INVESTIGATIONS IN SPACE AND LABORATORY PLASMA PHYSICS ARISING FROM POLAR IONOSPHERE TOPSIDE SOUNDER OBSERVATIONS AND RELATED SATELLITE MEASUREMENTS OF PRECIPITATED PARTICLE FLUXES

by

STUART FOSTER JAGGER

A thesis submitted for the degree

of Doctor of Philosophy in

the University of London

The Blackett Laboratory, Imperial College

January, 1978

Have courage to alter what can be altered,
Fortitude to endure what cannot be,
And wisdom to tell the difference.

Buddhist Prayer

ABSTRACT

The research arises from, or is directly related, to the interpretation of variable frequency, pulsed radar signals recorded by ionospheric topside sounder satellites. The work is in two parts. In the first section the distribution and extent of the polar exospheric plasma is investigated and an attempt is made to determine the source of the polar ionisation. The second part is concerned with a laboratory experiment to examine certain aspects of the propagation characteristics of plasma waves of the kind thought to be responsible for some of the 'resonance' signals on topside ionograms.

Selected electron density data, derived from Alouette I and II and from ISIS I satellite topside ionograms, are used to describe the distribution of plasma in the topside polar ionosphere and its temporal and spatial variations. It is shown that there is a characteristic density pattern which is dominant at all phases of the solar sunspot cycle and that the electron densities often exhibit a solar cycle modulation much greater than observed at mid-latitudes. The response of the distribution to the eccentrically-rotating geomagnetic pole is examined and two main effects are derived. Evidence is given for a wave of high electron density which co-rotates with the L-pole and is indicative of a possible high-latitude heat source. In addition, the universal time variation of the latitude of polar electron density gradients is used to infer a pronounced meridional motion of the plasma distribution in response to the rotation of the earth.

Then, using data from satellite-borne, low-energy particle detectors, an attempt is made to relate the observed densities and the measured fluxes of energetic electrons. Assuming suitable models for the polar neutral atmosphere composition, density and temperature, and using recent laboratory measurements for the range of energetic electrons in matter, calculations are performed to determine whether the observed values of electron

density can be maintained solely by collisional ionisation of neutral particles by energetic electrons precipitated into the polar region. It is shown that this is the case in the high density plasma 'ring' situated in the auroral region but at higher latitudes, near the geomagnetic pole, other production processes are necessary. Upward fluxes of particles above the F-region peak (principally of O^+ ions) are also important at polar latitudes and an attempt to measure these fluxes is described briefly.

The so-called 'resonances' on topside-sounder data have been interpreted in terms of the propagation of plasma waves in the ionosphere at frequencies close to certain characteristic frequencies of the plasma, notably, the electron plasma frequency, f_N, the electron upper-hybrid frequency, f_T, and the electron gyroharmonics, nf_H (n = 1, 2, 3, ...). Experiments conducted in a laboratory plasma to determine the propagation characteristics of plasma waves near the cyclotron harmonics (the Bernstein modes), are presented. Measurements which show the ability of these waves to yield information on the plasma parameters, electron density, N_e, and the electron temperature T_e, are presented and a possible effect of the ion 'sheath' examined. Their role in space plasma diagnostics in future projects is assessed.

Acknowledgements

It is a pleasure to acknowledge the assistance of my supervisor, Dr. J. O. Thomas, who initiated much of the work which follows and to whom I am deeply indebted for constant advice and encouragement.

I wish to acknowledge the assistance of Miss Susan Baker and Miss Colette Robertson. The former for typing this thesis and for data handling, the latter for her very efficient proof reading.

I would also like to acknowledge the fact that this work was made possible by a postgraduate grant from the Scientific Research Council and by data supplied by the Canadian Defence and Telecommunications Establishment.

Finally, I would like to thank my mother for help of a kind difficult to express in words.

CONTENTS

INTRODUCTION

There are now a variety of techniques available for the investigation of planetary plasma environments. Basically, these fall into two classes: remote sensing methods utilising radio waves received at or transmitted from the earth, and direct or in-situ observations from satellites or rockets located within the planetary plasma. Table I.1 forms a summary of the most important of these experiments and some of the parameters measured.

Through the combined application of such methods, the characteristics of planetary ionospheres and their controlling processes are gradually becoming known. The extent and composition of a planetary ionosphere is a function of many factors including the composition of the neutral atmosphere, and the planet's proximity to the sun. In addition, the interaction between the planet and the solar wind appears to be determined by the magnitude of its internal magnetic field and solar wind properties at the planet's orbit. Selected physical properties and typical parameters of the neutral and charged particle environments, for planets for which a plasma envelope has been detected, are shown in Table I.2. Bauer (1973) has recently reviewed present knowledge of planetary ionospheres.

Initially through routine ground-based sounding and more recently from satellites, the terrestrial ionosphere has been under scrutiny for over half a century (Beynon and Ratcliffe, 1974*). Surprisingly few of Chapman's original ideas on ionospheric formation at midlatitudes have changed, in that the D, E and F1 regions closely resemble simple Chapman layers. They are thought to be produced by solar ionising radiations in the X-ray and extreme ultra-violet wavelength range. The controlling processes are dominated by photochemical equilibrium. The F_2 layer is still not completely understood but seems to be best explained by the combined action of a linear loss process (with a loss coefficient, β, which

*This reference relates to a special issue of the Journal of Atmospheric and Terrestrial Physics containing review papers concerned with the history and present knowledge of the ionosphere.

Table I.1 Investigations of Planetary Plasma Envelopes

Technique	Experiment	Measurements
Remote sensing radio techniques	Faraday Rotation	Total electron content $\int N_e \, dh$
	Radio Occultation	Neutral and charged particle densities and temperatures as a function of altitude $N_e(h)$, $N_i(h)$, $(T_i+T_e)(h)$, $N_n(h)$, $T_n(h)$
	Ionospheric sounding (a) Ground-based bottomside sounding	(a) $N_e(h)$ in altitude range from ground to altitude of F2 peak density
	(b) Topside Sounding	(b) $N_e(h)$ from altitude of F2 peak density to altitude of satellite
	Incoherent Scatter	N_e, N_i T_i, T_e, m_i, ν_i, ν_e, ν_{in} as a function of altitude from 0 to a few x 1000 km
Direct or in-situ observations	Langmuir and allied direct current probes	N_e, T_e, N_i, T_i, m_i, $\underline{E}$
	Electron or ion mass spectrometers	N_e, N_i, T_e, T_i (or particle energy), m_i
	Interpretation of plasma 'resonance' signals and radio-frequency probes	N_e, N_i, T_e, T_i, m_i

Symbols are explained in Nomenclature (Part 1).

decreases with altitude giving an increasing electron density) and plasma diffusion. The F_2 region is also affected by dynamic processes of the neutral air. Incoherent scatter radars and members of the Atmospheric Explorer family of satellites are currently investigating the dynamic coupling of charged and neutral particle distributions and the complex ion chemistry of the low ionosphere.

However, the situation is much more complex at high latitudes and ground-based sounding has hitherto given very little insight into polar ionospheric structure and behaviour. In these regions factors additional to those at mid-latitudes must be considered. Here the magnetic field lines are swept back in the antisunward direction by the solar wind to form a tail-like structure. Unlike those at mid-latitudes they do not form closed loops (Russell,1972). Hence processes normally confined to the higher altitudes of the magnetosphere affect the topside F_2-region. Thus the ionospheric plasma is affected by the solar-wind-driven magnetospheric convection and the associated electric fields. This region is also subject to the effects of incoming energetic particles which have their origin in the solar wind and can reach the ionosphere, either directly at the so-called clefts or magnetic neutral points, or via the geomagnetic tail. It is thought that such particles may play an important role in maintenance of the polar ionosphere during the long winter night when appreciable production by solar photoionisation is absent. These and other processes give the polar ionosphere great temporal and spatial variability which a widely-spaced network of ground stations is unable to resolve. Only satellite or combined satellite and ground-based investigations are capable of providing a clear picture of the ionospheric microstructure.

The regime of magnetosphere-ionosphere-neutral atmosphere interaction is now a topic of much current interest and is under study from a variety of satellite and ground-based experiments. In particular, such satellites as Atmospheric Explorer C and E at lower altitudes and the ionospheric

TABLE I.2

Comparison data for planets with known ionospheres						
		Earth [1]	Mercury	Venus	Mars	Jupiter
Mean radius (km)		6371	2439	6050	3390	69500
[2] Average distance from Sun (AU)		1.00	0.39	0.72	1.52	5.20
[3] Mass of planet (m_P/m_E)		1.00	0.06	0.82	0.11	318.0
Acceleration due to gravity (cm s^{-2})		980	370	877	367	2500
[4] Magnetic moment (M_P/M_E)		1.0	6.4×10^{-4}	$<1.0 \times 10^{-4}$	$\sim 3.1 \times 10^{-4}$ (Mars 4)	$1.7 \times 10^{+4}$
[5] Solar flux (I_P/I_E)		1.00	6.65	1.90	0.43	3.7×10^{-2}
Neutral atmosphere	Constituents	O O_2 N_2 NO	He (confirmed) very tenuous	CO_2 (major) H (exosphere) CO He O (minor)	CO_2 (major) H (exosphere) O_2 N_2 Ar (minor)	H_2 (major) He CH_4 (minor)
	Neutral temperature (°K)	~1000	Near surface ~500°K (day)	<400	180 ±20	600 ±200
Ionosphere	Altitudinal extent (km)	~80-1000	Very near surface <200	100-300 (day) 100-1000 (night)	90-250 (day) No ionosphere observed at night	0-3000
	Peak electron density (electrons cm^{-3})	$\sim 10^6$ (day) $\sim 10^5$ (night)	$\lesssim 1 \times 10^3$ (day)	a few $\times 10^5$ (day) 10^4 (night)	$1\text{-}2 \times 10^5$ (day) $<5 \times 10^3$ (night - upper limit)	a few $\times 10^5$
	Height of ionisation peak (km)	200-400	-	~140 (day) ~120 (night)	140 (day)	multi-peaked 500-1500
	[6] Ion species	O^+ H^+ He (F-region) O_2^+ NO^+ (E-region) H_3O^+ $(H_2O)_n$ (D-region)	-	He^+ (major) CO_2^+ (minor)	O_2^+ CO_2^+	H^+ most likely H_3^+ CH_3^+ possible
	Topside plasma scale height (km)	~150	-	13 near peak day 725 (at 200 km - day) 30 near peak night 1500 at 800 km - night	35-40 (day)	675 ±300
	Plasma temperature (°K)	$T_i \sim 1200$ $T_e \sim 1500$ (500 km)	-	~500 (exosphere) ~5000 ($T_e + T_i$)	$T_i \sim 160$ ±8	$T_e + T_i \sim$ 900 ±400

Table I.2 (cont.)

Notes

1. Conditions given for the earth are typical of quiet-time mid-latitudes.
2. The distance from the sun is in Astronomical Units (AU) where 1 AU = 1.5×10^{11} m.
3. The mass of a planet, m_p, is normalised to that of the earth, m_E, given by 6.0×10^{24} kg.
4. The magnetic moment of a planet, M_p, is normalised to that of the earth, M_E, given by 8.1×10^{25} gauss cm^3.
5. The solar flux intensity outside a planetary atmosphere, I_p, is normalised to that observed in the terrestrial exosphere, I_E, given by 2 erg cm^{-2} s^{-1} (in the wavelength range 1310-270 Å).
6. Trace constituents (e.g. sulphur, phosphine), which in some cases may form a very important chain in the ion and neutral chemistry, have not been included.

observatories, ISIS I and II, at higher altitudes are monitoring global variations of these interactions. These satellites are comprehensively equipped to sense many parameters including charged particle densities and temperatures and fluxes of energetic particles. The versatility of incoherent scatter radars in high latitude studies has been impressively demonstrated by the facility at Chatanika, Alaska, (Rishbeth, 1976) and a further high latitude station, EISCAT, will shortly be observing in Scandanavia (Beynon, 1974). Ground radars will close an important gap in satellite observations in that they will sample diurnal variations of the ionospheric and magnetospheric plasma. These observations will clearly increase understanding of the polar ionospheric regions merely by increasing data coverage of such an inaccessible region.

One purpose of the present work is to obtain a comprehensive picture of the distribution of F-region and topside ionisation over the winter pole and to examine the possibility that such a distribution can be maintained by collisional ionisation effects of incoming particles. An investigation of this type is appropriate at this time since adequate data coverage is just now becoming available. The survey of the electron density distribution utilises seven years of observations from the series of topside-sounding satellites (Table I.3), Alouette I, Alouette II and ISIS I.

In such experiments essentially the delay time, between the transmission of a short RF pulse and its return to the satellite from its reflection point in the ionosphere, is measured as a function of frequency, f. These data are usually shown in a photographic format called an ionogram. A schematic example is shown in Figure I.1. Here the time delay, t, has been converted to a virtual depth of reflection, h', calculated by assuming the wave packet propagates at the speed of light, c, so the ionogram shows h'(f), where $h'(f) = \frac{1}{2} ct(f)$. The three normal electromagnetic traces (corresponding to the O, X and Z modes) are clearly seen. These are well understood in terms of cold plasma or Appleton-Hartree theory

Table I.3 IONOSPHERIC TOPSIDE-SOUNDING SATELLITES

		ALOUETTE I	EXPLORER XX	ALOUETTE II	ISIS I	ISIS II
Launch date		29.9.62	25.8.64	29.11.65	30.1.69	1.4.71
Perigee (km)		996	866	502	574	1358
Apogee (km)		1032	1010	2982	3522	1428
Period (min)		105.4	103.8	121.4	128.2	113.6
Inclination (°)		80.46	79.9	79.8	88.4	88.1
Sounding	Frequency (MHz)	.45 to 11.8	6 fixed frequencies: 7.22 5.47 3.72 2.85 2.0 1.5	.2 to 13.5	.1 to 20	.1 to 20
	Transmitter power (w)	100	8 to 45	300	400	400
Receiver (bandwidth)		400Hz-10kHz		50Hz-30kHz	50Hz-30kHz	50Hz-30kHz
Additional experiments		Particle spectrometer	Particle spectrometer	Particle spectrometer; Langmuir probe	Soft particle spectrometer; Ion mass spectrometer; Langmuir probes; Ion probe	

and yield information on the electron density profile with altitude. Also prominent are the so-called 'resonance' signals, or short delay echoes. These signals, not predicted by cold-plasma theory, appear at certain characteristic frequencies of the plasma.

