary units). The peak value is $|\chi(0,0)|$.

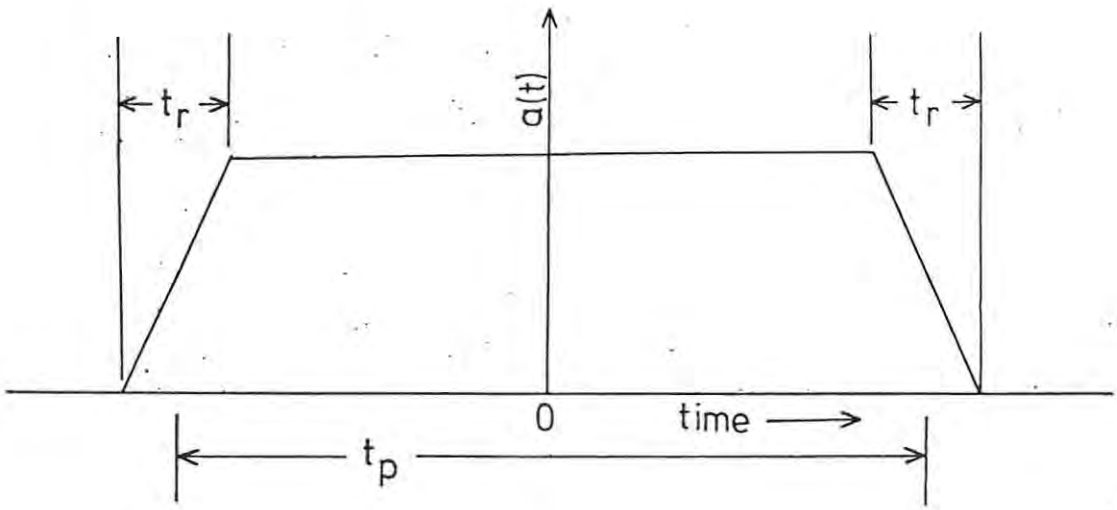
(ii)Discussion of accuracy:

Signal envelope: From equ.(2.11) it can be seen that the root mean squared range error estimate σ_τ is inversely dependent on the r.m.s. bandwidth β_o. However, β_o is infinite for the rectangular modulation expressed in equ.(2.20), inferring zero error in range measurement. In practice, it is possible neither to transmit nor to receive signal modulations with infinitely sharp rise times, and a more realistic model must be found. Following [5], an envelope with finite rise and fall times t_r and duration t_p is chosen, illustrated in fig.(2.6).

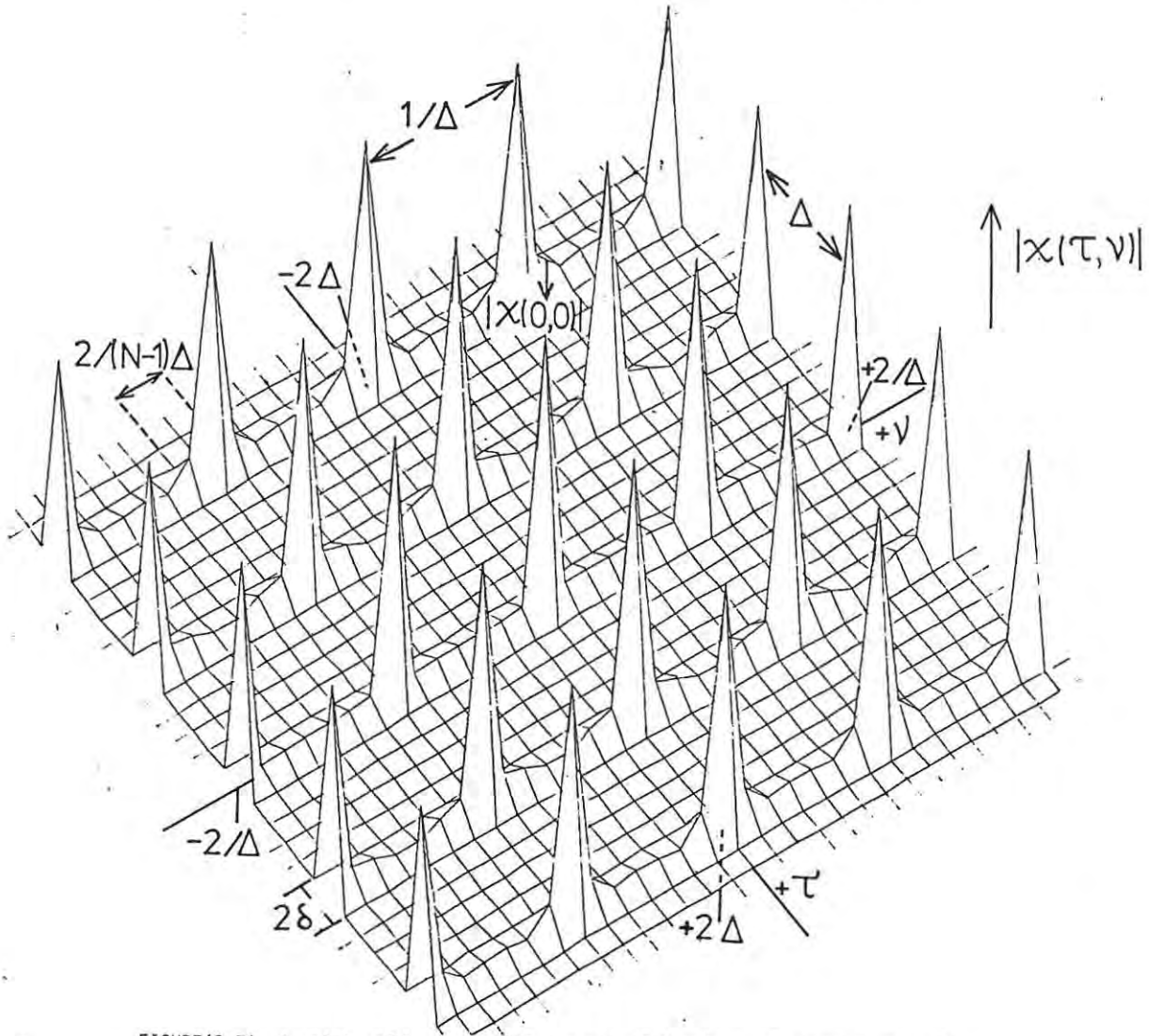
Short pulse, no frequency modulation: A short pulse of this type is commonly used in the average non-digital ionosonde, and is thus worthy of investigation. Since the rise time of such a pulse is largely determined by the receiver bandwidth, $t_p = t_r = 5 * 10^{-5}$ s is chosen, resulting in a triangular pulse that falls as soon as it has peaked. The factor $2E/N_o$ is arbitrarily chosen to equal unity. Thus, from [5], Chapter 5, equs.(85) and (88), and equs. (2.11) and (2.12):

rms duration t_o	=	$\pi(2/5)^{1/2}t_r$	= $9.93 * 10^{-5}$ s
rms bandwidth β_o	=	$3^{1/2}/t_r$	= $3.46 * 10^4$ radians
rms time error	=	σ_τ	= $2.89 * 10^{-5}$ s
rms frequency error	=	σ_ν	= $1.01 * 10^4$ Hz

Thus the time accuracy is comparable with the rise time t_r, which is acceptable, but the error in the frequency measurements is much larger than the values normally expected which are of the order of a few tenths of Hz. Thus, a single pulse is a very poor modulation for the evaluation of Doppler velocities in ionospheric work.



FIGURE(2.6) A realistic pulse modulation model with finite rise and fall times t_r and nominal duration t_p.



FIGURE(2.7) A plot of $|\chi(\tau, \nu)|$ for a set of N pulses of length δ and separation Δ, for N = 20 (arbitrary units). The central peak is at $|\chi(0,0)|$.

Long pulse, linear freqency modulation: This is the modulation used by a standard commercial chirpsounder. One can now choose the parameters $(2E/N_o) = 1$, $t_r = 5 * 10^{-5}$ as before, and $t_p = 1s$, $k = 5 * 10^4$ Hz s^{-1} (typical). Because the range-velocity coupling factor α is now non-zero, one must use equations (2.9) and (2.10) to evaluate the rms errors. It is thus necessary to evaluate the factor $(1 - \alpha^2/\beta_o t_o)$. It has been shown ([5], Chapter 5, equ.(102)) that the rms bandwidth for this modulation is given by:

$$\beta_o^2 = 2/t_r t_p + (\pi k t_p)^2/3 \qquad (2.24)$$

The first term is the contribution due to the rise time of the envelope, while the second is due to the linear frequency modulation. It is not common practice in chirpsounding to rely in any way on the rise time of the modulation to establish range, so one can approximate equ.(2.24) to:

$$\beta_o^2 \simeq (\pi k t_p)^2/3$$

Also, from [5], equs. (99) and (100), for $t_p >> t_r$,

$$t_o^2 = \pi^2 t_p^2/3 \quad \text{and} \quad \alpha^2 = k^2 t_o^4$$

from which one can evaluate the factor $(1 - \alpha^2/\beta_o^2 t_o^2) = 0$. Thus, from equs. (2.9) and (2.10), both the rms errors σ_τ and σ_ν are infinite. This result expresses mathematically the fact that if the expected velocities and ranges of the targets are both unknown and unbounded, it is not possible to make any precise measurement of either. This is due to the so-called "range-Doppler ambiguity" associated with linear F.M. In practice, these parameters are not unbounded, and in particular, the assumption in standard chirpsounding is that the velocity = 0. However, it has been shown [6],[7]

that even at mid-latitudes, sunrise effects can cause Doppler frequency shifts in excess of 1 Hz, which cause range errors of 3 km or more, depending on the FM rate k. It is thus seen that a short pulse can give good range estimates, with rms errors of the order of the rise time or better, while the range estimates of the long pulse with linear F.M. are in error by an amount proportional to the (unknown) velocity of the target. However, if this velocity can be assumed to be small, good range estimates can be achieved. Both modulations are inadequate for the estimation of velocities.

Velocity estimates: Pulse modulation. In order to improve the rms velocity error, it is necessary, from equ.(2.12), to increase the rms duration t_o. This can be achieved by repeating the pulse one or more times. For a single pulse, $t_o = (2\pi)^2 t_r^2/10$. For N pulses (Appendix C):

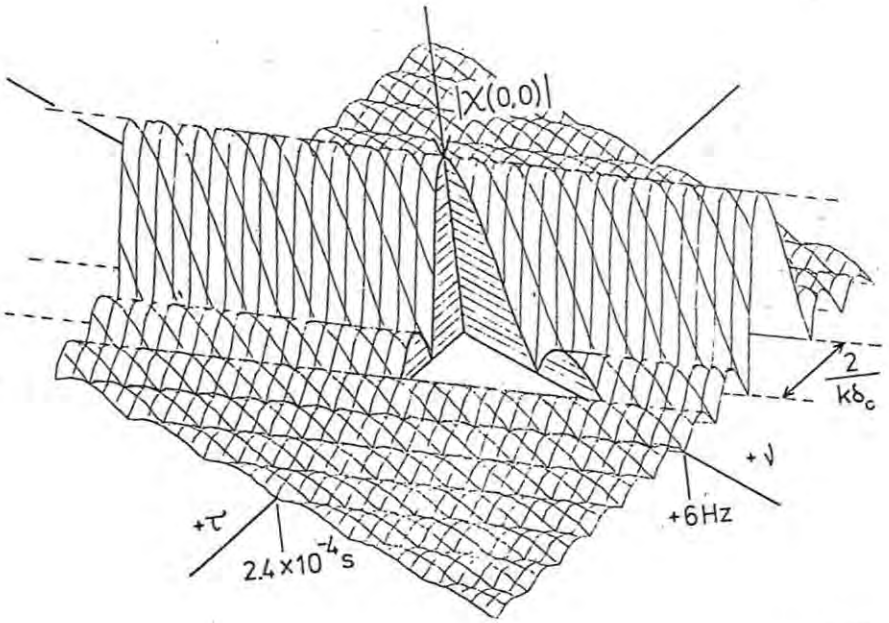
$$t_o = (2\pi)^2 t_r^2/10 + (N-1)(N+1)\Delta_p^2/12$$

where Δ_p is the pulse separation. Usually, $\Delta_p >> t_r$ and $N > 2$, so one can approximate:

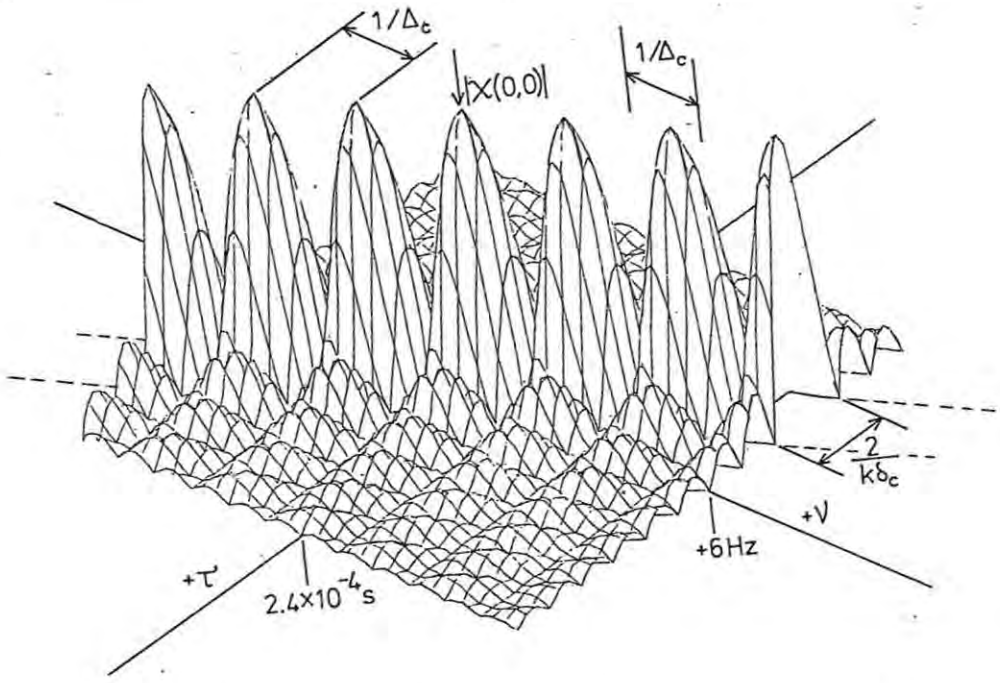
$$t_o \simeq 2\pi N\Delta_p/(12)^{1/2}$$

Thus, the rms bandwidth can be increased by increasing N or Δ_p, with a corresponding decrease in the rms velocity error σ_v. However, this tactic does introduce ambiguities, which have to be dealt with. An ambiguity diagram for a set of 20 pulses is shown in fig.(2.7) and will be discussed more fully under the topic of resolution.

Chirp modulation. In the same way, a long duration chirp signal can be made to achieve good velocity accuracy by repeating the modulation in the time domain. Fig.(2.8) shows the ambiguity function for a single chirp modulation, while (2.9) is the diagram for two identical chirp modulations separated by time Δ_c. In this case Δ_c has been chosen equal to the



FIGURE(2.8) A plot of $|\chi(\tau,\nu)|$ near the origin for a single chirp modulated pulse of length $\delta_c = 0.5s$ and chirp rate $k = 5*10^4$ Hz s^{-1}, showing the area under the intersection with the principal axes (arbitrary units).



FIGURE(2.9) A plot of $|\chi(\tau,\nu)|$ for two identical chirp modulations of length δ_c each = 0.5s, chirp rate $k = 5*10^4$ Hz s^{-1} and separation $\Delta_c = \delta_c$ = 0.5s (adjacent), near the point (0,0) (arbitrary units).

modulation length, i.e. the two modulations are adjacent. It is seen that the diagonal knife-edge that is characteristic of chirp modulation becomes serrated, falling to zero at regular intervals along its length. These could be referred to as "quantised ambiguities" whose projections onto the ν axis are separated by by the inverse of the modulation separation Δ_c. The projections onto the τ axis are given by $\tau_n = -\nu_n/k$. This device does not significantly affect the factor $(1 - \alpha^2/\beta_o^2 t_o^2)$ in equs. (2.9) and (2.10) which determine the range-Doppler coupling; however the quantisation of the ambiguity function now means that a specific velocity measurement takes on a set of discrete ambiguous values rather than a continuum as before. One can then make use of additional knowledge about the behavior of the ionosphere, or other independent knowledge, to discriminate between the ambiguities. Thus, the velocity is determined, and likewise the range, through the relation $\tau_n = -\nu_n/k$, so one can apply the formulas (2.11) and (2.12) for independent error estimates. It should be noted that a similar quantisation of the chirp ambiguity diagram can be achieved by transmitting two chirp modulations simultaneously, but with a <u>frequency</u> increment. A fuller discussion of this alternative is given in Chapter 3.

<u>Accuracy comparisons</u>: It is now possible to compare the accuracy capablities of pulse and chirp modulations. For equal range accuracy, σ_τ(pulse) = σ_ν (chirp). Thus from equ. (2.11):

$$2N_o/E(\text{pulse})\beta_o^2(\text{pulse}) = 2N_o/E(\text{chirp})\beta_o^2(\text{chirp})$$

By choosing as a pulse modulation a string of N pulses with realistic rise times t_r and, in the FM case, a pair of identical chirp modulations of length δ_c as described in the last section, it is possible to relate the energies E(pulse) and E(chirp) to the peak powers P(pulse) and P(chirp) and we can derive (appendix B):

$$P(\text{chirp}) = P(\text{pulse})\ 12N/t_r\pi^2k^2\delta_c^3$$

N = number of pulses

t_r = rise time of pulses

k = chirp rate

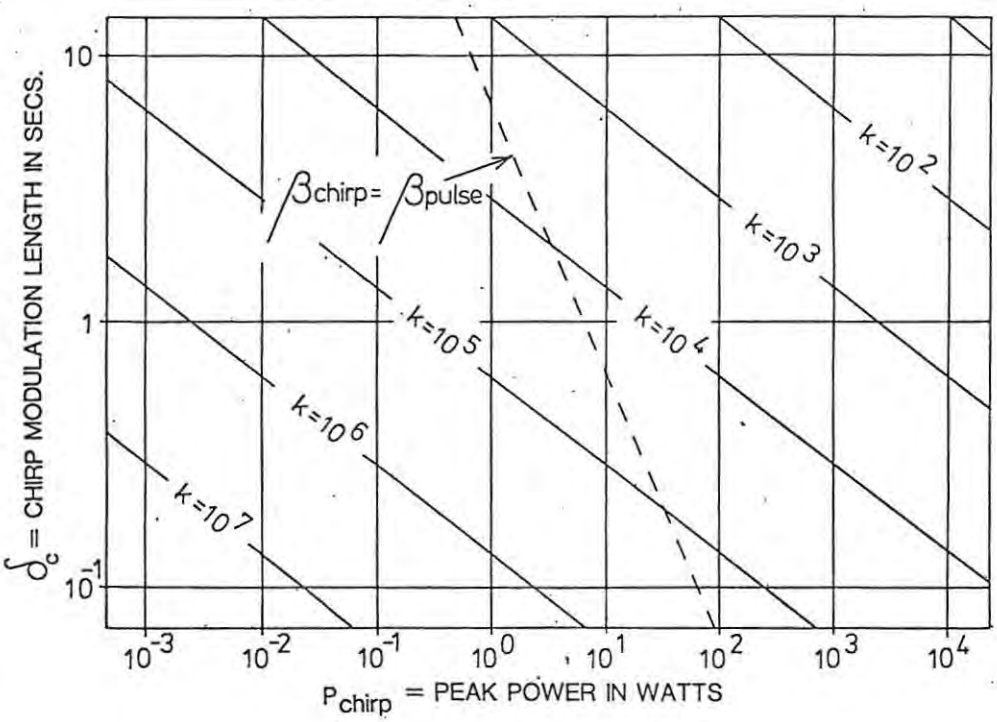
δ_c = individual chirp modulation length

This expression relates the peak chirp power necessary to achieve equivalent range accuracy to that of a string of N pulses. Choosing typical values of N = 10 pulses, $t_r = 5 * 10^{-5}$ s, P(pulse) = 10^4 watts, it is possible to plot the chirp power P(chirp) as a function of the length δ_c for various values of k. This is illustrated in fig. (2.10). As an example, for $k = 5 * 10^4$ Hz s^{-1} and δ_c = .5 s (typical), a peak chirp power of 9.5 watts will achieve equivalent range accuracy to a string of 10 pulses of peak power 10^4 watts. Thus, chirp modulation provides a means to simultaneously increase the duration and rms bandwidth of the signal while maintaining low peak power, and is thus useful where peak power is a limiting factor. The dotted line in fig.(2.10) is the locus of points for which the rms bandwidths for the two modulations are equal, thus points on this line refer to two modulations that are equally efficient in terms of occupancy of the RF spectrum.

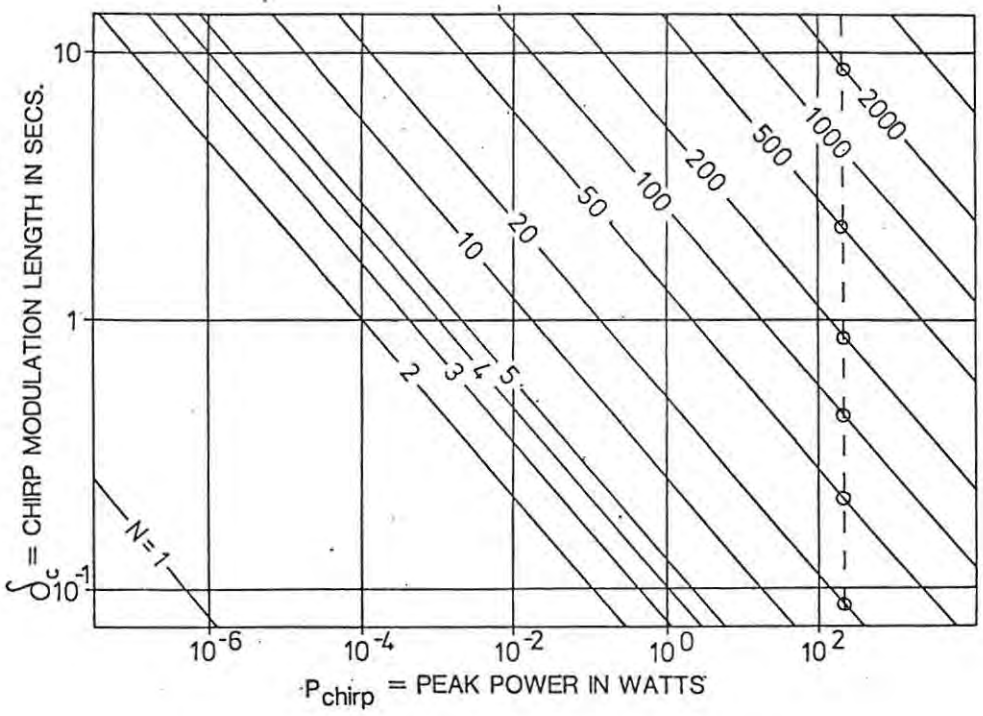
Similarly, for equal velocity accuracy, σ_ν(pulse) = σ_ν(chirp), so from equ. (2.12):

$$2N_o/E(\text{pulse})t_o^2(\text{pulse}) = 2N_o/E(\text{chirp})t_o^2(\text{chirp})$$

from which we can derive (appendix B):



FIGURE(2.10) The relationship beween the chirp modulation length δ_c, peak power P(chirp) and chirp rate k having equivalent range accuracy to a set of 10 pulses of risetime $5*10^{-5}$ s and peak power 10^4 watts.



FIGURE(2.11) The relationship between the chirp modulation length δ_c, peak power P(chirp) and the number of pulses N for equivalent velocity accuracy. Pulse risetime $t_r = 5*10^{-5}$ s, peak pulse power P(pulse) = 10^4 watts.

$$P(\text{chirp}) = P(\text{pulse})\ 4Nt_r^3/10\delta_c^3 + t_r N(N-1)(N+1)\Delta_p^2/3\delta_c^3$$

Δ_p = interpulse spacing.

In this case , $t_r = 5*10^{-5}$, $\Delta_p = 10^{-2}$, P(pulse) = 10^4 watts is chosen, and the variation of P(chirp) against δ_c is illustrated for various values of the number of pulses N, in fig.(2.11). As an example, if P(chirp) = 10 watts, δ_c = .5 s, it is seen that approximately 42 pulses are needed to achieve equivalent velocity accuracy to a dual chirp modulation of total duration 1 s. The intersection of the dotted line with the lines of constant N gives the values of δ_c and P(chirp) for which t_o(pulse) = t_o(chirp), for that value of N.

Summary: A single short pulse with no chirp component can have sufficient bandwidth to provide good range accuracy. However, by its very nature, the short pulse has short time duration and thus poor velocity accuracy. If the pulse is extended, the duration increases at the expense of bandwidth, so velocity accuracy can only be obtained at the expense of range accuracy. By repeating the short pulse in the time domain, bandwidth (and thus range accuracy) is maintained while duration (and thus velocity accuracy) is extended. However, the necessity for long inter-pulse periods of silence demands high peak powers to ensure sufficient total signal energy. Chirp modulation provides a means to simultaneously increase the duration and bandwidth of a single pulse while keeping peak powers low. It is necessary however to repeat the modulation in the time domain to provide discrete ambiguities which can be resolved by independent knowledge.

(iii)Discussion of resolution: In the last section on measurement accuracy, it was emphasized that the theory applied to a single "point" target. In this section, the problem of distinguishing between two targets reflecting

simultaneously is dealt with. In the ionospheric case simultaneous reflection is more the exception than the rule. Although one has often to deal with "overlapping" ordinary and extraordinary mode echoes, it is possible to distinguish between these modes by independent means (e.g. from a knowledge of the shape of the polarisation ellipses at the station). However, simultaneous reception of echo energy from two different reflection points is possible under conditions of (i) travelling disturbances, (ii) multiple reflections, and (iii) "spread F" conditions, so this section will attempt to compare the resolution properties of the two modulations. For this purpose, the ambiguity diagram is an invaluable aid. However, in using the ambiguity diagram for comparison, some account must be taken of the probability that ambiguous responses will be encountered. The total ambiguity, as has been shown, is constant for all modulations, thus for the purpose of comparison, the ambiguity diagram should be weighted with the two dimensional probability density of the parameters τ and ν. It is thus possible to define a "comparative combined resolution constant" γ that equals the volume under the weighted ambiguity function χ, as follows:

$$\gamma = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\chi(\tau,\nu)| \cdot P(\tau,\nu)\, d\tau\, d\nu}{4E^2} \qquad (2.25)$$

where $P(\tau,\nu)$ is the probability density function expressing the likelihood of occurrence of the "mismatches" τ and ν, and has the property

$$\int_{-\infty}^{\infty} P(\tau,\nu)\, d\tau\, d\nu = 1$$

The function $P(\tau,\nu)$ will depend on many factors, such as geomagnetic latitude, and season. Quantitative values for $P(\tau,\nu)$ for the ionosphere do not, to the best of the author's knowledge, exist. Thus the comparisons that follow must, of necessity, be tenuously quantitative.

Train of N rectangular pulses: By way of illustration, an ambiguity diagram for a set of 20 rectangular pulses of length δ_p separated by Δ_p is shown in fig.(2.7). The nature of the ambiguities is seen to be discrete, and their separation in the τ and ν directions bear a mutually inverse relationship. If one choses Δ_p = 6.67 ms corresponding to a range ambiguity of 1000 km through equ (2.13), one can say with reasonable safety that such ambiguous reflections are highly improbable. The corresponding ambiguities in the ν direction are given by $1/\Delta_p$ = 150 Hz. It is also improbable that Doppler frequencies of this magnitude would be experienced in the ionospheric case. Thus, with reasonable confidence, one can weight all the ambiguities, with exception of the central maximum, with the probability zero. The combined range/Doppler ambiguity can then be reduced to the volume under the central maximum only.

The denominator in equ.(2.25) normalises the amplitude to unity at (0,0), and the widths of the base of the central peak in the τ and ν directions are $2\Delta_p$ and $2/(N-1)\Delta_p$ respectively. The volume γ can thus be approximated by a cone of height unity and base area $\pi\delta_p(1/(N-1)\ \Delta_p)$. Thus

$$\gamma(\text{N pulses}) \simeq \pi\delta_p/3(N-1)\Delta_p \qquad (2.26)$$

For typical values $\delta_p = 10^{-4}$s, N = 10, $\Delta_p = 10^{-2}$, one has

$$\gamma(\text{N pulses}) = 1.16*10^{-3}$$

Two identical chirp modulations: In fig.(2.8), the ambiguity diagram for a single chirp modulation is re-plotted on an absolute scale to show the behaviour near $|\chi(0,0)|$ for a modulation of length δ_c = .5 s and chirp rate k = $5*10^4$ Hz s^{-1}. It is seen that the widths of the areas under the intersections of the "knife-edge" with the τ and ν axes are extremely

narrow, implying good range and velocity resolution when the other parameter is known. In the ionospheric case, the two parameters must be resolved simultaneously and the combined ambiguity is given by the volume under the knife-edge, which is constant. This is known as the range-Doppler ambiguity and is generally considered a drawback of this type of modulation.

Fig.(2.9) plots the ambiguity diagram for two identical adjacent chirp modulations of length δ_c = .5 s, $k = 5*10^4$ s^{-1} and separation $\Delta_c = \delta_c$ = 0.5 s. It is seen that a target with a Doppler shift of +2 Hz and differential range delay $\tau = -.4*10^{-4}$s will be indistinguishable from the target for which the filter is matched. Once again, for fair comparison, one should evaluate the volume γ, however the ambiguities are now more closely spaced than for the multiple pulse case and an absence of quantitative knowledge of the function $P(\tau,\nu)$ makes evaluation of γ difficult. It is nevertheless of interest to evaluate the volume under the central peak. The separations and widths of the "bumps" projected onto the ν axis are given by $1/\Delta_c = 1/\delta_c = 2$ Hz in this case. The width of the intersection of the knife-edge with the τ axis, at the base, is $2/k\delta_c$ ([2], pg. 3-19). Once again, the volume can be approximated by a cone:

$$\gamma(2 \text{ chirp}) = \pi(1/2\Delta_c)(2/2k\delta_c)/3 = \pi/6k\delta_c\Delta_c \qquad (2.27)$$

For reasonable values $\delta_c = \Delta_c = .5$ s, $k = 5*10^4$

$$\gamma(2 \text{ chirp}) = 4.18*10^{-5}$$

If one calculates the ratio $\gamma(10 \text{ pulses})/\gamma(2 \text{ chirp}) \simeq 28$ it is seen that one can include 28 ambiguities for which $P(\tau,\nu)$ could be set = 1, and zero elsewhere, and achieve equivalent combined resolution to that for 10 pulses. The widths of the ambigiuties in the ν direction are 2Hz in this example, so 28 ambiguities span a range in ν of $\pm$28 Hz. Thus, if one assumes a shape

for the function $P(\tau,\nu)$ that has the following form:

$$P(\tau,\nu) = \begin{cases} 1, & |\nu| <= 28 \text{ Hz} \\ 0, & |\nu| > 28 \text{ Hz} \end{cases}$$

then the two modulations will have equivalent values of γ. At locations for which the limits on ν are less that $\pm$28 Hz, γ(2 chirp) will be less than γ(pulse). At most geomagnetic latitudes, the limits on ν are probably as low as $\pm$3 Hz.

(iv) Long modulations: Up till now, it has been assumed that there was no interference between the transmitter and the receiver. In the case of short pulses, the receiver can be gated off during the (short) duration of the transmitted modulation. However, for modulations that are long in comparison with the expected delay of the echo, some method must be used to avoid desensitisation of the receiver by the transmitter signal. In oblique incidence applications this is achieved by physical separation of the transmitter and receiver, so that the "ground wave" is attenuated by distance. However, in vertical incidence sounding, the transmitter and receiver are co-located. The problem is solved by switching the transmitted modulation on and off to allow certain periods for reception. The form of the switching signal has been the subject of a separate publication by the author [8] and will not be repeated here, except to note that the switching waveform should have the following properties:

(i)50% duty cycle

(ii)Period comparable with the integration time of the matched filter

(iii)Auto-correlation function should be constant for delays not equal to integral multiples of the period.

The first of these points is worth a review because it bears directly on the comparisons just made. Assume a complex modulation u(t) of constant

amplitude, modulated by a switching function g(t) which can take on the values 1 or 0. The duty cycle of g(t) = a. Then the probability that the signal will be transmitted will be $p_T = a$, $0 < a < 1$. The signal can only be received when the transmitter is off, or g(t) = 0. Thus the probability that <u>any</u> signal can be received is $p_R = 1 - a$. The probability that the original signal will be received after some random delay will be $p = p_T p_R = a(1 - a) = a - a^2$. This has a maximum value of .25 for a = .5. Thus one will receive on average a maximum of only 25% of the energy that would have been transmitted had no switching taken place, and the duty cycle that achieves this maximum is 50%.

Points (ii) and (iii) ensure that the switching waveform has minimum effect upon the ambiguity function. Thus, the fact that long modulations need to be switched can be ignored for most purposes except to note that the average received power will only be 1/4 of that had the full unswitched modulation been transmitted and received.

REFERENCES: CHAPTER 2

(1) Nathanson F.E., "Radar Design Principles", McGraw-Hill Book Co., Library of Congress Card No. 79-80973, 1969.

(2) Skolnik M.I., "Radar Handbook", McGraw-Hill Book Co., Library of Congress Card No. 69-13615, 1970.

(3) Gething P.J.D., "Radio Direction Finding", Peter Peregrinus Ltd., ISBN 0 901223 71 9, 1978.

(4) Price R. and Hofstetter E.M., Bounds on the volume and height distributions of the ambiguity function, I.E.E.E. Transactions on Information Theory, IT-11, 207-214 (1965).

(5) Burdic W.S., "Radar Signal Analysis", Prentice-Hall Inc., Library of Congress Card No. 68-12906, 1968.

(6) Mackay J.R., Honours Project, Rhodes University, (1982).

(7) Wright J.W. and Pitteway M.L.V., Journal of Geophysical Research, 87, 1589 (1982).

(8) Poole A.W.V., I.E.E.E. Transactions on Antennas and Propagation, AP-27, 480 (1979).

(9) Woodward P.M., Radar Ambiguity Analysis, IRE Tech. Note 731, Royal Radar Establishment (1967).

CHAPTER 3

CHIRPSOUNDING: THEORY AND APPLICATION

Contents:

CHAPTER 3

CHIRPSOUNDING: THEORY AND APPLICATION

3.1 Introduction:

In the last chapter, it was mentioned that the ambiguity function $\chi(\tau,\nu)$ could be thought of as the output of a filter matched to some expected delay and Doppler frequency τ_R and ν_R respectively. Sections 3.2 to 3.5 deal with the theory behind the implementation of such a filter for the measurement of the modulation parameters τ and ν. Section 3.6 deals with modulation independent parameters.