From such data the trough-ring-cavity structure originally suggested by Thomas and Andrews (1970) is shown to persist, in the F region topside ionosphere, during periods of high solar sunspot number. With large amounts of data available the variations of this structure are analysed in an attempt to isolate processes which control the ionisation infrastructure. Thus densities and boundaries are shown to vary systematically with sunspot number and universal time.

Only recently have observations of fluxes of energetic particles produced measurements at energies relevant to topside and F-region physics, i.e. ∿100 eV. These results have been reduced to the form of energy spectra (largely from satellites detailed in the Appendix) and used with data regarding the penetration of electrons into air, in a study to determine the levels of ionisation produced solely by collisional ionisation effects of these particles. Comparisons between the observed and calculated ionospheres are then made.

Early work showed that interpretation of high-latitude ionograms, such as that shown in Figure I.1, is greatly enhanced by use of the 'resonance' signals. These short-delayed echoes occur at the characteristic frequencies of the plasma - the electron plasma frequency f_N, the electron upper-hybrid frequency, f_T, and the harmonics of the electron gyrofrequency, nf_H (n = 1, 2,...). MacAfee (1968, 1969a,b) has attributed these observations to the reflection of plasma waves, predicted by the hot plasma theory (Bernstein, 1958), from points near to the satellite. Recent work shows that the wavelength of a wave excited in the magnetoplasma is parametric in electron density, N_e, and temperature, T_e, and much current interest is centred on possible applications for planetary plasma diagnostics. Thus plasma wave experiments have been performed with

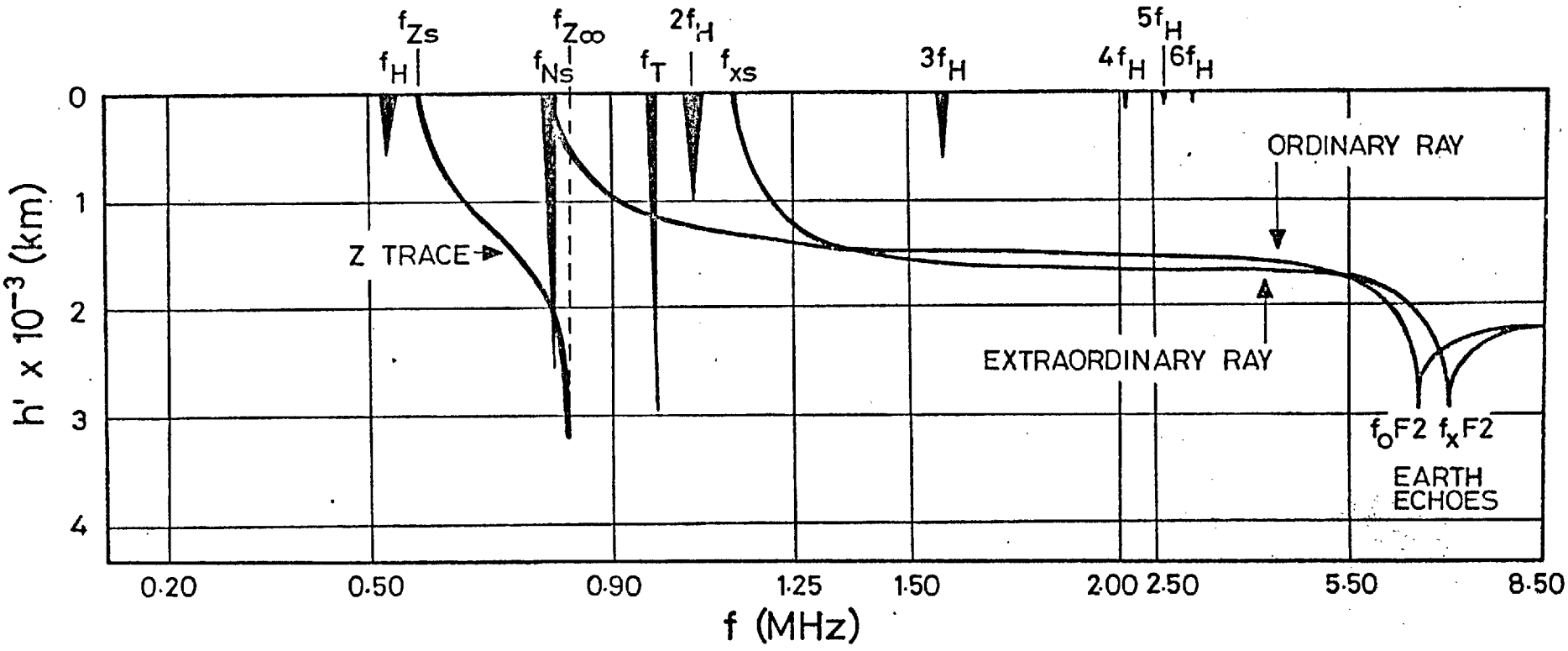


Figure I.1 A schematic Alouette II, high-latitude ionogram with typical scales of virtual height, h', and sounding frequency, f. The traces for the ordinary, extraordinary and Z propagation modes are shown. 'Resonances' or short-delay echoes at the satellite altitude, attributed to propagation of various classes of plasma wave, are clearly visible at the electron plasma frequency, f_N, the electron upper-hybrid frequency, f_T, and harmonics of the electron cyclotron frequency, nf_H (n = 1, 2, ...).

the sounder of the ISIS I satellite (Warnock, et al., 1970). Propagation experiments in space are now being attempted with the GEOS payload and will form a part of the AMPS Skylab programme.

Work in the second part of this thesis describes a laboratory examination of the propagation characteristics of one class of plasma wave, those propagating at harmonics of the electron-gyrofrequency (nf_H) or Bernstein modes. It is shown that the small scale length of these waves may make them especially suitable for the measurement of N_e and T_e in plasma irregularities, or in an ionosphere of irregular structure, such as Jupiter's (Fjeldbo et al., 1975).

A discrepancy between theory and experiment is investigated and explained in terms of an ion sheath around the antennae. This may lead to measurements of sheath thickness. Finally the advantages of the use of such waves in space over traditional diagnostic methods are considered.

References - Introduction

Bauer, S.J., Physics of Planetary Ionospheres, Springer-Verlag, Berlin, 1973.

Bernstein, I.B., Phys. Rev. 109, 10, 1957.

Beynon, W.J.G., Contemp. Phys. 15, 329, 1974.

Beynon, W.J.G. and J.A. Ratcliffe, Fifty Years of the Ionosphere, Special Issue J.A.T.P. 36 (No. 12), 1974.

Fjeldbo, G., A. Kliore, B. Seidel, D. Sweetman and P. Woicestyn, Paper presented at the IAU Jupiter Colloquium, University of Arizona, May 1975.

McAfee, J.R., J.G.R. 73, 5577, 1968.

McAfee, J.R., J.G.R. 74, 802, 1969a.

McAfee, J.R., J.G.R. 74, 6403, 1969b.

Rishbeth, H., Contemp. Phys. 17, 483, 1976.

Russell,C.T., In Critical Problems of Magnetospheric Physics, Ed. E.R. Dyer, 1972.

Thomas, J.O. and M.K. Andrews, Planet. Space Sci. 17, 433, 1969.

Warnock, J.M., J.R. McAfee and T.L. Thompson, J.G.R. 75, 7272, 1970.

PART I

THE POLAR EXOSPHERIC PLASMA

Preamble to Part 1

The first part of this work is concerned with a theoretical and experimental investigation of the polar topside ionosphere. It is made up of three chapters, of which the first contains a comprehensive summary of current knowledge of the electron density distribution and its controlling processes. This is necessarily detailed, due to the absence of any recent similar survey.

From this chapter two topics are chosen and considered further in Chapters 2 and 3. The former is made up of an attempt to extend present ideas on the electron density distribution, through a large-scale survey of topside-sounder satellite data. The final chapter of part 1 addresses itself to the question of the source of the polar ionosphere in winter and assesses the contribution to the total ionisation from incoming precipitating particles.

Central to this work is a knowledge of electron spectra. A study of all known observations was made which is very extensive and because of the detail involved this is included as a self-contained survey forming the Appendix of this section.

Nomenclature (Part 1)

$\underline{B}$	magnetic induction
CGL	corrected geomagnetic latitude
CGT	corrected geomagnetic time
e	modulus of electron charge
$\underline{E}$	electric field
F	flux of ions or electrons
f_H	electron gyrofrequency
f_oF2	ordinary penetration frequencies of the F2 layer
f_xF2	extraordinary penetration frequency of the F2 layer
g	acceleration due to gravity
h	altitude above the earth
H	measured plasma scale height = $-\frac{1}{N}\frac{dN}{dh}$
H_p	$= kT_p/m_ig$, scale height of plasma in diffusive equilibrium
h_mF2	height of occurrence of the maximum ionospheric electron density
JF02	the tabulated Alosyn parameter
k	chemical rate constant or Boltzmann's constant
L	magnetic L-shell or loss of ionisation
LT	local solar time
m_i	ion mass
MLT	magnetic local time
N	total ion or electron density
N_e, n_e	electron density
N_i, n_i	species ion density
N_mF2	maximum electron density of the F2 layer
N_s	electron density at the satellite
Q	production of ionisation
R	smoothed Zürich sunspot number
R_E	radius of the earth

t	time
T_e	electron temperature
T_i	ion temperature
T_p	$= T_e + T_i$
u, v	particle velocity (subscript i - ion and e - electron)
UT	universal time
z	atmospheric depth
β	electron loss coefficient or dipole tilt angle
λ	geographic latitude
Λ	invariant latitude
Λ'	generalised invariant latitude
ν_{in}	ion-neutral collision frequency
χ	solar zenith angle

CHAPTER 1

THE POLAR TOPSIDE IONOSPHERE

1.1 Polar F-region Structure

Before the physical processes active in the polar F-region ionosphere can be ascertained, it is important to obtain as precise and detailed a picture as possible of the spatial distribution of ionisation there and of its temporal variations. A number of methods are now available for this purpose and it is appropriate, as an introduction, to examine some of what has been learnt from the application of several of the most widely-used of these techniques.

1.1.1 Early ground-based observations

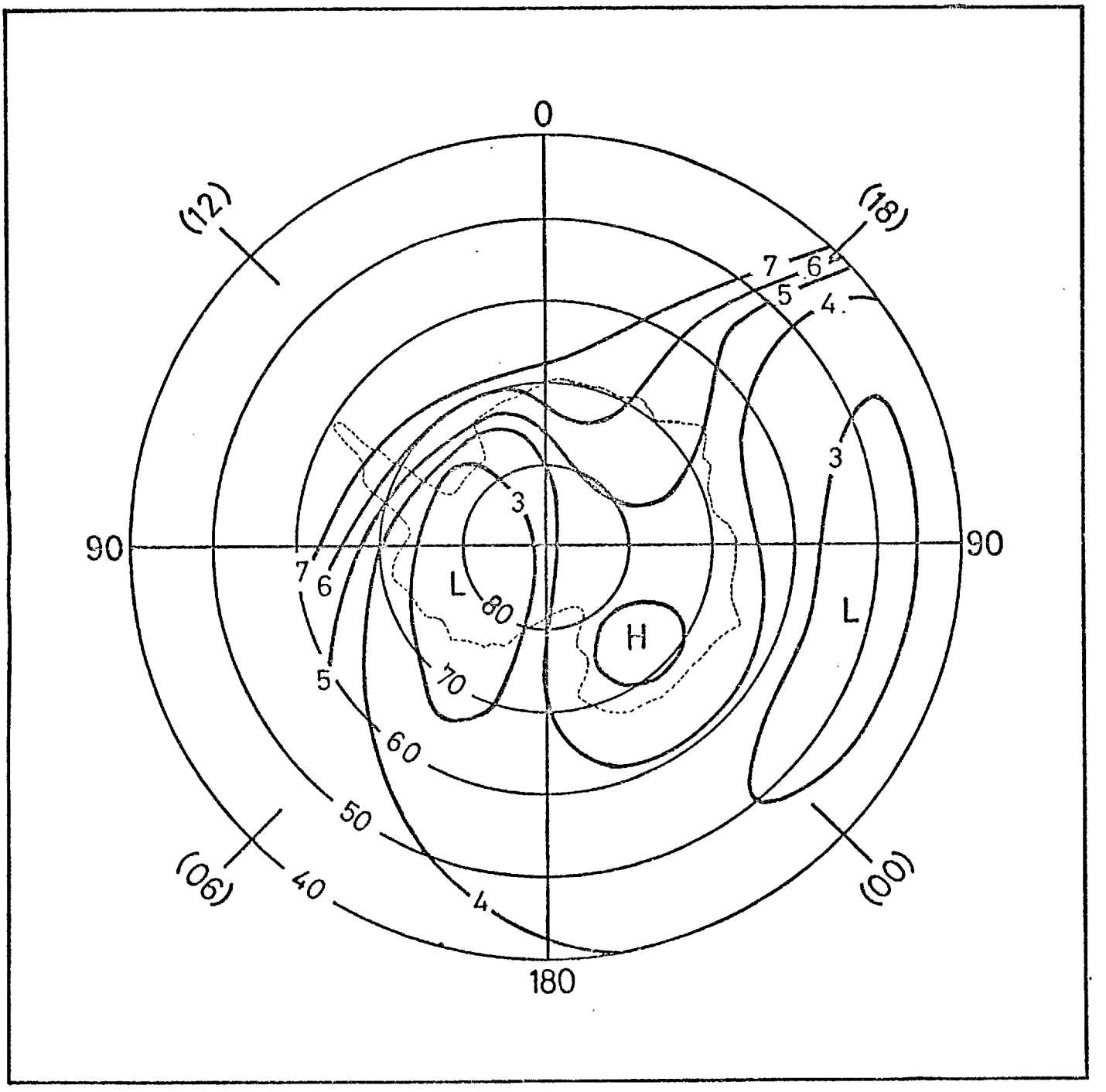
Reference to ground-based methods generally means investigation of the ionosphere by a vertical incidence, variable-frequency, pulsed radio sounder (ionosonde). The application of this technique suffers the inherent disadvantage of wide spatial separation of stations, and this is often compounded by the practice of presenting single station data in the form of hourly or monthly averages (Bellchambers et al., 1962). Hence, temporal and spatial resolution of ionospheric features are both quite poor (no better than 15 minutes and a few thousand kilometres respectively), and it is now known that relatively closely-spaced observations are necessary to study the basic polar ionospheric structure ($\sim\frac{1}{2}°$ of latitude, Buchau et al., 1970).

However, a large effort was made during International Geophysical Year, 1957-58, (IGY), when, especially in the Antarctic, many additional stations were established. Reviews of IGY observations in the Antarctic have been given for both quiet (Penndorf, 1965) and disturbed (Patton et al., 1965) conditions. In general, polar data revealed an ionosphere which was much more irregular than that found at mid-latitudes. When

data from many Antarctic stations were collected and instantaneous contours of f_oF2 (the F2-layer ordinary ray penetration frequency) constructed, the observed pattern, of which a typical example is shown in Figure 1.1, was interpreted in terms of a 'tongue' of higher density plasma extending over the polar cap from the sun-lit side to the dark hemisphere (Sato and Rourke, 1964). Similar investigations in the north polar region showed analogous behaviour (Sato, 1959; Hill, 1963). The observed electron density exceeded that predicted by normal photoionisation processes and a horizontal drift of ionisation, from mid- to high-latitudes, was invoked by Sato and Rourke to explain this excess. The drift theories were based on the $\underline{E} \wedge \underline{B}$ interaction, where $\underline{E}$ and $\underline{B}$ are the ionospheric electric and magnetic fields, and predicted drift speeds of $\sim$100 m s^{-1}.

Many workers have associated the high electron densities observed at polar latitudes with a corpuscular source of ionisation, similar to that carrying energy to the aurora, and have noted the connection between optical phenomena and electron density enhancements. In addition corpuscular ionisation is widely held to be responsible for the unusual diurnal variation, at a single station, (i.e. universal time control) of many ionospheric parameters, such as f_oF2 and h_mF2, the F2-layer ordinary ray critical frequency and the height of the F2 peak electron density respectively. This effect has been noted in the Arctic by Dieminger (1959) and at the South Pole station by Knecht (1959). Dieminger noted a longitude dependence in the time of occurrence of maximum value of critical frequency but did not point out that this occurred everywhere at 2000 UT. Knecht (1959) performed a harmonic analysis of winter IGY data and found that, although variations in solar zenith angle were negligible, a peak occurred in the diurnal variation of the maximum of the F2-layer electron density, N_mF2, at 0700 UT. Duncan (1962) has examined data from many stations and has concluded that diurnal variations of N_mF2 were such that maxima occurred at approximately 0600 UT in the south and at approximately 1900 UT in the

Figure 1.1 (After Sato and Rourke, 1964). A synoptic map of Antarctic monthly median critical frequency, f_0F2 (in MHz), at constant universal time (1500 UT), on a grid of geographic latitude and longitude. Also shown, in parentheses, are local times. The letters L and H refer to areas of low and high values of f_0F2 respectively. The map, for winter 1957, clearly illustrates the 'tongue' structure referred to in the text.



north polar region. Recently, much attention has been paid to these effects. Oguti and Marubashi (1966) examined f-plots from many Arctic and Antarctic stations and noted, for stations in the corrected geomagnetic latitude range 74-78° only, a characteristic enhancement in ordinary ray F-layer penetration frequency. This feature occurred at a time when the eccentric magnetic pole was towards the sun, i.e. at geomagnetic noon.

Hence, despite their disadvantages, ground-station observations using ionosondes revealed a number of interesting and important facts concerning the polar F-region. They revealed the unique universal time control of the polar F-layer but above all the variability and complex nature of the data showed that techniques of much greater resolution, such as those described below, were required before the polar F-region structure could be specified.

1.1.2 Early topside-sounder observations

The launch of the first topside-sounder satellite, Alouette I, in September 1962, gave a great stimulus to research on the polar F-region. As soundings of the electron density, from the satellite orbit to F-region peak, were recorded every 18 s, spatial resolution was greatly increased and recordings of electron density approximately every 100 kms (equivalent to $\sim$1° of latitude), providing almost continuous coverage of the electron density distribution, became possible.

First results at a minimum epoch of the solar sunspot cycle showed that the ionosphere at high altitudes and latitudes exhibited large horizontal density gradients. A marked depression of density was observed at the satellite during local night at about 60° latitude, which disappeared towards noon (Thomas and Sader, 1964). This feature subsequently was observed at the F-region peak by Muldrew (1965), who examined plots of the extraordinary ray F-region penetration frequency (f_xF2), in the total electron content measurements from the S66 beacon

satellite e.g. Liszka (1967) and in the bottomside of the F-layer by Sharp (1966) using a satellite-borne ion trap. This is the so-called trough.

Poleward of this depression, Thomas et al. (1966) often found a sharp increase of electron density, sometimes to values greater than those found at mid-latitudes. This feature they termed the 'auroral cliff' on the nightside and the 'polar peak' on the dayside. Typically a doubling of density within a 2° change of latitude was often observed, with just as sudden a decrease on the poleward side. Also a region of low electron density (sometimes $\sim 10^3$ cm^{-3}) appeared eccentrically centred around the geomagnetic pole. This regular feature of the polar F-region is widely known as the 'polar cavity'. Figure 1.2 shows a typical pass of Atmospheric Explorer-C over the north polar region giving examples of the features noted above. Thomas et al. (1966) collected topside-sounder satellite data from the winter of 1962-3 and the summer of 1963 and made an initial examination of the above features, their movements with geomagnetic activity, local and universal time. They showed that the distribution of electron density was similar at all altitudes between the F-layer electron density maximum and the satellite altitude (1000 km), although features were more pronounced at high altitudes. The trough was observed both in summer and winter and appeared to move to lower latitudes with the approach of local dawn. In addition, especially in summer, the trough tended to fill towards noon where it proved difficult to distinguish. In all cases it was most clear in the post-midnight sector. The 'cliff' was also seen both in summer and winter and a particular survey in the winter of 1962 and the summer of 1963, for local times between 2100 and 0600, showed it to be present on 88% and 85% of all observations respectively. During the summer, observations of the 'cavity' were rare and Thomas et al. proposed that, under constant illumination from the sun, ionisation tended to fill the low electron density region. All these features were noted to align along lines of invariant latitude (McIlwain, 1960), and moved

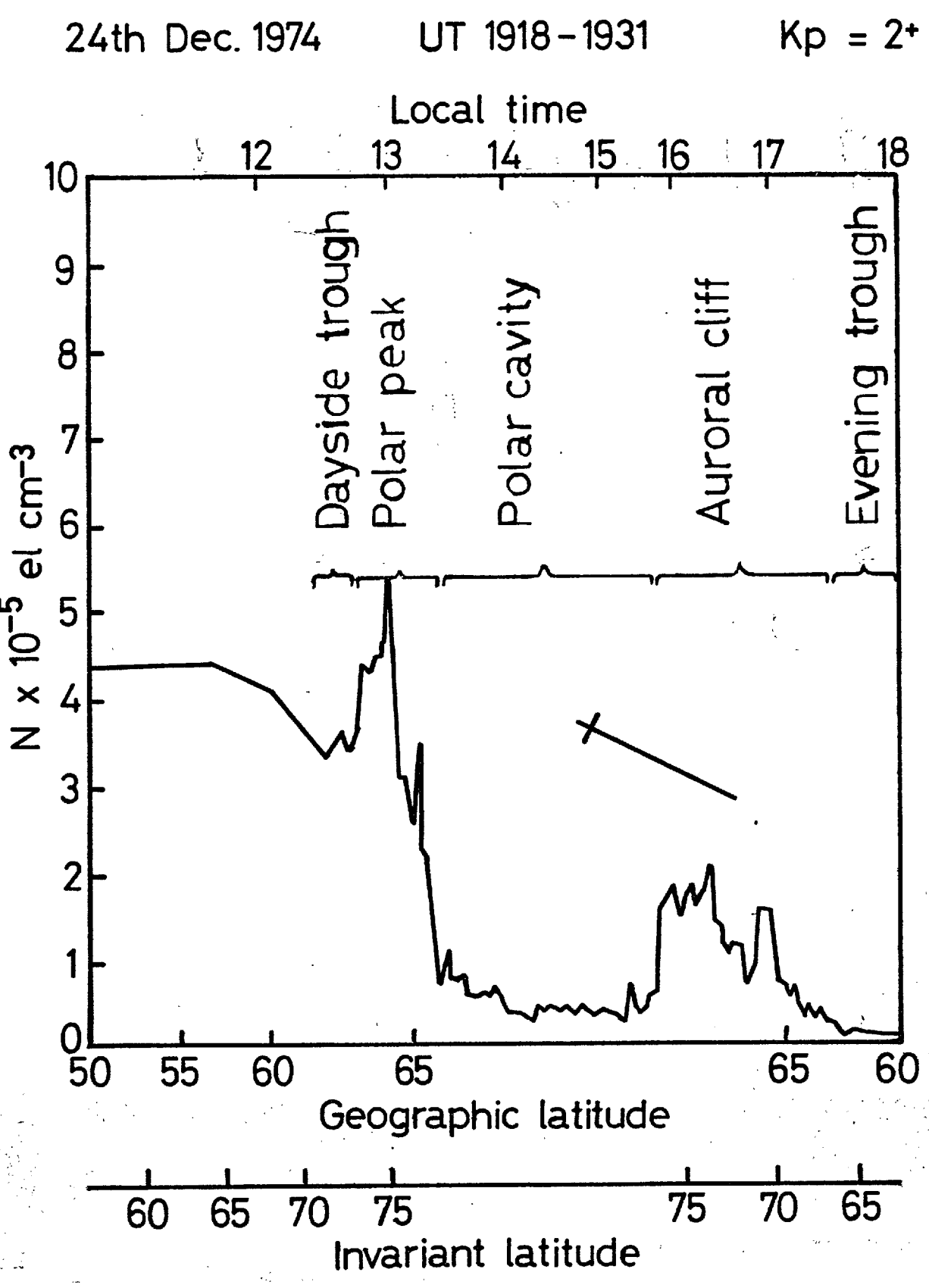


Figure 1.2 Measurements of the Atmospheric Explorer-C electron density probe, for pass 4707 of the satellite over the north polar region during winter 1974. The altitude of the satellite varied from 282 to 272 km while these data were collected. Prominent F-region features, first identified by the topside sounder, Alouette I, are labelled.

equatorward with increasing geomagnetic activity,as indicated by the Kp index. Such behaviour seemed to indicate geomagnetic or outer magnetospheric control of the polar plasma distribution. Further support is lent by the fact that peaks at 1000 km were seen to appear around geomagnetic sunrise, with an occurrence frequency maximum near geomagnetic noon and very few peaks near geomagnetic mid-night. The latter corresponded to the UT effect of Duncan (1962),showing peak densities in the north polar topside to occur about 1800 UT.

The fine structure resolved, by the Alouette I satellite, for the first time, had a spatial scale of a few hundred kilometres. This is just above the limit of the resolution of the satellite and is clearly on too small a scale to be discernible using relatively widely-spaced ground observations. However, such features may become masked, even in satellite data, if indiscriminate averaging of ionospheric parameters over time and Kp is attempted. Figure 1.3 illustrates this by showing two respresentations of the latitudinal variation of electron density at 400 km in summer. Even on the nightside the trough and cliff are not visible in the average plot due to smoothing over varying temporal and magnetic conditions. On the majority of single pass plots of latitudinal electron density variations these features are clearly seen.

The large number of variables affecting the polar plasma distribution places a constraint on the amount of data required if any attempt is to be made to isolate the active processes and their effects. Clearly large amounts of data are required and Nishida (1967) attempted to fulfill this need. From Alouette I data, using 2331 autumn passes (1962, 1963), 403 winter passes (1963, 1964) and 337 summer passes (1963, 1964) he attempted to obtain a synoptic model of the high latitude F-region for geomagnetically quiet times ($Kp \leq 3^{o}$). To do this, however, he averaged electron density data in bins (dimensions - 4° of corrected geomagnetic latitude by 2 hours of magnetic local time).

Although these data bins were as small as the data would allow, for sensible averaging and standard deviations of the sample, it was inevitable that some of the fine structure would be submerged by universal time and geomagnetic variations. Nishida presented his results in three forms:

(1) Polar plots of density contours with CGL (corrected geomagnetic latitude)/CGT coordinates for 950, 600 and 350 km altitude.

(2) CGT (corrected geomagnetic time) variations of density for lines of constant CGL (50°, 74°, 64°, 56°).

(3) Meridional profiles of density at 0800, 1200, 1800, 2400 CGT.

Of the features distinguished by Thomas et al. (1966), Nishida claimed to see the trough at 60° CGL and the polar peak. The polar peak appeared as an isolated peak at 1300 CGT and 76°N CGL but it was not observed at all altitudes, being more pronounced at high altitudes. However, no polar depletion was seen and on single passes, Nishida found the auroral cliff to be small in magnitude and occurrence frequency. Thus it appeared to have been smoothed out by the averaging process.

In order to avoid losing the basic infrastructure in the averaging process, Thomas and Andrews (1968, 1969) and Andrews and Thomas (1969), hereinafter referred to as TA, performed a large scale investigation of the north polar topside ionosphere using the electron density distribution *found along single tracks of the satellite, Alouette I*, at the F-layer peak. Again data from the summer and winter of 1962 and 1963 were used. They found that the winter density distribution along a satellite track was repeatable from orbit to orbit for periods of up to ten hours. Also it was noticed that passes fell into two clear groups illustrated by Figure 1.4. These two categories seemed to depend on the angle between the satellite orbit and the Earth-Sun line and were:

(1) Those passes showing two density enhancements separated by a region of depleted density.

(2) Those passes showing only a single extended enhancement.