3.2 Relation to ambiguity function

The ambiguity function for any modulation is given by equ.(2.12):

$$\chi(\tau,\nu) = \int_{-\infty}^{\infty} u(t)u(t+\tau)e^{-j2\pi\nu t}dt \qquad (3.1)$$

The real signal is given by equ.(A3), Appendix A. :

$$x(t) = 1/2\ u(t)e^{j2\pi f_o t} + u^*(t)e^{-j2\pi f_o t} \qquad (3.2)$$

It is not difficult to combine (3.1) and (3.2) to give:

$$\chi(\tau,\nu) = 2\int_{-\infty}^{\infty} x(t)x(t+\tau)e^{-j2\pi\nu t}dt \qquad (3.3)$$

It is known that $\chi(\tau,\nu)$ is the output of the matched filter and is a maximum when $\tau = \nu = 0$. However, because of the range-Doppler ambiguity of linear FM, the response at (0,0) will be indistinguishable from the response when $\tau = -\nu/k$. Thus:

$$\chi(0,0)_{chirp} = \chi_{max} = 2\int_{-\infty}^{\infty} x(t)x(t-\tau_R)e^{-j2\pi\nu t}dt \qquad (3.4)$$

where $\nu = +k\tau_R$.

The right hand side of equ.(3.4) is seen to be proportional to the Fourier transform of the product of the transmitted signal x(t) and the reflected signal $(t-\tau_R)$. Thus, in standard non-digital chirpsounding, the product $[x(t)x(t-\tau_R)]$ is formed by mixing in an RF receiver, and the output is spectrum analysed. The delay τ_R can then be deduced from the relation $\tau_R = f_R/k$ where f_R is the value of the frequency for which the transform is a maximum.

As was shown in Chapter 2, this method of determining τ_R is subject to range-Doppler ambiguity, and the assumption is always made that the velocity of the target is small or zero. In digital sounding, the velocity is an important measurable quantity, and for the reasons given in Chapter 2, the ambiguity diagram of fig.(2.9) is more appropriate. However, the correspondence between the mathematical matched filter of fig.(2.9) and the practical implementation is not as distinct as the case for a single modulation, so the analogy is not pursued further. In an attempt to derive a relationship that will allow evaluation of the velocity it is necessary to find an analytical expression for an ionospherically reflected chirp signal.

3.3 Review of ionospheric terms

Up till now, all discussion concerning radar has assumed that the probing signal has travelled at the speed of light through free space, been reflected, and returned along the same path. For reflection from some point in the ionosphere, the passage of the signal can be described in terms of three paths, two of which are conceptual. Since these terms will be used in the following discussion, they are briefly reviewed here.

Real range: The real range is the actual path along which a signal of frequency f travels to the point of reflection and will be denoted by

h(f,t). The real path is not directly measurable by the means described in this monograph.

Phase range: The phase range $h^*(f)$ is the apparent range at which successive phase states of a monochromatic signal of frequency f are reflected, assuming a velocity equal to that of light, and is given by:

$$h^*(f,t) = \int_0^h \mu^* dh \tag{3.6}$$

where μ^* is the "phase refractive index" given by the Appleton-Hartree equation ([1], Chapter 6). The value of μ^* is 1 in free space (below the ionosphere) and is equal to zero at the reflection point.

Group range: The group range $h'(f,t)$ is the apparent range to which the sum of a set of monochromatic waves associated with a signal of finite bandwidth will appear to travel at the speed of light, and is also the apparent range at which the energy in the wave appears to be reflected. The group range is related to the phase range by:

$$h'(f,t) = \frac{\partial[f.h^*(f,t)]}{\partial f} = h^*(f,t) + f.\frac{\partial h^*(f,t)}{\partial f} \tag{3.7}$$

Also,

$$h'(f,t) = \int_0^h \mu' dh \tag{3.8}$$

where μ' is called the group refractive index and is related to μ^* by

$\mu' = 1/\mu^*$.

The following relation is also important:

$$h^*(f,t) < h(f,t) < h'(f,t) \tag{3.9}$$

3.4 An analytical expression for an ionospherically reflected chirp signal

For a linear FM or chirp signal, the complex modulation is given by:

$$u(t) = A_1 e^{j\pi kt^2} \tag{3.10}$$

A_1 = constant

From Appendix A, equ.(A.3), the real signal is:

$$x(t) = (1/2)[u(t)e^{j2\pi f_o t} + u^*(t)e^{-j2\pi f_o t}] \tag{3.11}$$

where f_o is the carrier frequency. Combining (3.10) and (3.11) gives an expression for the instantaneous transmitted signal:

$$x_T(t) = A_1 \cos[\varphi_T(t)] \tag{3.12}$$

$$\varphi_T = 2\pi[f_o t + kt^2/2]$$

The instantaneous frequency of $x_T(t)$ is given by:

$$f_T = (1/2\pi)\frac{d}{dt}\varphi_T(t)$$

Equation (3.12) can be expressed as the sum of fixed frequency Fourier components

$$x_T(t) = A_2 \int_{-\infty}^{\infty} \cos[2\pi ft - (\pi/k)(f - f_o)^2 + \pi/4]\, df$$

The ionosphere can be modelled in terms of the phase range h^* described in the last section. The phase range is a function of the frequency of the radio signal, and time, assuming that the ionosphere is varying with time. Thus:

$$h^* = h^*(f,t) \tag{3.14}$$

After reflection, each component frequency will be delayed by its own phase

delay $h^*(f,t)$. Thus the instantaneous received signal is:

$$x_R(t) = A_3 \int_{-\infty}^{\infty} \cos[\psi_R(f,t)]\, df$$

$$\psi_R(f,t) = 2\pi f(t - 2h^*(f,t)/c) - (\pi/k)(f - f_o)^2 + \pi/4 \qquad (3.15)$$

A_3 = amplitude, including attenuation.

This has solution, from the principle of stationary phase [(4), pg. 3-5],

$$x_R(t) = A_4 \cos[\psi_R(f_s,t) + \text{sign}\{\psi_R''(f_s)\}(\pi/4)] \qquad (3.16)$$

where f_s is the frequency such that

$$\frac{d\psi_R}{df}(f,t)_{f=f_s} = \psi_R'(f,t)_{f=f_s} = 0$$

Because dh^*/df is constrained to be always positive, inspection of the form of equ. (3.15) reveals that there is only one real value for f for which

$$\psi_R' = 0$$

whatever the form of $h^*(f)$.

$$\psi_R'(f,t) = 2\pi[t - \frac{2h^*}{c}(f,t) - \frac{2f\partial h^*}{c\partial f}(f,t)] - 2\pi[f - f_o]/k \qquad (3.17)$$

$$= 0 \text{ when } f = f_s.$$

$$\text{Thus } kt - 2kh^*(f,t) - 2f_s k \frac{\partial h^*}{\partial f}(f,t)_{f=f_s} - f_s + f_o = 0$$

$$f_s = f_o + kt - \frac{2k}{c}\{h^*(f,t) + f_s \frac{\partial h^*}{\partial f}(f,t)\}_{f=f_s}$$

The term in brackets is seen to be identical to the group height h' at the frequency f_s, from equ. (3.7).

Thus, $f_s = f_o + kt - 2kh'(f_s,t)/c$

$$= f_T(t) - 2kh'(f_s,t)/c \qquad (3.18)$$

The frequency $f_s(t)$ does have some physical meaning. It will be shown later that $f_s(t)$ is equal to the instantaneous received frequency $f_R(t)$ without any Doppler component due to temporal variations in h^*.

From (3.7) and (3.17),

$$\phi'_R(f,t) = 2\pi[t - 2h'(f,t)/c] - 2\pi(f - f_o)/k$$

Thus $$\phi''_R(f,t) = 2\pi[-\frac{2\partial h'}{c\partial f}(f,t)] - 2\pi/k$$

The condition that ϕ''_R be positive is that

$$\frac{\partial h'}{\partial f} < -1.5 \text{ km Hz}^{-1}$$

Gradients this steep will only be found very close to critical frequencies and in this region an analytical solution to the integral is probably not possible; however the condition can be ignored for all practical purposes. Thus, we can assume that ϕ''_R is always negative, i.e. sign $\phi''_R = -1$ and

$$x_R(t) = A_4\cos[2\pi f_s\{t - \frac{2h^*}{c}(f,t)_{f=f_s}\} - \pi(f_s - f_o)^2/k - \pi/4 + \pi/4]$$

Substituting for f_s from (3.18) gives:

$$x_R(t) = A_4\cos[\varphi_R(t)]$$

where $\varphi_R(t)$ is the instantaneous phase of the received signal given by:

$$\varphi_R(t) = 2\pi[f_o + kt^2/2 - 2f_s\frac{h^*}{c}(f,t)_{f=f_s} - 2k\{\frac{h'}{c}(f,t)\}^2_{f=f_s}]$$

$$= \varphi_T(t) - 2\pi[2f_s\frac{h^*}{c}(f,t)_{f=f_s} + 2k\{\frac{h'}{c}(f,t)\}^2_{f=f_s}] \qquad (3.19)$$

The instantaneous frequency of the received signal is

$$f_R(t) = 1/(2\pi)\frac{d[\varphi_R(t)]}{dt}$$

$$= f_o + kt - \frac{2d}{cdt}[f_s h^*(f,t)]_{f=f_s} - 4kh'\frac{dh'}{dt}/c^2$$

$$= f_T(t) - \frac{2[f_s\frac{\partial h^*}{\partial f}(f,t)\frac{df}{dt} + \frac{\partial\{f_s h^*(f,t)\}dt}{\partial t}]_{f=f_s}}{dt} - 4kh'\frac{dh'}{dt}/c^2$$

$$= f_T(t) - \frac{2[f_s\frac{\partial h^*}{\partial f}(f,t)\frac{df}{dt} + \{f_s\frac{\partial h^*}{\partial t}(f,t) + h^*(f,t)\frac{\partial f_s}{\partial t}\}]_{f=f_s}}{c} - 4kh'\frac{dh'}{dt}/c^2$$

Now $\frac{\partial f_s}{\partial t} \simeq \frac{df}{dt}_{f=f_s} = k - \frac{2kdh'}{c\,dt}$ from (3.18).

Thus

$$f_R(t) = f_T(t) - \frac{2[(k - \frac{2kdh'}{cdt})\{h^*(f,t) + f_s\frac{\partial h^*}{\partial f}(f,t)\} + f_s\frac{\partial h^*}{\partial t}(f,t)]_{f=f_s}}{c}$$

$$- 4kh'\frac{dh'}{dt}/c^2$$

$$= f_T(t) - \frac{2[(k - \frac{2kdh'}{cdt})\{h'(f,t) + f_s\frac{\partial h^*}{\partial t}(f,t)\}]_{f=f_s}}{c} - 4kh'\frac{dh'}{dt}/c^2$$

$$= f_T(t) - \frac{2kh'}{c}(f_s,t) - 2f_s\frac{v^*}{c}(f_s,t) \qquad (3.20)$$

where $h'(f_s,t)$ is the group range at frequency f_s
and $v^*(f_s,t)$ is the phase velocity or rate of change of phase range at the frequency f_s with time.

A comparison of (3.18) and (3.20) illustrates the point made just after equation (3.18), i.e.

$$f_R(t) = f_s(t) - 2f_s v^*(f_s,t)/c$$

The received signal $x_R(t)$ is "demodulated" by removing the chirp

part of the signal. In practice, this is realised by using a heterodyne receiver which performs the function of subtracting the instantaneous transmitted and received phases. The instantaneous difference signal is thus:

$$x_D(t) = A_5\cos[\varphi_T(t) - \varphi_R(t)]$$

$$= A_5\cos[\varphi_D(t)]$$

$$\varphi_D(t) = 2\pi[(2h^*(f,t)f_s/c)_{f=f_s} + 2k(h'(f_s,t)/c)^2 \qquad (3.21)$$

from equation (3.19).

The last term on the right hand side of equation (3.21) can be thought of as a small phase correction given by the product of the frequency $2kh'/c$ and a small time increment h'/c. For typical values $k=10^5$ Hz s^{-1} and h' = 300 km, this term evaluates to $.4\pi$ radians. Thus, while not insignificant in absolute measurements of $\varphi_D(t)$, in measurements involving comparisons of (i.e. differences in) two values of φ_D for which h' can be considered constant, this correction term will subtract out, and can be ignored.

The instantaneous difference frequency is given by

$$f_D(t) = (1/2\pi)[\frac{d}{dt}(\varphi_T(t) - \varphi_R(t))]$$

$$= (1/2\pi)[\frac{d}{dt}(\varphi_D(t))]$$

$$= f_T(t) - f_R(t)$$

$$= 2kh'(f_s,t)/c + 2f_s v^*(f_s,t)/c \qquad (3.22)$$

from equation (3.20).

Equations (3.21) and (3.22) have two important consequences:

(1) From equation (3.21) it is seen that the phase of the demodulated "difference" signal $x_D(t)$ behaves, instantaneously, like the phase of a fixed frequency signal of frequency f_s, reflected from a phase range $h^*(f_s)$, plus a small, slowly varying correction that can be considered constant.

(2) From equation (3.22), the instantaneous frequency of the difference signal is equal to the sum of two terms, one proportional to the group range h', the other representing a Doppler contribution proportional to the frequency f_s and the phase velocity of the target. The second term is much more sensitive to changes in phase range with time than the first term. In standard chirpsounding, $v^*(f_s,t)$ is assumed to be zero, and the group range is evaluated simply from equation (3.22) by measurement of the frequency f_D through spectrum analysis. However, as was pointed out in Chapter 2, this assumption is by no means always valid. Equation (3.20) can be seen as a mathematical representation of the rôle of the range-Doppler ambiguity in attempts at practical evaluation of these parameters.

It can further be seen from equ.(3.20) that the "range" parameter τ used in the previous chapter in the discussion of ambiguity functions must be identified with the group delay of the probing signal, while the "velocity" parameter ν must be identified with the phase range velocity. Correspondingly, equations (2.7) and (2.8) can be re-written, for the ionospheric case, as:

$$\text{Range} = R = h'(f_s,t) = c\tau/2 \tag{3.23}$$

$$\text{Velocity} = V = v^*(f_s,t) = c\nu/2f_o \tag{3.24}$$

The phase range $h^*(f,t)$ is seen to be a function of the frequency of the probing signal, and time. In chirpsounding, the frequency is itself a

function of time, and this can lead to confusion, particularly when dealing with derivatives of h^* with respect to either parameter. Figs. (3.1),(3.2) and (3.3) attempt to clarify the relationships between the phase range, the group range, and the phase velocity when all are functions of both frequency and time, for a chirp signal.

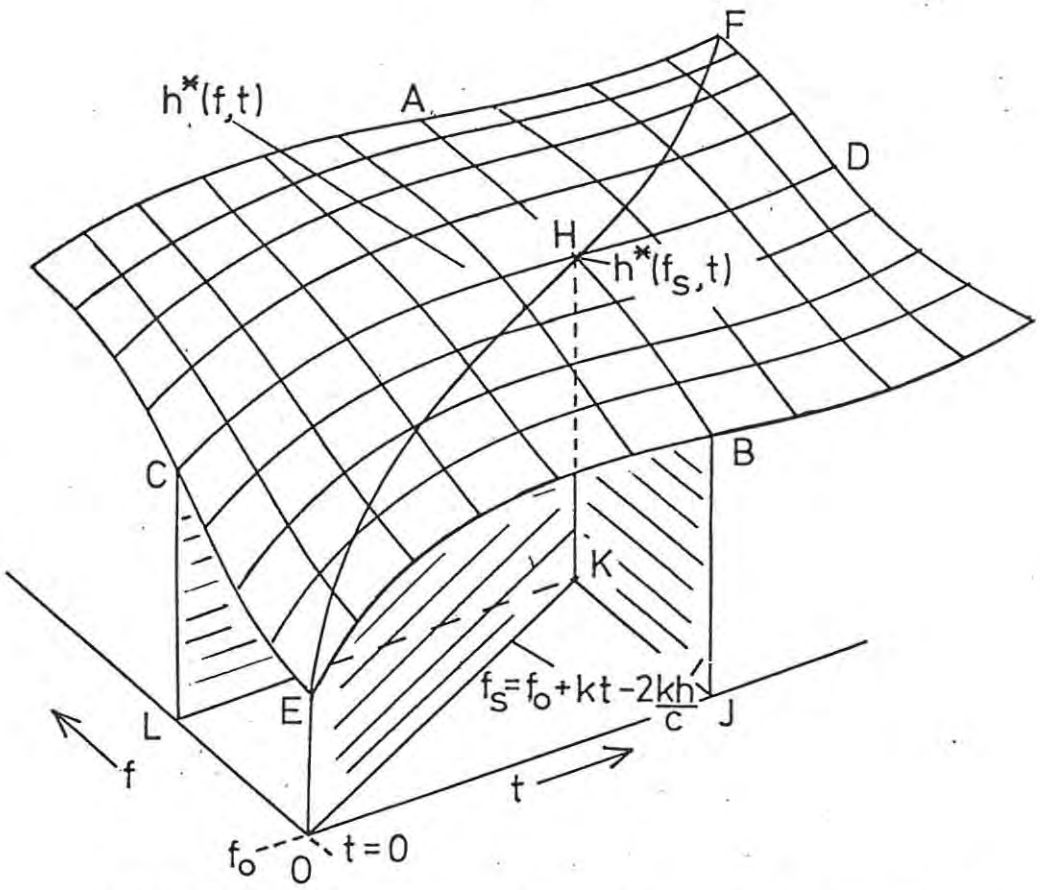
The surface ACEBDF in fig.(3.1) is a representation of the general bi-variant phase range function $h^*(f,t)$ as a function of frequency and time. The line OK is the locus of the instantaneous received chirp frequency which is a linear function of time given by $f_s = f_o + kt - 2kh'/c$. The intersection of a plane through OK, perpendicular to the plane LOJ, with this surface is given by EHF, which is the locus of values of the phase range $h^*(f_s,t)$ which contribute successively to the chirp reflection. The line CHD represents the locus of values of h^* at constant frequency. Figs.(3.1) and (3.2) illustrate that the tangent to this line at the point H gives the phase velocity $v^*(f_s,t)$ which, in general, will also be a slowly varying function of frequency and time. Similarly, the line AHB is the locus of values of h^* at constant time. The slope of the curve AHB at the point H is equal to the partial derivative of h^* with respect to frequency and can be used to evaluate the group range h' according to the relationship:

$$h'(f_s,t) = h^*(f_s,t) + f_s \cdot \frac{\partial h^*}{\partial f}(f_s,t)$$

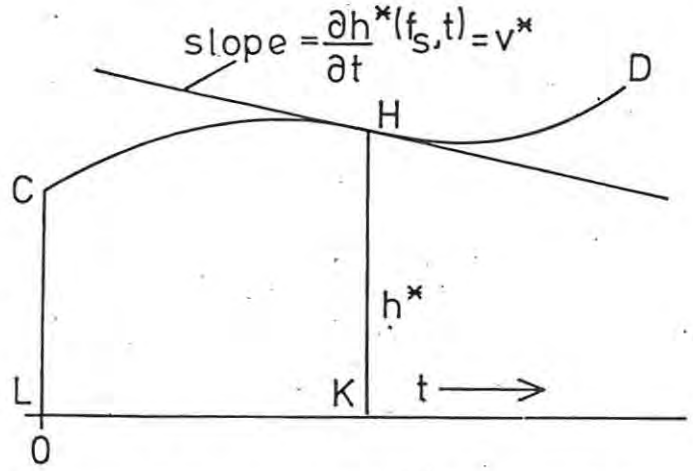
A construction illustrating this relation is shown in fig.(3.3).

3.5 Dual chirp modulation

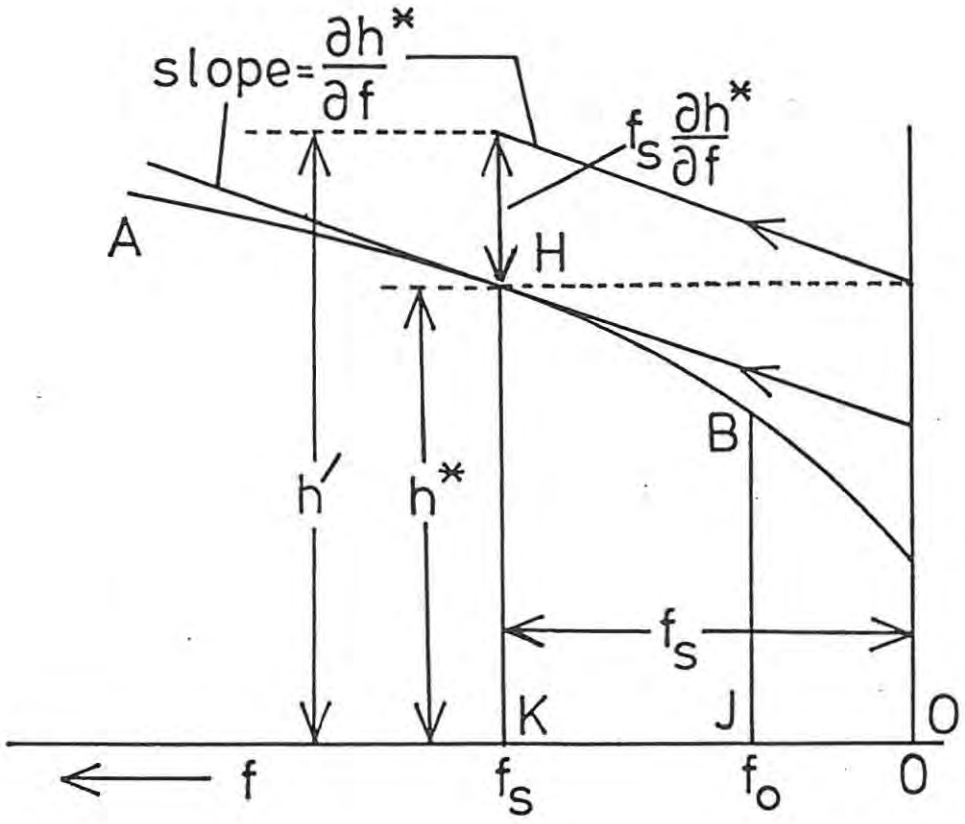
It has been shown that the use of a single chirp modulation leads to confusion of the range and Doppler variables. In Chapter 2, it was further shown that the use of dual modulation led to quantisation of the ambiguity function. This section will deal with the manner in which these ambiguous



FIGURE(3.1) A diagrammatic representation of the variation of the phase range h*(f,t) as a function of frequency and time. The line EHF is the locus of values of h* that will contribute to the reflection of a chirp signal of instantaneous frequency $f_s = f_o + kt - 2kh'/c$.



FIGURE(3.2) An example of the variation of h*(f,t) at constant frequency. the slope gives the phase range velocity v^*.



FIGURE(3.3) An example of the variation of h*(f,t) at constant time, from which the group range can be constructed as shown.

values can be distinguished.

Time diversity: Consider two identical chirp modulations separated by time Δt. For the first modulation, choose $f = f_o$ when $t = 0$, for the second, $f = f_o$ when $t = \Delta t$. It is now helpful to make the approximation $f_s = f_o$ when $t = 0$. Then from equ.(3.21), the phases of the difference frequencies at the start of the modulations are given by:

$$[\varphi_D(0)]_1 = 2\pi[(2h^*(f_o,0)f_o/c) + 2k(h'(f_o,0)/c)^2]$$

$$[\varphi_D(0)]_2 = 2\pi[(2h^*(f_o, \Delta t)f_o/c) + 2k(h'(f_o, \Delta t)/c))^2] \quad \ldots\ldots (3.25)$$

If one makes the reasonable approximation that h' remains constant with time, the difference in the initial phases is given by:

$$\Delta\varphi = [\varphi_D(\Delta t)]_2 - [\varphi_D(0)]_1 \quad (3.26)$$

$$= (4\pi f_o/c)[h^*(f_o,\Delta t) - h^*(f_o,0)]$$

Thus: $\Delta\varphi/\Delta t = (4\pi f_o/c)[h^*(f_o,\Delta t) - h^*(f_o,0)]/\Delta t$

The expression in square brackets is seen to be an approximation to the partial derivative of the phase range at $f = f_o$ with respect to time, or $v^*(f_o,0)$.

Thus: $$v^*(f_o,0) \simeq (c/4\pi f_o)\,[\Delta\varphi/\Delta t] \quad (3.27)$$

where $\Delta\varphi$ = difference between initial phases

Δt = modulation spacing

Equation (3.27) provides a means to evaluate v^* that is independent of the range delay, provided $\Delta\varphi$ can be evaluated.

Frequency diversity: The above discussion concerned the use of two chirp modulations with identical start frequencies and different start times, enabling an evaluation of the phase velocity. In a completely analogous fashion, one can employ two modulations with identical start times and different start frequencies, which could be expected to provide an independent evaluation of the group range of the reflection point.

Consider two chirp modultations with identical start times $t = 0$ and frequencies f_o and $f_o + \Delta f$. Then, analogous to equs.(3.25),

$$[\varphi_D(0)]_1 = 2\pi[2h^*(f_o,0)f_o/c + 2k(h'(f_o,0)/c)^2]$$

$$[\varphi_D(0)]_2 = 2\pi[2h^*(f_o+\Delta f,0)(f_o+\Delta f) + 2k\{h'(f_o+\Delta f,0)/c\}^2]$$

from which:

$$\Delta\varphi \simeq 2\pi[(2/c)\{(f_o + \Delta f).h^*(f_o + \Delta f,0) - f_o.h^*(f_o,0)\}]$$

Thus: $\Delta\varphi/\Delta f \simeq (4\pi/c)[(f_o + \Delta f).h^*(f_o + \Delta f,0) - f_o.h^*(f_o + \Delta f,0)]/\Delta f$

where the expression in brackets is an approximation to the partial derivative of the product $f.h^*$ with respect to f, which from equ.(3.7) is the group range $h'(f,t)$.

Thus, $$h'(f,t) \simeq (c/4\pi)[\Delta\varphi/\Delta f] \qquad (3.28)$$

where once again, $\Delta\varphi$ is the unknown to be evaluated.

Ambiguities: Equations (3.27) and (3.28) provide the means to evaluate either the group range or the phase velocity independently, by measurement of the difference in the initial phase $\Delta\varphi$. However, such an angular evaluation always carries with it an inherent 2π ambiguity. Using equs. (2.13) and (2.14), the relations (3.27) and (3.28) can be re-written:

$$v^*(f_o)_n = (c/4\pi f_o)[n\,\Delta\varphi/\Delta t] = (c/4\pi f_o)\nu_n \qquad (3.29)$$

$$h'(f_o,t)_n = (c/4\pi)[n\Delta\varphi/\Delta f] = (c/4\pi)\,\tau_n \qquad (3.30)$$

$\nu_n = n\Delta\varphi/\Delta t = n^{th}$ Doppler frequency ambiguity (3.31a)

$\tau_n = n\Delta\varphi/\Delta f = n^{th}$ group delay ambiguity (3.31b)

$n = ...-2,-1,0,1,2,....$

Fig.(3.4) shows once again the ambiguity diagram for two chirp modulations of identical frequency but separated in time by Δt. The separation of the projections of the ambiguities onto the ν axis is $1/\Delta t$, in confirmation of equ.(3.31a). Note that the loci of points for which $|\chi(\tau,\nu)|$ is zero are straight lines running parallel to the τ axis.

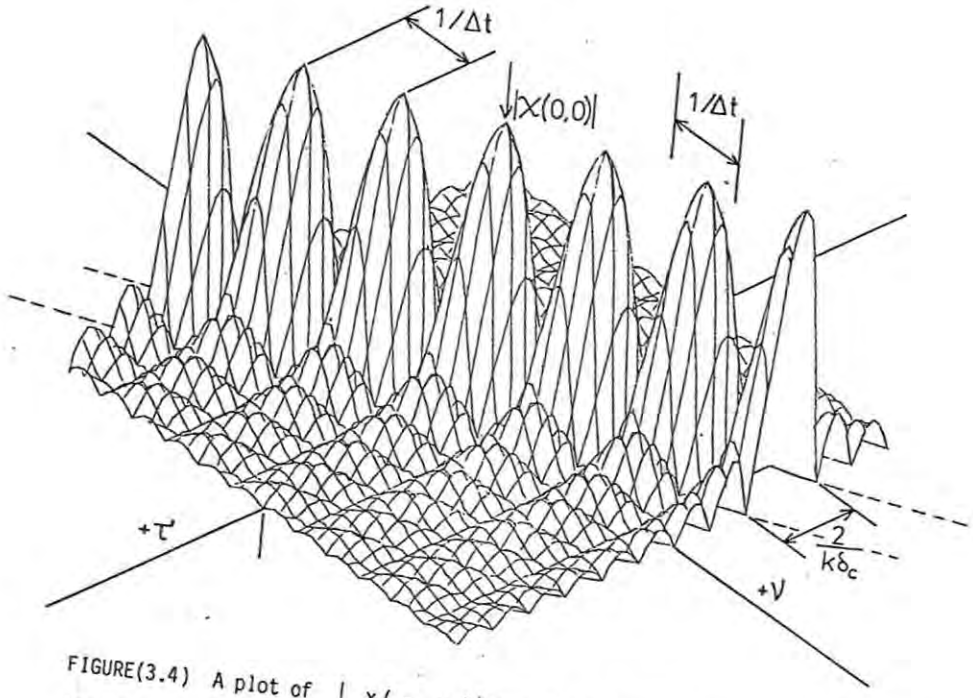
For comparison, fig.(3.5) shows the ambiguity diagram for two chirp modulations transmitted at the same time, but with a frequency increment Δf between them. The separation of the projections of the ambiguities onto the τ axis is $1/\Delta f$, in confirmation of equ.(3.31b). This time the loci of $|\chi(\tau,\nu)| = 0$ are straight lines running parallel to the ν axis.

Determination of the phase angle $\varphi_D(0)$: The phase $\varphi_D(0)$ is the phase of the difference signal at the receiver output at time $t = 0$. From equ.(3.22), the instantaneous difference frequency is given by:

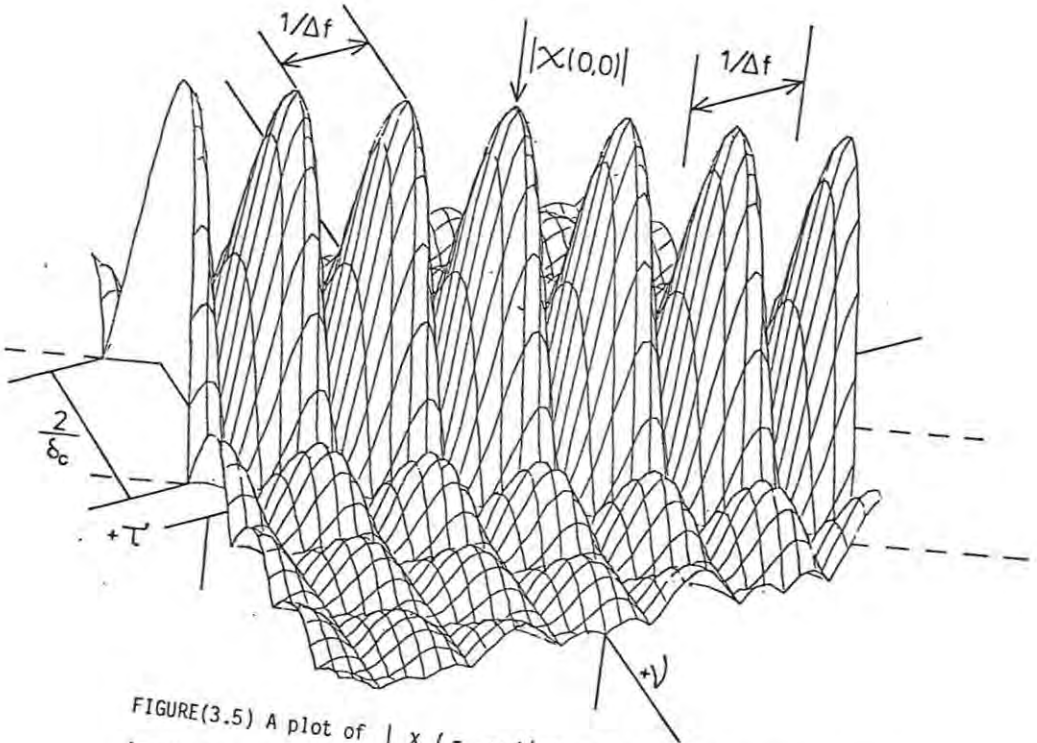
$$f_D(t) = (2k/c)h'(f_s,t) + (2f_s/c)v^*(f_s,t)$$

For the purpose of illustration, one can assume that h' and v^* are slowly varying functions of time, and thus consider f_D to be constant with respect to time. In this case, the difference signal can be expressed as

$$x_D(t) = A_3\cos[2\pi f_D t + \varphi_D(0)] \qquad (3.33)$$



FIGURE(3.4) A plot of $|\chi(\tau,\nu)|$ near (0,0) for two chirp modulations of length δ_c and identical start frequency f_o, separated in time by an amount $\Delta_c = \delta_c = 0.5s$. The chirp rate $k = 5*10^4$ Hz s^{-1}.



FIGURE(3.5) A plot of $|\chi(\tau,\nu)|$ near (0,0) for two chirp modulations of length δ_c, transmitted simultaneously, separated by a frequency increment $\Delta f = 5*10^4$ s^{-1}. The chirp rate $k = 5*10^4$ Hz s^{-1}.

In general, $x_D(t)$ will have a finite duration which can be well approximated by the duration of the chirp modulation δ_c. The phase and amplitude spectra $\Phi(f)$ and $X(f)$ for such an idealised signal $x_D(t)$ are shown in fig.(3.6). Because the time origin has been chosen at the start of the signal instead of centrally as is usual with transform analysis, the phases of the frequency components f that are not equal to f_D vary linearly as:

$$\Phi(f) = \Phi(f_D) + \pi(f_n - f_D)/\delta_c - \pi[\{1 - \text{sign}(\sin[\pi(f_n - f_o)/\delta_c])\}/2$$

where $\Phi(f_D)$ is the phase of the difference signal, $\varphi_D(0)$.