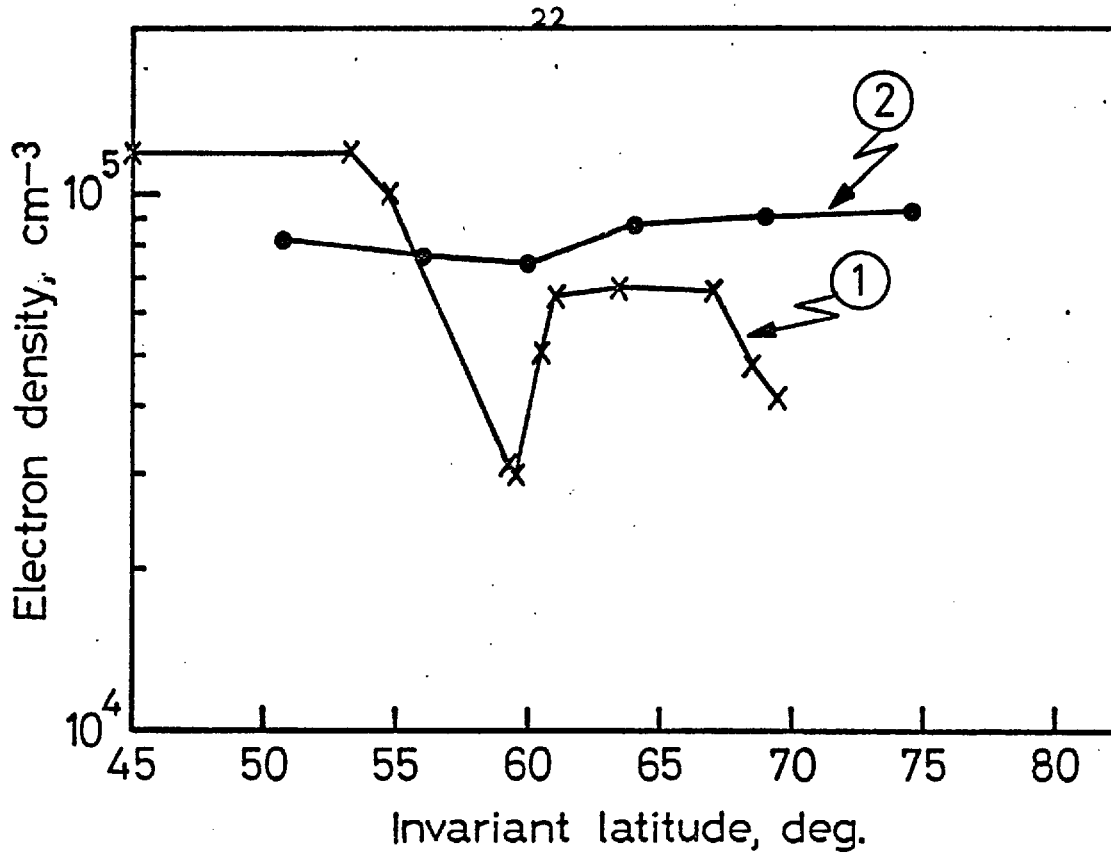


Figure 1.3 Effect of averaging Alouette I electron density data over time and magnetic activity. (1), from Thomas et al. (1966), shows the latitudinal variation of electron density at 400 km for an individual nightside pass on 7th June 1963 (Kp = 5⁻). (2), after Thomas et al. 1965, shows the average latitudinal variation of electron density at the same altitude for all 1963 summer passes in the 2100-0500 hr local time sector. The averaging process included all data, irrespective of universal time and level of magnetic activity. The trough and cliff, clearly visible on the single pass, have disappeared in the averaging procedure.

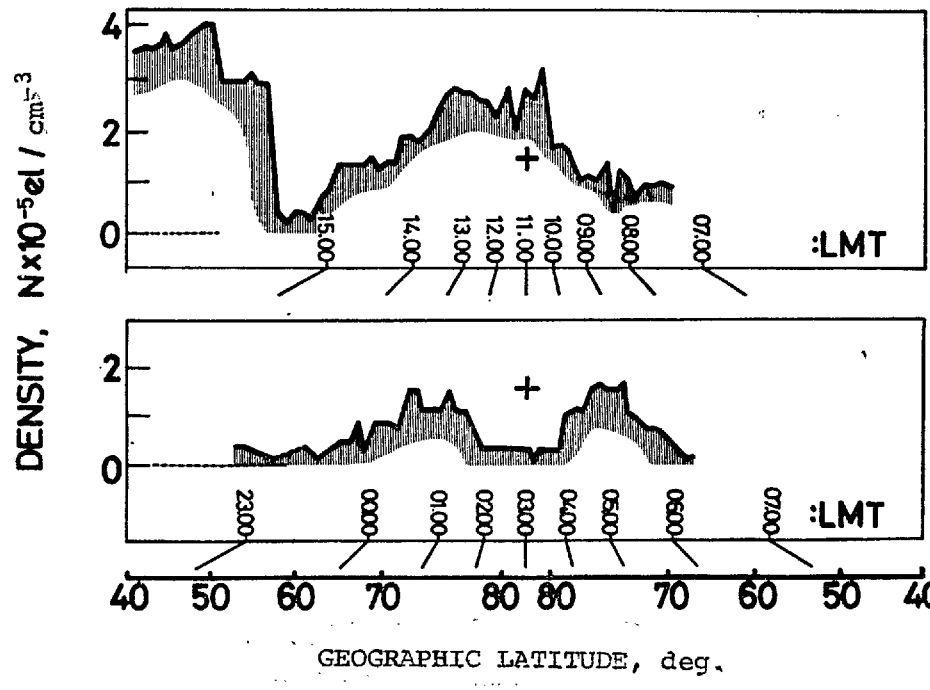


Figure 1.4 Typical examples of the two principal latitudinal electron density patterns observed by the Alouette I satellite near solar minimum. The local time of the satellite is also shown. The single- and double-peaked distributions are clearly evident.

Passes in category (1) outnumbered those in (2) by approximately 2:1.

The fact that patterns of density were repeatable and that these fell into two categories seemed to suggest a stable density distribution, independent, to a first order, of longitude (i.e. universal time), which was stationary with respect to the Sun (i.e. dependent on local time). From these considerations they therefore proposed the model for the top-side winter electron density distribution shown schematically in Figure 1.5. (In this figure the density of dots schematically represents electron density, i.e. the denser the dots, the greater the electron density.) *Thomas and Andrews visualised the plasma distribution to consist of a ring of enhanced plasma density, containing the auroral oval (Feldstein and Starkov, 1967), surrounding a polar cavity of depleted density. Equatorward of this plasma 'ring', the trough is situated.* In the morning hours, at trough latitudes, the sun begins to influence the density distribution and the trough begins to fill with ionisation, probably solar-produced, until around noon it becomes entirely filled. Hence the trough appears rather as a crescent-shaped area of depleted density around the nightside of the earth. This whole structure is eccentrically centred on the geomagnetic pole and the Earth rotates beneath it.

Also shown in Figure 1.5 are two typical satellite tracks which fall into the categories (1) and (2) noted above. The transition from category (1) to (2) can arise through slight movements of the structure, through changes in Kp and universal time. These can cause the satellite to sample densities in the polar cap or solely within the density enhancement region.

From their data Thomas and Andrews investigated many variations in this structure. From 276 observations of the low-latitude wall and 183 observations of the high-latitude wall, they fixed statistical boundaries of the enhancement region and investigated the equatorward movements of the low latitude wall with geomagnetic activity, as represented by the Kp

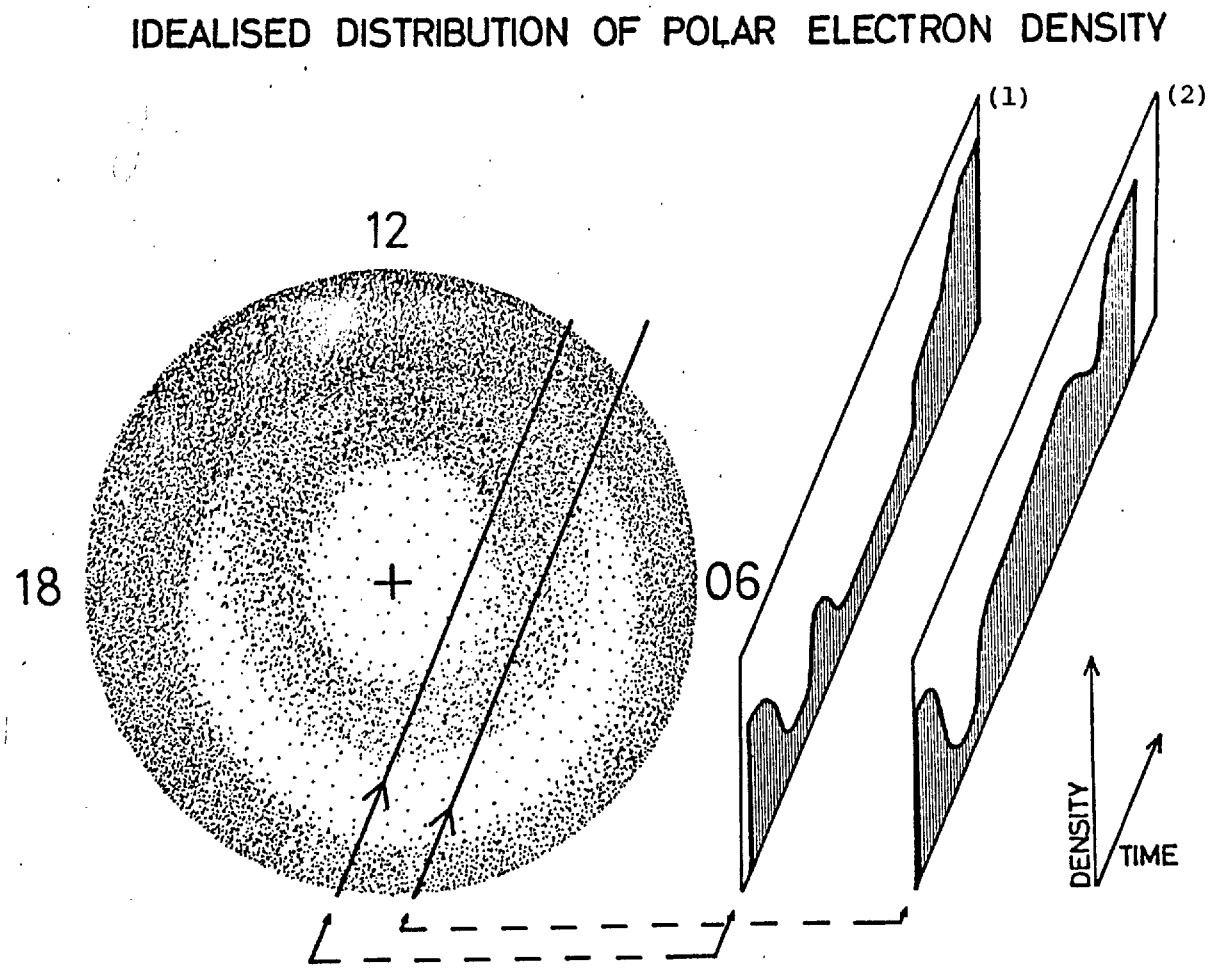


Figure 1.5 (After Thomas and Andrews, 1969) A schematic view of the model electron density distribution proposed by Thomas and Andrews. The grid is in geographic latitude and local time and the darker shading represents greater electron density. The two latitudinal electron density profiles most often seen by the Alouette I satellite are shown and the schematic way in which they can arise. If the whole distribution moved slightly with respect to geographic pole, as would occur if the structure followed geomagnetic control, then the satellite would sample densities solely in the plasma ring and not in the cavity.

index at the time of observation and that observed up to 30 hours before and 24 hours after. No time-lagged dependence on Kp of any significance was discovered. Summer investigations showed very little structure and there was little evidence of the presence of peaks and troughs. The influence of photoionisation was dominant at this time although the polar domain of the ionosphere was still recognisable from that found at mid-latitudes by the disturbed nature of the ionograms recorded there. TA also found that their ideas could be reconciled with earlier ground-based work. The so-called tongue of ionisation (Sato and Rourke, 1964), which bore a distinct resemblance to the structure observed by Nishida, is consistent with large-scale spatial averaging of the ring structure. The polar cavity merely disappears during this process. Diurnal variations identical to ground-based results were discovered for response to geomagnetic control i.e. a peak in enhancement occurrence was again found near 1800 UT - geomagnetic noon.

Other topside-sounding satellite investigations found similar results. Sato and Colin (1969) used electron density data, from the Alouette I satellite, at an altitude of 1000 km. They examined 1200 passes at various times in 1962, 1963 and 1964 and plotted these on an invariant latitude-local magnetic time grid. These showed that the polar region could be divided into three regions where electron density peaks occurred depending on time, season and the level of magnetic activity. The first region was between 75° and 80° Λ (invariant latitude) and from 0300 - 1900 magnetic local time and corresponded most closely to the ring position of TA. Here, regardless of season and magnetic activity, electron density peaks appeared and satellite passes showed two enhancements when they crossed this latitude belt at either side of the pole. Winter disturbed conditions produced some electron density peaks in two more regions found, on the nightside only, at invariant latitudes from 60-70° and around dawn and dusk at invariant latitudes >80°. In summer

there was very different behaviour and electron density peaks (of magnitude approximately twice winter peak densities) were found in all the domains outlined above, irrespective of the level of magnetic activity. Some shift of the peaks was noted with change of season and they were seen more often during magnetic day. Hence the peaks in Sato and Colin's region from 75-80° Λ appeared to coincide closely with TA's ring structure, but additional enhancements of electron density were noted, especially during magnetically disturbed conditions.

Hagg (1967) has used resonance phenomena on Alouette II ionograms to detect a high altitude (>2000 km) depletion region of electron density, ($\lesssim 10^2$ cm^{-3}) poleward of the expected equatorward wall of the trough. Significantly, there was no noticeable recovery or enhancement of electron density poleward of the trough wall, indicating that the polar ring structure did not extend to those altitudes.

1.1.3 Recent observations

Following TA's work many techniques were used to further probe the topside ionospheric structure. More topside-sounding satellites were launched (viz Isis I, Isis II) and more sophisticated ground-based methods, such as incoherent scatter, have been developed. Many of these observations have been obtained during 1968-70 when the solar cycle was close to its 'peak'.

Ground observations TA's findings led Bowman (1969), Pike (1970,1971b) and Deshpande (1973) to re-examine ground station records for the epoch 1958-1964. Bowman used ionograms obtained at three Antarctic stations to measure velocity and the spatial dimensions of the trough. These agreed very well with topside-sounder observations and were consistent with the Earth rotating under a stationary plasma distribution. Pike (1970) used IGY Antarctic F-layer ordinary ray penetration frequencies for June 1958 and largely confirmed Duncan's (1962) results.

At the four UTs;00, 06, 12 and 18 hours Pike produced individual polar plots of the peak electron density of the F-layer and f_oF2 on a grid of corrected geomagnetic latitude and local time. For Antarctic data he examined only the 06 UT plots showed any order and this indicated a ring and cavity structure similar to that of TA. Pike (1971b) has found similar behaviour in the Arctic during December 1958. Here, as expected, only the 18 UT plot ordered the data well and only when a polar plot of f_oF2 was made for 18 UT a distribution appeared which bore a close resemblance to both the tongue and ring-cavity structure of Sato and Rourke (1964) and TA respectively. A polar plot of f_oF2 for the Arctic, December 1958, is shown in Figure 1.6 with the features mentioned marked. Thus ground-based observations at solar maximum showed many features common to the TA structure of Figure 1.5. All these workers attributed the magnetic dayside enhancement and the anomalous diurnal variation to the entry of solar wind particles down the cleft region of the magnetosphere. Entry would appear to be easiest when the angle between the geomagnetic axis and Earth-Sun line is a minimum, at geomagnetic noon. Deshpande (1973) examined this non-local noon enhancement in median f_oF2 obtained from several stations and attempted to follow it, and thus the cleft region, through season and solar cycle, from 1958 to 1964. He reported that for this period the enhancement showed the following properties:

(1) It decreased with decreasing solar cycle.

(2) It shifted poleward with decreasing solar cycle.

(3) It shifted to lower latitudes during winter.

(4) It was of greatest magnitude during equinox.

A flying laboratory A novel investigative method utilised an airborne observatory (Buchau et al., 1970). An aircraft, equipped with a variety of optical and radio sounding probes, recently carried out flights into both the day and night sectors of the auroral ionosphere.

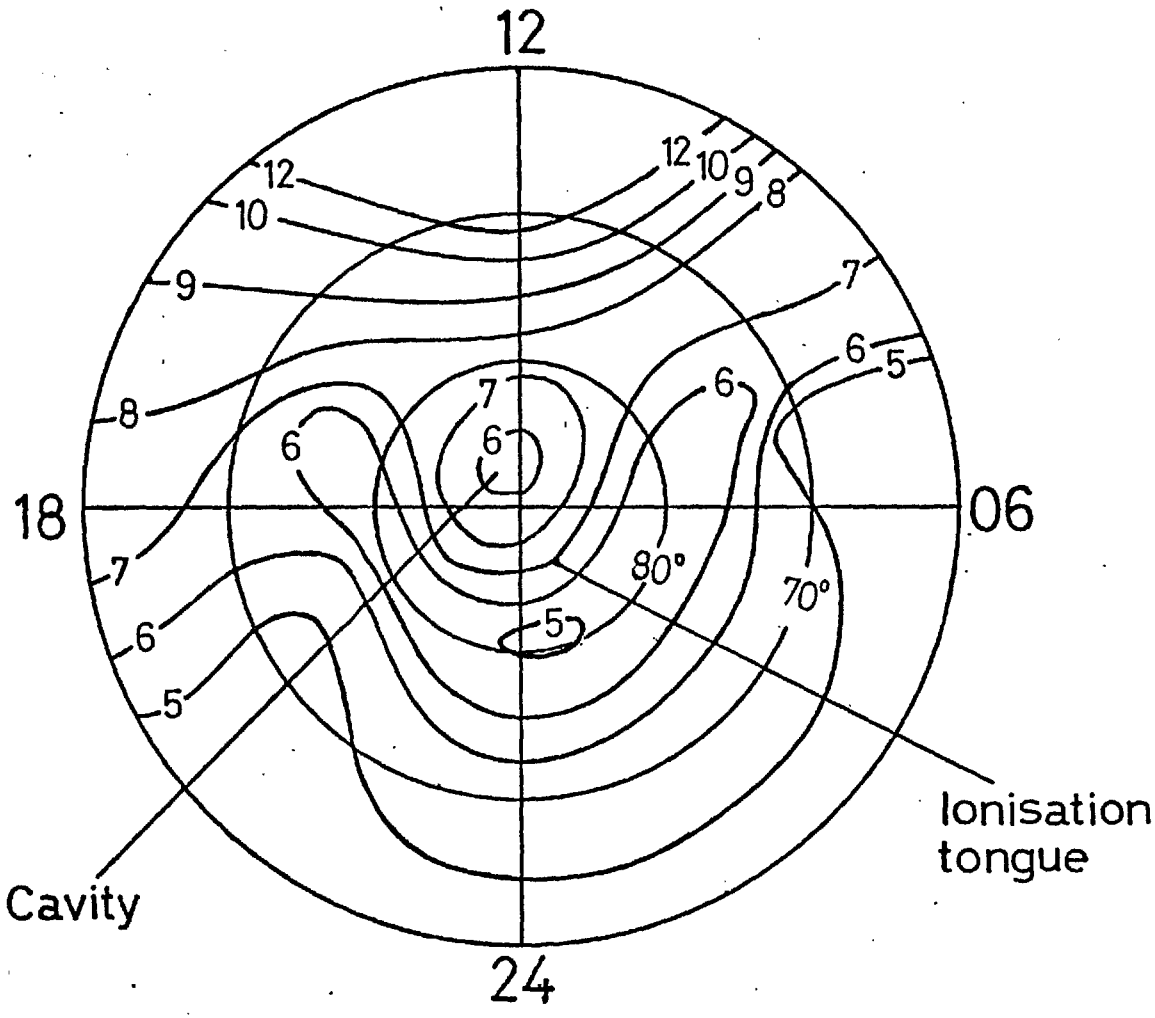


Figure 1.6 (After Pike, 1971) Features of the F-region peak plasma distribution displayed on a polar plot of corrected geomagnetic latitude and time (CGL/CGT). The contours and figures refer to values of median f_oF2 obtained in the Arctic in December 1958. A tongue of enhanced density stretches from the dayside across the pole to the nightside. The possibility of a trough-ring-cavity structure is suggested by the low values of critical frequency around the geomagnetic pole.

Flights were planned with reference to the auroral oval and it proved possible to either contract or expand the time of observation of auroral phenomena by flying with the motion of the oval or against it. (In the Earth's frame of reference.) Both local time and meridional scans of the auroral region (70-85° CGL) were completed during 1967-1970. Unfortunately, at this time, only results for the dayside (i.e. the 07-21 hr. corrected geomagnetic time sector) of the Earth have been presented in any great detail. Thus Pike (1971a, 1972a, 1972b), reporting results from the ionospheric sounder aboard the aircraft, described a zone, approximately 2 - 4° wide, in the bottomside F-region, of diffuse and irregular nature, where the stratification into F_1 and F_2 layers did not exist. This region, termed the F-layer Irregularity Zone (FLIZ), is a regular feature of both the summer and winter ionosphere. Levels of peak density within the zone are enhanced in winter only to $\sim 2 - 4$ 10^5 cm^{-3}, at a lower base height of 200 - 300 km altitude. In summer there is no density enhancement but the FLIZ merely appears as an irregular layer. Poleward of this zone Pike often observed a region of depleted density in the bottomside. In winter this, however, did not appear to be a regular structure and was not seen at all in summer. From 18 crossings of this zone, Pike (1972b) found equatorward motion of the FLIZ and an apparent increase in width of the zone, with increasing geomagnetic activity. From coincident Alouette I electron density data he was able to relate this bottomside feature very closely with the dayside sector of the topside plasma ring. Despite only occasional observations of the cavity it appears that the topside structure, proposed by TA, maps into the bottomside without any great modification even at solar maximum, the solar epoch of Pike's observations.

Wagner et al. (1973) have used combined ground and aircraft-borne ionosondes to obtain bottomside winter ionograms in all local time sectors. Unlike early experimenters they were able to use closely-spaced ground

stations. For very quiet equinox and winter conditions they presented their observations as contours of electron density on a CGT/CGL grid. Only data at constant longitude were analysed and the contours obtained clearly showed a distribution which was stationary with respect to the sun. The trough was pronounced during the morning hours and was positioned at ∿60-65° CGL. Poleward of this, a wall of ionisation appeared at 72°. The density enhancement, in this investigation, appeared as a plateau of ionisation extending equatorward from the auroral oval. Only on two flights out of 4 which passed over 80°N was the polar cap clearly visible. However, on some occasions the ring structure is found at latitudes >80° and this may have coincided with the flights mentioned above, especially as these were at very quiet periods.

In-situ data One other method often used for the investigation of the ionospheric plasma is the Langmuir Probe. This samples the electron density *in situ* i.e. in the vicinity of the satellite. Examples have been carried on such satellites as Ariel 3, Explorer 22 and Isis I. Miller (1970, 1974) used Explorer 22 and Isis I instruments to provide data on the trough at both high and low altitudes. Two new facts which arose out of this work were that with increasing sunspot number, R, the density in the trough was seen to increase. Also he found that the trough in electron density, at high altitudes (∿2500km), was easily seen even on the dayside where, at low altitudes, it was difficult to resolve.

Ariel 3, however, provided some very interesting data concerning the overall distribution of electron density in the polar topside. Launched in 1967, into an almost circular orbit, above the F-layer peak at 5-600 km, in situ electron densities were obtained every 1.7° of latitude (210 km) during the summer of 1967 and the winter of 1968. Hopkins (1974) analysed these data and initially carried out a 'Nishida-type' averaging using a grid for the polar regions above 60° latitude

consisting of 126 boxes of varying dimensions. It was hoped that such a scaling would be sufficiently fine to reveal any polar ionospheric structure (Hopkins, private communication). Hopkins (1974), for summer and winter, presented his observations in the form of constant UT plots of electron density, i.e. for each of four UT regions a polar plot of the electron density variation with local time and latitude was shown. Very few features were evident in this analysis so Hopkins (1974) re-examined his data using an individual pass approach. The individual passes indicated a ring-cavity distribution even at solar maximum, thus confirming, through independent techniques, the Alouette I observations in the topside region of the ionosphere. Hopkins (ibid) discovered that the distribution described by Figure 1.5 is regularly present, in winter, in both Arctic and Antarctic ionospheres. However, it appears to be much more varied and less pronounced in the southern hemisphere, where the cavity is very small ($\sim$ a few degrees). In the summer, at the maximum epoch of the solar cycle, both Arctic and Antarctic ionospheres were very featureless. The polar cavity seemed to be filled with ionisation, in agreement with solar minimum results of Thomas et al., (1966) and the trough was rarely visible.

Very recent experiments With satellites reliance must be made on the precession of their orbit for diurnal coverage of densities. For the topside sounders this period is approximately three months and thus any data giving 24-hour local time coverage will have an inbuilt seasonal variation. To overcome this difficulty Whitteker et al. (1976) have selected a time when the three topside-sounding satellites Alouette II, Isis I and Isis II, were passing over the north polar region almost simultaneously. Since their orbits were dissimilar, the local time coverage afforded by each satellite was very different. Where there were gaps in the local time coverage, conveniently placed ground stations were used to supplement the data. The time chosen was an April day, 1971,

at ∿0440 UT. With measurements from the three satellites and from ground station data, an attempt was made to construct almost instantaneous contours of topside electron density over the entire polar region, for both F-region peak altitude and for 1400 km altitude. Unfortunately this coincidence occurred at equinox during local morning and solar ionisation would play an important part, most of the polar region being sunlit. These contours display a tongue of enhanced plasma extending, across the polar cap, from the day to the nightside, similar to that observed by Sato and Rourke (1964) and Pike (1971). Accompanying this enhancement there is an adjacent low density plasma region. This could be interpreted in terms of the well-known trough region. It would appear that the observations of Whitteker et al. (1976), on that occasion, showed no polar cavity but it is difficult to assess whether this was due to photoionisation, convection of plasma into the polar cap or merely large spacing of observations obscurring the cavity feature. In a further examination of 32 individual Isis I and Isis II winter passes through the cleft region of the magnetosphere Whitteker (1976) observed a clear isolated enhancement on 22 occasions, in the same latitude region as the topside polar ring and the bottomside FLIZ. The enhancement of density increased with altitude and fell into a broad region 3 - 4° wide and slightly poleward of the cusp/cleft region. The polar cap depletion was also evident but during a magnetic substorm was seen to fill with ionisation. The latter result could explain the intermittent visibility of the trough on many occasions. If observation is made close to a period when Kp was high then the polar depletion may still be recovering from the influx of ionisation.

In addition to these observations, ground-based techniques are again coming more into use, especially the incoherent scatter method. For example, Taylor (1973) has seen the trough pass overhead at Malvern (51°N) during a geomagnetic storm, and investigated the characteristics of the

wall of the electron density enhancement. He found it to be closely field-aligned. Only one such facility exists at high latitudes; that at Chatinika, but this has clearly demonstrated the usefulness of the incoherent scatter technique in the polar region. The advantage of this method lies in their ability to provide information on a large number of parameters over a wide range of altitudes.

1.1.4 Ionospheric temperature and composition

Of available ground-based experiments only incoherent scatter can provide reliable data on ionospheric temperature and composition. Thus, until recently, in-situ satellite- and rocket-borne ion mass spectrometers and Langmuir probes have been the principal source of such measurements. These reveal similar behaviour to electron density in that,in the polar regions,ionospheric temperature and ion densities exihibit a high degree of irregularity and large latitudinal gradients.