Provided the signal contains no frequency components above the upper limit f_Q called the Nyquist frequency, and provided $T < 1/2f_Q$, it is well known that $x_D(t)$ can be completely represented by a set of N samples equally spaced by an amount T. Correspondingly, the spectrum of $x_D(t)$ can be completely represented by a set of samples in the frequency domain, given by the discrete Fourier transform of $x_D(t)$ as follows:

$$X_n = \sum_{k=1}^{N-1} x_k e^{-j2\pi f_n t_k} \quad n = 0,1,2....N-1$$

where $x_k = x(t_k)$ is the k^{th} discrete time sample

$X_n = X(f_n)$ is the n^{th} discrete frequency sample

t_k = time at which sample x_k is taken

f_n = frequency at which sample X_n is taken

Since $t = 0$ is taken at the start of the signal, $t_k = kT$. It can be shown that the f_n are given by $f_n = n/NT$

Thus:

$$X(n/NT) = \sum_{k=0}^{N-1} x(kt)e^{-j2\pi nk/N} \quad n = 0,1,2.....N-1 \qquad (3.34)$$

In general f_D will not coincide with any of the discrete sample points $f_n = n/NT$, as illustrated in fig.(3.6). If the sample at f_n is chosen as representative of f_D, it will in general not be equivalent to the component at f_D either in amplitude or phase. The true amplitude $X(f_D)$ can be retreived from the samples by interpolation involving the convolution of the samples with the complex "window" used to truncate $x_D(t)$, but such an amplitude only has direct meaning when $x_D(t)$ has some known form such as that of equ.(3.33). Usually $x_D(t)$ will not be of constant amplitude or frequency during the measurement interval and a more meaningful parameter is the total energy in the signal, given by:

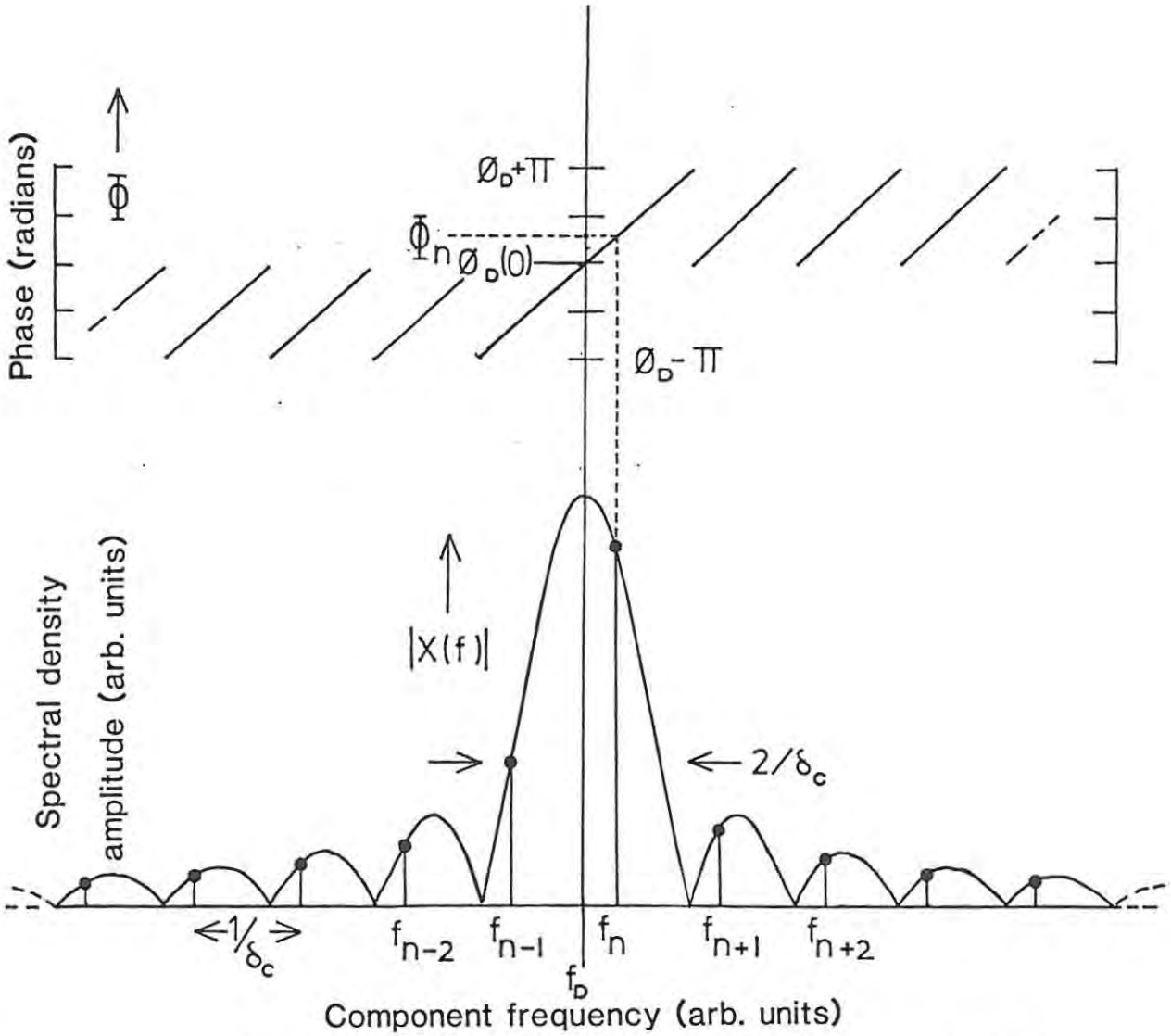
$$E = (1/2)\int_{-\infty}^{\infty}|X(f)|^2\,df = (NT/2)\sum_{n=0}^{N-1}|X(n/NT)|^2 \qquad (3.35)$$

Likewise, the phase $\Phi(f_D) = \varphi_D(0)$ of the ideal signal $x_D(t)$ can be obtained by interpolation. However, from equs.(3.27) and (3.28) it is seen that v^* and h' are evaluated not from $\varphi_D(0)$ but rather from $\Delta\varphi$ which is, from equ.(3.26), given by <u>differences</u> in measured values of $\varphi_D(0)$. In general, these differences can be well approximated by differences in the phases of a representative frequency component f_n, i.e. equ.(3.26) can be re-written:

$$\Delta\varphi = (\Phi_n)_2 - (\Phi_n)_1 \qquad (3.36)$$

where Φn can be any frequency component, but is usually selected to be the one with the highest amplitude to ensure good signal-to-noise ratio.

<u>Fast Fourier Transform (FFT)</u>: In the preceding section it was shown that the independent estimates of v^* and h' could be made by a measurement of the change in the initial phases of the difference signal $x_D(t)$ obtained from the two chirp modulations, which would be, respectively, time or frequency



FIGURE(3.6) A plot of the amplitude and phase spectra for a truncated signal of fixed frequency f_D and duration δ_c, when the origin of time is chosen to coincide with the start of the signal.

diverse. Furthermore, this change could be well approximated by the difference in the phases of a single discrete Fourier component of $x_D(t)$, provided that the amplitude was sufficiently large.

The fast Fourier transform (henceforeward referred to as the FFT) is an algorithm developed to evaluate the discrete Fourier components f_n from a set of N equally spaced samples of the time domain signal, x_k. The relationship on which the algorithm is based is expressed in equ.(3.34). Details of the derivation [2] of the FFT are by now well-known and will not be repeated here. However, it is important to note the following points about the FFT algorithm:

(i) Given a set of N samples of the difference signal $x_D(t)$ spaced T apart, the FFT will provide a set of N discrete complex spectral densities spaced 1/NT apart, that is, N real and N imaginary amplitudes. The first N/2 complex points represent the $X_n(+f)$, $n = 0,1,\ldots(N/2)-1$, while the next N/2 points represent $X_n(-f)$, $n = -(N/2),-(N/2)-1,\ldots-1$. A good illustration of this is given in [2], pages 136 and 142. The complex amplitude $X_n(f)$ can be expressed as:

$$X_n(f) = R_n(f) + jI_n(f) \tag{3.37}$$

where R_n and I_n are the real and imaginary amplitudes produced by the FFT. If $x_D(t)$ is real, then:

$$R_n(f) = R_n(-f) \text{ and } I_n(f) = -I_n(-f) \tag{3.38}$$

The phase of the n^{th} component is given by ([3], pg 14):

$$\Phi_n = \tan^{-1}[-j\{X_n(f) - X_n(-f)\}/\{X_n(f) + X_n(-f)\}]$$

Substitution from (3.37) and (3.38) gives:

$$\Phi_n = \tan^{-1}[I_n(+f)/R_n(+f)] \tag{3.39}$$

Also, the power spectrum is given by:

$$P_n(f) = X(f).X^*(f) = R_n(f)^2 + I_n(f)^2 \qquad (3.40)$$

If the n^{th} real component is expressed in the form:

$$x_n(t) = A_n\cos[2\pi f_n t] + B_n[2\pi f_n t]$$

the phase is given by:

$$\Phi_n = \tan^{-1}[-B_n/A_n]$$

Thus $A_n = R_n(+f) = R_n(-f)$

$$B_n = -I_n(+f) = I_n(-f)$$

This relationship between the FFT amplitudes R(+f) and I(+f) and the real amplitudes A and B has been emphasized deliberately because confusion about the correct sign of the imaginary amplitude results in confusion about the sign of Φ, which is central to the measurement of many parameters.

(ii) The zero of time in FFT algorithms is taken at the start of the signal so that the phase Φ_n is the phase at this time. However, since it is always phase differences that are used, the time at which the phase is evaluated is immaterial.

3.6 Modulation independent measurables:

Sections 3.2 to 3.5 deal with the measurement of the range and Doppler parameters τ and ν for which the accuracy and resolution are functions of the modulation u(t). This section will deal with the other measurable signal parameters that do not depend on u(t) and are thus common to all radars. The measurement of these parameters is characterised by spatially diverse sampling of the signal.

Antennas: As with τ and ν, the measurement of the spatially dependent parameters will involve signal phase comparisons. The device which makes possible the sampling of the signal as a function of space is the antenna. A correctly chosen antenna will exhibit two properties which will affect the phase of the received signal, namely, its position and its orientation. Following Wright [4], the antenna chosen in this application is the so-called "short dipole", where the term "short" refers to the fact that the dipole is electrically short in comparison with the shortest expected signal wavelength. Further discussion of the antennas and alternatives is given in Chapter 4.

Arrival angle: The arrival angle of the signal in one plane can be calculated from the measured signal phase difference on two antennas that are spatially separated but have the same orientation, as illustrated in fig.(3.7). If $(\varphi_R)_1$ and $(\varphi_R)_2$ are the instantaneous phases on antennas 1 and 2 respectively, and θ is the arrival angle as shown, then:

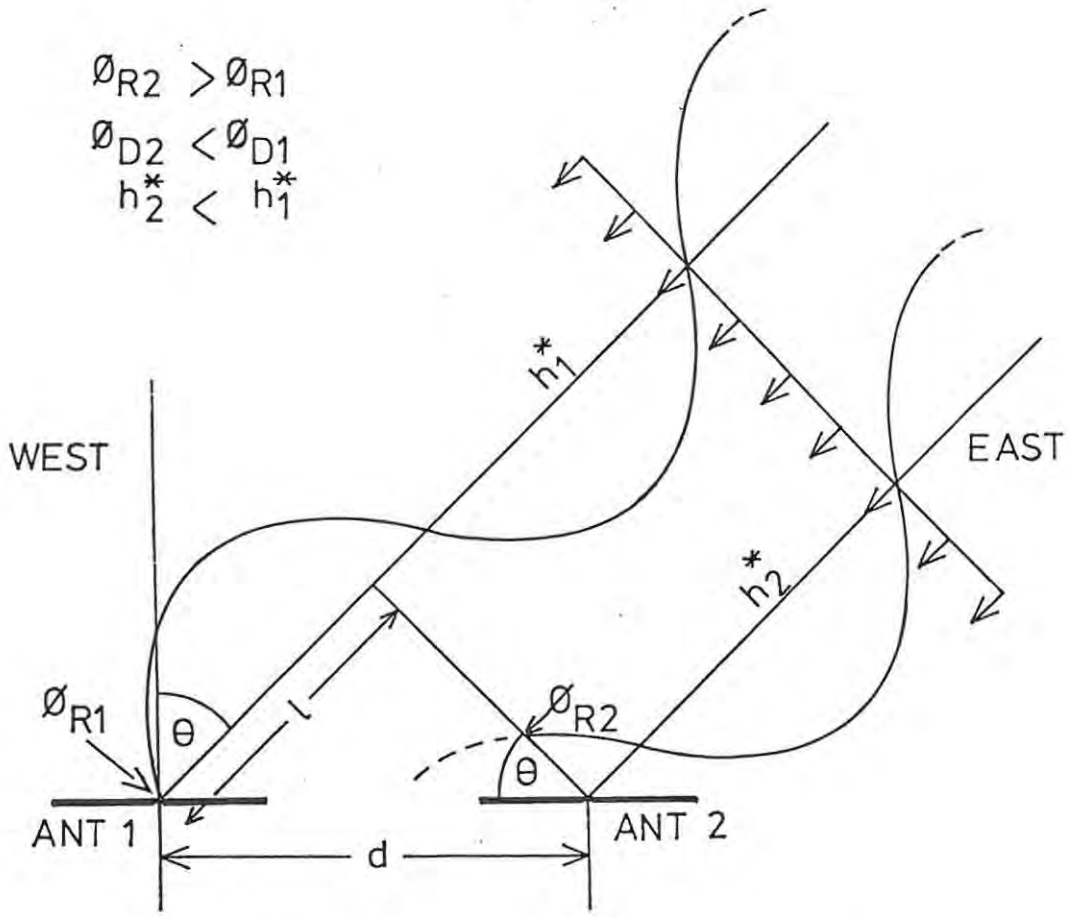
$$1/\lambda = [(\varphi_R)_1 - (\varphi_R)_2]/2\pi \quad \text{where} \quad \lambda = c/f_o$$

and $1/d = \sin[\theta]$, whence:

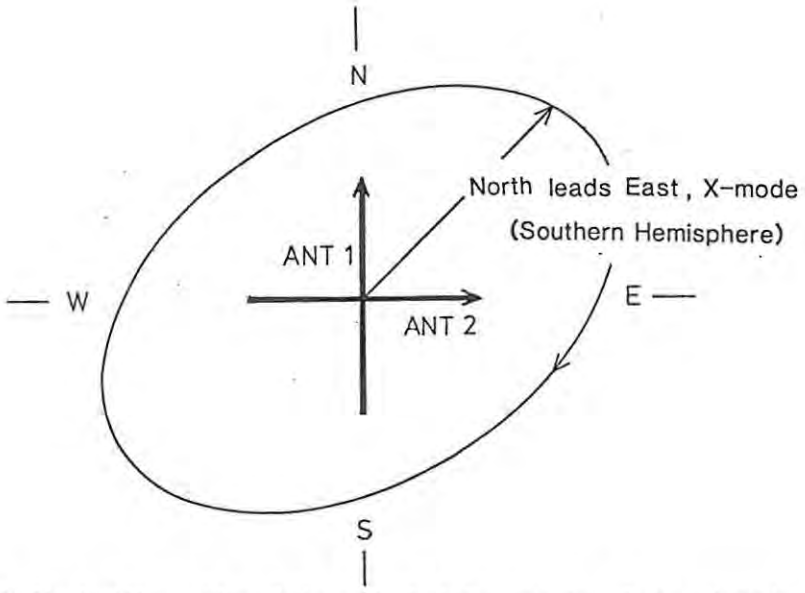
$$\sin[\theta] = c[(\varphi_R)_2 - (\varphi_R)_1]/2\pi f_o d$$

The phase path h^*_2 from the transmitter to antenna 2 is shorter than h^*_1, but $(\varphi_R)_2$ is seen to be greater than $(\varphi_R)_1$. Thus the instantaneous phase on the antenna varies in a negative sense with h^*. However, from just before equation (3.21), it is seen that the phase of the signal at the receiver output is formed by subtracting φ_R from the instantaneous transmitted phase φ_T, so that:

$$\Delta\varphi = \varphi_D(0)_2 - \varphi_D(0)_1$$



FIGURE(3.7) An illustration of the relationship between the antenna phases φ_R, the receiver output phases φ_D and the phase range h^*, for a non-vertically propagating reflected wave.



FIGURE(3.8) An illustration of the anti-clockwise rotation of the electric vector at the ground for a downcoming extraordinary ray reflection in the Southern Hemisphere. The phase on antenna 2 lags the phase on antenna 1. For the receiver phases the reverse is true.

$$\Delta\varphi = \varphi_T - (\varphi_R)_2 - [\varphi_T - (\varphi_R)_1]$$

$$= -[(\varphi_R)_2 - (\varphi_R)_1]$$

Therefore $\sin\theta = -c\Delta\varphi/2\pi f_o d$ (3.41)

As before, the phase difference $\Delta\varphi$ can be approximated well in practice by the difference in the phases of the discrete Fourier spectral amplitudes obtained via the FFT.

Another way to look at the relationship between the sign of the phase difference $\Delta\varphi$ and the direction of arrival is to note from equ.(3.21) that $\varphi_D(0)$ varies in a positive sense with h^*. Thus, if $\varphi_D(0)_E - \varphi_D(0)_W$ is negative, $h_E^* - h_W^*$ must be negative, thus $h_E^* < h_W^*$ and the ray must come from the East.

To obtain the full 3-dimensional arrival angle, this procedure should be repeated in three orthogonal directions; however, since half the possible arrival angles are excluded by the earth's surface, only two orthogonal antenna pairs are required.

Polarisation: In an analogous way to the determination of the arrival angle, if two antennas have the same position, but are orientated differently, it is possible to determine the sense of polarisation of the downcoming wave, as illustrated in fig.(3.8). In general, the ordinary and extraordinary rays will, on arrival at the ground from vertical incidence, have elliptical polarisations with opposite senses and differently orientated major axes ([1], pg. 51). The difference in the sense of rotation of the electric vector for the two modes enables one to differentiate between them by means of phase comparisons on two orthogonally orientated antennas. In the Southern Hemisphere, in the horizontal plane of the antennas, the electric vector rotates from East towards North for the ordinary ray, and from North

towards East for the extraordinary ray, looking down in the direction of ray propagation. The reverse is true for the Northern Hemisphere, while at the magnetic equator the two polarisations are linear, but not parallel.

Fig.(3.8) illustrates schematically the right-handed rotation of the electric vector for the extraordinary ray in the Southern Hemisphere, where the observer is looking down at the earth in the direction of the wave normal. The electric signal on the West-East oriented antenna will lead the signal on the South-North antenna by a value near $+\pi/2$, thus:

$$\Psi = (\varphi_R)_{EW} - (\varphi_R)_{NS}$$

According to this definition, the polarisation mode is related to the sign of the angle Ψ according to:

$$\text{sign } [\Psi] = \begin{cases} \text{positive, r.h. polarisation (X)} \\ \text{negative, l.h. polarisation (O)} \end{cases}$$

As before, because the receiver phases change in the opposite sense to the antenna phases, the receiver phase difference for two receivers connected to an imaginary crossed antenna pair, is given by:

$$\varphi_{OX} = (\varphi_D')_{EW} - (\varphi_D')_{NS}$$

$$= -[(\varphi_R)_{EW} - (\varphi_R)_{NS}]$$

$$= -\Psi$$

Thus $$\text{sign } [\varphi_{OX}] = \begin{cases} \text{negative, l.h. polarisation (O)} \\ \text{positive, r.h. polarisation (X)} \end{cases}$$

We see that the angle is positive if the rotation of the vector is clockwise (r.h. polarisation). If we define the orientation angle of the two antennas as being $\varphi_{OR} = \pi/2$, it is possible to think of the polarisation as

a "rate of change of phase with antenna orientation" by defining the dimensionless parameter:

$$P = \frac{\varphi_{OX}}{\varphi_{OR}}$$

The relationship between the components of the electric vector and and φ_{OX} can be derived as follows. If the electric field is represented by:

$$E_{NS} = A_{NS}\cos[\omega t + (\varphi_R)_{NS}] \qquad (3.42)$$

$$E_{EW} = A_{EW}\cos[\omega t + (\varphi_R)_{EW}] \qquad (3.43)$$

then the polarisation angle can be defined as

$$\Psi = (\varphi_R)_{EW} - (\varphi_R)_{NS}$$

Now $\varphi_{OX} = -\Psi$

By shifting the phase in equations (3.42) and (3.43) by subtracting $(\varphi_R)_{EW}$, we get:

$$E_{NS} = A_{NS}\cos(\omega t - \Psi)$$

$$E_{NS} = A_{NS}\cos(\omega t + \varphi_{OX}) \qquad (3.44)$$

$$E_{EW} = A_{EW}\cos(\omega t) \qquad (3.45)$$

3.7 Errors in the measurement of $\varphi_D(0)$.

In sections 3.5 and 3.6 it was shown that the quantities h', v^*, the arrival angle and the polarisation can be evaluated by measurement of the difference in the phases of echo signals that are in some respect diverse. Since such measurement will always involve some unwanted noise component, it is of interest to investigate the effect of such noise on the phase of the signal.

Let us assume that we wish to evaluate some parameter by means of

measuring the change in the phase of i^{th} component of the spectrum of the difference signal. We thus need to know what the expected error in the phase φ_i is likely to be.

The i^{th} component of the signal can be expressed as:

$$E_S(i) = S_i\cos(2\pi f_i t + \varphi_i) \qquad (3.46)$$

Similarly, we can visualise an arbitrary noise component at the same frequency given by:

$$E_N(i) = N_i\cos(2\pi f_i t + \varphi_i + \varepsilon)$$

where ε is the difference in phase between the signal and the noise component. Bearing in mind that every measurement of the phase φ_i will have a $\pm 2n\pi$ ambiguity associated with it, we can, without loss of generality, rotate both vectors so that $\varphi_i = 0$. Thus, dropping the subscript i,

$$E_S = S\cos(2\pi ft)$$

$$E_N = N\cos(2\pi ft + \varepsilon)$$

The resultant is given by:

$$E_{S+N} = S\cos(2\pi ft) + N\cos(2\pi ft + \varepsilon)$$

Using standard trigonometric functions it can be shown that:

$$E_{S+N} = A\cos(2\pi ft + \theta) \qquad (3.47)$$

where $A = [S^2 + N^2 + 2SN\cos(\varepsilon)]^{1/2}$

and $\theta = \tan^{-1}\left[\dfrac{N\sin(\varepsilon)}{S + N\cos(\varepsilon)}\right]$

From a comparison of equations (3.46) and (3.47) it is seen that θ is the error introduced into the phase of E_S by the noise component. If we

define the ratio of the noise voltage to the signal voltage as X = N/S, we have:

$$\theta = \tan^{-1}\left[\frac{X\sin(\varepsilon)}{1 + X\cos(\varepsilon)}\right] \qquad (3.48)$$

Note that for $\varepsilon = \pm\pi/2$, $\theta = \tan^{-1}[\pm X]$, thus the function $\theta/[\tan^{-1}(X)]$ will be normalised to equal ± 1 when $\varepsilon = \pm\pi/2$.

This function is plotted in figure (3.9) for different values of the ratio X. Two interesting limiting values are evident:

(i) X = 1 $\qquad \theta(X=1) = \tan^{-1}\dfrac{\sin(\varepsilon)}{1 + \cos(\varepsilon)}$

$$= \tan^{-1}[\tan(\varepsilon/2)]$$

$$= \varepsilon/2$$

(ii) $X \to 0 \qquad \lim_{X \to 0} \dfrac{X\sin(\varepsilon)}{1 + X\cos(\varepsilon)} \to X\sin(\varepsilon)$

However, for small angles, $X\sin(\varepsilon) \to \tan[X\sin(\varepsilon)]$, thus $\theta \to X\sin(\varepsilon)$. Also, $X \to \tan(X) \to \tan^{-1}(X)$

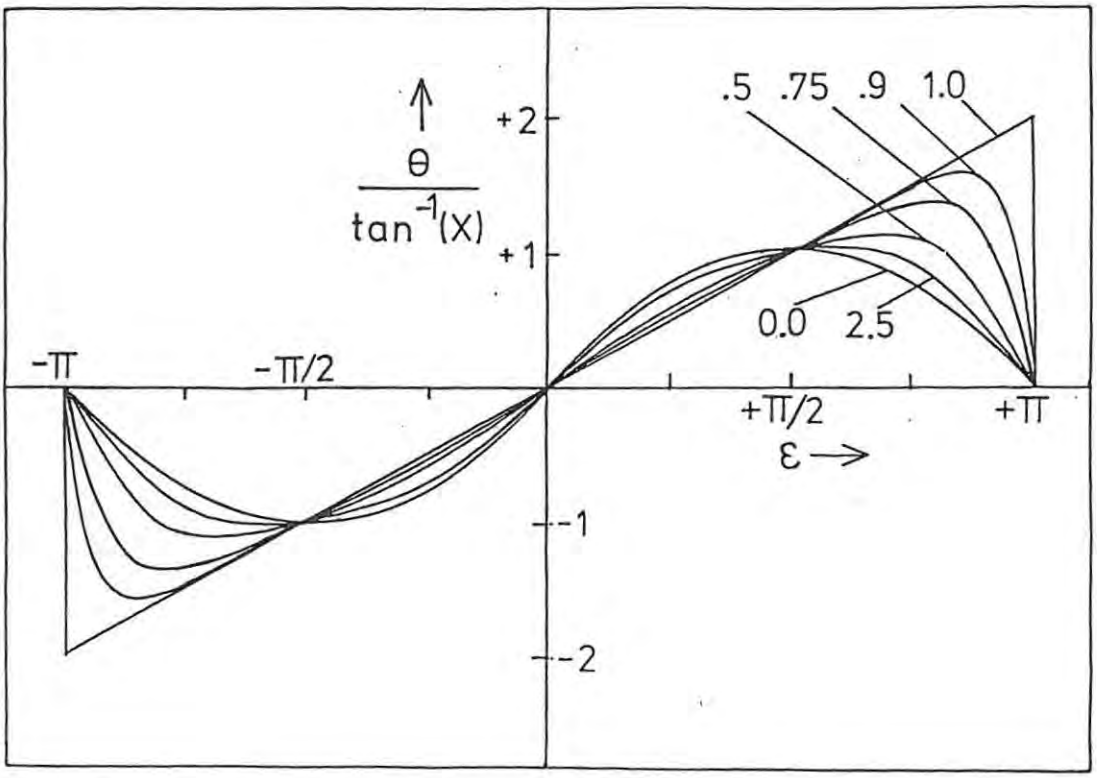
Therefore, $$\frac{\theta}{\tan^{-1}(X)} \to \frac{\theta}{X} \to \sin(\varepsilon)$$

Both these results are seen in figure (3.9). The graphs for N/S >1 are also interesting, but not of importance to this discussion.

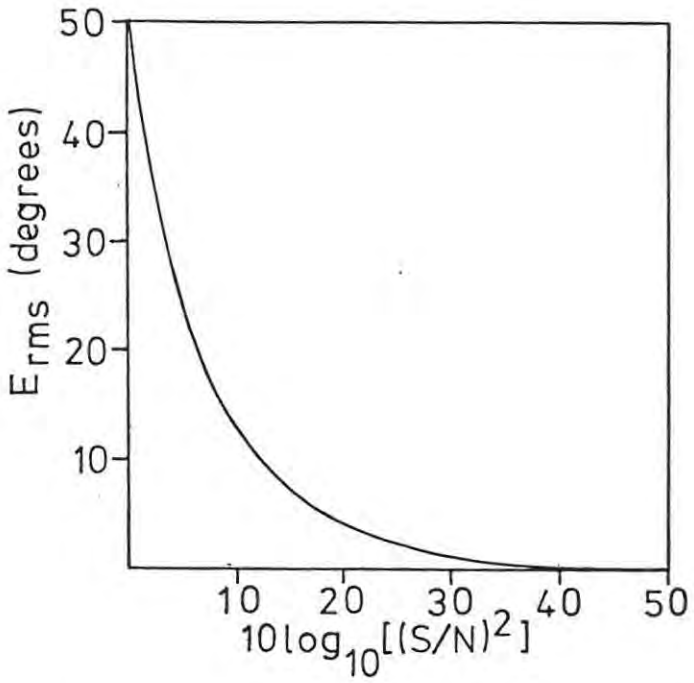
Knowing the distribution of θ as a function of ε for given X, it is possible to evaluate a root-mean-square expectation value for θ which we can call the rms error E_{rms}. This is defined as

$$E_{rms}(X) = (1/2\pi).\left[\int_{-\infty}^{\infty} \theta^2(\varepsilon, X)d\varepsilon\right]^{1/2}$$

By substituting for θ from equation (3.48) and integrating



FIGURE(3.9) The variation of the error in the phase measurement θ as a function of the phase difference ε between the signal and the noise components. X is the ratio of the noise to the signal voltage (N/S).



FIGURE(3.10) The variation of the root mean square value of θ , averaged over all ε .

numerically, E_{rms} has been evaluated as a function of X. This relationship is shown in figure (3.10), where X is expressed in the more usual form in dB, i.e. $10\log_{10}(1/X)^2$ where $(1/X)^2$ is called the signal-to-noise ratio.

3.8 Summary:

It is of interest to collect the results of Chapter 3 as follows:

phase range velocity	$= v^*$	$= [c/4\pi f_o](\Delta\varphi/\Delta t)$	(3.49)
group range	$= h'$	$= [c/4\pi](\Delta\varphi/\Delta f)$	(3.50)
sin arrival angle	$= \sin$	$= [-c/2\pi f_o](\Delta\varphi/d)$	(3.51)
polarisation sense	$= P$	$= -\text{sign}[\Delta\varphi/\varphi_{OR}]$	(3.52)

It is seen that all four of the above quantities are evaluated from a measured phase difference. One can express this phase difference as a function of four parameters as follows:

$$\Delta\varphi = \Delta\varphi(\Delta t, \Delta f, d, \varphi_{OR})$$

where Δt = separation of the two modulations in time

Δf = separation of the two modulations in frequency

d = separation of the two modulations in space

φ_{OR} = separation of the two modulations in orientation

There is therefore a unifying aspect to the above results in that they can all be evaluated from the partial derivative of the phase with respect to one of the four parameters, with all other parameters being held constant. The phase difference $\Delta\varphi$ can in all cases be well approximated by the difference in the phases of a discrete Fourier component of the signal.

3.9 A 3-CELL SOUNDING STRUCTURE

Comprehensive descriptions of a sounding structure for the measurement of

the above parameters for a pulse ionosonde have been given by Bibl and Reinisch [5], and Wright and Pitteway [(6),(7)]. This section describes a structure for chirpsounding analogous to that of Wright and Pitteway. The total unified structure to measure all the parameters will be referred to as a "sounding", and in this case, a "sounding" will consist of 3 "cells", each of duration T. This is illustrated in Fig (3.11). During each cell two of the four possible receiving antennas are connected to the twin phase-matched receivers as indicated. The antenna configuration and phase analysis is based on that described by Wright and Pitteway [7], and use is made of their concept of virtual crossed antenna pair at the centre of the array.

Definitions: (i) the measured phase angles are φ_{W1}, φ_{E1}, φ_{W2}, φ_{N2}, φ_{N3} and φ_{S3} where the first subscript refers to the antenna to which the receiver is coupled and the second to the cell during which it is measured.

(ii) the phase differences which we hope to evaluate are

φ_{EW} = angle of arrival phase difference on E/W receiver outputs.

φ_{NS} = angle of arrival phase difference on N/S receiver outputs.

φ_{DP} = "Doppler" phase difference on West receiver over 1 cell.

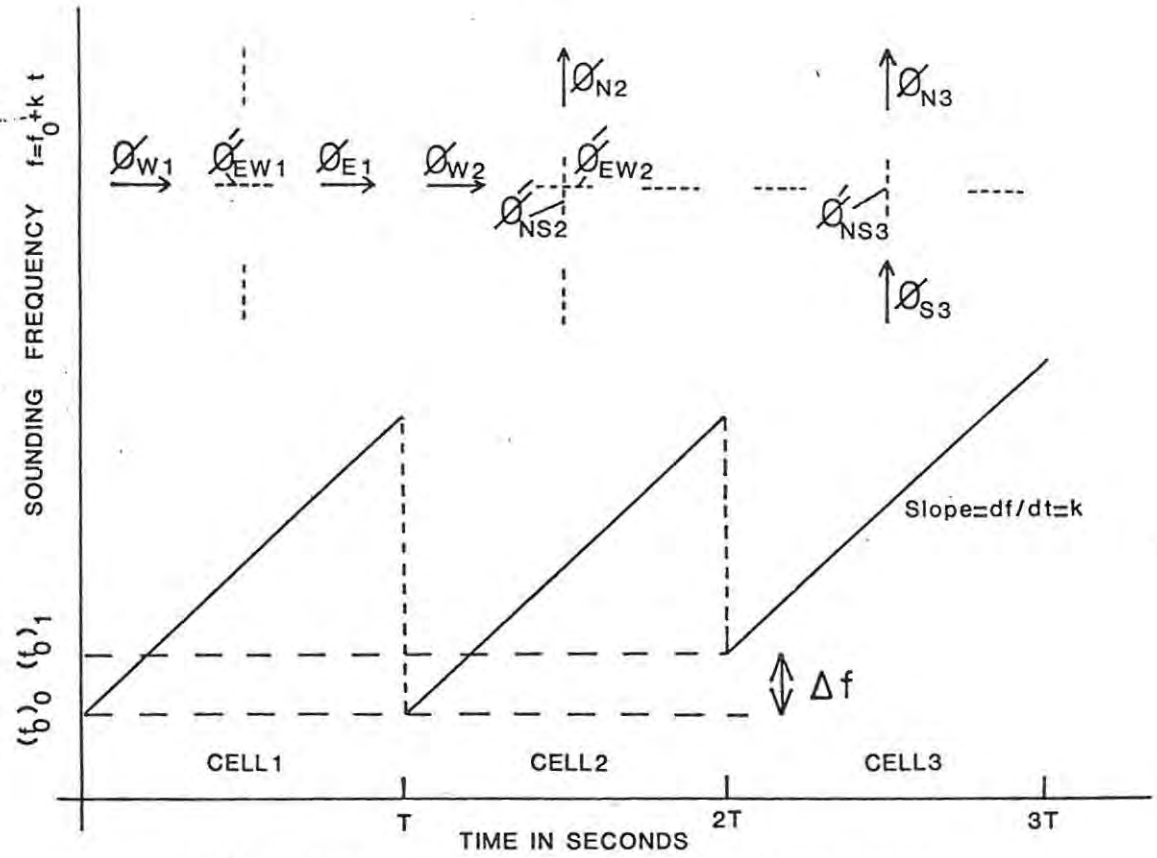
φ_{OX} = polarisation phase difference between receivers connected to orthogonal virtual antennas at the centre of the array during cell 2.

φ_{V} = "virtual height" phase difference on North receiver output caused by a cell start frequency (f_o) increment (Δf) between cells 2 and 3.

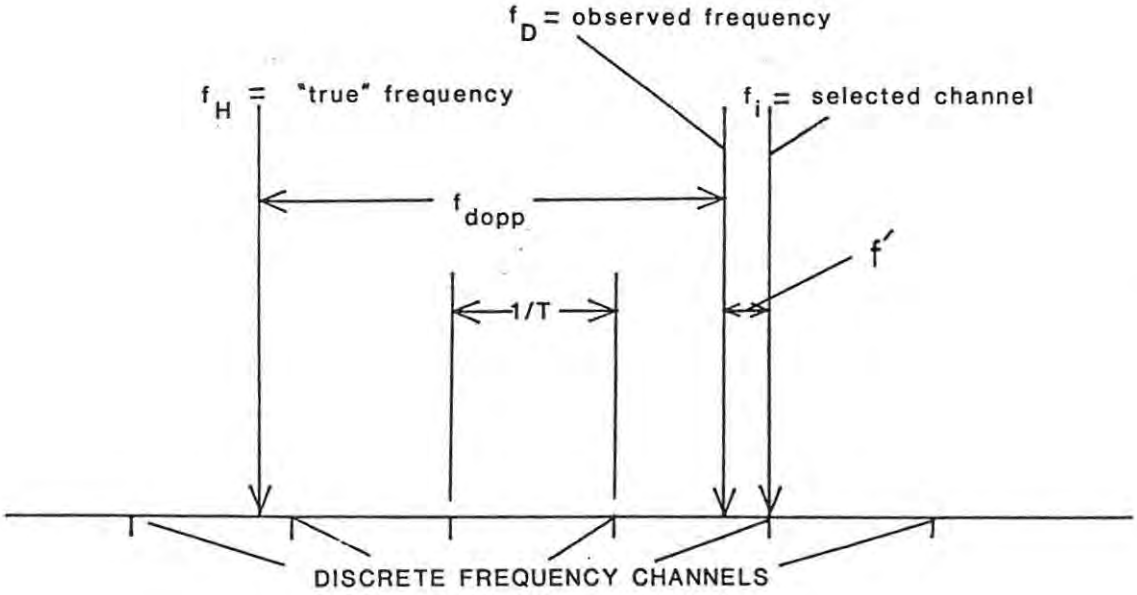
(iii) the phases of receiver outputs connected to virtual antennas used as an intermediate step in the evaluations of (ii) are φ'_{EW1}, φ'_{NS3}, φ'_{EW2} and φ'_{NS2} where the prime denotes "virtual", the first and second subscripts denote the orientation and the third denotes the cell number. The phase differences in (ii) are defined as follows:

$$\varphi_{EW} = \varphi_{W1} - \varphi_{E1} \tag{3.53}$$

$$\varphi_{NS} = \varphi_{S3} - \varphi_{N3} \tag{3.54}$$



FIGURE(3.11) A plot of the sounding frequency f as a function of time, and (above) the antenna connections during a 3-cell sounding structure.



FIGURE(3.12) The relationship between the frequencies f_H, f_D, f_i, f_{dopp}, and f' used in the h'-v^* separation algorithm.

$$\varphi_{DP} = \varphi_{W2} - \varphi_{W1} \qquad (3.55)$$

$$\varphi_{DP} + \varphi_V = \varphi_{N3} - \varphi_{N2}$$

The virtual receiver phases can be evaluated as follows:

$$\varphi'_{EW1} = (\varphi_{W1} + \varphi_{E1})/2$$

$$\varphi'_{NS3} = (\varphi_{N3} + \varphi_{S3})/2$$

Thus

$$\varphi_{EW2} = \varphi_{EW1} + \varphi_{DP}$$

$$\varphi_{NS2} = \varphi_{NS3} - (\varphi_{DP} + \varphi_V)$$

and

$$\varphi_{OX} = \varphi_{EW2} - \varphi_{NS2}$$

$$= (\varphi_{W2} - \varphi_{N2}) - (\varphi_{NS} + \varphi_{EW})/2 \qquad (3.56)$$

$$\varphi_V = \varphi_{N3} - \varphi_{N2} - (\varphi_{W2} - \varphi_{W1}) \qquad (3.57)$$

Procedure

At the end of cell 1, a discrete power spectrum of the energy present at one receiver output during that cell is computed by means of a hardware Fast Fourier Transform and all spectral "channels" that exceed a certain amplitude, up to a maximum of 16, are noted and the complex amplitudes recorded. At the ends of cells 2 and 3, only the complex amplitudes corresponding to the channels selected at the end of cell 1 are recorded. The frequencies of these channels are the set f_i.

Then, the phases φ_{EW}, φ_{NS}, φ_{DP}, φ_{OX} and φ_V are evaluated from eqns (3.53-3.57), and adjusted to lie in the range $-\pi < \varphi < \pi$. The two angles of arrival can be calculated directly from φ_{EW} and φ_{NS} according to equation (3.51) (Wright and Pitteway [7]) and the sign of φ_{OX} will determine the polarisation. The evaluation of h' and v^* is more involved and deserves detailed explanation.

The frequency resolution of the FFT and thus the channel spacing, is 1/T where T is the cell duration. The selected channels represent the closest discrete estimates to the frequencies f_D of eqn (3.22). Eqn (3.22) can be re-expressed as:

$$f_D = f_H + f_{dopp} \tag{3.58}$$

where $$f_H = 2kh'/c \tag{3.59}$$

and $$f_{dopp} = 2 f_o v^*/c \tag{3.60}$$

If f_i is the closest discrete spectral estimate to f_D in frequency, then the relationship between f_i , f', f_D , f_H and f_{dopp} are illustrated in Fig (3.12). The frequency f' is, by virtue of the assumption of "nearest estimate", restricted to

$$f' < 1/2T \tag{3.61}$$

Our task is to establish values for f_H and f_{dopp}, from which h' and v^* can be inferred from eqns (3.59) and (3.60).

The group range h' is evaluated by measuring the change in the phase of the discrete frequency f_i as the cell start frequency f_o is incremented by Δf, according to eqn (3.28). This will usually involve an initially unknown number N of full revolutions of 2π plus some "fractional" part which is identical to φ_V.

Thus $$h' = N.h_I + h_F \tag{3.62}$$

where $$h_I = c/2.\Delta f \tag{3.63}$$

and $$h_F = \varphi_V.c/(4\pi.\Delta f) \tag{3.64}$$

Correspondingly, from eqn (3.59)

$$f_H = Nf_{HI} + f_{HF} \tag{3.65}$$

where $f_{HI} = 2kh_I/c = k/\Delta f$ (3.66)

and $f_{HF} = 2kh_F/c = \varphi_V.k/(2\pi.\ \Delta f)$ (3.67)

Thus $f_i = f_D + f'$

$$= f_H + f_{dopp} + f' \tag{3.68}$$

$$= Nf_{HI} + f_{HF} + f_{dopp} + f'$$

$$N = (f_i - f_{HF})/f_{HI} - (f_{dopp} + f')/f_{HI} \tag{3.69}$$

It can be seen that

$$N = \text{integral part of } [(f_i - f_{HF})/f_{HI}] \tag{3.70}$$

provided that $(f_{dopp} + f')/f_{HI} < 1/2$

or $f_{dopp} \quad f_{HI}/2 - f'$ (3.71)

This sets the condition for correct group range evaluation free from "2π" ambiguity. The quantity f_{HI} is a conceptual element of frequency corresponding to a group range element that would cause a shift of 2π in the phase of a signal incremented by Δf, and is inversely proportional to Δf (see eqn. (3.66)). From eqns. (3.71) and (3.66) it can be seen that reducing Δf, for example, will extend the range over which f_{dopp} can exist without ambiguity, but at the expense of group path resolution which is proportional to f_{HI}. Evaluating f_{HI} for typical numerical values, we have for $k = 10^5$ Hz and $f = 5 \times 10^3$ Hz, $f_{HI} = 20$ Hz. For $f_o = 8 \times 10^6$ Hz, we have from eqns. (3.60) and (3.71) that the maximum unambiguous Doppler velocity is approximately 188 m/s. During substorm activity at high latitudes, where values of v^* in excess of 500 m/s have been observed on

analogue chirp ionograms, f' may have to be reduced to as low as 1 kHz to measure Doppler shifts successfully.

We can now evaluate f_H from (3.65) and thus h' from (3.59).

In the same way, the Doppler contribution to f_D can be expressed as the sum of an integral number M of full revolutions of 2π per T seconds, plus a fractional part, as follows

$$f_{dopp} = M/T + \varphi_{DP}/(2\pi T) \qquad (3.72)$$

From (3.68), $f_{dopp} = f_i - f_H - f'$

Thus, $M = Tf_i - Tf_H - \varphi_{DP}/2\pi - Tf'$

Once again, M = integral part of $[Tf_i - Tf_H - \varphi_{DP}/2]$ (3.73)

provided $Tf' < 1/2$

This condition is met by the definition of f' (eqn (3.61)). Thus f_{dopp} can be evaluated from (3.72) and v^* from (3.60).

Further discussion of group height/phase velocity separation is given in section 5.2.

REFERENCES: CHAPTER 3:

(1) Budden K.G., "Radio Waves in the Ionosphere", Cambridge University Press (1961).

(2) Brigham O.E., "The Fast Fourier Transform", Prentice-Hall, ISBN 0-13-307496-X (1974).

(3) Champeney D.C., "Fourier Transforms and Their Physical Applications", Academic Press Inc., (London) Ltd., ISBN 0-12-16745-9 (1973).

(4) Skolnik M.I., "Radar Handbook", McGraw-Hill Book Co., Library of Congress Card No. 69-13615 (1970).

(5) Bibl K. and B.W. Reinisch, The universal digital ionosonde, Radio Science, 13, 519-530 (1978).

(6) Wright J.W. and M.L.V. Pitteway, Real time data acquisition and interpretation capabilities of the Dynasonde 1. Data acquisition and real time display, Radio Science, 14, 815-825 (1979).

(7) Wright J.W. and M.L.V. Pitteway, Real time data acquisition and interpretation capabilities of the Dynasonde 2. Determination of the magnetoionic mode and echolocation using a small spaced receiving array, Radio Science, 14, 827-835 (1979).

CHAPTER 4

HARDWARE MODIFICATIONS

Contents:

CHAPTER 4

HARDWARE MODIFICATIONS

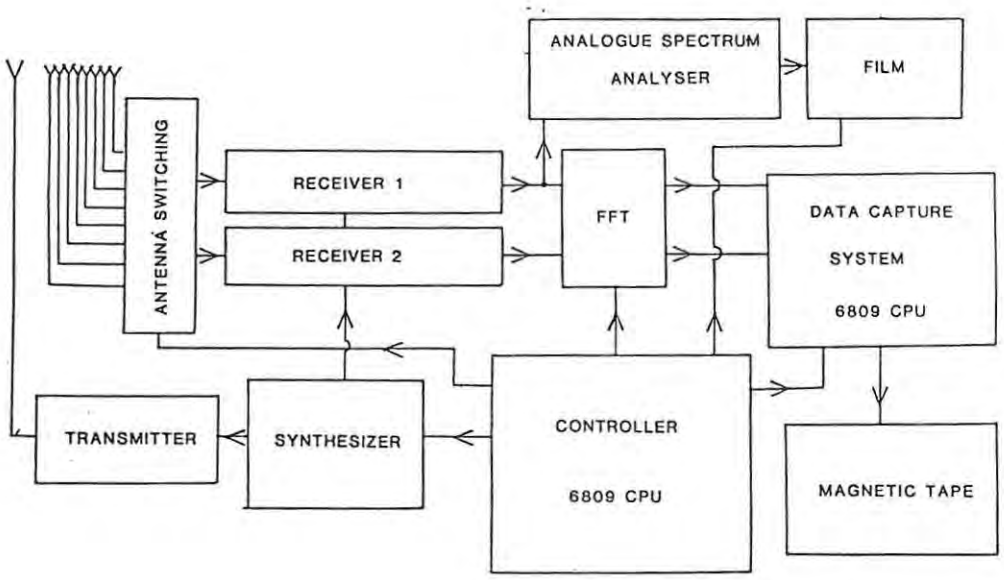
4.1 Introduction:

The basic "VOS-1 Vertical/Oblique Chirpsounder" as supplied by the manufacturers, Barry Research Corp. of California (now BR Communications), is not a "digital" device in the sense alluded to in the last part of Chapter 1. In order to implement the theory outlined in Chapter 3, extensive hardware changes to the analogue system had to be made. Where these hardware changes involved the use of microcomputers, software had also to be designed and written. It should be pointed out that this was an enormous task shared by workers other than the author. Although all the modifications will be reported in this chapter, the reader will be referred to the separate publications for details where appropriate. Figure 4.1 is a block diagram of the modified system, and will form the basis for the sectionalised description of the modifications that follow.

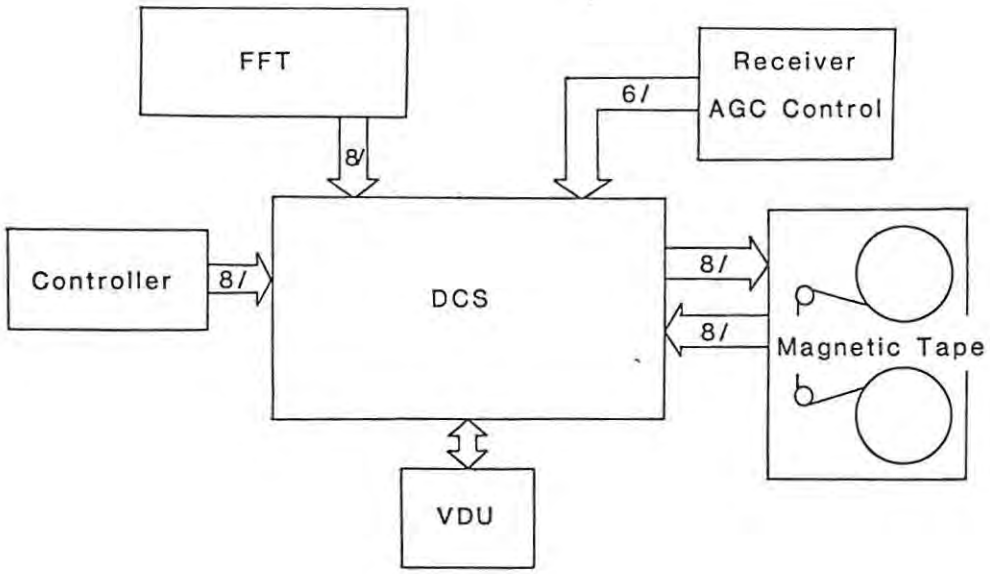
Ionogram structure: In order to facilitate the explanations, it is necessary to define the terms used in descriptions of the structure of a digital ionogram:

(i) A "cell" is the fundamental measurement unit consisting of a single chirp modulation, as expressed in equ.(2.15). Associated with each cell will be a set of data points which can be compared with those of other cells.

(ii) A "sounding" is a group of cells whose unifying feature is the similarity of their sounding frequency, while other parameters such as antenna connections will in general be different. The idea is that data from such a sounding "set" can be can be intercompared on the basis of



FIGURE(4.1) A block diagram of the modified digital chirpsounder.



FIGURE(4.2) A block diagram illustrating the DCS communication paths.

equs.(3.49), (3.50), (3.51) and (3.52) to give information about the ionosphere at that specific sounding frequency.

(iii) An "ionogram" is a collection of soundings whose characteristic frequencies will increase in some ordered fashion to provide information about the ionosphere over that range of sounding freqencies.

4.2 Modifications - other workers:

This section deals with modifications to the analogue system that were either purchased, or built by other workers, and for which details of the hardware/software involved can be sought in other texts. However, the reasons for the changes and a brief review of their main features are included here for completeness.

(i) Receivers: At the end of the last chapter it was emphasised that the four measurable quantities h', v^*, θ and P could be evaluated from the comparison of phase measurements that were respectively frequency, time, spatially or orientationally diverse. Also, the diversity due to the remaining parameters should be zero. This implies that, except for those pertaining to the evaluation of v^*, the phase measurements should all be made simultaneously. Since one receiver can only provide the means to evaluate one phase at a time, the need for more than one receiver is apparent. Excluding v^*, and remembering that the arrival angle θ has to be evaluated in two planes, there are four quantities to be measured, h', θ_{SN}, θ_{WE}, and P. Since each requires a phase difference, this implies five phases to compute the four unknowns. To economise on measurement time, the most time-efficient procedure would thus be to employ five matched receivers to make the five simultaneous measurements. However, because of the expense involved, the Rhodes system has adopted a compromise of two matched receivers, which were purchased from the manufacturers [5], and replaced the

single channel receiver originally supplied.

Furthermore, the original receiver had a continuously variable analogue Automatic Gain Control (AGC), which would seriously influence attempts to measure the total energy [equ.(3.35)]. The matched receivers contain digitally switchable attenuators that can vary the gain in 2-db steps over a total range of 110 db. The control logic for these attenuators is described in [1].

A further problem initially encountered with the matched receivers was the lack of provision in the original design for "gain weighting" of the signal. Because of the necessity to switch the receivers between the receive and transmit states (see "long modulations" - Chapter [2]), excessive noise is generated if the switching is not done "slowly". After consultation with the manufacturers, a gain-weighting module was included in each receiver to reduce this noise. The reader is referred to [2] for details regarding the shape of the weighting waveform.

(ii) Antenna switching: Because of the economic restraint on associating a separate receiver with each measureable quantity, it is necessary to switch the two receivers to the antennae appropriate to that measurement. A two-pole, eight-way R.F. switch was designed for this purpose [1], switched by the controller.

(iii) Controller: Control of the original equipment [3] was characterised by the inflexibility of the front panel switch selection of the different sweep parameters such as start time, sweep rate etc. To improve flexibility, and in order to extend the scope of the sounder control to encompass new areas such as antenna switching, transmitter filter switching etc., a Southwest Technical Products Corp. 6809 based microcomputer has been developed as a controller. The hardware and software design of this controller has been an

extensive endeavour and is the subject of a separate thesis, to which the reader is referred [1].

(iv) Fast Fourier Transform Analyser: As mentioned in Chapter 3, the phase differences $\Delta\varphi$ used in the evaluation of the measureable quantities are well approximated by differences in the phases of discrete Fourier frequency components. The FFT analyser performs the task of simultaneously sampling and transforming the time-varying signals from the dual phase-matched receivers, thus providing two sets of discrete complex amplitudes from which the phases can be calculated, according to equ.(3.39). In addition, the analyser computes the power spectrum according to equ.(3.40), which is used by the Data Capture System to discriminate signal components from noise. If the input consists of two sets of N real points each, the output will contain N/2 cosine, N/2 sine and N/2 power spectral amplitudes for each input channel, making a total of 3N data points. Because of the seeming lack of availability of commercial equipment to do this task at reasonable cost, it was decided to design and construct the FFT Analyser at Rhodes. This successful endeavour is the subject of an independent thesis [4].

4.3 Data Capture System (DCS):

General: Fundamental to the design philosophy of the DCS has been the idea that, apart from noise rejection, no real time analysis of the data would be effected by the system. Thus, as the name implies, the DCS has the task of collecting data from various sources and ensuring their storage in some ordered fashion, for later offline analysis by powerful mainframe computers. For this reason, reel-to-reel magnetic tape has been chosen as the bulk storage medium, since the ability to re-read magnetic tape seems to be universal, unlike, for instance, cartridge tape. The flow of data to and from the DCS is illustrated in fig. 4.2.

Tasks (i) Collection of FFT-provided data

Data discrimination: As was noted in [4.2(iii)] above, the FFT analyser will produce 3N data points per cell. Most of these will be noise points, and need rejection for efficient use of the data storage medium. In general, the difference frequency f_D, corresponding to an echo, will be approximately constant with respect to time, so echoes will be characteristically impulsive in the frequency domain, forming a basis for noise rejection.

Amplitude criterion: One of the hidden advantages of chirp modulation is that noise which is impulsive (in the time domain) and which is most troublesome to the echo recognition algorithms of pulsed ionosondes, has a flat frequency spectrum and thus contributes more or less equally to all the power spectral estimates of the FFT. In contrast, an echo will be impulsive in the frequency domain, making possible the use of a simple amplitude criterion for echo recognition, which a random noise spectral component is statistically unlikely to exceed. Once this level is set, the common (digital) AGC of the receivers is trusted to maintain the average energy below this level.

The discrimination is done on the basis of the amplitude of the power spectrum of one receiver channel. As was noted in 4.2(iv), the power spectra for both receiver channels areis calculated and provided by the FFT analyser along with the sine and cosine amplitudes. Although the main purpose of the power spectrum calculation is to allow real time discrimination of the data, the power spectrum data are nevertheless recorded with the sine and cosine amplitudes. Since the power spectrum can be re-evaluated from the sine and cosine amplitudes via equ.(3.40), this amounts to some redundancy in the data; however, such redundancy can be used to check on system performance.

As a further safeguard against excessive data capture in noisy RF environments, the DCS limits the number of "echoes" recorded per cell to 15.

This limit is judged to be sufficient to allow analysis of genuine multi-echo structures as would be encountered, for instance, during occurrences of spread-F, or travelling disturbances. Furthermore, if the number of cells per sounding for the ionogram exceeds 1, the DCS will only record the complex amplitudes during the second and subsequent cells for those Fourier components that exceeded the amplitude criterion during the first cell.

Collection of data: After the echo points have been selected, they are sorted into order of decreasing amplitude and stored in that order. Each echo will correspond to a single discrete Fourier component, and the power spectra and the complex amplitudes for that component measured on both receiver channels are recorded. Each echo is stored on tape as a single record, which also contains the DCS-generated information about the cell number and sounding number, based on the cell/sounding information passed by the controller at the ionogram start. Details of the data structure are given in Appendix E.

Tasks (ii) Collection of Controller-provided data

Associated with every ionogram will be a set of parameters providing identification and information regarding structure. This information is provided by the controller and falls into two categories:

(a) Global data which apply to the whole ionogram and include:

Station site identification

Time/date information

Basic chirp rate (k)

No. of cells/sounding

Cell length (δ)

FFT sample rate

Receiver local oscillator offset

Timing shift

This global information is read from the controller by the DCS at the ionogram start and is transferred in the first block of data to magnetic tape.

(b) Cell data which change from cell to cell and include:

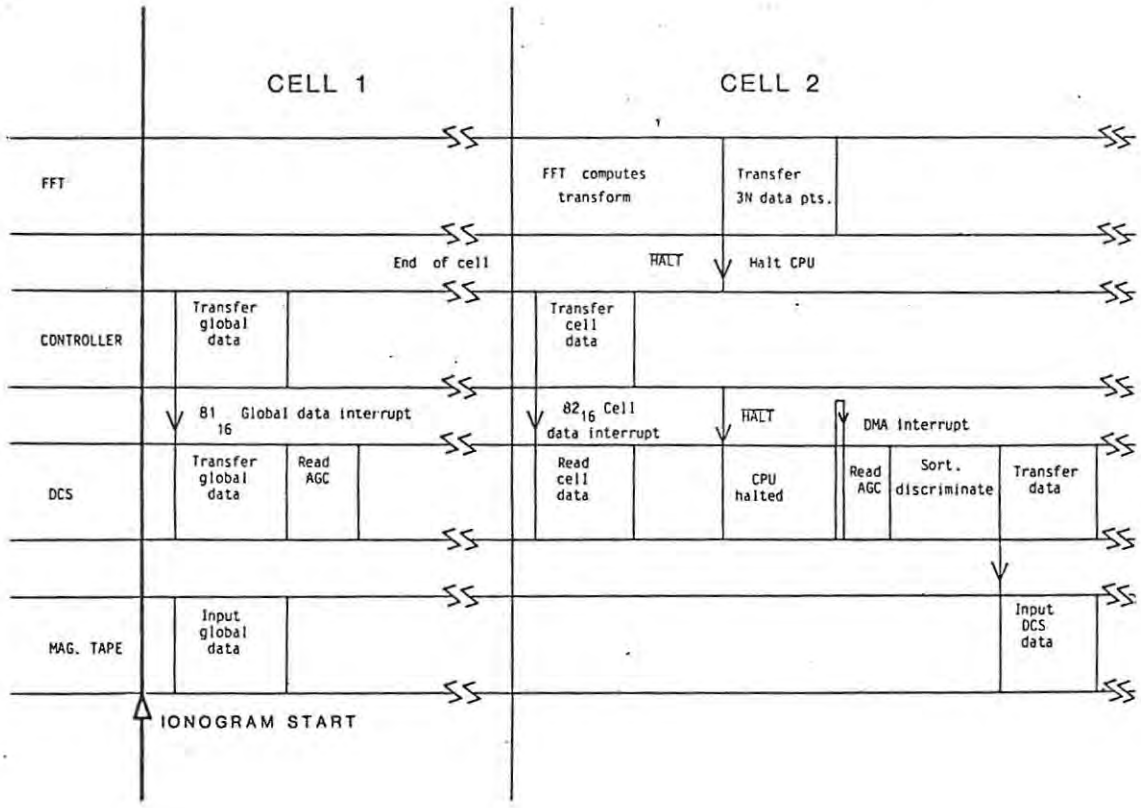
Antenna connections (N,E,S,W)

Cell start frequency (f_o)

Referring to fig.(4.3), it is seen that these data are read in just after the start of each cell. However, the FFT data for the current cell will of necessity only become available during the next cell, so the cell data must be stored temporarily for the duration of the cell. They are then associated with the correct echo data and transferred to magnetic tape during the next cell.

Tasks (iii) Collection of AGC data: The controller is programmed to update the receiver digital AGC at the end of each cell, keeping the gain constant during the cell. The controller reads the AGC information just after the start of the cell (see fig.[4.3]), but like the antenna and frequency information, it is stored temporarily for the duration of the cell. It is transferred to magnetic tape as part of the individual echo records, whose structure is discussed in more detail in Appendix E.

Tasks (iv) Test routines: As a diagnostic aid, some test routines are included in the DCS software to enable real time monitoring of some features of ionospheric behaviour. The tests can be run simultaneously with the routine data collection programme. A brief description of these tests is given in Appendix D. For a more detailed description the reader is referred to the DCS test manual. All communication to and from the test routines is steered via the Visual Display Unit (VDU).



FIGURE(4.3) A timing diagram showing the times at which data are transferred during the first two cells of an "ionogram".

Tasks (v) Magnetic tape Read/Write: Much of the DCS software is devoted to communication with the magnetic tape drive, in the first instance, so that data can be written to the tape. As an additional diagnostic aid, there are various commands that will allow manipulation of the tape, in particular the reading of one or more blocks of data. Communication is, as before, via the VDU. Details of these commands are given in the DCS manual.

Data structure: The data are structured into records of length equal to 32 characters, of which 64 are collected to form a 2048 character block. The first 9 records of an ionogram contain text and numbers which describe the global ionogram parameters, while each subsequent record refers to one spectral component or "echo" that has exceeded the amplitude criterion. Details of the data structure are given in Appendix E.

Software considerations

(a) Language: The programming language FORTH was used to write the DCS software routines, this being considered a good compromise between speed on the one hand and computational flexibility on the other. Where FORTH was too slow to accomplish tasks needing speed, such as magnetic tape transfers, assembly language routines were written which interfaced easily to the main programme. The source text for FORTH programmes is not particularly readable, and has not been included here. Once more, the reader is referred to the DCS manual for source listings of the DCS software.

(b) Programme structure: The language FORTH was specifically designed to be used as an interactive control-oriented language. When "up-and-running", FORTH is continually monitoring the VDU, awaiting the input of some previously defined command that is to be executed immediately, or the input of code defining some new command for later execution. In order to

preserve this attractive capability, rather than dedicate the entire FORTH "machine" to DCS tasks, the software has been designed so that the three major external data sources request servicing by means of interrupts. These are:

D.M.A. interrupt (FFT data)

Controller interrupt (global and cell parameters)

Magnetic tape status interrupt

A timing diagram showing the relationship between the DCS tasks and the interrupts is shown in figure (4.3), while flow diagrams of the interrupt service routines are given in Appendix F. Other functions such as the various tests, and all magnetic tape control commands (e.g. REWIND, Read one block etc.) that are optional, are run by the operator from the terminal. For this reason, the programme can handle the data capture tasks in parallel with the tests.

REFERENCES: CHAPTER 4

(1) Evans G.P., "Computer Control of a Barry Research Chirpsounder", M.Sc. Thesis, Rhodes University, (in process) November 1983.

(2) Barry Research Corp., (now BR Communications) 4050A Manual, "Amplifier T/R", T/R Control.

(3) Barry Research Corp., (now BR Communications) 1015 Manual, "Logchirp Control".

(4) Fisher J.S., M.Sc Thesis, "A Real Time Fast Fourier Transform Analyser", M.Sc. Thesis, Rhodes University, 1979.

(5) BR Communications, Sunnyvale, California, U.S.A.

CHAPTER 5

INITIAL RESULTS AND CONCLUDING REMARKS

Contents:

CHAPTER 5

INITIAL RESULTS

5.1 A two stage process

In section 4.3 it was mentioned that a fundamental feature of the hardware modifications to the chirpsounder is that all analysis is done offline. Thus, during the course of an experiment, all data are stored on magnetic tape, for later analysis on a large mainframe computer that has access to large amounts of storage and a useful graphplotting peripheral. All the results reported in this chapter have been produced in this fashion from the two stage process illustrated in figure (5.1).

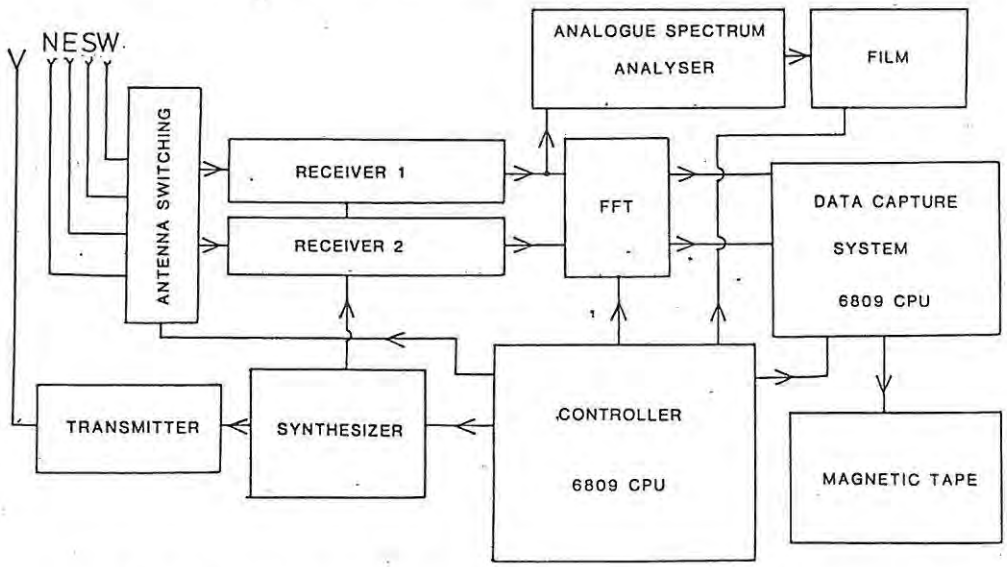
To accomplish the second stage of the process, many FORTRAN language programs have been written, for instance, to do the calculations necessary for the 3-cell sounding structure described in section (3.9). Although much of the effort involved in this project has gone into the development of these programmes, they are not reported here in detail.

5.2 Receiver calibration

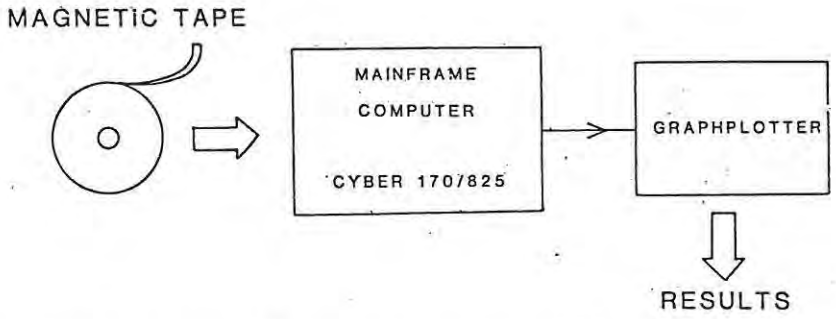
General: It was mentioned in Chapter 3 that all the measurable parameters could be evaluated from phase comparisons with the appropriate diversity. For all measurements involving spatial diversity, it is necessary to compare two phases simultaneously, which implies the use of two antennas, two cables and two receivers. It is thus essential to know to what extent the use of these separate hardware items introduces errors into the phase measurements.

If we consider two signals of identical phase present at the antenna inputs, then ideally the difference between the phases of the receiver output signals should be zero. One can expect that the error due to hardware inconsistencies will be a function of several variables, i.e.

A 2-STAGE PROCESS



Stage 1: A block diagram of the digital ionosphere sounder.
Raw data are stored on magnetic tape



Stage 2: The data magnetic tape is processed offline.

$$E_H = \Psi(f, f_D, A, \Delta A)$$

where f = sounding (received) frequency

f_D = difference frequency at receiver output

A = amplitude of signal

ΔA = signal amplitude difference between receivers

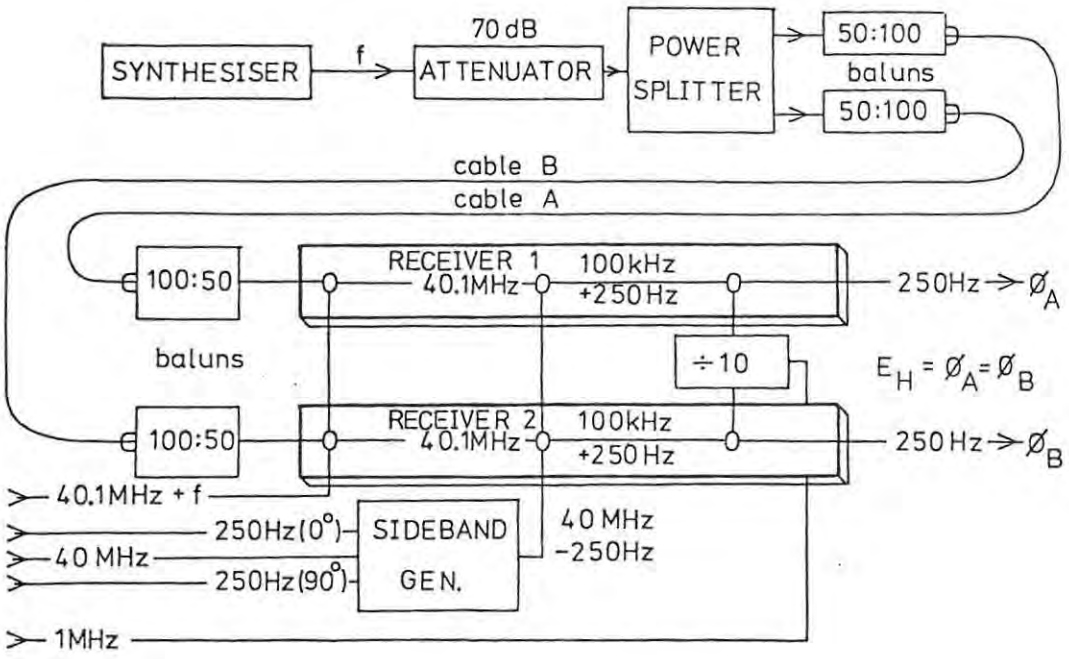
Several tests were done to get a "first estimate" of the error E_H, and it was found that the only significant variable was the sounding frequency f. This is fortunate, for to measure and attempt to implement a four dimensional calibration vector would have been an almost impossible task. In fact, it will be shown that even the error due to f is small and can be ignored in most cases.