For all latitudes both electron (T_e) and ion (T_i) temperatures increase with altitude although observed gradients are small, amounting to $\sim$ a few °K km^{-1} at high latitudes. At mid-latitudes near the height of the F-layer peak typical measured values of the T_e and T_i are $\sim$1500 and 1000°K respectively. Here diurnal variations are large with changes of several hundreds of degrees throughout the local day. Furthermore electron and ion temperatures increase with increasing latitude and maximise in the 60-70° geomagnetic latitude region,roughly coincident with the auroral region. These temperature enhancements are strongly affected by geomagnetic activity and are thought to be caused by electric field joule heating. Values of T_e and T_i as high as 2500-3500 and $\sim$1500°K are not uncommon. Temperatures around the geomagnetic pole are lower than these, but still exceed mid-latitude values. (Typically $T_e \sim$ 2000°K). At polar latitudes there is little observed variation throughout the local day. A detailed review of temperature measurements from space has been presented by Willmore (1974).

The existence of two differing ionospheric regimes, at high and low latitudes, first noted in the electron density data, is also apparent in satellite ion density measurements (Taylor and Walsh, 1972). Two distinct ion distributions are found and these occur in separate latitude regions. Mid-latitude observations indicate that at altitudes near the peak of the F-layer the principal ionic constituent is oxygen (O^+), with nitrogen (N^+), hydrogen(H^+) and helium (He^+) minor contributents to the total ion density. In accordance with the diffusive equilibrium theory of Angerami and Thomas (1964), the concentration of O^+ decreases with altitude and the lighter ions H^+ and He^+ become more important. At about 900 km altitude, H^+ and He^+ are the dominant ions, with O^+ and N^+ minor constituents.

The high-latitude distribution is somewhat different. Diffusive equilibrium seems to break down and in the topside near 900 km the concentrations of H^+ and He^+ are much less than those at mid-latitudes, while the density of the heavier ions remains largely unchanged. The result is that O^+ continues to be the dominant ion to very great heights ($\sim$1500 km).

Latitudinal variations of individual ion species densities give clear evidence for an abrupt transition between these two distributions. At night, when the sun's influence is absent, latitudinal gradients of O^+ are small. However, at $\sim$60° geomagnetic latitude, there is a sudden decrease in the abundance of the light ions H^+ and He^+. This steep gradient seems to be closely connected to the low-latitude wall of the electron density trough and the plasmapause (Taylor, 1972).

More recent data suggest that the picture may be more complex than the one presented here (Grebowsky et al., 1974), with NO^+ an important F-region ion in narrow latitude regions. However, the simplified developing picture views the polar F-region and topside as predominantly consisting of hot O^+.

1.2 The Physics of the Polar F-region

1.2.1 The continuity equation

The distribution of ionisation in the ionosphere is described by the well-known continuity equation (1.1). This is outlined below with, in Table 1.1, processes and parameters which are thought to play some part in determining polar F-region structure.

$$\frac{\partial N}{\partial t} = Q - L - \nabla.(N\underline{v}) \qquad 1.1$$

where $\underline{v} = \underline{v}_d + \underline{v}_e + \underline{v}_n$

and $\underline{v}_d$ = diffusion velocity

$\underline{v}_e$ = velocity due to electromagnetic drift

$\underline{v}_n$ = velocity imparted by motions of the neutral atmosphere.

Table 1.1

Term	Q	L	$\nabla.(N\underline{v})$
Process	Production	Loss	Movement
Possible cause	Solar uv ionisation. Particle precipitation	Dissociative recombination $X^+ + Y \rightarrow X + Y^+$ Recombination $X^+ + e^- \rightarrow X$	Polar wind Neutral air winds Magnetospheric convection Local heating
Influential factors	← Neutral atmosphere - composition and thermospheric structure →		
	Ionisation cross sections. Spectrum of incoming radiation or particles	Ion and electron temperatures Electric fields Rate constants	Electric fields and convection patterns. Electron and ion temperatures and gradients. Strength of heating source

1.2.2 The neutral atmosphere

As outlined in Table 1.1 parameters describing the neutral atmosphere enter all three terms of the electron continuity equation. The production of ionisation, whether by photoionisation or corpuscular bombardment, depends on the neutral constituents and their densities, which, in the polar regions, are not well-known. Recently, however, Reber and Hedin (1974) examined high-latitude data for N_2, O and He from the OGO 6 neutral mass spectrometer and deduced marked differences from the earlier mid-latitude models of Jacchia (1965, 1971). The deviations took the form of enhancements of N_2 and O, with a coincident minimum of the He concentration, situated $\sim$70° Λ at certain universal times. (In the southern hemisphere maximum (N_2) at approximately 0800 UT.)
This effect has been built into a recently formulated global model of neutral composition (Hedin et al., 1974) and is illustrated in Figure 1.7 for the northern hemisphere. The 'bulge' of enhanced densities is clear at 1930 UT and is thought to be due to a cyclic variation of thermospheric heating caused by particle precipitation. Such enhancements of neutral composition appear to be related to the so-called F-region universal time effect (maxima of f_oF2, occur at high latitude stations at $\sim$0600 UT and $\sim$1800 UT in the Antarctic and Arctic respectively (Duncan, 1962, 1969; Pike, 1970)).

One explanation of the F-layer universal time effect is based on the raising and lowering of the layer by ion drifts, induced by drag of the neutral air winds (King et al., 1971). Calculations of the effects of neutral air winds have involved global wind patterns, such as that shown in Figure 1.8, produced by the solar diurnal temperature bulge (Rishbeth, 1972). Recently, observations have shown marked deviations from this pattern, often with meridional equatorward winds $\sim$400 m s^{-1} at 200 km on the dayside (Stoffregen, 1972). These observations are consistent with an additional high-latitude heating source with an energy input comparable with solar input. This is possibly joule heating, through electric field effects,

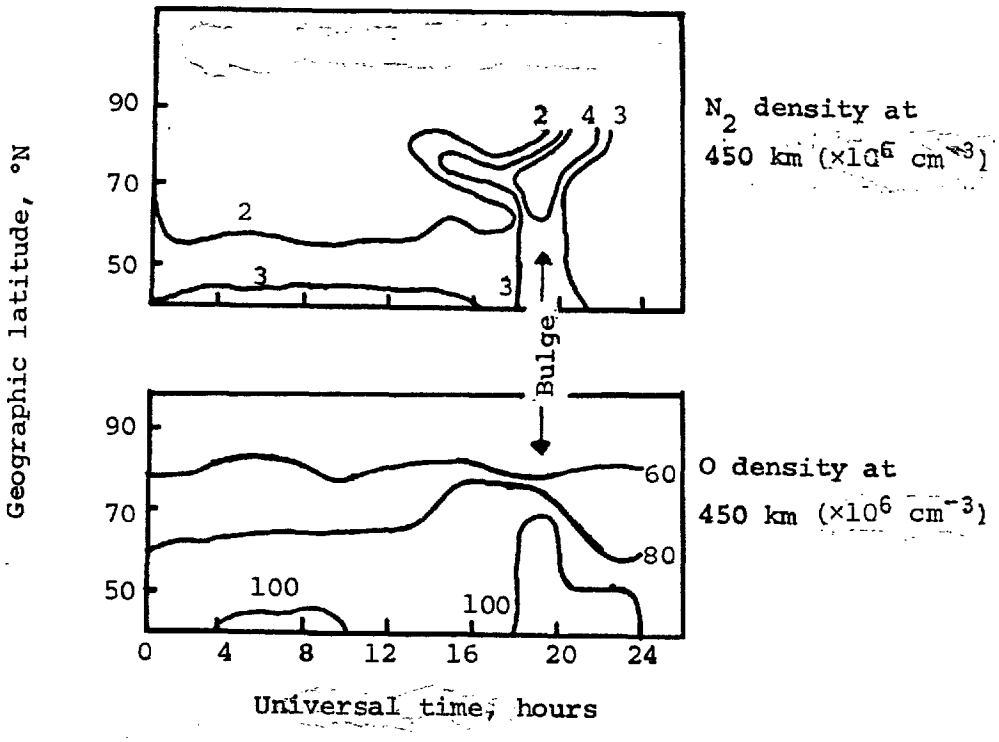


Figure 1.7 (From Reber and Hedin, 1974). Contours of constant molecular nitrogen and atomic oxygen densities measured in the northern hemisphere by the OGO 6 neutral mass spectrometer on 15-16th October 1969. The results are for a constant local time (1300 LT) and are shown as a function of universal time (UT), i.e. longitude. The UT dependent bulge of densities is evident, and for this local time appears at ~1900 UT.

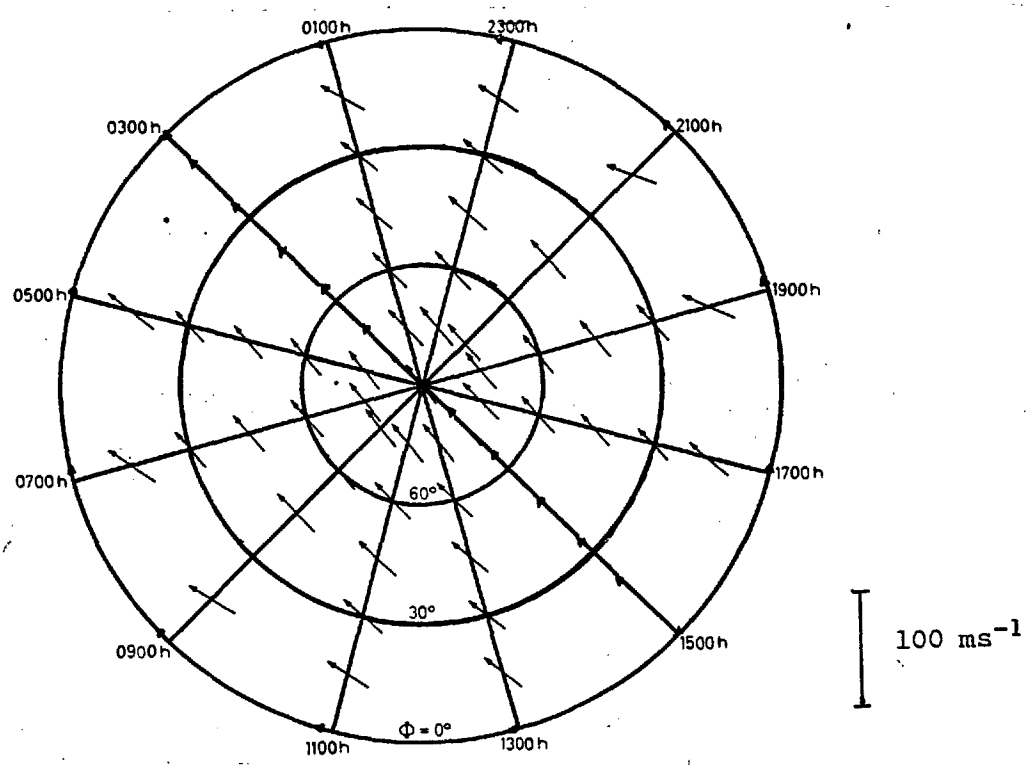


Figure 1.8 (From Rishbeth, 1972) The calculated system of neutral air winds at 300 km arising from solar thermospheric heating in the northern hemisphere. The grid is one of geographic latitude and local time and the pole is at the centre with the outer circle representing the equator. The computations are appropriate to a Chapman layer with a maximum ion density of 10^6 cm^{-3} and a scale height of 80 km.

or soft particle precipitation, which deposit ∿30% of their energy in the form of heat. This heating produces vertical expansion of the thermosphere, composition changes and vertical winds,which clearly could drag ionisation up field lines (Hays et al., 1974) The role of neutral air winds thus requires further examination at high latitudes. Here interactions with the plasma may be very different from those at midlatitudes,where winds explain many phenomena. It is also unclear whether the neutral air winds at high latitudes drive motions of the plasma or if the reverse is true and convective plasma motions pull the neutral air with them. This is a very complex problem and has been treated in a qualitative way by Rishbeth (1977) and quantitatively by Maeda (1976) for model ionisation distributions. A full treatment of the problem is awaited.

1.2.3 The polar wind

During periods of low magnetic activity, magnetic flux tubes poleward of the plasmapause participate in the general magnetospheric convection. Over a small area of the polar cap, the field tubes are open to the distant tail of the magnetosphere, while those at lower latitudes form closed loops.

The polar wind theory arose through the need to explain the depletion of the light ions, H^+ and He^+, principally on the open field lines, at high latitutdes (section 1.1.4). It proposes a field-aligned flow of these ions on open flux tubes,from the ionosphere into the magnetospheric tail. Plasma properties found in the tail region suggest that, for steady-state conditions, opposing pressure gradients are not possible, and fluxes of $\sim 2\text{-}6 \times 10^8$ ions cm^{-2} s^{-1} can exist (Banks and Holzer, 1968, 1969). Although polar wind calculations were originally restricted to open field lines, there are now strong reasons to believe that 'polar wind' type flows extend to low latitudes,where field tubes are sometimes open and sometimes closed. However, these steady-state conditions do not exist and ion flow speeds and fluxes will not attain such high

values. Closed field lines allow the build-up of opposing pressure gradients in the flux tube which can eventually cut off flow.

Recently, the calculations of Banks and Holzer (ibid) have been reviewed to include energy considerations and the effect of the flows on the thermal structure of the ion distributions (Banks et al., 1975; Raitt et al., 1975). Their conclusions are largely unaltered from those of Banks and Holzer, except that it now seems unlikely that flows attain the previously-predicted supersonic speeds. Typical ion density distributions are shown in Figure 1.9, taken from Raitt et al. (1975). The separate curves for hydrogen refer to different flow velocities at 3000 km. The shaded region indicates the range of possible O^+ densities; the lower boundary corresponding to H^+ profile (a) and the upper boundary to (h). A neutral atmosphere, taken from Banks and Kockarts (1973), with an assumed exospheric temperature of 1000°K, was used.

Examination of these profiles shows that the polar wind is largely a minority ion phenomenon. The distribution of the major ion, O^+, is largely unaltered below 1000 km. Thus it would appear that, to observe polar wind effects, the distribution of plasma above this altitude must be examined. Banks and Doupnik (1974) have followed this tenet with data from the Alouette II satellite, above 2000 km, and largely confirmed it. They detected thermal proton flows within the plasmasphere from analysis of topside-sounder scale height data. At low altitudes, where O^+ is dominant, they believed that unambiguous identification of flow effects was not possible, due to the masking effects of other processes.