It can therefore be assumed that

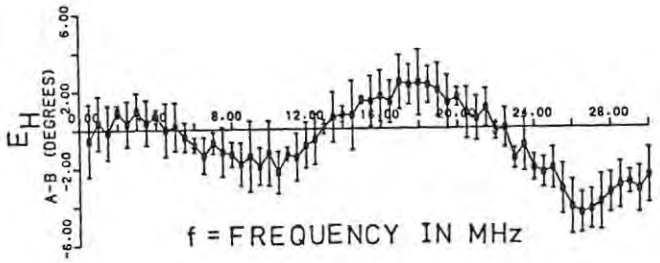
$$E_H = \Psi(f, f_D, A, \Delta A) \simeq \Psi(f)$$

Figure (5.2) shows the experimental setup used to measure E_H as a function of f. Since the receiver bandpass is 0 - 500 Hz, a nominal frequency f_D = 250 Hz was chosen to be in the centre of the bandpass. A major difficulty was the synthesis of a coherent parallel test input frequency (f - 250 Hz) over the range .5 - 30 MHz. This was overcome by offsetting the fixed local oscillator frequency of 40.0 MHz by 250 Hz and using the signal of frequency f directly from the synthesizer. The author is grateful to Geoff Evans for his design and construction of the single sideband generator that was used for this purpose.

For these tests, a 10-cell sounding was defined which sampled the same frequency 10 times with a cell length of .5 s. An "ionogram" was then defined which ran from .5 to 30 MHz in .5 MHz increments. In order to activate the synthesizer transmit signal (frequency f) during the receive portions of the T/R switching pattern, the "T/R SYNC GATE" input to the synthesizer was grounded.



FIGURE(5.2) The experimental set-up used for the calibration of the receivers and the vertical incidence receive antenna cables.



FIGURE(5.3) A plot of the phase difference $\varphi_A - \varphi_B$ for the two receivers only, for identical signals at the input, as a function of frequency.

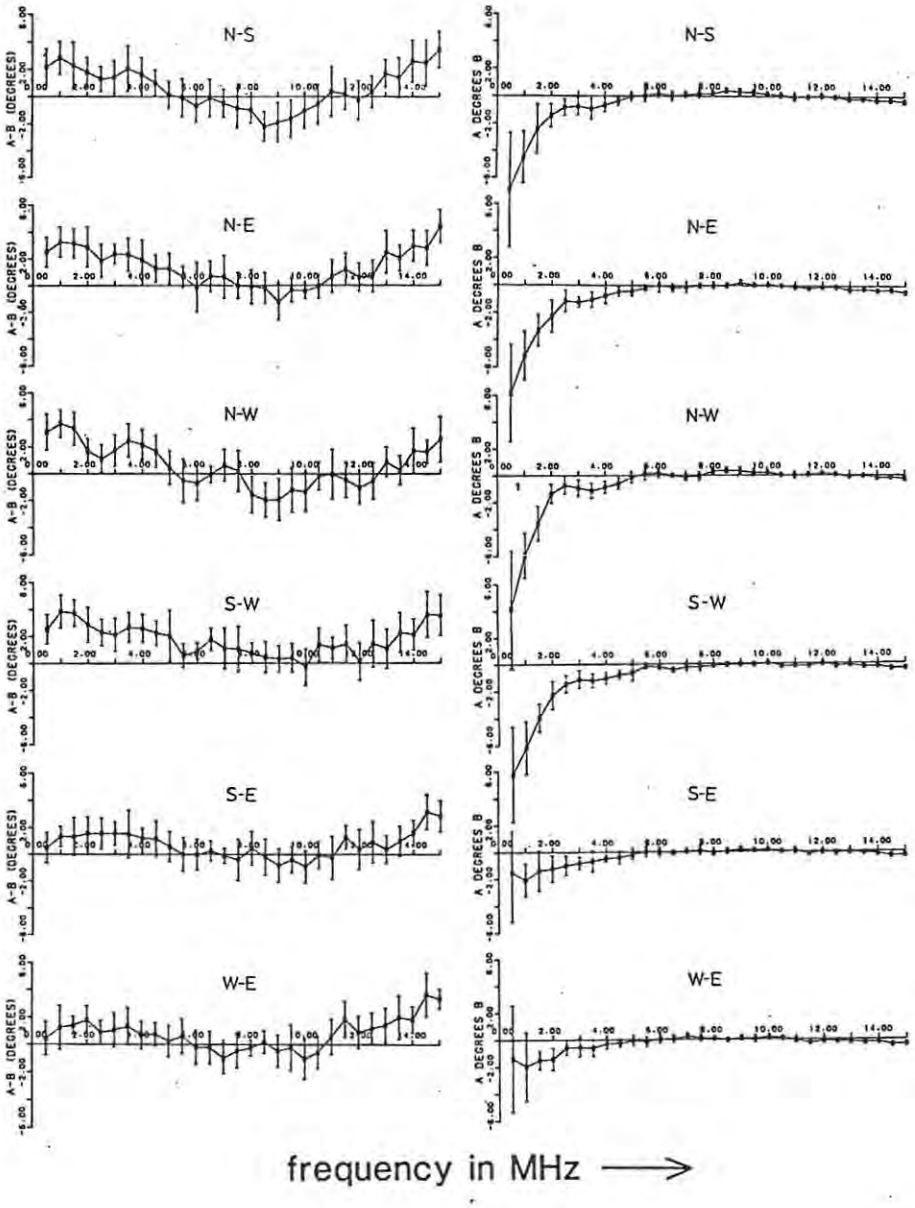
Use was made of the fact that the sinc function frequency domain representation of the receiver output overlapped at least two discrete Fourier analysis channels in the FFT analyser, so that at least 20 evaluations of $E_H = \varphi_A - \varphi_B$ could be made per frequency increment. These results could then be used to find an average value for E_H, as well as an rms deviation. The results shown in figures (5.3) and (5.4) show straight line segments joining the average values of E_H, while the error bars indicate the $\pm$ rms deviation in the usual way.

The scatter in the values of E_H is due to quantisation error. The accuracy of the input and output data of the FFT analyser is limited to 8 bits, which implies a minimum error of 1 part in 256 in the data.

Receivers only: To evaluate the dependence of E_H on the receivers only, the circuit elements consisting of the 4 baluns, and the two cables A and B in figure (5.2) were replaced by two short cables of identical length, directly connecting the power splitter to the two receivers. The results are shown in figure (5.3). It is seen that the average value of E_H is comparable to the rms quantisation error of the measurement over most of the range. For most real experiments, it will usually be found that both E_H and the quantisation error are small compared with errors brought about by noise (see section 3.7), and can probably be ignored.

Vertical incidence cables As was shown in section (3.6), angle of arrival measurements and O/X mode discrimination are made possible by the use of a spaced antenna array. This array in turn requires the use of cables, which must be of the same electrical length to avoid the introduction of systematic phase errors.

Pairs of cables were measured using the system illustrated in figure (5.2). Each cable was named, viz. N,E,S,W, which, if compared in pairs, would require 12 sets of graphs. However, since half the set would just be



FIGURE(5.4) The phase difference between the two receiver outputs when identical signals are input via the cables as indicated.

the same pair reversed, only 6 graphs are presented in figure (5.4). On the left hand side, E_H is shown as before, with the specific cable pair being indicated next to each graph. On the right hand side, the error E_H is transformed, using equation (3.41), to evaluate the error in the angle of arrival measurement that would be caused by the primary error E_H. If this error is E_θ , then:

$$E_\theta = \arcsin \left[\frac{\lambda E_H}{2\pi d}\right] \qquad (5.1)$$

where d = distance beween antennas

$\lambda = c/f$

Since the vertical incidence sounding freqency f will rarely exceed 15 MHz, the graphs are only plotted up to that frequency. The graphs all exhibit much the same form, which is consistent with the shape of the graph for the receivers only, indicating that the cables are electrically equivalent. There seems to be some additional periodicity in the data, which is thought to be some artefact of the experimental setup and is probably not of concern in a real experiment. The graphs on the right hand side indicate that the modulus of the error in the angle of arrival evaluation falls rapidly to less than 2 degrees at frequencies higher than 3 MHz. The poorer behaviour below 3 MHz can be countered by using larger antenna spacings, a fact evidenced by equation (5.1). For this reason, it will later be recommended that a larger antenna array be used for frequencies below 3 MHz.

Summary: In general it has been shown that the systematic errors introduced by the combination of the vertical incidence antenna cables and the receivers are of the same order as the quantisation error due to the finite bit length of the FFT analyser, and, even under the very best of signal-to-noise ratios (in excess of 20 dB), less than the error due to noise.

Further, measurements to establish this error in the range .5 - 15 MHz show that the error is of the order of the uncertainty in its measurement. For these reasons, no attempt has been made in the test results which follow to compensate for E_H. It is suggested that any experiment which requires extreme accuracy be designed to eliminate E_H by reversing the receivers within the sounding structure.

5.3 Tests of h'-v^* separation

Fixed frequency experiment. At the end of section 3.9, a procedure is described which will enable the separation of group height h' and phase velocity v^*. In figure 3.12, the relationship between the following five frequencies is illustrated:

f_D = observed difference frequency

f_i = nearest Fourier estimate to f_D

f' = separation $f_i - f_D$

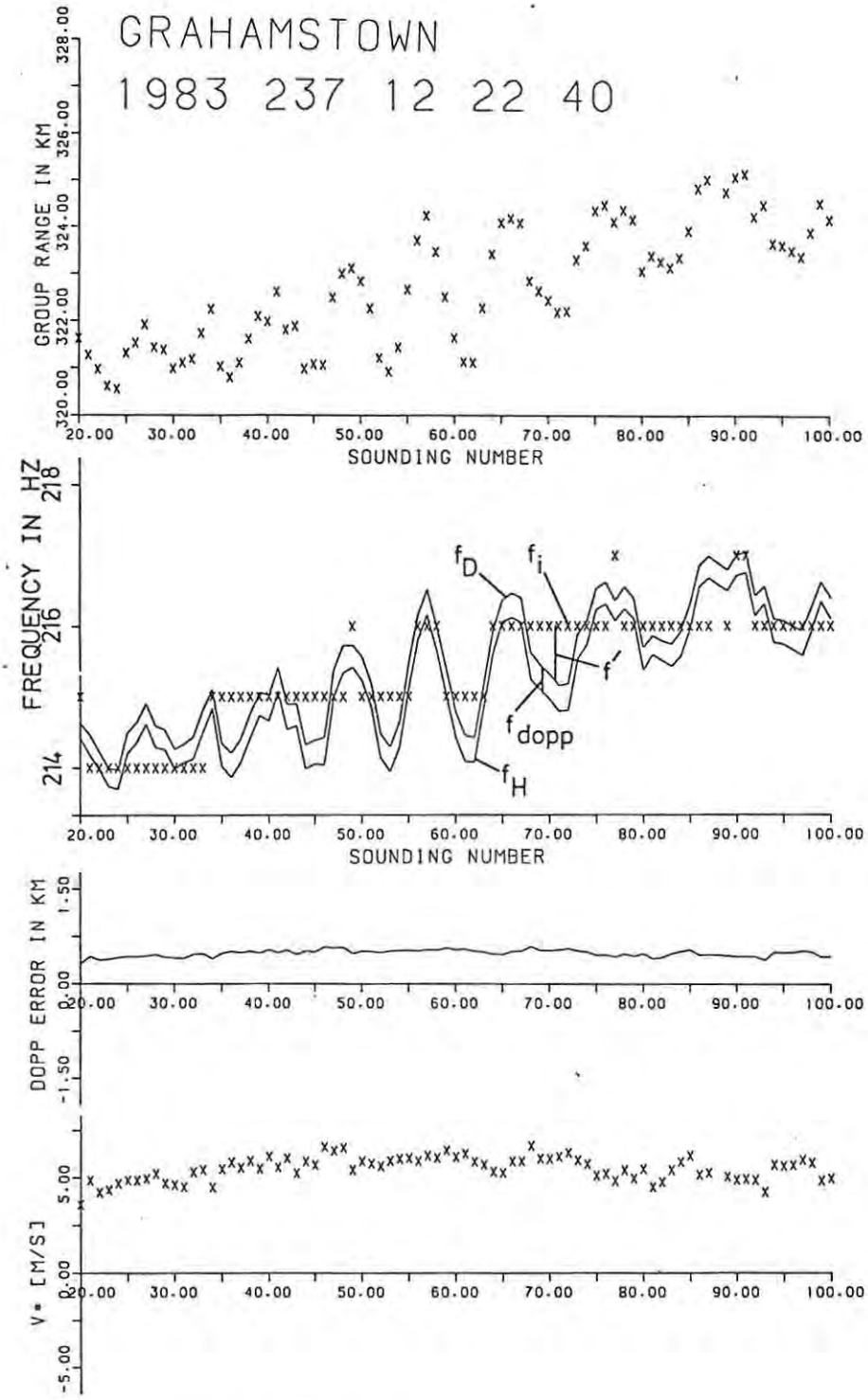
f_{dopp} = Doppler frequency shift

f_H = imaginary frequency due to group range only, = $f_D - f_{dopp}$

In figure (5.5), the results of a real "fixed frequency" experiment are shown. The input parameters for this sounding are shown in table (5.1). On this occasion, damped sinusoidal oscillations of the group range h' were present, which well serve to illustrate the relationship between the five frequencies mentioned above in a real situation.

The top graph shows the group range h' as a function of time or sounding number. In the next graph the measured values of f_i and the calculated values of the other frequencies are shown on the same time scale. The h' values are calculated from the f_H values according to equation (3.59).

The third graph is the error in group range that would occur if the



FIGURE(5.5) Results of a fixed frequency experiment used to illustrate the relationship between the variables involved in the h'-v^* separation algorithm.

Parameter	Value
Overall (average) sweep rate	0 kHz s^{-1}
Basic chirp rate k	100 kHz s^{-1}
Cell start frequency f_o	8.8 MHz
No. of cells/sounding	3
Cell length T	.5 s
Sounding duration	1.5 s
Frequency increment between soundings	0 Hz
Group range frequency increment	5 kHz

Table (5.1) Input parameters for the fixed frequency experiments of figures (5.5), (5.7), (5.10) and (5.13).

Parameter	Value
Overall (average) sweep rate	50 kHz s^{-1}
Basic chirp rate k	100 kHz s^{-1}
Cell start frequency f_o	increases linearly
No. of cells/sounding	3
Cell length T	.5 s
Sounding duration	1.5 s
Frequency increment between soundings	75 Hz
Group range frequency increment	5 kHz

Table (5.2) Input parameters for the swept frequency experiments of figures (5.6), (5.9), and (5.12).

Parameter	Value
Overall (average) sweep rate	100 kHz s^{-1}
Basic chirp rate k	100 kHz s^{-1}
Cell start frequency f_o	increases linearly
No. of cells/sounding	3
Cell length T	.5 s
Sounding duration	1.5 s
Frequency increment between soundings	75 Hz
Group range frequency increment	5 kHz

Table (5.3) Input parameters for the swept frequency experiments of figures (5.16)(a)-(j).

Doppler effect were ignored. This is equivalent to calculating h' from the f_D values rather than the f_H values using equation (3.59). In this example, the error is less than 750 m, which would probably not be of importance to observers of synoptic parameters, justifying to some extent the standard chirp sounding assumption that $v^* = 0$. However, in the next example, it will be shown that v^* can become significantly large.

The last graph shows the phase velocity v^* calculated from f_{dopp} using equation (3.60).

In order to test the algorithm, one notes that the primary phase measurements are made during the 3-cell sounding on the discrete Fourier estimate f_i. From these phases, the frequencies f_H and f_{dopp} can be calculated according to the algorithm described in section (3.9), equations (3.62) and (3.69). In figure (5.5), the "observed" difference frequency f_D is then calculated as the sum of f_H and f_{dopp}, according to equation (3.55). Since the frequency f_D is formed from calculations based on the phase of the discrete component f_i without regard for the actual value of f_i, the value of f_D can be compared with f_i to test the consistency of the algorithm. For the algorithm to be successful, f_D should lie within $\pm$ 1/2 times the inverse of the cell length (equation (3.58)), in this case, $\pm$ 1 Hz. This is seen to be the case throughout the course of this experiment.

There are three factors which could cause f' to exceed 1/2T:

(i) low signal to noise ratio could cause the evaluation of f_H and/or f_{dopp} to be in error by an amount sufficient to violate the condition

(ii) evaluation of the group range phase φ_V assumes that the Doppler phase φ_{DP} remains constant during the course of the sounding (equation (3.55)). If dv^*/dt is appreciably non-zero, this leads to an incorrect evaluation of f_H and thus a violation of the criterion.

(iii) over-lapping O- and X-mode echoes (i.e. for which the values of f_i are the same) can result in near-linear polarisation of the echo energy, whose component in the direction of the receive antennas used for the $h' - v^*$ phase measurement may be small. The amplitude of this component will depend on the relative phases and amplitudes of the elliptical O- and X-mode polarisations, which will be random. A small amplitude will result in temporary apparent reduction in signal to noise ratio, leading to errors such as those described in (i) above.

Swept frequency experiment: Table (5.2) shows the input parameters for this experiment, which is illustrated in figure (5.6). This swept frequency ionogram was chosen specifically because the X-mode reflections show a large positive trend in v^* as one approaches f_xF2. The top graph shows the normal h' vs f curve for both the O- and the X-modes. In the next curve, only v_x^* is plotted. The trumpet-shaped curves show the limits of the $\pm\pi$ ambiguity. Points lying inside these limits in this example have M=0 in equation (3.73), while points lying outside have M=1.

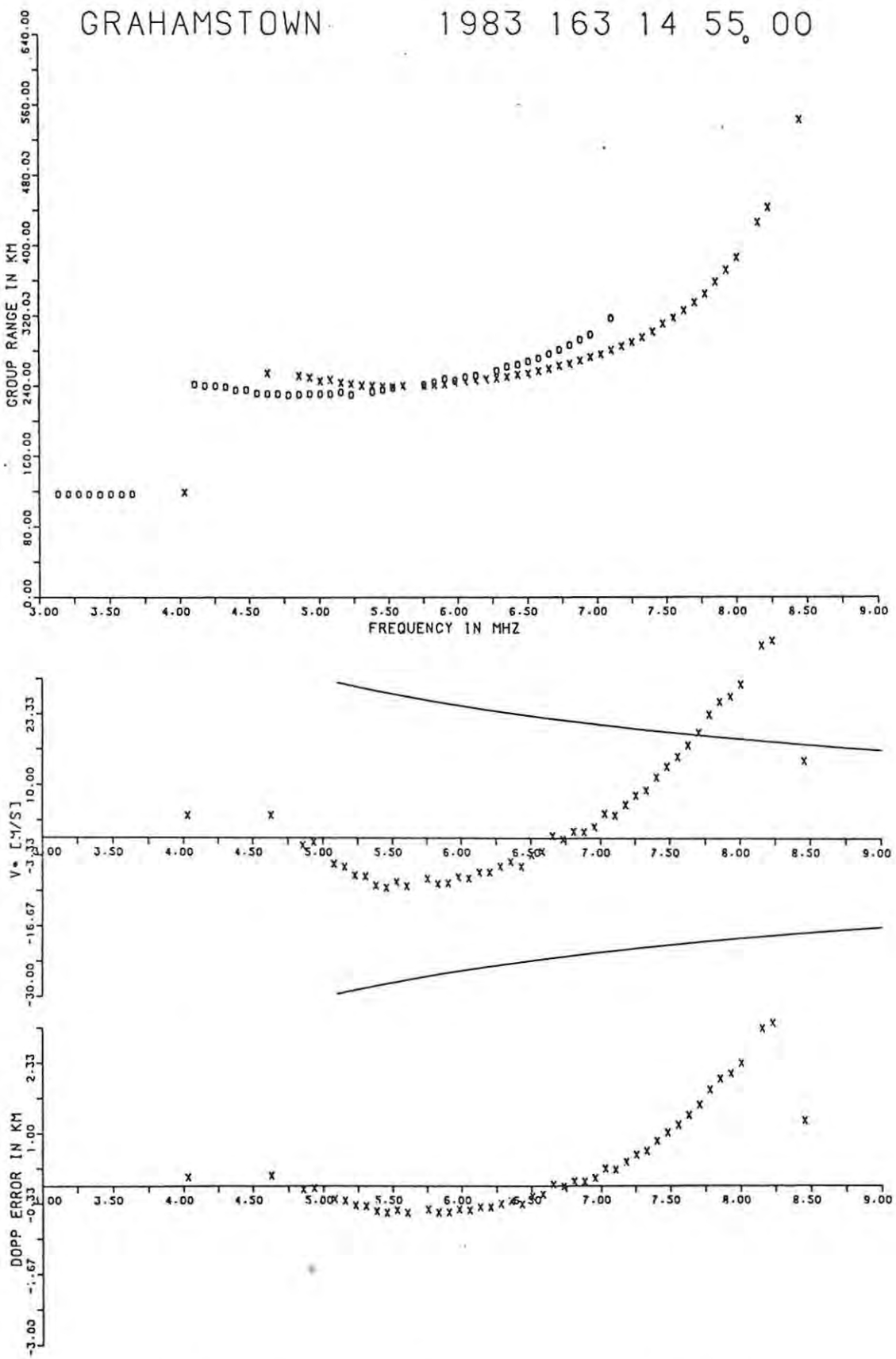
The example illustrates two points:

(i) the algorithm can successfully overcome 2π ambiguities in the v^* measurement, up to the limiting value of v^* specified by equations (3.71) and (3.60). This limiting value is ± 188 m s^{-1} at 8.0 Mhz, for this example.

(ii) the third graph shows the error in the virtual height measurement that would be caused by assuming that $v_x^*=0$. This error exceeds 3 km near f_xF2, a fact that would be of importance to observers interested in N-h profile analysis.

5.4 Magnetoionic mode identification

In section 3.6, it was shown that the sense of rotation of the



FIGURE(5.6) A swept frequency experiment in which the phase velocity for the X-mode reflection (v^*_x) exceeds the bounds of the $\pm 2\pi$ ambiguity level (indicated by the trumpet shaped curves).

elliptically polarised electric vector could be found by establishing the sign of phase difference of the signal measured on two orthogonal antennas. The measured angle φ_{OX} is identical with the angle φ_{OX} equation (3.56) .

Fixed frequency Figure (5.7) shows the result of a fixed frequency experiment, using the same parameters as table (5.1), but with a start frequency f_o = 8.6 Mhz. The upper graph shows the values of the angle φ_{OX} as a function of time, with the mode identified by the symbols in the usual way. The average value of the X-mode values was calculated to be $+59.9^o$, while the O-mode average was -108.9^o. A striking feature of these data is the consistency of the values obtained as a function of time.

Less consistent are the angles between the major semi-axes of the ellipses and the E-W direction, shown in the lower graph. These were evaluated according to the formula:

$$\varphi_{SA} = \frac{1}{2}\tan^{-1}\frac{2A_{NS}A_{EW}\cos(\varphi_{OX})}{A_{EW}^2 - A_{NS}^2} - \frac{\pi}{4}(1 - V)$$

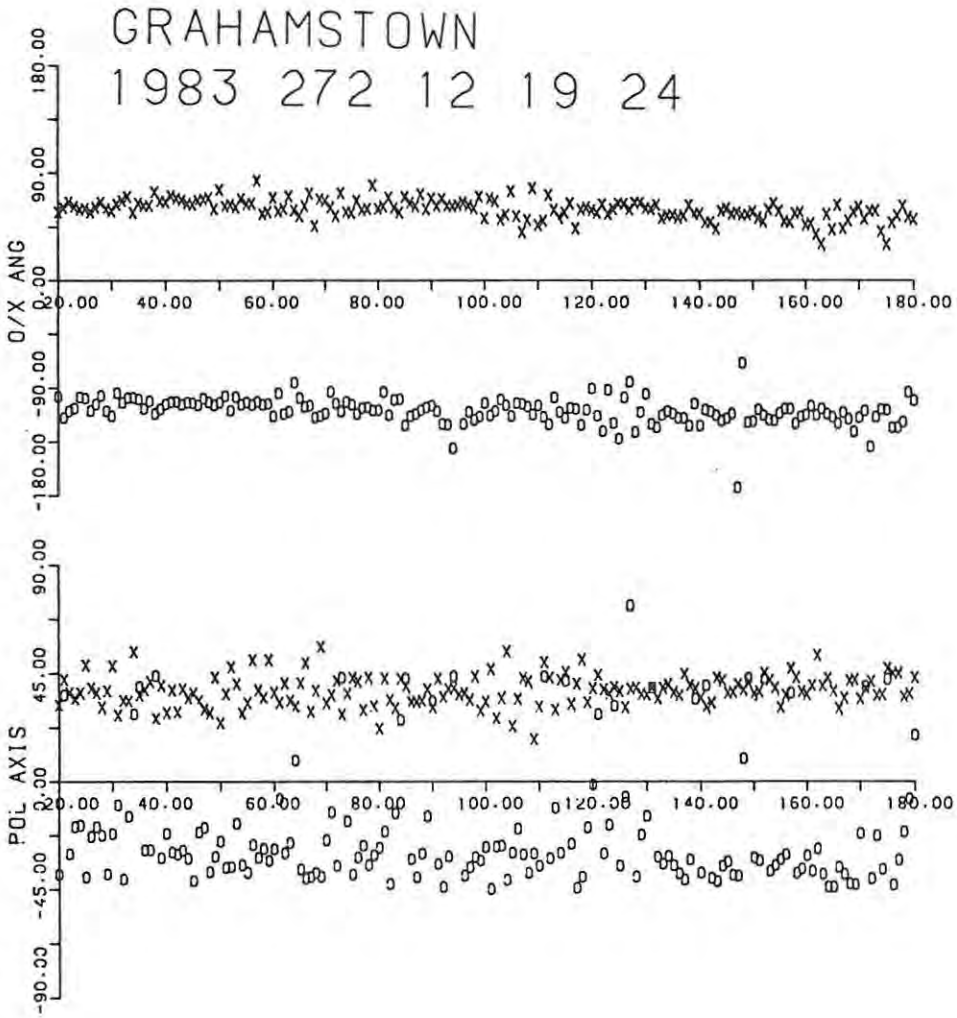
$$V = \begin{cases} +1, & A_{EW} >= A_{NS} \\ -1, & A_{EW} < A_{NS} \end{cases}$$

The values of φ_{SA} were adjusted to lie in the range

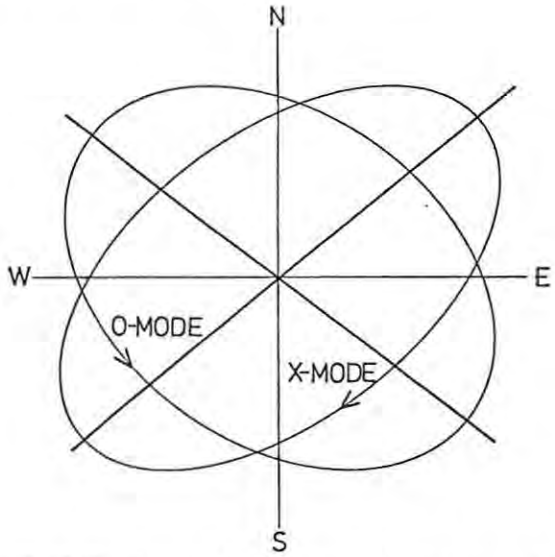
$$-\pi/2 < \varphi_{SA} <= +\pi/2$$

by subtracting $\pi/2$ from all values $> \pi/2$. The increased scatter on these points is no doubt due to the fact that the polarisation is nearly circular, making the value φ_{SA} sensitive to small amplitude variations.

By evaluating average values for A_{EW} and A_{NS}, and with the known average value of φ_{OX}, the locus of the point of the electric vector could be plotted using equations (3.44) and (3.45). This is illustrated in figure (5.8). It is thought that this may be the first time that the shape of the polarisation ellipses have been plotted for a Southern Hemisphere station.



FIGURE(5.7) A fixed frequency experiment used to illustrate the variation in the polarisation angle (top) and the inclination of the major semi-axis of the ellipse as a function of time.

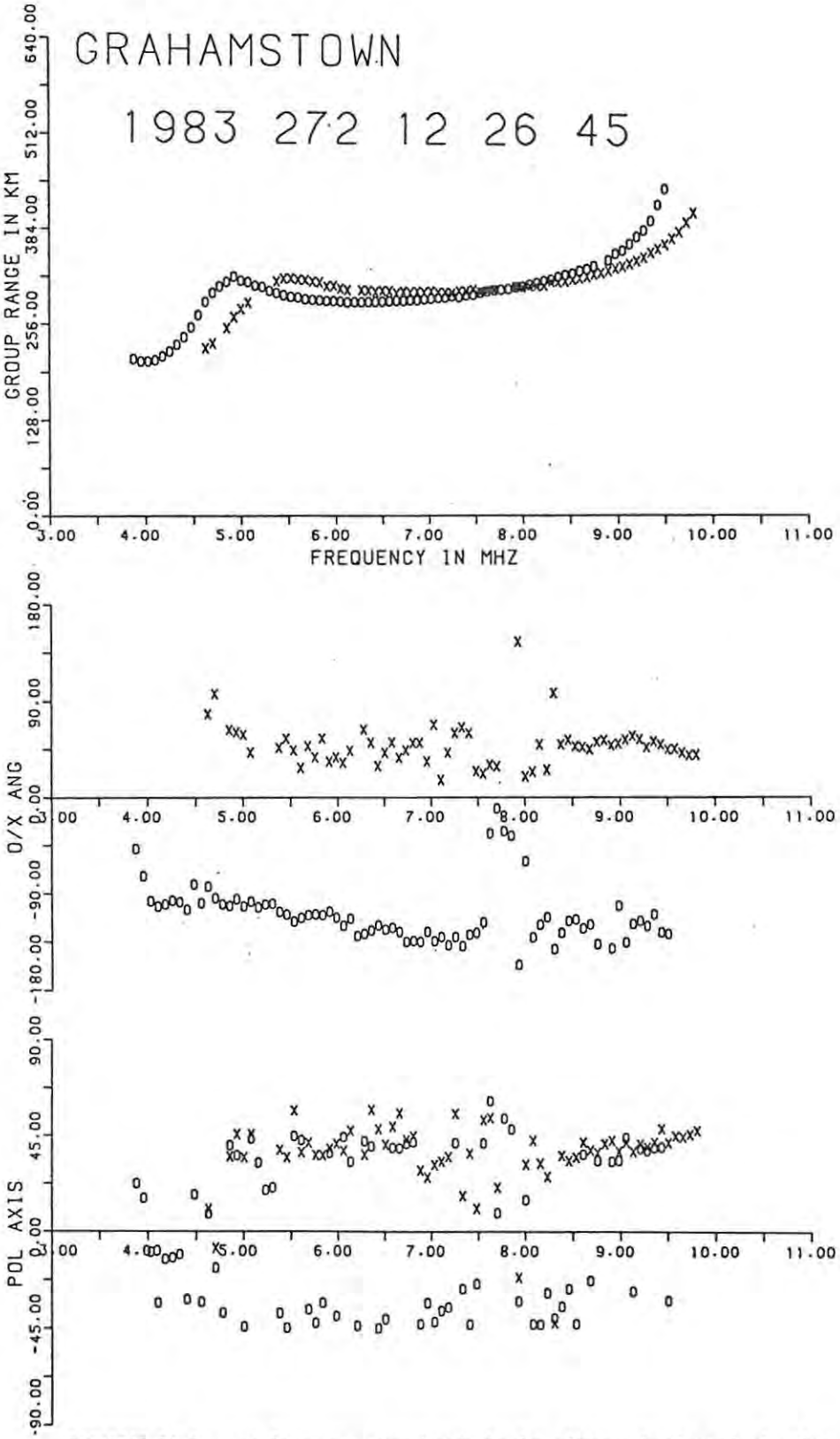


FIGURE(5.8) A plot of the shape of the average polarisation ellipses that were present during the course of the experiment shown in figure (5.7).

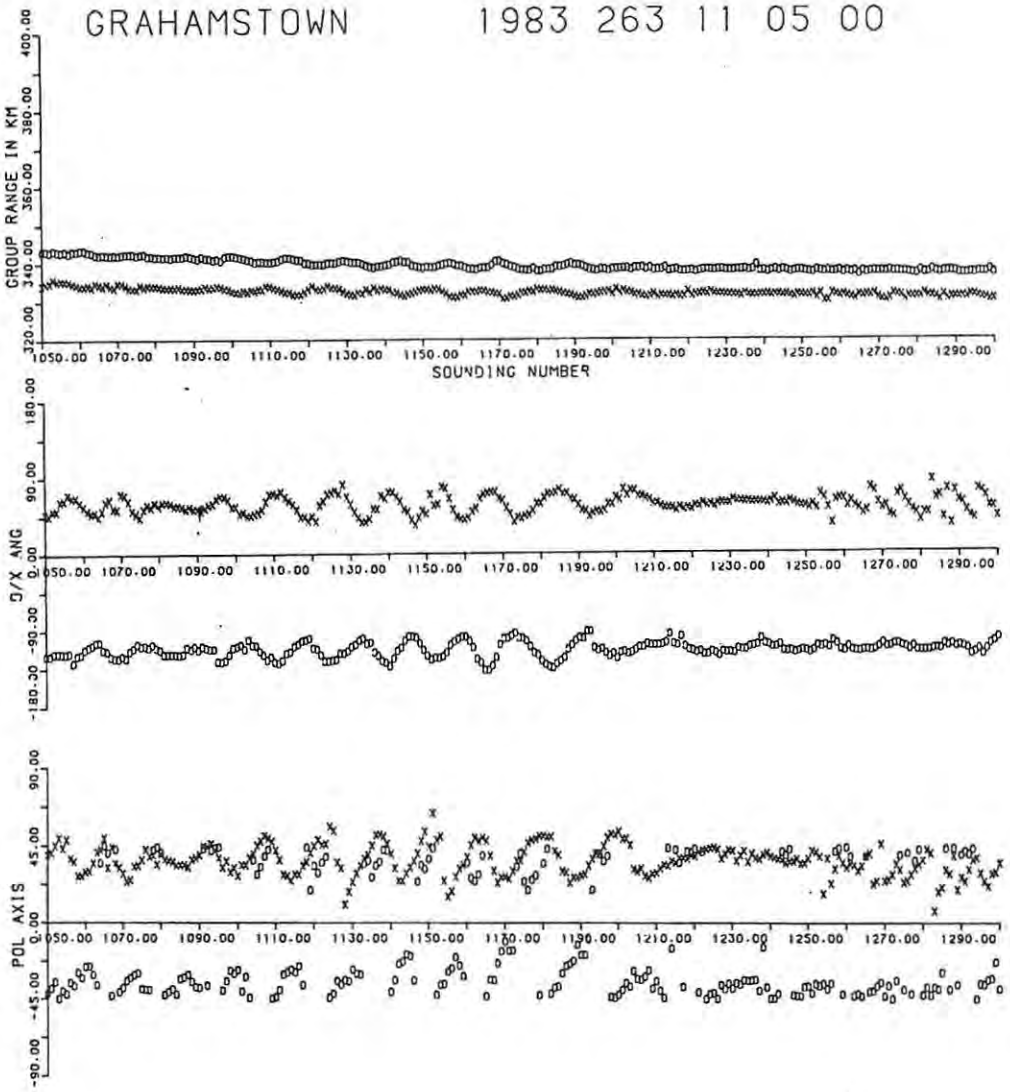
Swept frequency In order to test the stability of the angle φ_{OX} measured between the two orthogonal antennas as a function of frequency, the results are shown for a typical swept frequency ionogram in figure (5.9). The input parameters for this experiment are shown in table (5.2). Represented in the diagram are, from top to bottom, the group range, the polarisation angle, and the inclination of the major semi-axis with the E-W direction. It is of interest to note that the O-mode polarisation angle shows a steady tendency towards more negative values as the frequency increases from 4.0 to 7.5 MHz, with the rate of change being particularly steep near f_oE at 3.6 MHz. The X-mode polarisation angle shows a definite structure between 8.4 MHz and f_xF2, which might indicate that the polarisation ellipse owes its shape more to the morphology of the plasma near the point of reflection than that near the exit point from the ionosphere, the so-called "limiting polarisation" ([1],section 19.11). Near f = 7.8 MHz, where the O- and the X-modes have the same group range (strictly, for which the expression $(2kh' + 2f_ov^*)/c$ is constant in equ. (3.22)), the two polarisations are confused and cannot be resolved using the methods described in Chapter 3. Some recent work by Paul [2] describes an iterative method whereby the two modes can be separated for pulsed digital sounders. A similar procedure is feasible for digital chirp ionograms as well.

A remarkable feature of the semi-axis inclination data is that for the O-ray, the inclination seems to swap from that characteristic of the O-ray to that for the X-ray, while the reverse is only true for two data points near f = 8.0 MHz. This phenomenon is reported without comment except to note the promise of an interesting future line of investigation.

Mode leakage Figure (5.10) shows similar data for a fixed frequency



FIGURE(5.9) A swept frequency experiment showing changes in the polarisation angle and axis inclination as a function of frequency.



FIGURE(5.10) An illustration of a fixed frequency experiment in which the O- and the X-modes are almost overlapping. Leakage of the one mode into the other causes the symmetric "wobbles" as a function of time.

experiment at f = 8.8 MHz, the remainder of the parameters being the same as those for table (5.1). Referring to figure (3.6), the separation of the central maximum of the spectral component from the second sidelobe in frequency is 4Hz, for a cell length of .5 s. At the basic chirp rate k = 10^5 Hz s^{-1}, this is equivalent to 7.5 km of group height, which is the approximate separation of the two modes at the top of figure (5.10). Thus, each of these modes can be expected to contain some contribution from the second sidelobe of the other. The ratio of the energy in the second sidelobe to that for the central maximum is approximately 18 dB for a time series that has the unwindowed "sync" function spectral distribution shown in figure (3.6). The sideband energy of the one mode can be considered to be an interfering signal at the same frequency, with arbitrary phase, as described in the error analysis of section (3.7). The phase relationship between the real and the interfering signals will depend on the difference in the phase paths h_o^* and h_x^*. In the example shown, these two phase paths are changing at different rates, and this slowly varying phase relationship causes the resulting polarisation ellipses to "wobble". Thus, both the polarisation angle and the ellipse inclination angle will deviate either side of the uncontaminated mean position with a period equal to the time taken for the phase path difference $h_o^* - h_x^*$ to change by one wavelength at the sounding frequency. The slow build-up and decay of the oscillation amplitude is produced by the increase and decrease in the amplitude of the interfering signal as the nulls in the spectrum pass over the real signal frequency as the two modes move closer together. In the example shown, the roughly equal amplitudes of the two modes cause them to leak into each other in the same proportion, which explains the symmetry of the oscillations.

The oscillations of the group range data are not as easy to explain. Errors in the group range can only be caused by errors in the angle φ_V used

in equation (3.67). This angle is, however, the difference between two contaminated angles in which the effect of the contamination would subtract out, unless

$$[h_o^*(f) - h_x^*(f)]_{f=f_o} = [h_o^*(f) - h_x^*(f)]_{f=f_o+\Delta f}$$

where Δf is the frequency increment used in the group range evaluation. It is thus clear that in this example, the phase path difference $h_o^* - h_o^*$ is not only a function of time, but also of frequency.

This example has been shown in order to illustrate two points:

(i) all interesting cyclic behaviour in some measured ionospheric parameter should be scrutinised for leakage of energy from nearby echoes before any geophysical explanation is sought,

(ii) it might be necessary in some applications to reduce the amount of leakage by applying some form of window to the time series, or, in the frequency domain, by a complex convolution of the signal spectrum with the spectrum of the chosen window. This in general will reduce the range resolution by broadening the width of the central maximum, but the trade-off may be worth the reduction in leakage.

5.5 Angle of arrival

The orientation of the returning echo ray path, because of its three dimensional spatial character, proves difficult to illustrate two dimensionally. For this reason, the data have been presented in three ways:

(i) as plots of the arrival angles θ_{NS} and θ_{EW} described in section 3.7

(ii) as plots of zenith angle and bearing which can be calculated from φ_{NS} and φ_{EW} according to:

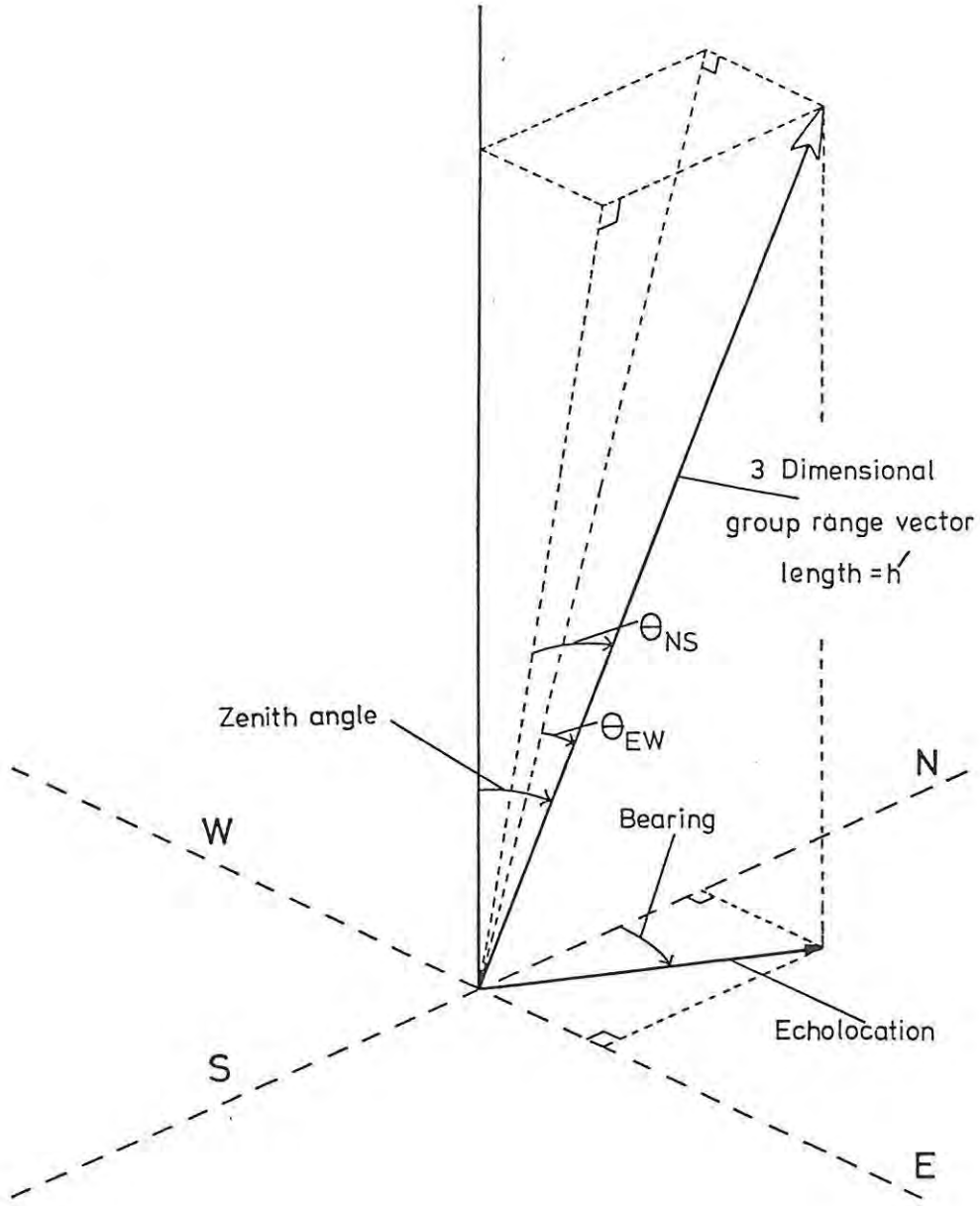
$$\text{zenith angle} \quad Z = \arccos\left[(1 - \sin^2\theta_{NS} - \sin^2\theta_{EW})^{1/2}\right]$$

$$\text{bearing} \quad B = \tan^{-1}\left[\frac{\sin\theta_{EW}}{\sin\theta_{NS}}\right]$$

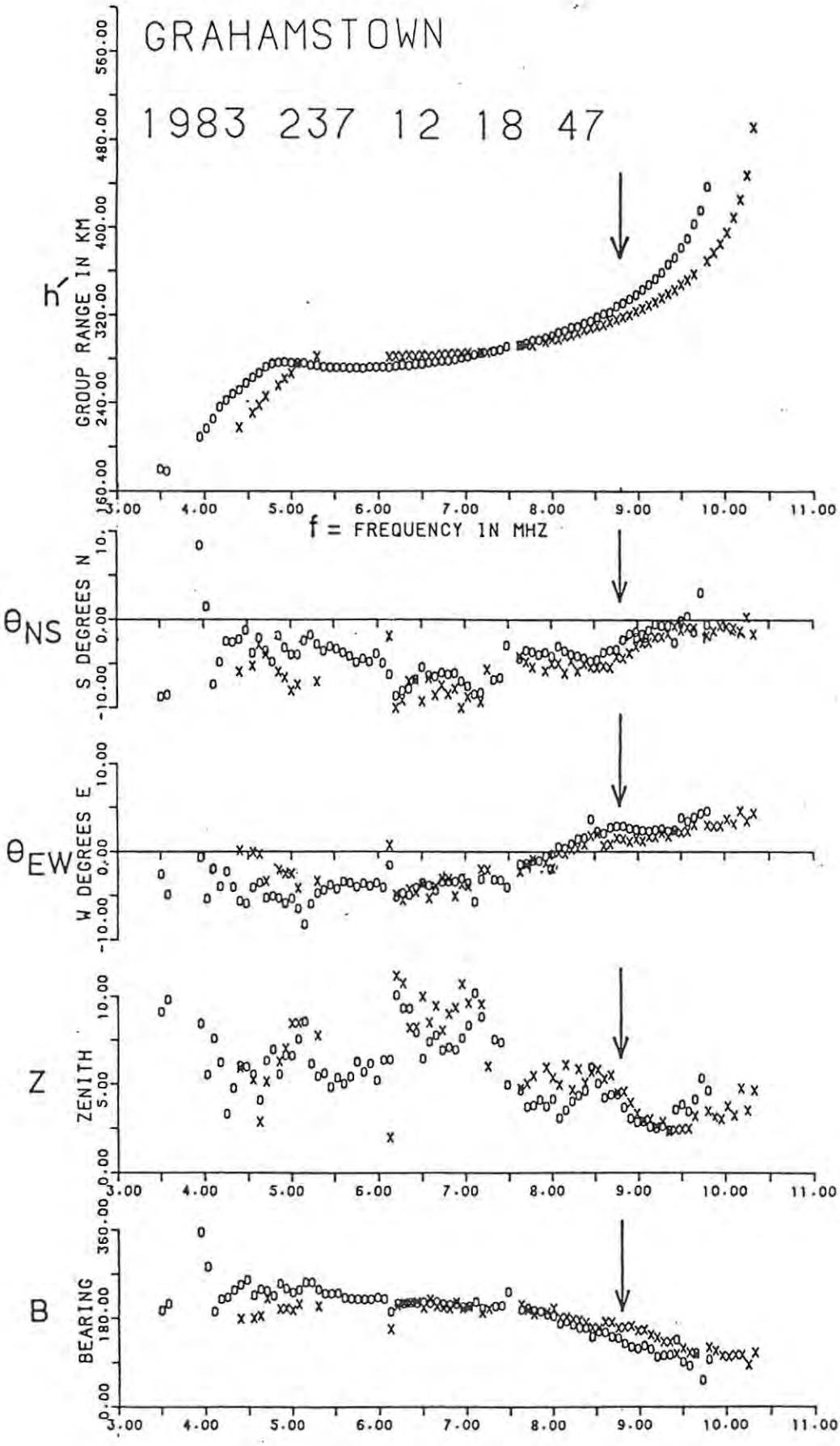
(iii) as an "echolocation" plot, a system used by Wright and Pitteway [3]. In this format, the co-ordinates of the echolocation are are the projection of the point of the group range vector onto a horizontal plane. The relationship between these variables is illustrated in figure (5.11).

Swept frequency Figure (5.12) shows the variation of these parameters for a typical quiet ionogram. The input parameters are shown in table (5.2). The echolocation plot for this experiment is shown at the top of figure (5.14). An interesting point to note about the variation in the orientation of the group range vector is that the X-mode variables do not necessarily follow those for the O-mode. It is known that for sounding frequencies $f_r >> f_H$, the gyrofrequency, the plasma frequency at the reflection level for the X-mode is approximately equal to that for the O-mode, less half the gyrofrequency. It might then be expected that the X-mode would follow that for the O-mode with a separation in f_r of $f_H/2$. This differing behaviour of the arrival angle of the two modes has been noticed often on similar experiments and is most marked during the passage of travelling disturbances when the contours of equal ion density are curved.

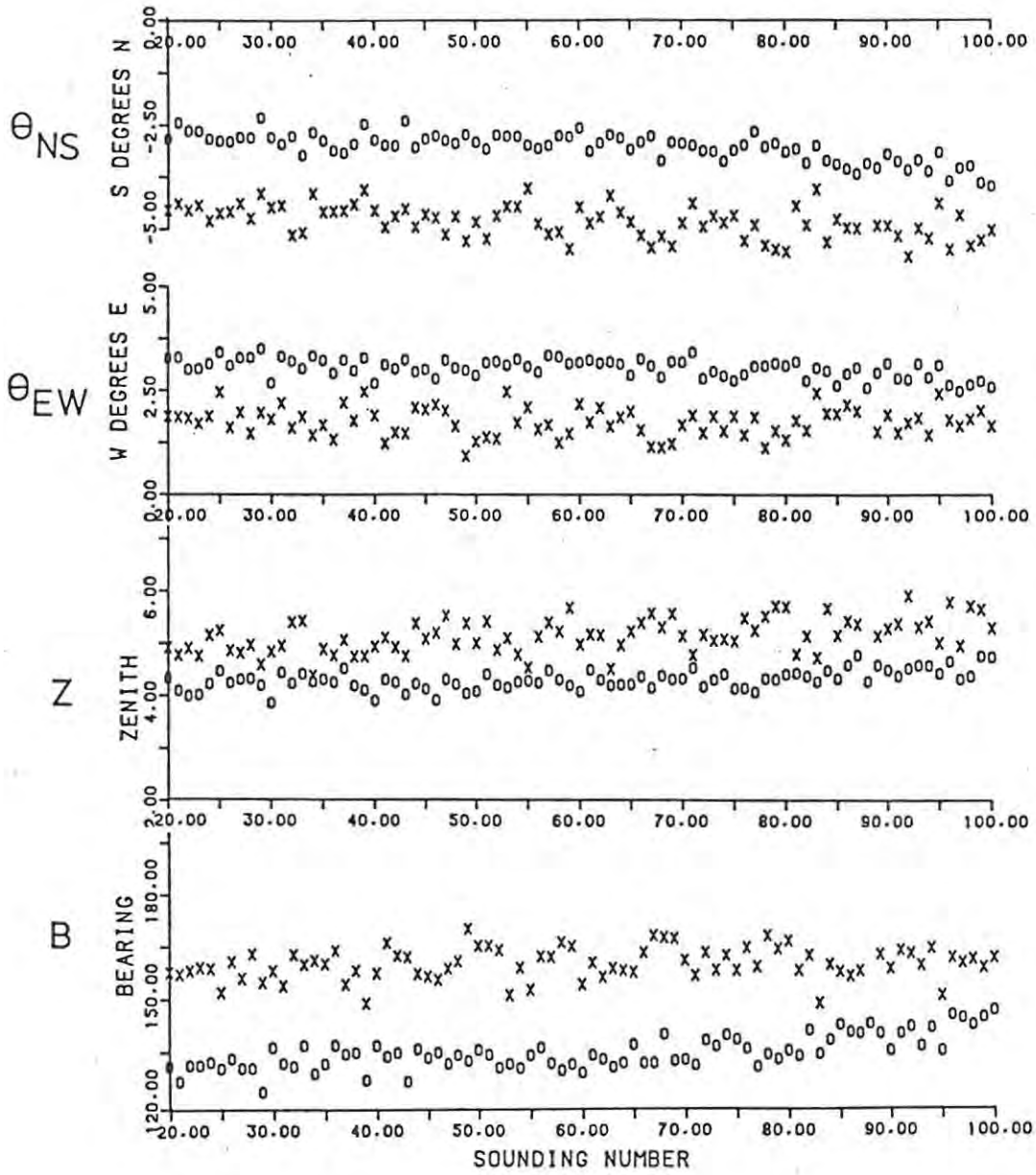
Fixed frequency The arrows in figure (5.12) indicate the frequency at which a fixed frequency experiment was run to assess the consistency of the arrival angle measurements as a function of time. The rest of the input parameters are shown in table (5.1) The results are shown in figure (5.13), and the lower half of figure (5.14). There appear to be periodicities in the data which cannot be explained in terms of the mode leakage described in section 5.3.



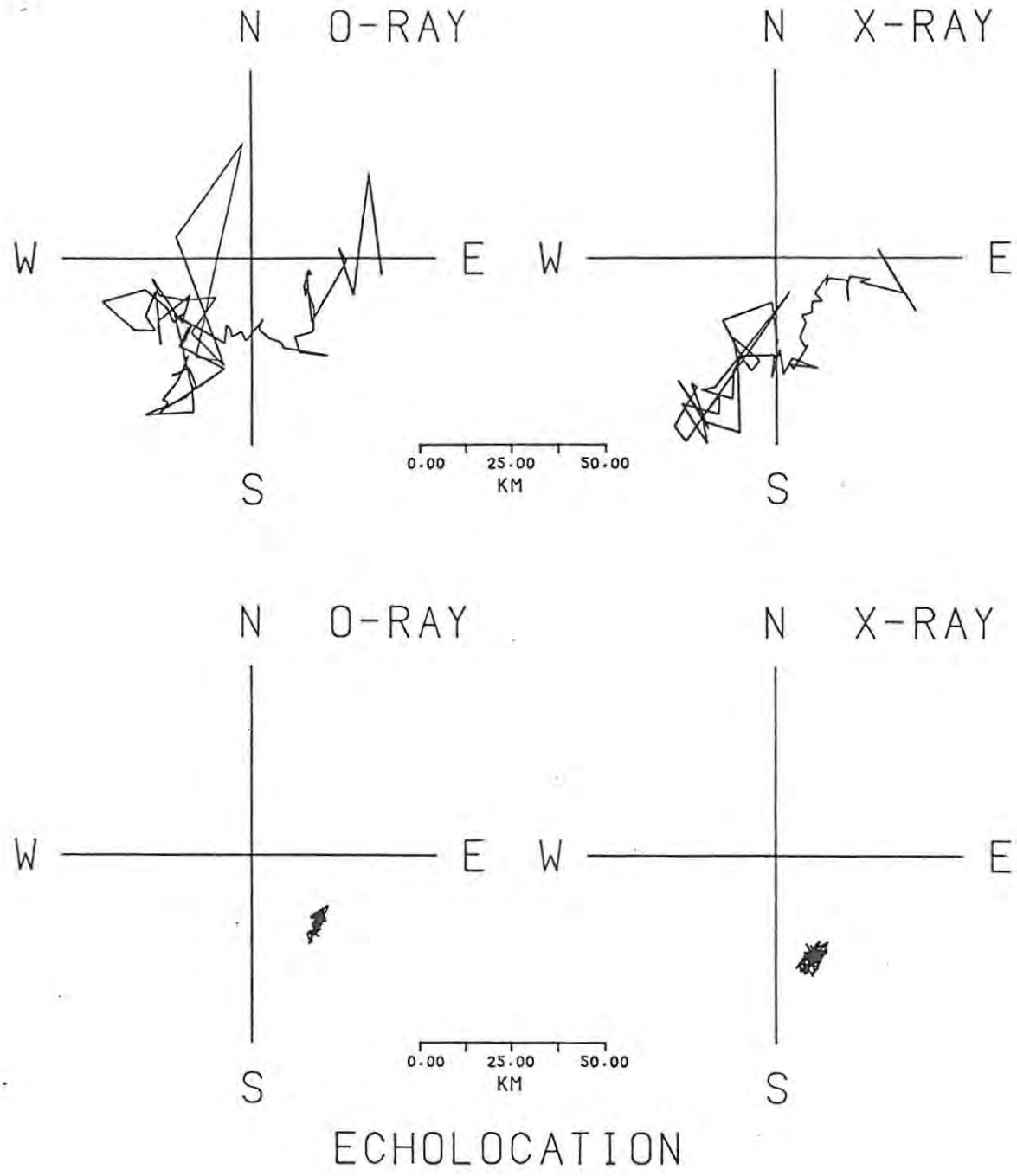
FIGURE(5.11) The relationship between the arrival angles θ_{NS} and θ_{EW}, the zenith angle Z, the bearing B and the echolocation for a non-vertical ray.



FIGURE(5.12) A swept frequency experiment showing typical behaviour of the arrival angles as a function of frequency. Arrows indicate the frequency of a fixed frequency experiment, illustrated in figure (5.13).



FIGURE(5.13) Fine structure in the arrival angle parameters during a fixed frequency experiment.



FIGURE(5.14) Echolocation plots for the experiments illustrated in figures (5.12) at the top, and (5.13) at the bottom.

A conservative estimate of the ratio of signal to noise power would be 20 dB. From figure (3.10), the expected rms error due to noise will be approximately $E_{rms} = 4^{o}$. From this an estimate of the rms error in the arrival angle θ can be obtained by adapting equation (5.1):

$$E = \arcsin \frac{\lambda E_{rms}}{2\pi d}$$

where λ is the freespace wavelength at 8.8 MHz and d = 30 m is the antenna spacing. This gives $E = .72^{o}$, so that approximately 60 $^{o}/o$ of the values can be expected to lie between $\pm .72^{o}$ of the true value. This result is consistent with the scatter seen in figure (5.13).

Amplitude measurements: The received signal power from an ionospheric reflection can be obtained from the radar equation ([4], pg.1-4, equ. (1)):

$$P_r = \frac{P_t . G_t . \sigma . A_r}{4\pi R^2 \; 4\pi R^2} \qquad (5.2)$$

where P_t = transmitter power

G_t = gain of transmit antenna

R = range of target

σ = ionospheric cross-section

A_r = effective aperture of receive antenna

In general, measurements of P_r will be used to evaluate some aspect of the quantity σ, this being the only variable that relates to the behaviour of the ionosphere. However, use of (5.2) to determine σ will depend on a knowledge of the variations of G_t and A_r as functions of frequency and ray direction, which are in general not known and at best would be difficult to determine. If the information required about σ can be expressed as a

comparison of two values of P_r for which G_t and A_r could be considered constant, then such relative values of P_r can be obtained, with some approximation, from the discrete power spectral components of the difference signal x_D given by equation (3.40).

Some aspects of the use of discrete values of the power spectrum P_n to evaluate P_r should be borne in mind:

(i) the P_n are only samples of the power spectral density of the signal, and should be integrated over the appropriate frequency range to give the total energy in the signal E according to equation (3.35).

(ii) some account of the gain setting of the receiver will have to be taken.

In figure (5.15), the amplitude of the discrete Fourier components of the difference signal $x_D(t)$ is expressed in decibels as the ratio of the measured discrete power spectral value to the minimum value that would pass the DCS amplitude criterion with the receiver at maximum gain, i.e. :

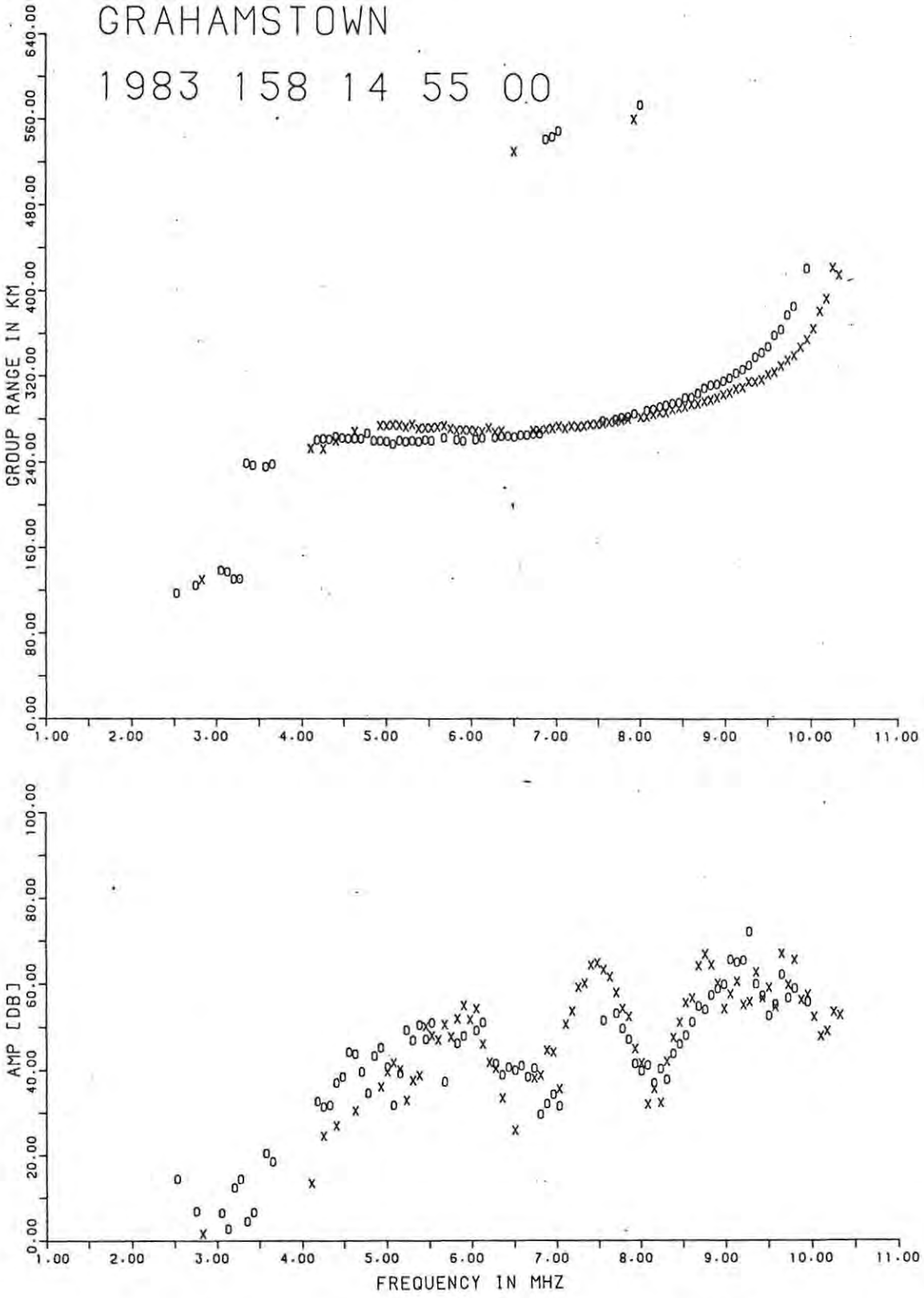
$$\text{Amplitude A} = 10 \log_{10}(P_i/P_{min}) + \text{AGC}$$

$$\text{AGC} = \text{receiver attenuation in decibels}$$

Figure (5.15) illustrates two points:

(i) There is a steady increase in signal strength as the frequency increases, indicating the extent to which low frequency antenna inefficiencies affect the received power through the influence of G_t and A_r in equation (5.2)

(ii) There is an interesting sinusoidal variation in the signal strength which is most probably a time rather than a frequency related phenomenon, but which is certainly due to the influence of σ in equation (5.2). The value of σ could change, for instance, by alternate focussing



FIGURE(5.15) A plot of the group range and amplitude for a swept frequency ionogram. The amplitude data shows pronounced focussing effects.

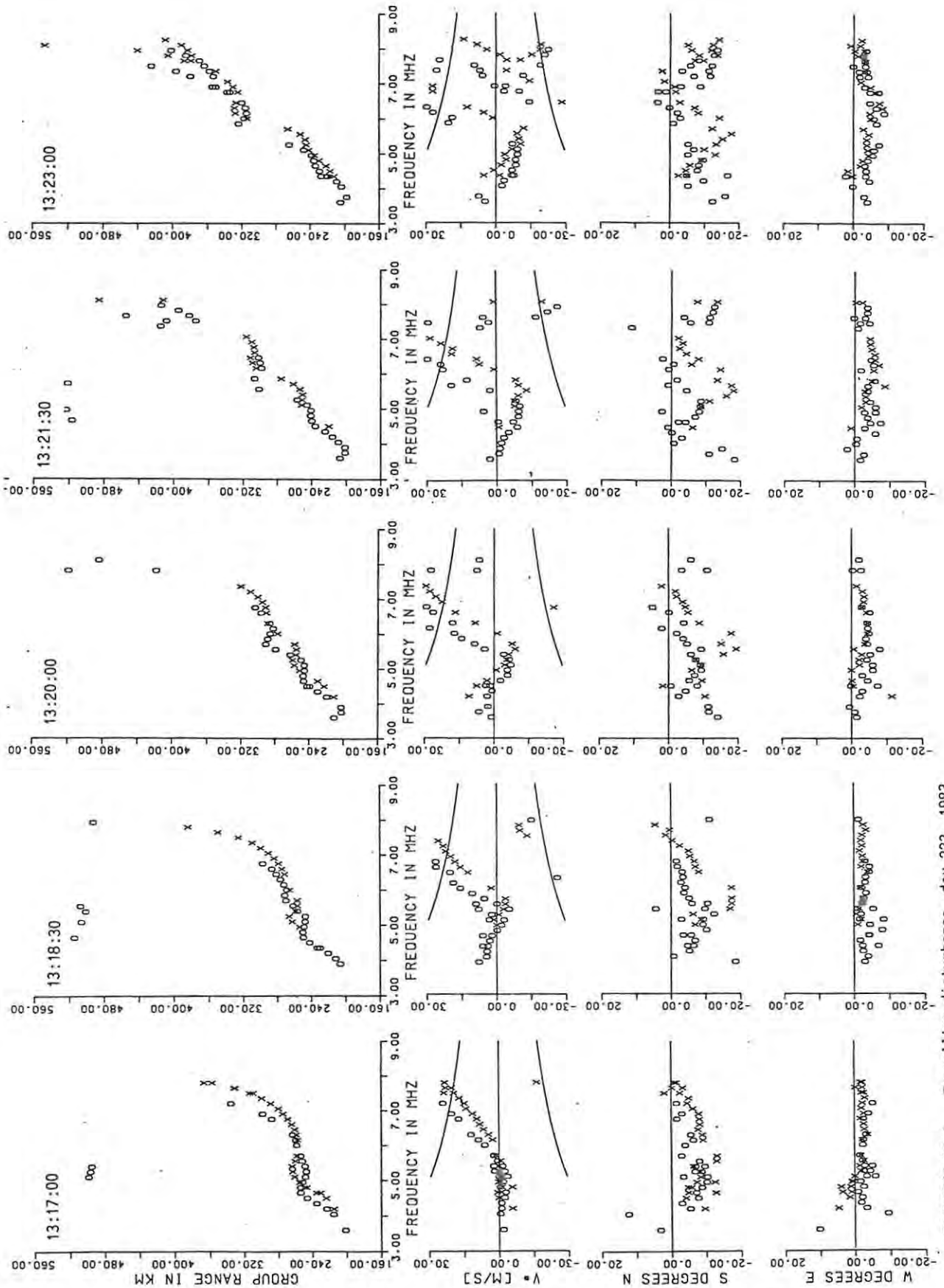
and de-focussing of the reflected energy by curved contours of equal electron density at the reflection level. The difference in the phases of the plots for the O- and X-modes shows that the effect is at least to some extent dependant on the plasma frequency.

Finally, the author's brief experience with the polarisation ellipses has shown that both the shape and the inclination of the ellipses can change with time and frequency. Thus, the total energy in the signal should be based on measurements made on orthogonal antennas, since the component of the amplitude on a single receiving antenna will depend on the ellipse orientation.