However, it has been suggested (Thomas and Andrews, 1970) that small flow effects should be visible in the altitude range 400 - 700 km, within the compass of Alouette I. The source of H^+ ions, which partake in polar wind flow, is an identical flux of O^+ ions from the F-region peak, where they are produced either by solar or particle effects. At altitudes just above the F2 peak the O^+ ions are converted to H^+ ions through the reaction

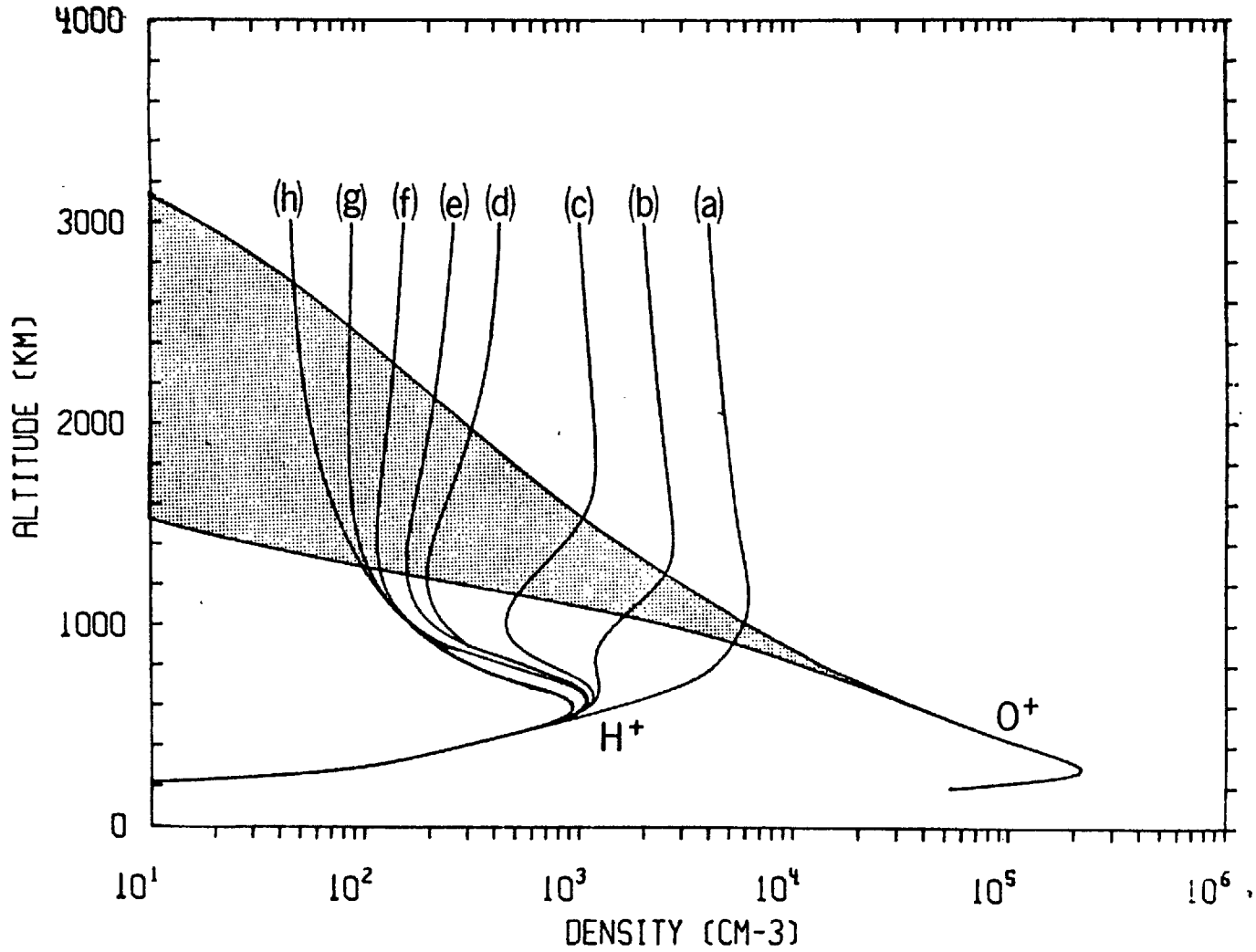


Figure 1.9 (From Raitt et al., 1975) Typical H^+ and O^+ ion density profiles resulting from polar wind type flow of ions. The curves labelled (a)-(h) refer to increasing H^+ flow velocities at 3000 km [(a) .06, (b) .34, (c) .75, (d) 2.0, (e) 3.0, (f) 5.0, (g) 10.0, (h) 20.0 km s^{-1}]. The shaded region shows the extent of O^+ densities; the lower boundary corresponding to H^+ profile (a) and the upper boundary to (h).

$$H + O^+ \underset{k_b}{\overset{k_f}{\rightleftharpoons}} H^+ + O \qquad 1.2$$

where

$$k_f = \frac{9}{8}\,\frac{N(H^+)\,N(O^+)}{N(O)} \qquad k_b = \frac{9}{8}\,\frac{N(H^+)\,N(O)}{N(H)}$$

Such upward fluxes have been measured from incoherent scatter radars at midlatitudes (Schunk and Walker, 1972). The existence of larger flows at higher latitudes may be an important perturbation in Alouette I-measured electron density profiles.

This has been investigated by examining data from many polar passes of the satellite, recorded in 1962 and 1963. These data have been published in the form of variations, along the track of the satellite, of electron densities and scale heights at fixed altitudes (Thomas et al., 1966). The scale height, H, of the plasma, a derived parameter reflecting the electron density gradient, is defined as

$$H = -N\frac{dh}{dN} \qquad 1.3$$

The flow equations of Banks et al. (1975) can be simplified considerably by making the following approximations

1. Neglect of inertial terms. At low altitudes this is justified because of the low speeds encountered.
2. Consideration of a single ionic constituent such that $N_e = N(O^+)$ in the altitude range 400 - 700 km.

Thus at high latitudes the flux of oxygen ions is derived, assuming a single ion constituent, from the equation

$$u_i = \omega_D - D_i\left(\frac{1}{N_i}\frac{dN_i}{dh} + \frac{1}{T_i}\frac{dT_i}{dh} + \frac{1}{T_i}\frac{dT_e}{dh} + \frac{T_e}{T_i N_e}\frac{dN_e}{dh} + \frac{m_i g}{kT_i}\right) \qquad 1.4$$

where the subscripts i and e refer to ion and electron parameters respec-

tively and motion along a field line. In addition

D_i is the ion diffusion coefficient = $kT_i/m_i\nu_{in}$

k is the Boltzmann constant

ω_D is the ion drift velocity due to netural air winds.

Rearranging and putting $N_i = N_e = N$, $D' = D_i(1 + \frac{T_e}{T_i})$ and $\omega_D << u_i \rightarrow 0$

$$Nu_i = -ND'\left[\frac{1}{N}\frac{dN}{dh} + \frac{1}{T_e+T_i}\frac{d}{dh}(T_e+T_i) + \frac{m_i g}{k(T_e+T_i)}\right] \qquad 1.5$$

Where production and loss are small the ion continuity equation 1.1 for steady state conditions takes the form

$$\nabla.(N\underline{u}) = 0 \qquad 1.6$$

The solutions to this equation takes two forms; $N\underline{u} = 0$ or $N\underline{u}$ = constant = F. In these two situations 1.5 can be written in the form:

$$-\frac{1}{N}\frac{dN}{dh} = \frac{1}{T_e+T_i}\frac{d}{dh}(T_e+T_i) + \frac{m_i g}{k(T_e+T_i)}; \qquad N\underline{u} = 0$$

$$-\frac{1}{N}\frac{dN}{dh} = \frac{1}{T_e+T_i}\frac{d}{dh}(T_e+T_i) + \frac{m_i g}{k(T_e+T_i)} + \frac{F}{D'N}; \qquad N\underline{u} = F \qquad 1.7$$

The quantity given by the L.H.S. of equation 1.7 is freely available from topside sounder observations as the measured scale height of the plasma (equation 1.2). Equation 1.7 shows that the effect of an upward flux of oxygen ions, supporting a polar wind type flow, is to depress the measured plasma scale height below that found for the diffusive equilibrium conditions ($N\underline{u} = 0$).

Of the many Alouette I and II passes examined in a search for these affects, the general features of their scale height data indicate:

1. A sharp decrease in scale height at 900 km and above with

increasing latitude near 60° geomagnetic latitude.

2. At 500 km altitude a steady decline of scale height with increasing latitude, often followed by a sharp increase in, and at, latitudes above the auroral zone.

An example of such behaviour is shown in Figure 1.10. The electron density data show the typical enhancement structure, with the satellite sampling densities within the polar cap at altitudes $\gtrsim$75° Λ.

Scale height data reveal a very complex structure with localised troughs and enhancements. The plasmapause shown by the transition from light to heavy ions, is clearly seen at 900 km near $\Lambda \sim 62°$. At 500 km the scale height is found to be typical of an oxygen plasma in the trough region, being ~150 km. Such values predict $(T_e+T_i) \sim 3000°K$ for diffusion equilibrium conditions. However, in the plasma enhancement region values as high as 220 km are reached, indicating the presence of very high temperatures, i.e. $(T_e+T_i) \sim 4400°K$.

Polar cap scale height values at this altitude, however, fall to 100 km or less. Here assumption that a diffusive equilibrium distribution of an oxygen plasma still exists, even with very low temperatures this requires unrealistic temperature gradients $\frac{d}{dh}(T_e+T_i) \gtrsim 14°$ km^{-1}. Measured temperature gradients have not been found to exceed a few °K km^{-1} (Willmore, 1974).

Compositional changes such as a change of dominant ion to NO^+, as proposed by Schunk et al. (1976) can resolve the problem and the calculated temperature gradients are reduced to their former realistic values. However, the extent of the depressed region of scale height, shown in Figure 1.10, $\gtrsim$8° and such composition changes have not been observed over such a wide area (Grebowsky et al., 1974).

The only adequate explanation of these low scale heights seems to be the presence of an upward flux of ions. Using equation 1.7 and with $T_e+T_i \sim 3000°K$ and $\frac{d}{dh}(T_e+T_i) \sim 3°K$ km^{-1} scale heights ~100 km predict

the existence of an upward flux of oxygen ions of $\sim 3 \times 10^8$ cm^2 s^{-1}. This is typical of the value predicted by Banks and Holzer (ibid) necessary to support light ion outflow.

At the higher latitudes shown in Figure 1.10, however, the scale height at 500 km altitude recovers to values typical of diffusive equilibrium distribution of an oxygen plasma. Thus the flow has either disappeared or is masked by some other process. The electron density gradient is a very sensitive indicator of plasma temperature and it seems likely that the increase in plasma scale height at very high latitudes is due to some local heating in source analogous to that producing large values of scale height in the electron density enhancement region. Consideration of equations 1.7 indicate, that for an oxygen plasma, an enhancment of $(T_e+T_i) \sim 1000°K$ can introduce a positive scale height increment of $\sim$50 km.

Observations of high-latitude plasma temperatures from satellites and incoherent scatter radars have now shown that such temperature enhancements are possible (Willmore (ibid)). Therefore it would seem that these dominant effects act to conceal the indications of plasma flow at these altitudes. Ion fluxes will, therefore, only be visible in Alouette I data when temperature enhancements are not present,and investigations of plasma outflows must either be carried out from high-altitude satellites,where large latitudinal temperature gradients are rare, or be undertaken with simultaneous electron density and temperature data.

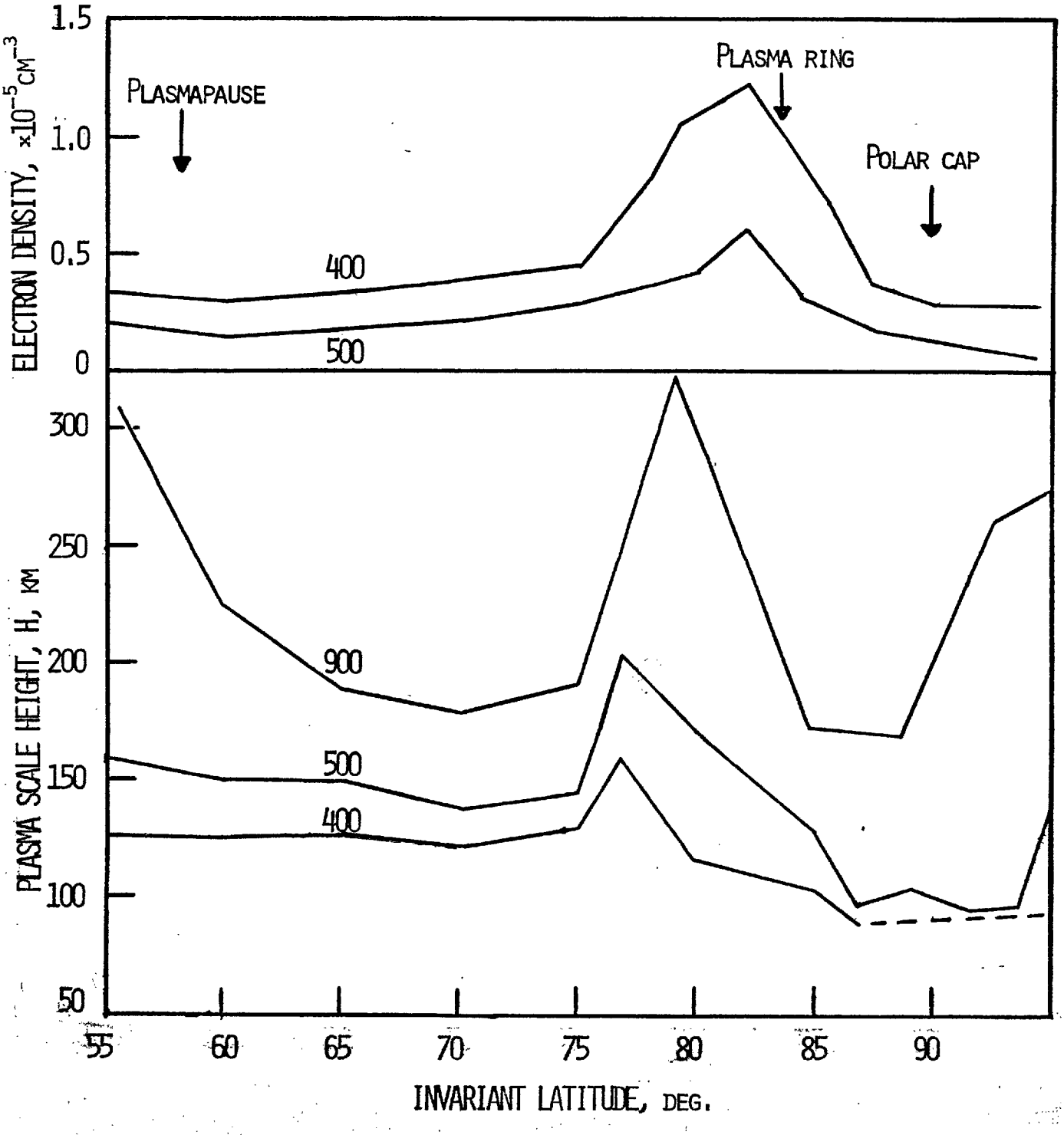


Figure 1.10 Electron density and scale height data from pass 208 of the satellite Alouette II, over the north polar region. The pass occurred during local morning with local time varying between ∿0700 and 1200 LT. The electron densities illustrate a structure typical of the trough-ring-cavity model. The scale height data in the polar cavity region show both depressed and enhanced values commensurate with field-aligned plasma flow and enhanced temperatures respectively. The figured curves refer to data at different altitudes.