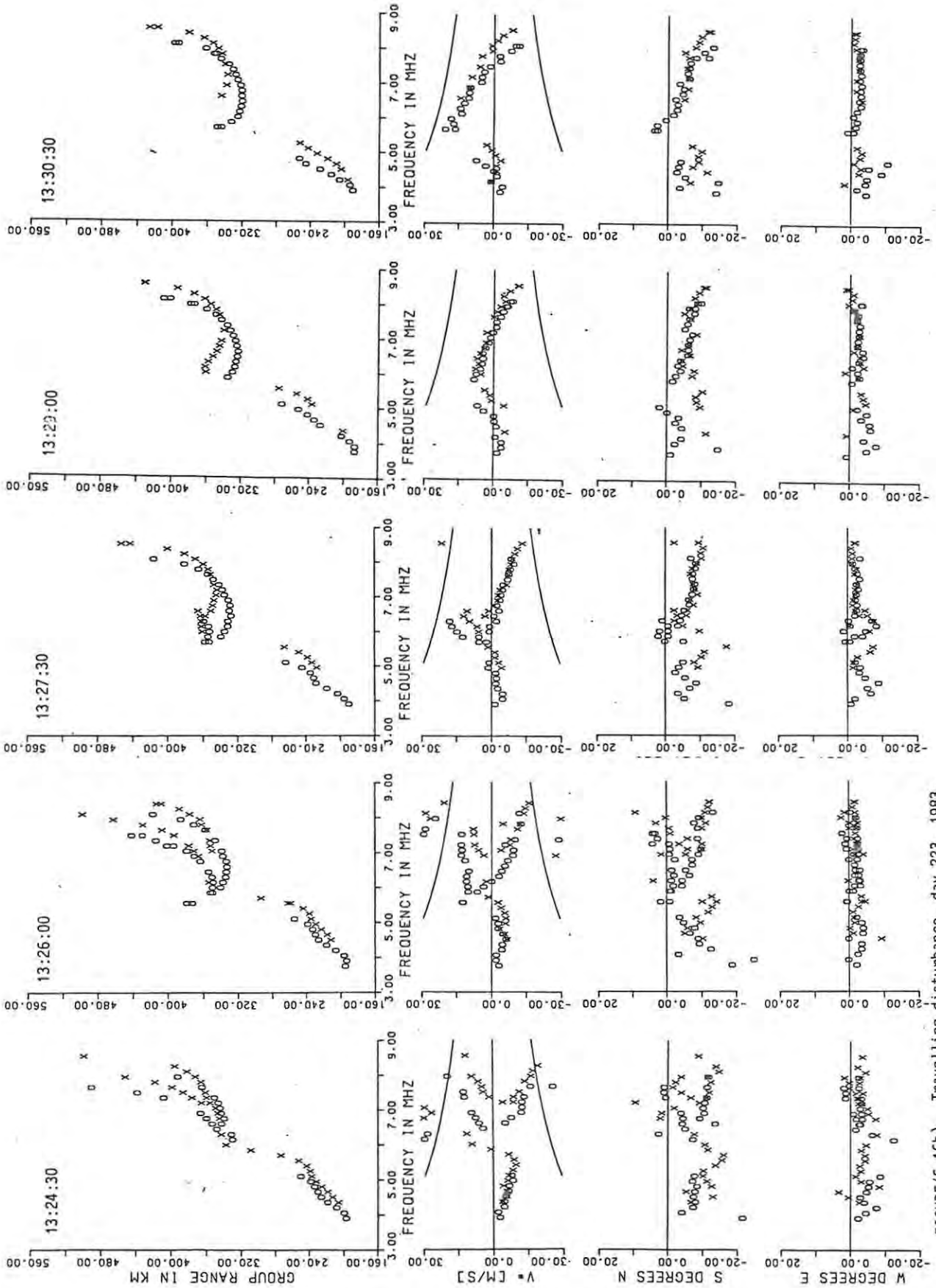
5.7 An interesting travelling ionospheric disturbance

As a further example of how the digital ionosonde can be used to supply useful information in a real sounding situation, the observation of the passage of a large travelling ionospheric disturbance is illustrated in figures 5.16(a) to (j) inclusive. The input parameters for each ionogram are shown in table (5.3). The ionosonde was programmed to perform a swept frequency sounding every 90 seconds between the frequencies of 3.0 and 11.0 MHz, although only the range 3.0-9.0 MHz is shown. This provided high time resolution of features of the event, however, since each ionogram took 80 seconds to record, there is some confusion of time- and frequency- related phenomena.

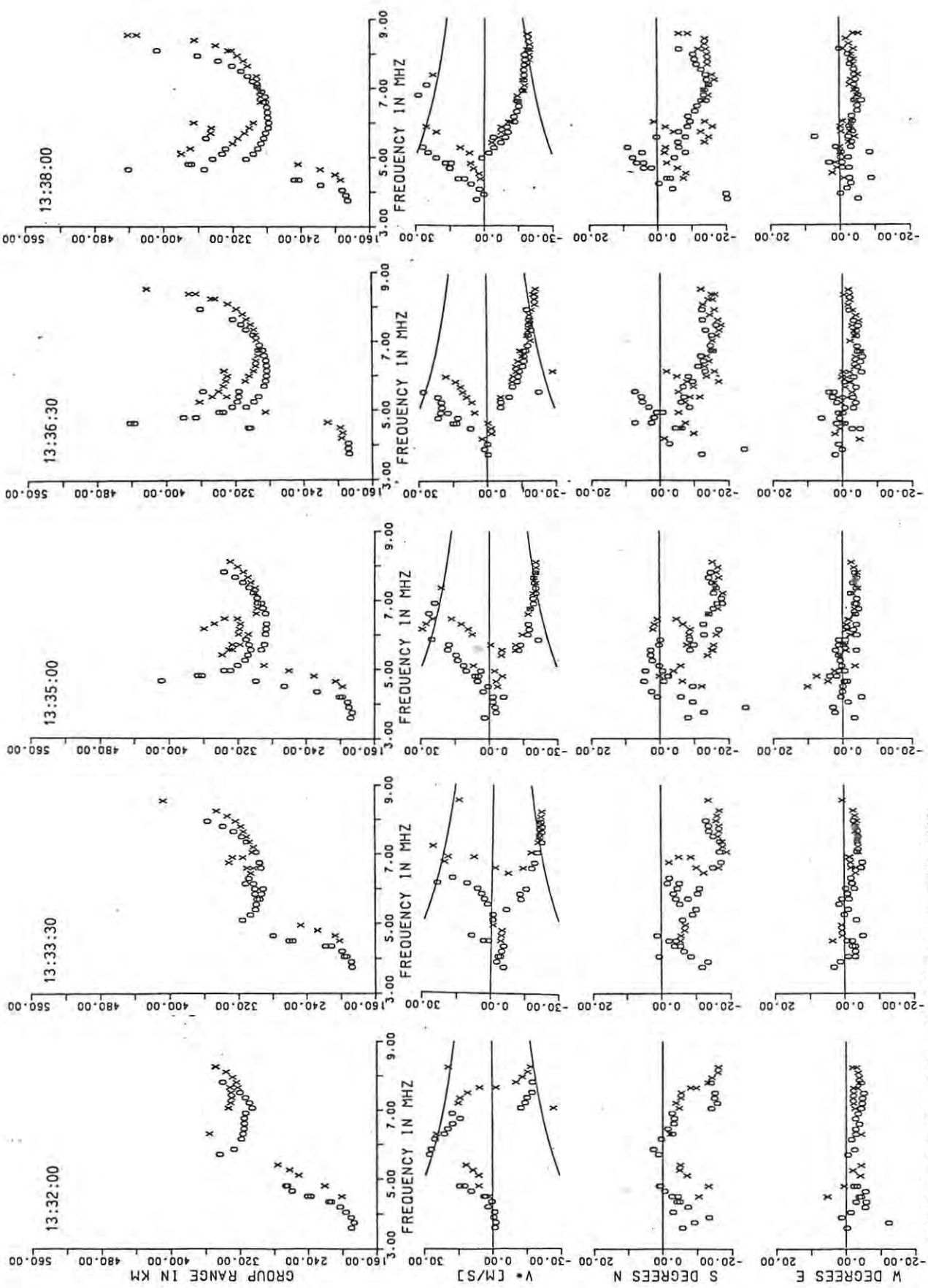
In figures 5.16 (a)-(c), the group range, phase velocity, and arrival angles in two directions are illustrated for the first 15 ionograms. For the same 15, the zenith angle, bearing, amplitude, polarisation angle and ellipse inclination angle are shown in figures 5.16 (d)-(f). Figure 5.16 (g) shows all the parameters for the remaining 2 ionograms. Figures 5.16 (h)-(j) show the echolocation plots for all 17 ionograms, where the sounding



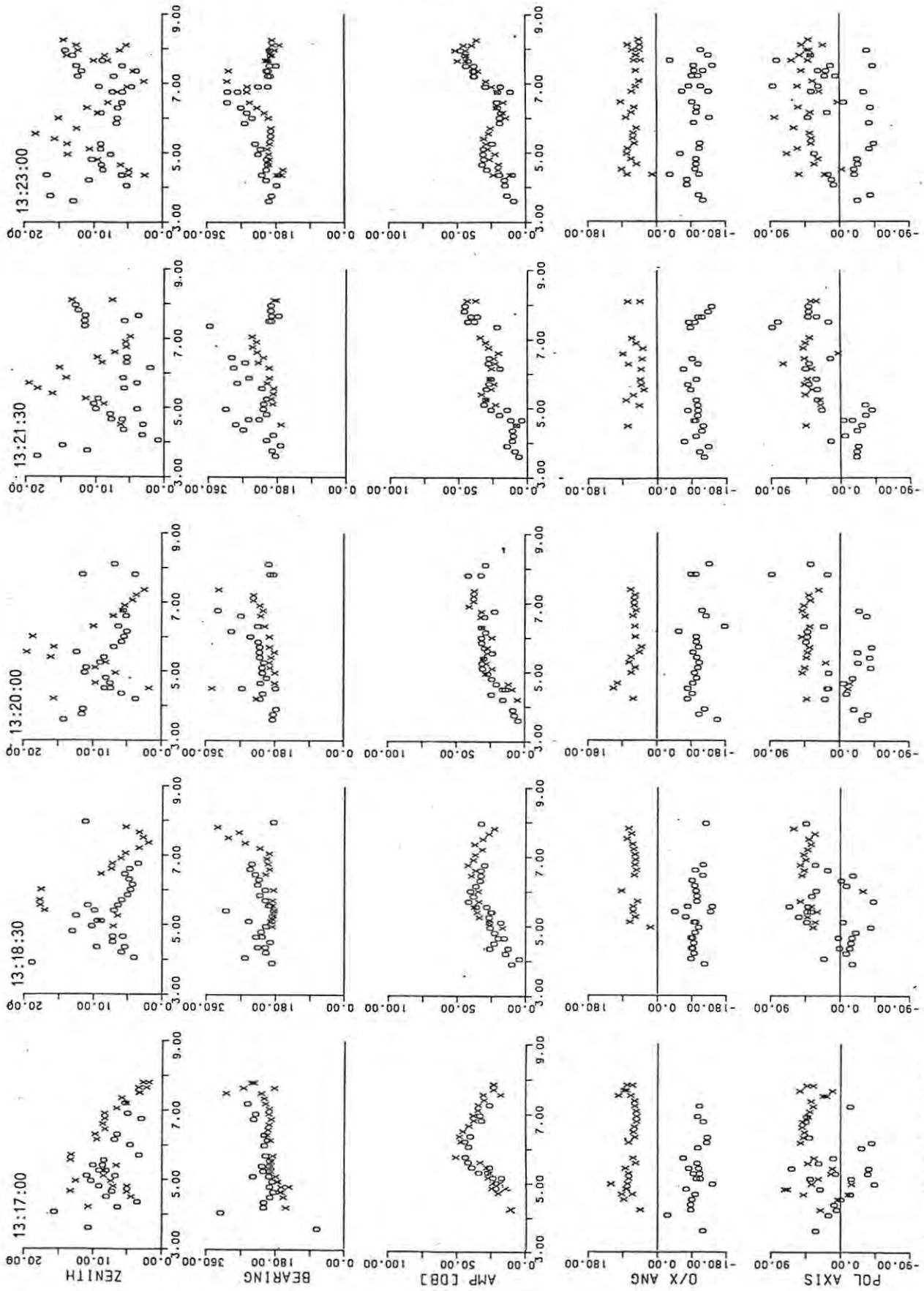
FIGURE(5.16a) Travelling disturbance, day 223, 1983.



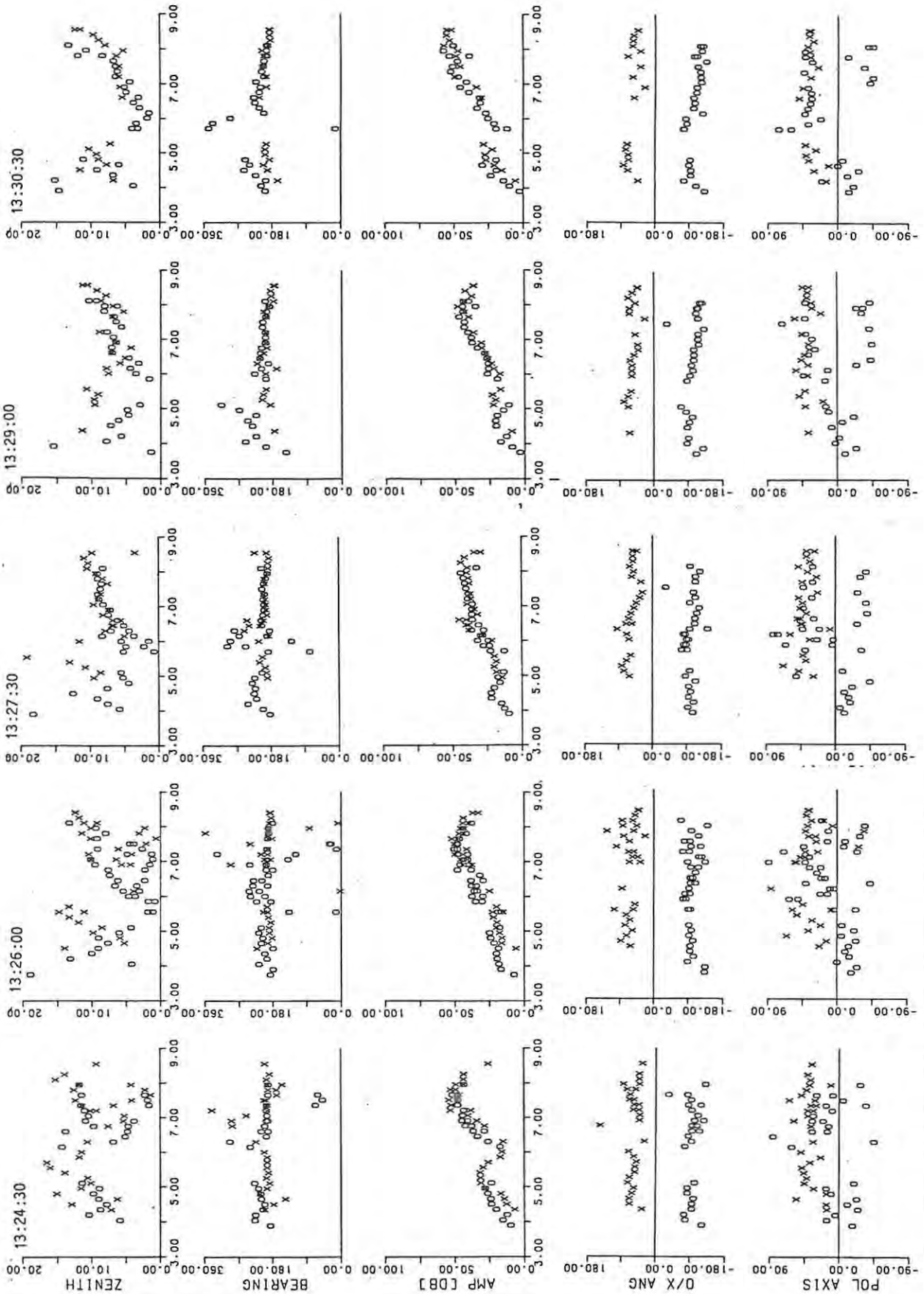
FIGURE(5.16b) Travelling disturbance, day 223, 1983.



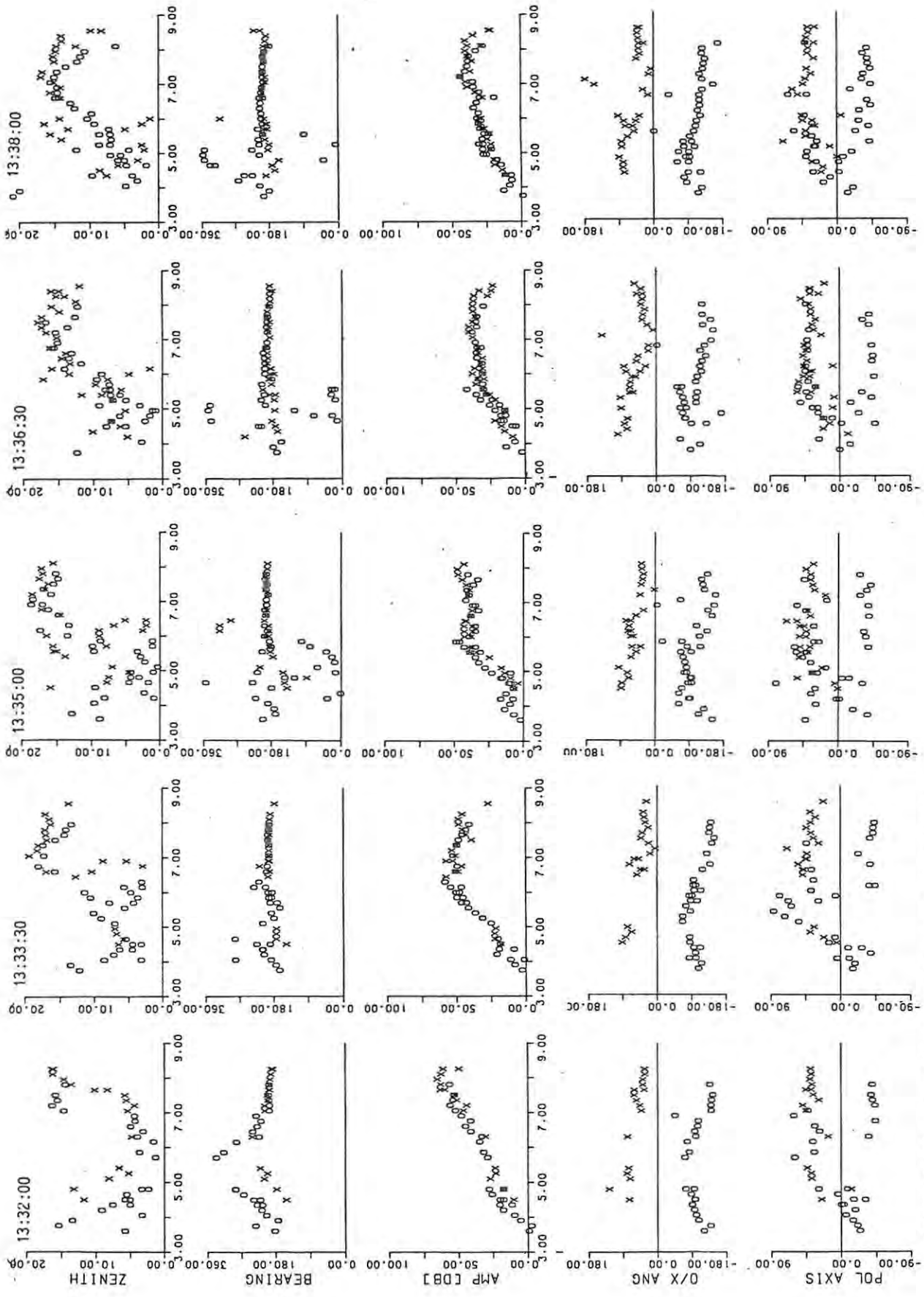
FIGURE(5.16c) Travelling disturbance, day 223, 1983.



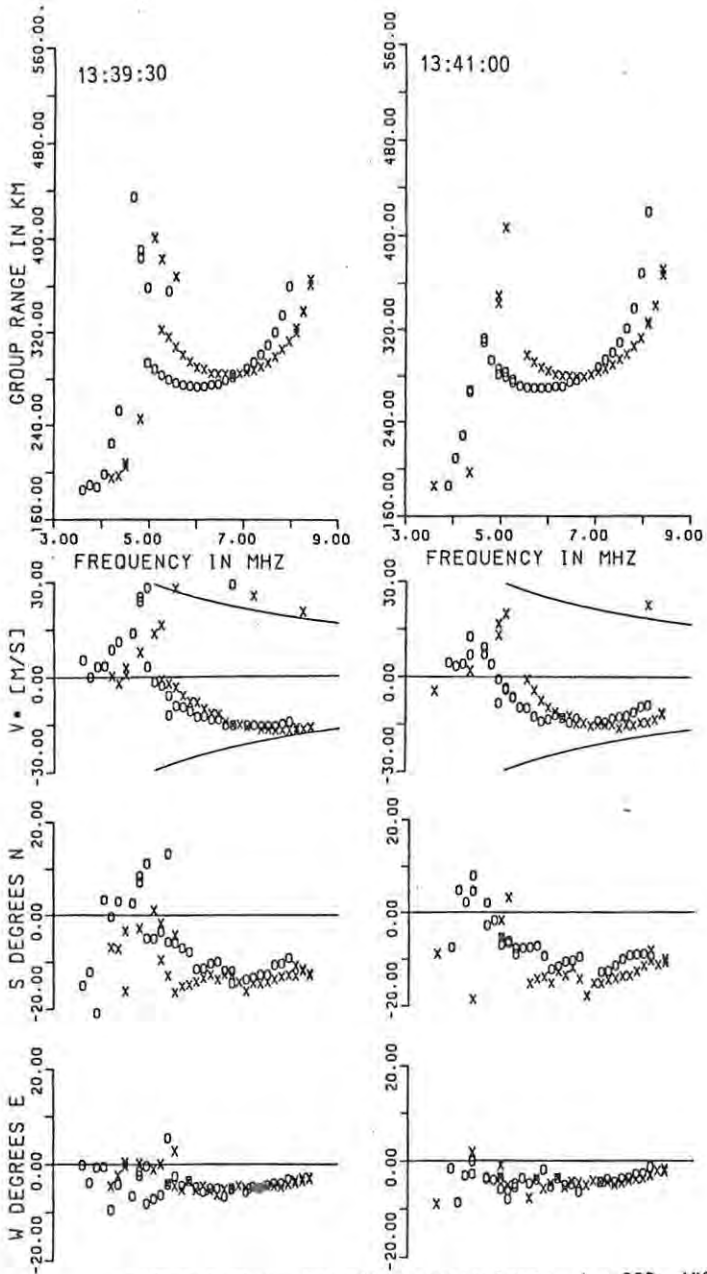
FIGURE(5.16d) Travelling disturbance, day 223, 1983.



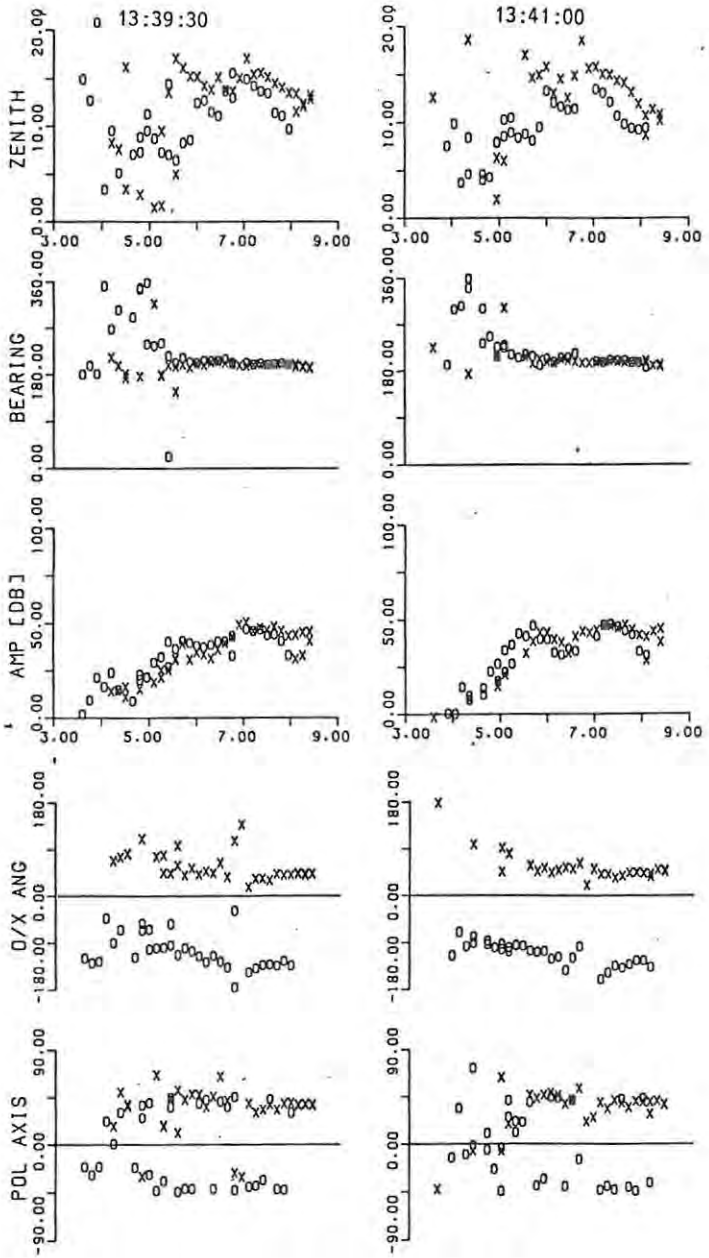
FIGURE(5.16e) Travelling disturbance, day 223, 1983.

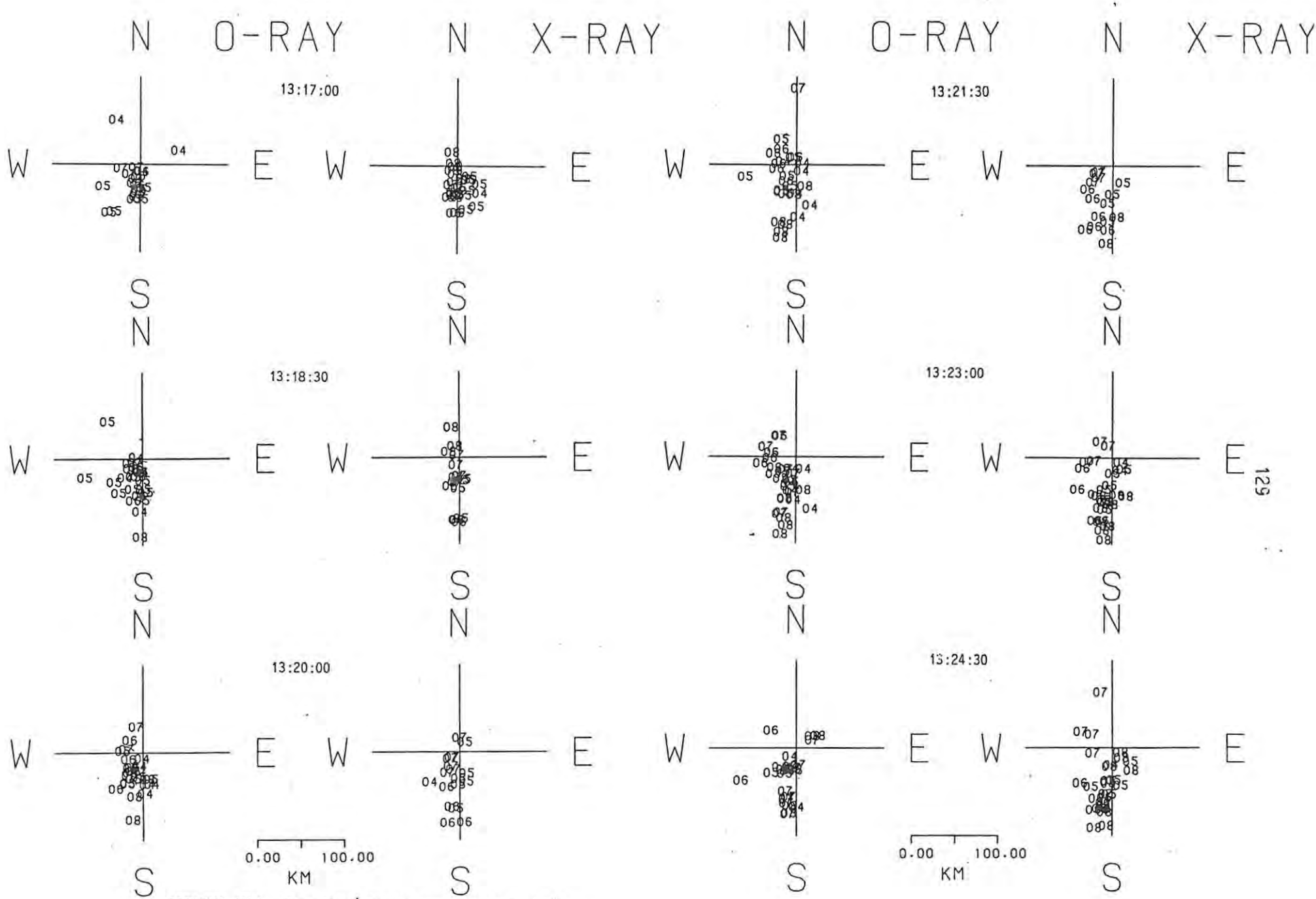


FIGURE(5.16f) Travelling disturbance, day 223, 1983.

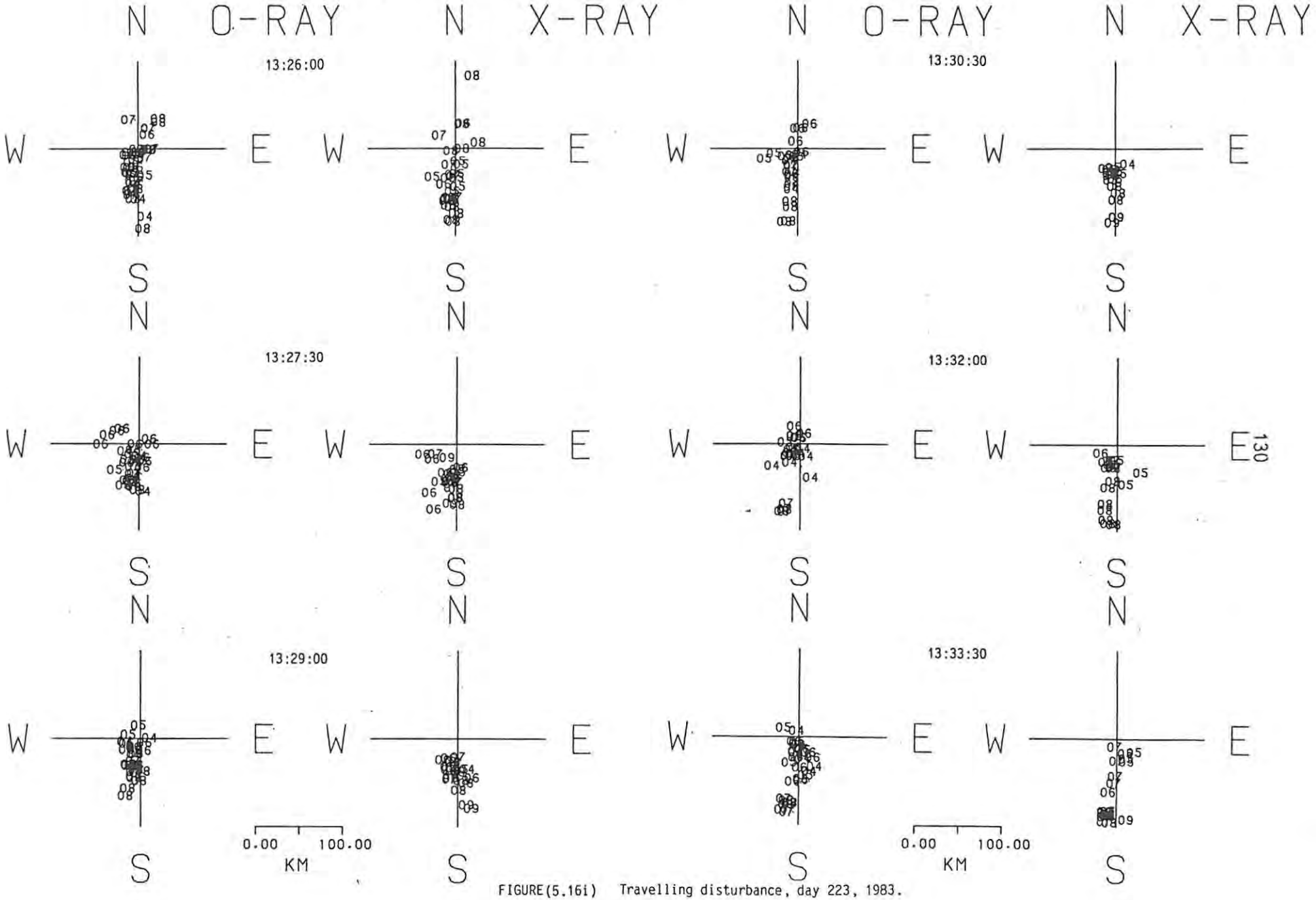


FIGURE(5.16g) Travelling disturbance, day 223, 1983.

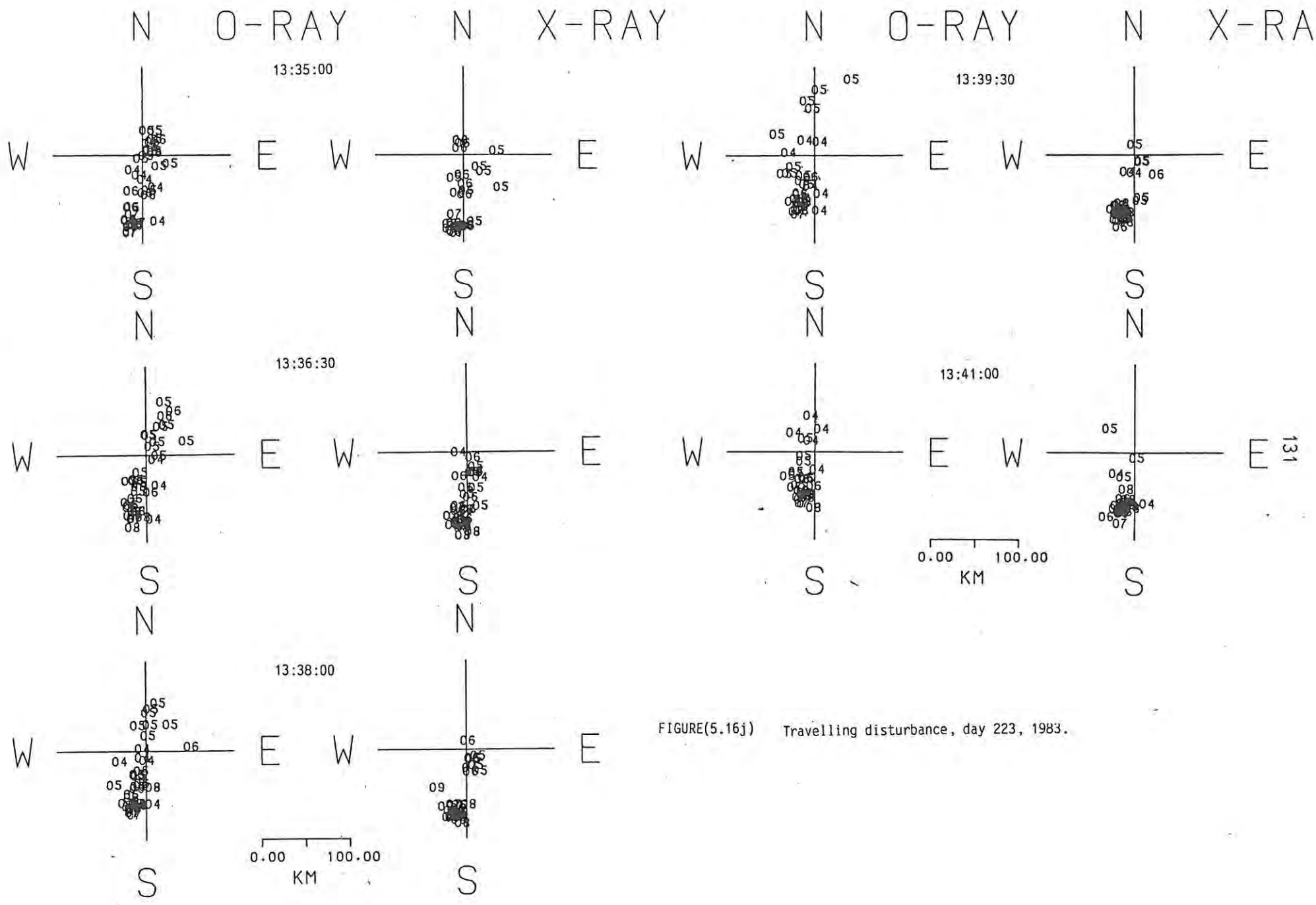




FIGURE(5.16h) Travelling disturbance, day 223, 1983.



FIGURE(5.16i) Travelling disturbance, day 223, 1983.



FIGURE(5.16j) Travelling disturbance, day 223, 1983.

frequency is denoted by a two digit decimal number, rounded to the nearest Mhz.

Some of the interesting features of this series of ionograms are as follows:

(i) At a given frequency, times when the phase velocity is positive (target receding) coincide with a northward deviation of the arrival angle, whilst negative phase velocities (target approaching) seem to favour a southward arrival angle. This indicates that the disturbances are moving from the south towards the north. The clearest examples of this are the N-S arrival angle and v^* data at 13:17:00 (figure 5.16 (a)), and at 13:30:30 (figure 5.16 (b)). This effect is also evident when more than one arrival angle is present at a particular frequency, as for 13:26:00 (figure 5.16 (b)).

(ii) the phase velocity v^* and the group velocity v' are related through [5] the equation:

$$v' = v^* + f \frac{\partial v^*}{\partial f}$$

At 13:17:00 (figure 5.16 (a)), an estimate of $\frac{\partial v^*}{\partial f}_x$ at 7.0 MHz can be obtained, being approximately (20 m s^{-1})(2 * 10^6 Hz). This gives:

$$(v'_x)_{f=7.0} = 15 + 7.0 * 10^6 * 10^{-5} = 85 \text{ m } s^{-1}$$

Between 13:17:00 and 13:23:00, a duration of 6 minutes or 360 seconds, the group height h'_x could be expected to rise by 85 * 360 * 10^{-3} = 30.6 km. This is indeed the case, confirming that the sign of v^* is evaluated in the correct sense.

(iii) the polarisation angle at 13:36:30 (figure 5.16 (f)) becomes double valued corresponding to the two arrival angles evident in figure

5.16 (c). There is thus evidence that the shape of the polarisation ellipse is affected by the distortion of the contours of equal electron density caused by the passage of the disturbance.

(iv) The echolocation plots of figures 5.16 (h)-(j) show the spread of the echoes to lie very much in the geographic North-South direction. Also noticeable is the general tendency towards the south for all the arrival angles.

5.8 Concluding remarks

The main thrust of this monograph has been in two regions, firstly to establish a theoretical background from which to develop a working digital chirpsounding system, and secondly to describe and demonstrate the performance of one such system and assess the success achieved. The first area is, of course, independent of any financial or time constraints, while the second has been very much influenced by the need to tailor the system from existing hardware within the framework of limited manpower, financial resources and development time. Despite these limitations, it is felt that a large measure of success has been achieved, and it is hoped that these efforts will point the way to improved versions of the digital chirpsounder. Indeed, at the time of writing, the National Institute for Telecommunications Research (NITR) of the Council for Scientific and Industrial Research (CSIR) has negotiated the purchase of an advanced instrument from the manufacturers [6], based on the experience acquired by the author and colleagues in the process of this research, and for which the author was pleased to act in the capacity of consultant. Even this system has been influenced to some extent by the expense that would be involved in designing anew "from the ground up", and has had to be designed around existing working modules.

Were it possible to begin from scratch, without regard to financial or

manpower limitations, the following are areas in which improvements could be made:

(i) In section 3.5 the theory was described from which the group range could be evaluated on the basis of simultaneous frequency diverse phase measurements. However, this would require the simultaneous transmission and reception of two parallel synthesized chirp waveforms. In practice, as described in section 3.9, the measurement is made with frequency and time diversity, and the time diverse contribution, which is independently evaluated, is subtracted out. This method is fundamentally unsatisfactory, since it requires that the second order derivative of the phase with respect to time be zero (section 5.3). Ideally, one would hope for hardware that would allow for the simultaneous Δf measurement described above. There are many ways in which this could be achieved, but at least two synthesizers would be a pre-requisite.

(ii) The FFT analyser currently employed in the Rhodes chirpsounder does not have the capability automatically to scale and record data that exceed the bounds of its arithmetic calculators. This severely limits the dynamic range of the device, and results in loss of data at both ends of the range. This is probably the single most important drawback of this sounder.

(iii) At present, automatic gain control of the receivers can only increment or decrement by one step (2 dB). Since the gain is required to be constant during the course of one sounding, to prevent the receivers from "following" the signal when overlapping echoes cause a low frequency beat signal, this means that the gain tracking of the receivers is inclined to be slow. A more sophisticated digital AGC feedback path allowing non-unit steps would be a recommendation.

(iv) The receiver bandpass of 500 Hz defines the limits of the range over which the difference frequency f_D can exist. However, this range can be

adjusted to cover any "window" of 500 Hz, provided the correct offset frequency is added to the receiver local oscillator signal. This requires the synthesis of an additional parallel chirp waveform containing the required offset. At present, this is achieved by generating a single sideband from the original signal by means of 90^{o} phase shifters and mixers, but a more satisfactory solution would be to employ an additional synthesizer. Both this feature and that mentioned in (i) above could be accomplished by the use of four parallel synthesizers, two to provide the transmit frequencies f and $f+\Delta f$, and two to provide the receiver local oscillator frequencies $f + f_{IF} + f_{window}$, and $f + \Delta f + f_{IF} + f_{window}$, where

f = nominal sounding frequency

Δf = group range offset frequency

f_I = first receiver intermediate frequency

f_{window} = required windowing offset

The use of four synthesizers is not as extravagant as it sounds since the technology is well developed and merely requires duplication.

(v) In section 5.2 it was mentioned that the efficiency of the current receiving antenna system is inadequate at frequencies below 3.0 MHz. This is not surprising, since no one antenna system of such necessarily simple design could be expected to cover the range .5 - 15 MHz (some 5 octaves). For this reason, another larger receiving antenna system is envisaged to operate below 3.0 MHz, but has not thus far been implemented due to lack of space at the field station. Similar arguments apply to the transmitting antenna system, which should ideally consist of a large inverted log periodic. This would involve the construction of four 100m masts which has obvious financial implications. Such improvements to the antenna systems would greatly increase the signal to noise ratio, resulting in more accurate

measurements.

In conclusion, the author wishes to restate the final paragraph of Chapter 1, namely, that this work has endeavoured to describe the theory and application of the FMCW technique to digital sounding, in order to advance our knowledge in this field. In addition, the author hopes that this effort will stimulate interest in the new techniques of digital sounding, and open the doors to a new area of research that have hitherto been closed.

REFERENCES: CHAPTER 5:

(1) Budden K.G., "Radio Waves in the Ionosphere", Cambridge University Press, 1961.

(2) Paul A.K., Numerical separation of overlapping ordinary and extraordinary echoes in digital ionograms, Radio Science, 18, 416, (1983).

(3) Wright J.W. and M.L.V. Pitteway, Real time data acquisition and interpretation capabilities of the Dynasonde 1. Data acquisition and real time display, Radio Science, 14, 815-825, (1979).

(4) Skolnik M.I., "Radar Handbook", McGraw-Hill Book Co., Library of Congress Card No. 69-13615, 1970.

(5) BR Communications, Sunnyvale, California, U.S.A.

APPENDIX A

DEFINITIONS

Real signal $x(t) = a(t)\cos(2\pi f_o t + \Psi(t))$ (A1)

Exponential form $s(t) = u(t)e^{j2\pi f_o t}$ (A2)

where $u(t) = a(t)e^{j\Psi(t)}$ is called the complex modulation

Thus, $x(t) = (1/2)\, u(t)e^{j2\pi f_o t} + u^*(t)e^{-j2\pi f_o t}$ (A3)

Fourier Transforms:

$$S(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft}dt \quad \text{(A4)}$$

$$s(t) = \int_{-\infty}^{\infty} S(f)e^{+j2\pi ft}df \quad \text{(A5)}$$

Total energy in signal:

$$E = \int_{-\infty}^{\infty} x^2(t)dt = \int_{-\infty}^{\infty} |X(f)|^2 df \quad \text{(A6)}$$

$$2E = \int_{-\infty}^{\infty} |u(t)|^2 dt = \int_{-\infty}^{\infty} |U(f)|^2 df \quad \text{(A7)}$$

Root-mean-square bandwidth:

$$\beta_o = \frac{(2\pi)^2 \int_{-\infty}^{\infty} f^2 |U(f)|^2 df}{\int_{\infty}^{\infty} |U(f)|^2 df} = \frac{\int_{-\infty}^{\infty} |u'(t)|^2 dt}{\int_{\infty}^{\infty} |u(t)|^2 dt} \quad \text{(A8)}$$

$u'(t) = du/dt$

Root-mean-square time duration:

$$t_o = (2\pi)^2 \frac{\int_{-\infty}^{\infty} t^2 |u(t)|^2 dt}{\int_{-\infty}^{\infty} |u(t)|^2 dt} = \frac{\int_{-\infty}^{\infty} |U'(f)|^2 df}{\int_{-\infty}^{\infty} |U(f)|^2 df} \qquad \text{(A9)}$$

$U'(f) = dU/df$

Time-phase coupling factor:

$$= \frac{2\pi \int_{-\infty}^{\infty} t\, \Psi'(t) |u(t)|^2 \, dt}{\int_{-\infty}^{\infty} |u(t)|^2 dt} \qquad \text{(A10)}$$

$\Psi(t)$ = complex modulation

$\Psi'(t) = d\Psi(t)/dt$

Effective bandwidth:

$$\beta_e = \frac{\left[\int_{-\infty}^{\infty} |U(f)|^2 df \right]^2}{\int_{-\infty}^{\infty} |U(f)|^4 df} \qquad \text{(A11)}$$

Effective duration:

$$t_e = \frac{\left[\int_{-\infty}^{\infty} |u(t)|^2 dt \right]^2}{\int_{-\infty}^{\infty} |u(t)|^4 dt} \qquad \text{(A12)}$$

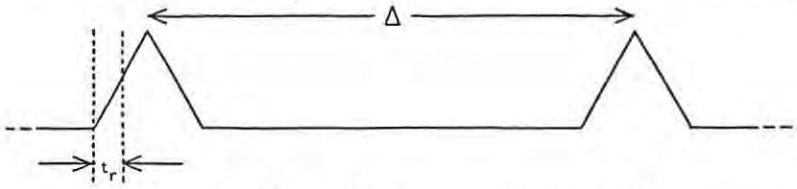
APPENDIX B

Calculation of the accuracy parameters β_o and t_o and the energy for:

(1)A string of N pulses

(2)Two adjacent chirp modulations

(1) A string of N triangular pulses, rise time = t_r, separation Δ.



$$t_o^2(\text{N pulses}) = (2\pi)^2\left[t_r^2/10 + \frac{(N-1)(N+1)}{12}\right]^2$$

(Reference: this thesis, Appendix C)

For a single pulse, $\beta_o^2(1\ \text{pulse}) = 3/t_r^2$ (Appendix C, Ref.[1], Section 5 equ. 88, $t_p=t_r$)

For N pulses, top and bottom of equ. 88 are multiplied by N, thus:

$$\beta_o^2(\text{N pulses}) = 3/t_r^2$$

Energy in N pulses:

For 1 pulse, $E(1\ \text{pulse}) = 1/2 \int_{-\infty}^{\infty} |u(t)|^2 dt$ [Ref. Appendix A, equ A7]

$$u(t) = \frac{+(a)}{t_r}(t + t_r) \qquad -t_r <= t <= 0$$

$$= \frac{-(a)}{t_r}(t - t_r) \qquad 0 <= t <= t_r$$

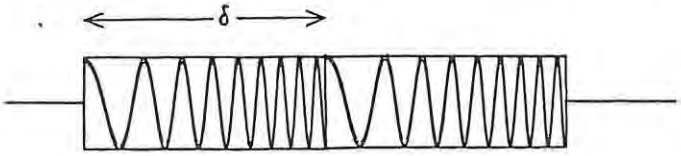
Now $a^2 = P(\text{pulse}) = \text{peak power}$

Therefore $E(1\ \text{pulse}) = (1/2).2.\int_0^\infty (a/t_r)^2 (t - t_r)^2\, dt$

$= a^2 t_r/3 = t_r P(1\ \text{pulse})/3$

For N pulses, $E(N\ \text{pulses}) = N t_r P(1\ \text{pulse})/3$

(2) Two identical adjacent chirp modulations, rate k kHz/s, length δ each, approximately rectangular.



For 1 modulation, length δ:

$$t_o^2(1\ \text{chirp}) = \pi^2\delta^2/3 \quad [\text{Ref:(1), equ. 86, } t_p >> t_r]$$

For 2 adjacent modulations, total length 2δ :

$$t_o^2(2\ \text{chirp}) = 4\pi^2\delta^2/3$$

For 1 modulation:

$$\beta_o^2(1\ \text{chirp}) = \pi^2 k^2 \delta^2/3 \quad [\text{Ref:(1), equ. 102, } 2/t_r t_t << \pi f_m{}^2/3,$$

$f_m = k t_p$]

For 2 modulations, multiply integral and W by 2 in equ. 101, and cancel:

$$\beta_o^2(2\ \text{chirp}) = \pi^2 k^2 \delta^2/3$$

Energy in 2 modulations:

$$E(2\ \text{chirp}) = 1/2 \int_{-\infty}^{\infty} |u(t)|^2\, dt = a^2/2 \int_0^{2\delta} dt = a^2\delta$$

$$= P(\text{chirp})$$

However, because of T/R limitation,

$$E(2\ \text{chirp}) = P(\text{chirp})/4$$

For equal range accuracy:

$$\sigma_t^2(N\ \text{pulses}) = \sigma_t^2(2\ \text{chirp})$$

$$N_o/2E(N\ \text{pulses})\beta_o^2(N\ \text{pulses}) = N_o/2E(2\ \text{chirp})\beta_o^2(2\ \text{chirp})$$

$$E(N\ \text{pulses})\beta_o^2(N\ \text{pulses}) = E(2\ \text{chirp})\beta_o^2(2\ \text{chirp})$$

$$[Nt_r P(\text{pulse})/3][3/t_r^2] = [\delta P(\text{chirp})/4][\pi^2 k^2 \delta^2/3]$$

$$P(\text{chirp}) = P(\text{pulse})\ .12N/(t_r\ \pi^2 k^2 \delta^3)$$

For $\beta_o(N\ \text{pulses}) = \beta_o(2\ \text{chirp})$

$$Nt_r P(\text{pulse})/3 = \delta P(\text{chirp})/4$$

$$P(\text{chirp}) = P(\text{pulse})[4Nt_r/3\ \delta].$$

For equal velocity accuracy

$$E(N\ \text{pulses})t_o^2(N\ \text{pulses}) = E(2\ \text{chirp})t_o^2(2\ \text{chirp})$$

$$\left[\frac{Nt_r P(\text{pulse})}{3}\right].(2\pi)^2.\left[\frac{t_r^2}{10} + \frac{(N-1)(N+1)\Delta^2}{12}\right] = \left[\frac{\delta P(\text{chirp})}{4}\right]\left[\frac{4\pi^2\delta^2}{3}\right]$$

$$P(\text{chirp}) = P(\text{pulse})(4Nt_r/\delta^3)\ \frac{t_r^2}{10} + \frac{(N-1)(N+1)\Delta^2}{12}$$

$$= P(\text{pulse})\ \frac{4Nt_r^3}{10\delta^3} + t_r\frac{N(N-1)(N+1)\Delta^2}{3\delta^3}$$

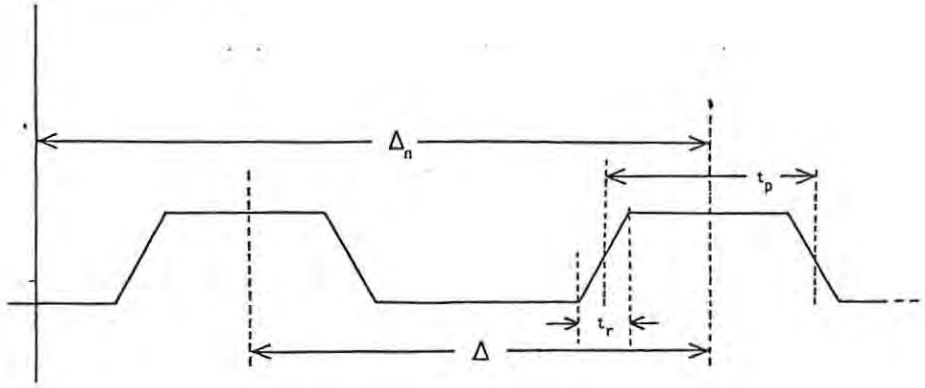
For $t_o(N\ \text{pulses}) = t_o(2\ \text{chirp})$

$$P(\text{chirp}) = P(\text{pulse})4Nt_r/3\delta\ .$$

APPENDIX C

Calculation of the RMS duration t_o of a string of N pulses of the form

$$u(t) = \sum_{n=1}^{N} u_n(t)$$



$$u_n(t) = \frac{A}{t_r}(t - \Delta_n + t_p/2 + t_r/2),$$

$$\Delta_n - t_p/2 - t_r/2 < t <= \Delta_n - t_p/2 + t_r/2$$

$$= A \quad , \quad \Delta_n - t_p/2 + t_r/2 < t <= \Delta_n + t_p/2 - t_r/2$$

$$= -\frac{A}{t_r}(t - \Delta_n - t_p/2 - t_r/2),$$

$$\Delta_n + t_p/2 - t_r/2 < t <= \Delta_n + t_p/2 + t_r/2$$

$$= 0 \quad , \text{ elsewhere.}$$

From A9,

$$t_o^2 = \frac{(2\pi)^2 \int_{-\infty}^{\infty} t^2 |u(t)|^2 \, dt}{\int_{-\infty}^{\infty} |u(t)|^2 \, dt}$$

$$= \frac{(2\pi)^2 \sum_{n=1}^{N} \int_{\Delta_n - t_p/2 - t_r/2}^{\Delta_n + t_p/2 + t_r/2} t^2 |u(t)|^2 dt}{2NE(1pulse)}$$

$$= (2\pi)^2 \sum_{n=1}^{N} I_n / 2NE(1pulse)$$

Divide I_n into 3 parts:

$$I_n = I_{n1} + I_{n2} + I_{n3}$$

then, putting $t_p/2+t_r/2=a$, $t_p/2-t_r/2=b$

$$I_n = (A/t_r)^2 \int_{\Delta_n - a}^{\Delta_n - b} t^2 (t - \Delta_n + a)^2 dt$$

$$+ A^2 \int_{\Delta_n - b}^{\Delta_n - a} t^2 \, dt$$

$$+ (A/t_r)^2 \int_{\Delta_n - a}^{\Delta_n + a} t^2 (t - \Delta_n - a)^2 dt$$

$$= A^2[(t_p^3 - t_p^2 t_r + t_p t_r^2)/12 - t_r^3/60 + \Delta_n^2\{t_p - t_r/3\}]$$

Now $2NE(1pulse) = A^2 N t_p[1 - t_r/3t_p]$ (ref. [1], sec.5, equ. 84)

$$t_o^2 = \sum_{n=1}^{N} I_n / [A^2 N(t_p - t_r/3)]$$

For a symmetrical set of pulses,

$$\Delta_n = (2n - 1 - N)/2, \ \Delta = \text{pulse separation}$$

$$\sum_{n=1}^{N} (\Delta_n)^2 = \Delta^2 N(N-1)(N+1)/12$$

Thus

$$t_o^2 = \frac{(2\pi)^2[t_p^3/12 - t_p^2 t_r/12 + t_p t_r^2/12 - t_r^3/60]}{t_p - t_r/3} + \Delta^2(N-1)(N+1)/12$$

If $t_p = t_r$

$$t_o^2 = (2\pi)^2 [t_r^2/10 + \Delta^2(N-1)(N+1)/12]$$

REFERENCE: Appendix C :

(1) Burdic W.S., "Radar Signal Analysis", Prentice-Hall Inc., Library of Congress Card No. 68-12906, 1968.

APPENDIX D

DATA CAPTURE SYSTEM (DCS): TESTS AND COMMANDS

General: When active, the DCS software is in a continuous command mode, monitoring serial port 1 for the input of some command. The data capture tasks are all handled under interrupts and are thus independant of, and invisible to, the command mode. In order to communicate such commands and receive a response, it is necessary to connect a Visual Display Unit (VDU) to port 1. A command is identified by a "FORTH word" which consists of a string of up to 32 ASCII characters (except space). The command is issued by merely typing that word at the terminal, followed by a carriage return.

The commands fall into 3 main groups:

(A) Tests: These are commands that have been specifically designed to allow some real time monitoring of the data, and involve a limited amount of analysis. All tests will continue until exited by means of depressing the space bar.

Command: ATEST

Display: Gives a visual display of the phase difference between the two receiver channels for the Fourier component of largest magnitude. The difference is displayed as a character displacement from zero at the centre of the screen, each single character representing $2\pi/80$ radians. The character displayed is the least significant hex digit of the Fourier component number.

Command: ITEST1

Display: Operates on the first cell only of a mulitcell sounding structure. A single character N is displayed representing the power spectrum amplitude expressed in the form $A = 2^N$. The ordinate of the character on the

screen represents the Fourier component number, the abscissa represents the frequency. All components that exceed the amplitude criterion are shown, the resulting display is thus similar in form to a conventional ionogram.

Command: ITEST2

Display: Same as for ITEST1, except that all cells are displayed.

Command: n TEST

Display: This command requires that a single digit number in the range 1-7 precede the command e.g. 4 TEST. The test will only work for a specific 3-cell sounding structure. All displays refer to the Fourier component of highest amplitude detected during cell 1.

1 TEST : Displays the amplitude of the power spectrum

2 TEST : Displays the Fourier component number

3 TEST : Displays the phase range velocity

4 TEST : Displays the East-West arrival angle

5 TEST : Displays the South-North arrival angle

6 TEST : Displays the polarisation mode

7 TEST : Displays all of 1-6 above

(B) Magnetic tape commands: These are commands that can be used to manipulate the magnetic tape drive. They should not be issued while an ionogram is in progress as they might conflict with the interrupt service routines which may themselves be accessing the drive.

MTRST : Reset the tape drive, cancelling all previous tape commands

MTR1BLK : Read 1 block of data from tape, output to VDU

MTRC : Read continuously from tape, output to VDU, stop on key

MTF1BLK : Move forward 1 block on tape

MTB1BLK : Move backward 1 block on tape

REW : Rewind to start of tape

REW* : Rewind to start of tape and place drive offline

MTWFM : Write file mark on tape

INITMT : Set up drive to write data in blocks of N characters in the continuous mode, until MTRST is issued. The command must be preceded by the number e.g. 800 INITMT.

?STATUS : Will query the current status of the tape drive. See the tape drive manuals for the interpretation of the status bytes output to the VDU.

(C) Other commands: Any FORTH word which has already been defined can be issued as a command by typing it at the VDU (with preceding numbers where appropriate). With some familiarity with the DCS software, the operator may use these commands for test purposes, as well as new commands which can be defined at the terminal. It is beyond the scope of this dissertation to enlarge upon these possibilities, and the reader is referred to the DCS test manual for further details.

APPENDIX E

MAGNETIC TAPE DATA STRUCTURE

The data is written to magnetic tape in blocks of 2048 ASCII coded characters. The block consists of 64 records of 32 characters each.

(A) Identification and structure: At the start of each ionogram, the first 9 records contain global identification and structure information as shown below. The blanks in the records correspond to spaces (ASCII code 20_{16}). Explanations are given in the column marked "comments".

Char. no.		1 8 16 24 32	comments
record no.	1	START	Always the same
	2	PPPPPPPPPPPPPPP	16 character place iden.
	3	19yy ddd hh:mm:ss	yy=year,ddd=day no.,etc.
	4	BASIC SWEEPRATE = ???? KHZ/SEC	????=4 digit decimal no.
	5	NO OF CELLS/SNDING = ??	?? = 2 " " "
	6	CELL LENGTH = 1/32 X 2**?? SECS	?? = 2 " " "
	7	FFT RATE = 2**?? HZ	?? = 2 " " "
	8	L.O. OFFSET = ????? HZ	?????=5 " " "
	9	TIMING SHIFT = ?????.?? MSEC	?????.?? 7 dig. dec. no.
	10	#??????????????????????????????	1st data record

(B) Data records: Each record from the 10th to the penultimate one inclusive will be a data record with the following format:

Char. no.	1 8 16 24 32
	#aaaaaaaabcdddeffggghhiijjkkllmm

```
#    : "Data" record identifier-always present
aaaaaaaa: 8 digit positive BCD number, frequency in Hz.
b       : 1 digit BCD no. in range 1-8, receiver A antenna connection
c       : 1   "    "  "   "   "    1-8,     "    B    "         "
ddd     : 3 digit unsigned hexadecimal no., sounding number
e       : 1   "      "         "       "  , cell number
ff      : 2   "      "         "       "  , AGC attenuation (*2 for dB)
ggg     : 3   "      "         "       "  , FFT component number (1=DC)
hh      : 2   "      "         "       "  , receiver A power spectrum amp.
ii      : 2   "      "         "       "  ,     "    B    "       "      "
jj      : 2   "    signed      "       "  , rec. A cosine amplitude
kk      : 2   "      "         "       "  ,  "   B    "       "
ll      : 2   "      "         "       "  ,  "   A  sine     "
mm      : 2   "      "         "       "  ,  "   B    "       "
```

(C) Finish record: At the end of the ionogram, the DCS will write one record containing the word "FINISH" as the first 6 characters, with the remaining 26 being padded out with spaces (ASCII code 20_{16}). The final block is then padded out with space filled records to make up the full 2048 characters.

APPENDIX F

FLOW DIAGRAMS

Flow diagrams for the interrupt servicing part of the DCS programme are shown in figs.(F.1) and (F.2). After an interrupt, the programme goes into a polling routine as shown fig.(F.1), to determine the interrupt origin. The three options are illustrated in more detail in fig.(F.2). The magnetic tape option M1-M2 is shown in fig.(F2.1) and shows the servicing of the status bytes which are output by the Mag. tape drive from time to time. Fig.(F2.2) shows the servicing of the controller interrupt. By arrangement with the controller software, the first byte of the controller data set is a "control" byte which determines one of four options as follows:

80_{16} = start of ionogram, global data, do not write to tape
81_{16} = start of ionogram, global data, write to tape
82_{16} = cell data
83_{16} = end of ionogram

If neither of these pairs of characters are encountered as the first bytes, no action is taken. Finally, fig.(F2.3) shows the procedure followed if the interrupt comes from the DMA interface. This interrupt is generated by the DMA interface at the end of the transfer of the FFT data. After reading the new AGC data, the DCS determines whether the FFT data refers to the first cell of a sounding. If so, the data is discriminated, sorted and stored. If the cell is not the first, the DCS will only record amplitudes for those FFT components that were selected in the first cell.

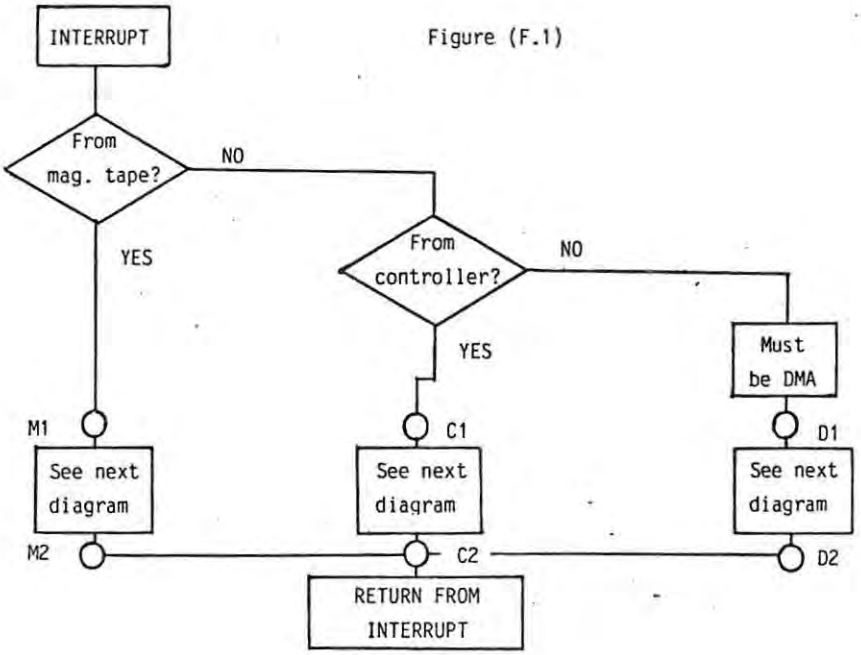


Figure (F.1)

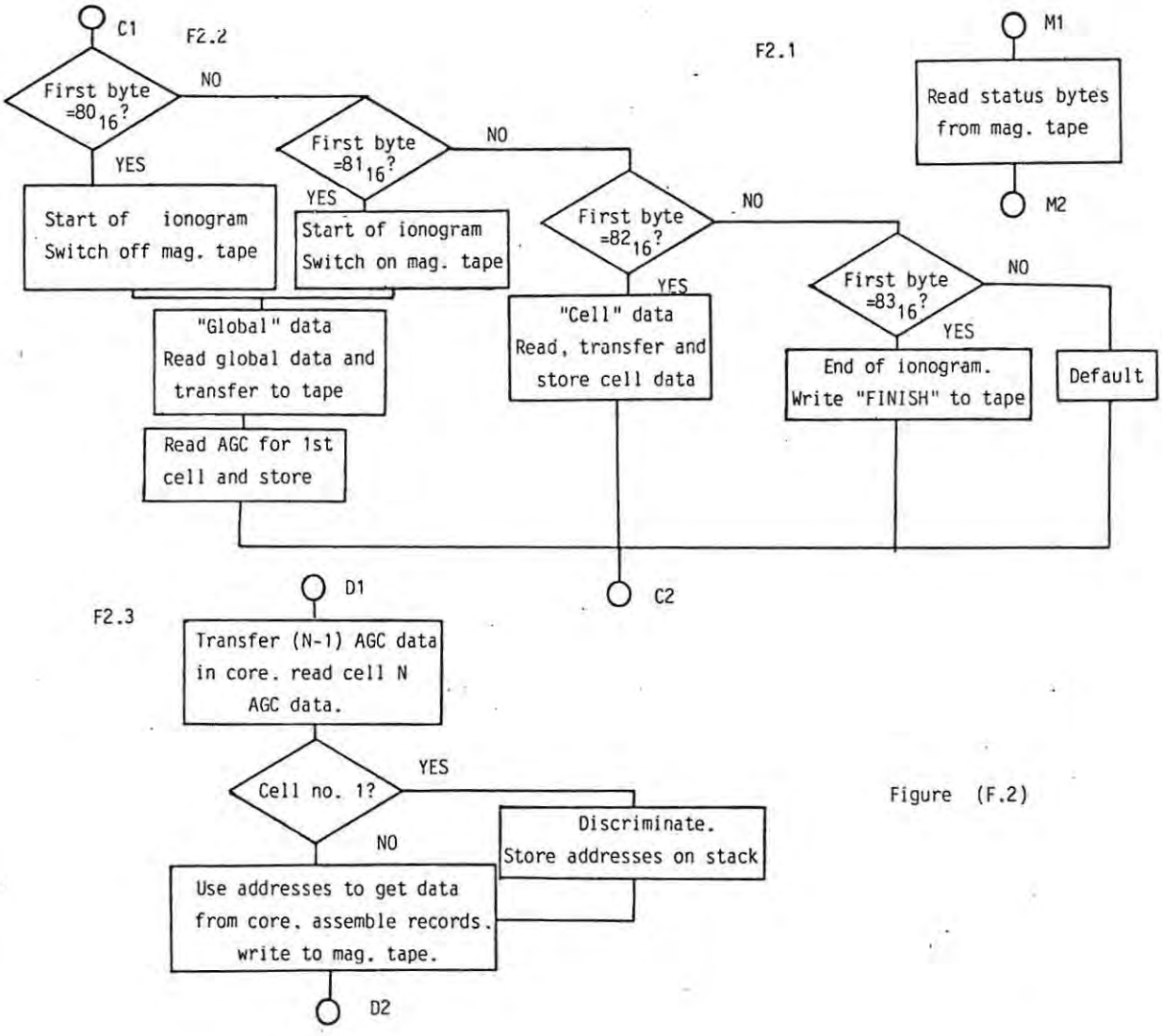


Figure (F.2)