1.2.4 Electric fields and magnetospheric convection

Much recent interest has been paid to the effects of magnetospheric electric fields and F-region processes associated with magnetospheric convection of plasma. These movements (Axford and Hines, 1961) are due to an interaction with the solar wind, the form of which remains unclear (Nishida and Obayashi, 1972). Associated with this convective flow is an electric field, $\underline{E}$, given by $\underline{v} \wedge \underline{B}$, where $\underline{v}$ is the convection velocity and $\underline{B}$ is the magnetic field. This field can be mapped down into the ionosphere by regarding the field lines as equipotentials. The resulting convection pattern, implied from magnetospheric models, is shown in Figure 1.11. Above ∿200 km altitude the plasma is 'frozen in' to magnetic field lines. Field lines originating in the region of the cleft on the dayside are open, convect over the polar cap to the tail of the magnetosphere, reconnect in the tail and then convect sunward as closed field lines until, again they reach the dayside magnetopause.

The pattern illustrated is necessarily idealised, and is the subject of much current investigation. Not included here is the drag of the neutral air motions on the ions (section 1.2.2). Many methods are now being used to detect polar F-region electric fields (e.g. from satellite-borne probes and barium cloud releases) and plasma drift velocities (using incoherent scatter and satellite instruments). Data at F-region altitudes have largely been obtained by satellites INJUN V, IMP 6 and OGO 6. Heppner (1972) has reviewed these measurements and Figure 1.12, taken from Heppner (ibid), illustrates typical magnitudes and variations seen in the dawn-dusk plane horizontal electric field data. The transition from sunward to antisunward flow is seen in electric field measurements as a change of sign of the horizontal component. Hence, in Figure 1.12 convection, ∿1 km s^{-1}, is antisunward from ∿73-84.3-75° ($E_x < 0$). Often very strong sunward convection and very large electric field values are observed

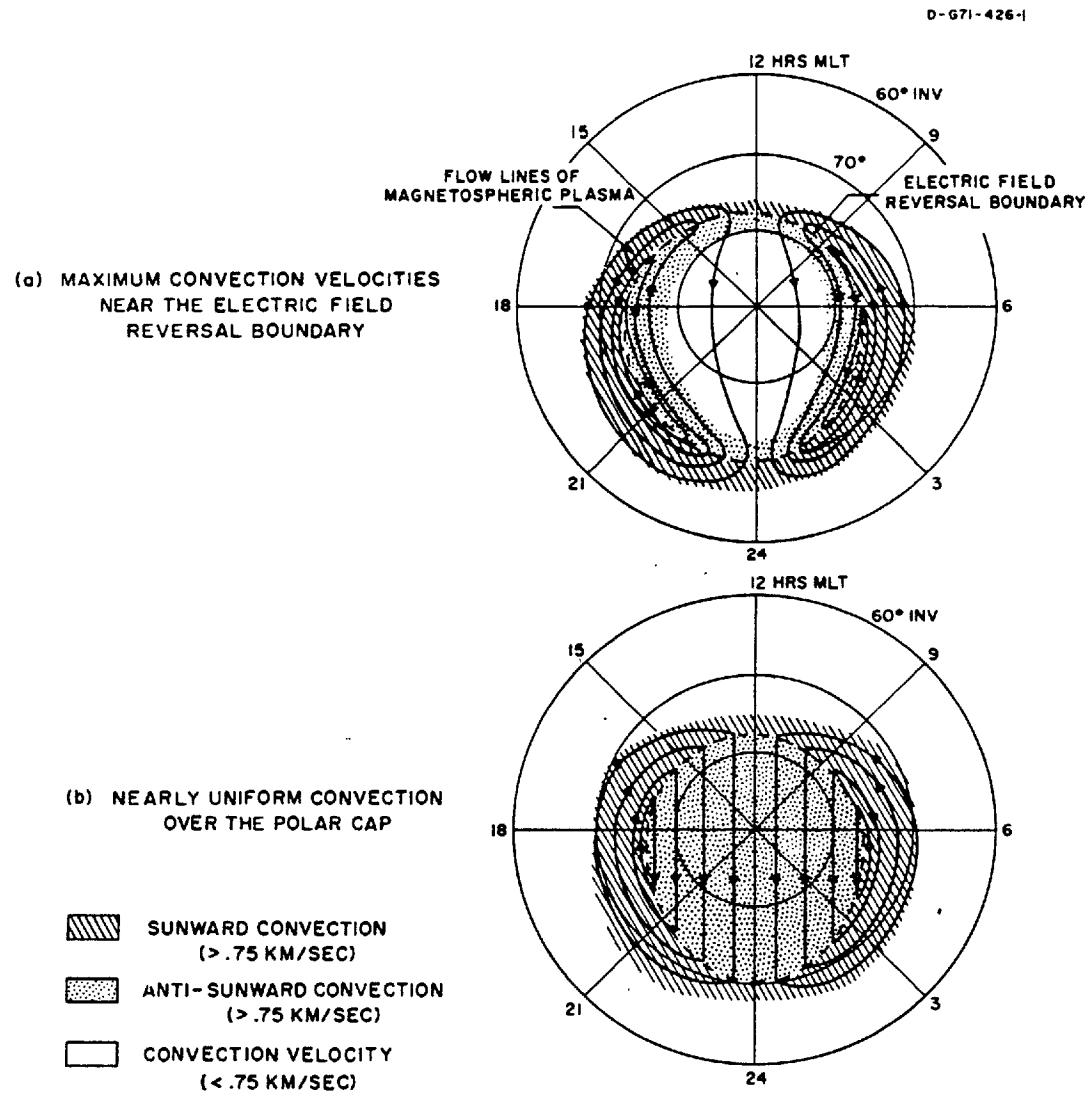


Figure 1.11 A schematic view of average F-region plasma convection at high latitudes derived from INJUN V and OGO 6 data. (From Gurnett, 1973)

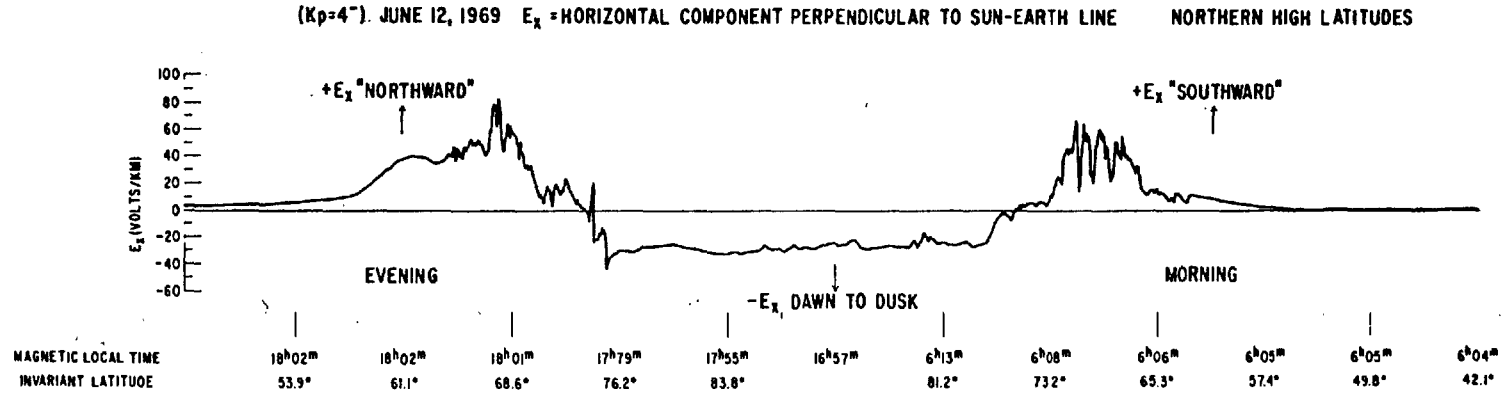


Figure 1.12 Measurements of the horizontal component of the electric field perpendicular to the Earth-Sun line (from Heppener, 1973).

in the auroral zone and in the transition region.

These large electric field values are now thought to have significant affects on ionospheric chemistry. Banks et al. (1974) have suggested that enhanced $\underline{E_\perp} \wedge \underline{B}$ drifts, due to enhanced electric field values, increases ion energy by amounts comparable with thermal energies. Hence,electric field energisation must be taken into account in termperature-dependent ionospheric reactions. Particularly, Banks et al. (1974), drew attention to the important F-region reaction;

$$O^+ + N_2 \xrightarrow{k} NO^+ + N$$

and gave a reaction rate for $T_{eff} > 750°K$ of $8.0 \times 10^{-14}(T_{eff}/300)^2 cm^3\ s^{-1}$ where T_{eff} includes electric field effects and is given by:

$$T_{eff} = T_n + .329\ E_\perp'^2\ °K \qquad 1.8$$

Here T_n is the neutral gas temperature and $E_\perp'$ is the effective electric field (in mVm^{-1}) measured in the moving frame of the neutral gas. Significant joule heating will occur if ion and neutral velocities are different. Thus any temperature-dependent reaction rate can be modified by electric fields. In addition to that mentioned above it is now known that $O_2 + O^+ \longrightarrow O_2^+ + O$ displays similar behaviour.

The effects of such electric field heating on the high latitude ionosphere, have been considered by Schunk et al. (1975, 1976). On the dayside,the coupled steady-state continuity, momentum and energy-balance equations for NO^+, O_2^+ and O^+ were solved and ion density profiles for a copious ionisation source computed. These results show that electric fields act so as to decrease the concentration of O^+ and electrons and increase those of NO^+ and O_2^+. For moderate electric field, $\sim$50 m V m^{-1}, peak densities are shifted to the F1 peak while for large fields, $\gtrsim$100 m V m^{-1}, NO^+ can become the dominant F-region ion and the F2 layer disappears.

In winter or during nightside hours Schunk et al. (ibid) have investigated ionospheric response to electric fields when no production source is active. By solving time-dependent equations, they demonstrated the possible formation of the nightside 'main' trough by a combination of zero production and decay of ionisation through enhanced recombination. They also investigated the possible involvement of the polar wind and enhanced recombination through increased N_2 densities and vibrationally excited N_2 temperatures. In combination, these destructive processes appear strong enough to account for the less-pronounced electron density troughs but are not capable of explaining the very deep troughs which are often seen. Such mechanisms have also been suggested as a possible cause of high-latitude troughs (Grebowsky et al., 1974). An example of such computations is shown in Figure 1.13 (from Schunk et al., 1975) showing expected dayside ion density profiles in the presence of an electric field of 100 m V m^{-1}.

The effect of convective motion of plasma-carrying flux tubes has been examined by Knudsen (1974) who followed the conditions in such a flux tube as it is convected through the dayside cleft region, over the polar cap, through the nightside soft precipitation zone and finally equatorward around the dawn side of the earth toward the sun.

On the basis of derived time constants Knudsen (ibid) proposes that the diffusion equilibrium state is inappropriate to the cleft F-region. Typically a field line takes time, τ_c, (Table 1.2) to cross the cleft which compares with the time constant required to establish a diffusive equilibrium layer at ~280 km of τ_d. Thus ionisation produced by soft precipitation will not have time to diffuse away from the altitude of production and a normal F-layer will not exist. However, the characteristic time associated with heating of the plasma is short (= τ_t) and there is sufficient time for heating to take effect. There is a short lag between application of the heat source and F-region response so that enhancement of T_e will occur at the poleward edge of the cleft. Thus the effect of passage, through the

Table 1.2

Time constants for processes in the polar ionosphere

Time Constant	Process	Expression	Magnitude (minutes)
τ_c	Convection time to cross drift	$\Delta L/v$	3
τ_d	Diffusion	$H^2/D\ (T_e + T_i/T_n)$	∿75 (at 280 km)
τ_t	Plasma heating	$\frac{C_e(\Delta T_e)}{K(\Delta T/\Delta L)} \sim \frac{N_{m}F_2 \times H_p^2 \times \frac{3}{2} K}{K_e.e}$	∿1
τ_i	Recombinative decay	β^{-1}	∿50
τ_p	Polar wind	$\frac{N_{m}F_2}{\phi(H^+)}$	∿420

cleft, of a flux tube, is to add ionisation at the altitude of production and lift the plasma distribution through thermal expansion. Any enhancements of ionisation at 1000 km seen in this region are the result largely of ionospheric expansion though, production is also active. At F-peak altitudes, conditions are such that only additional production mechanisms are most important. In addition, they should be shifted poleward of the precipitation region due to lag of ionospheric response. Such effects have been reported by Whitteker (1976) in coincident electron density and electron precipitation data from the ISIS I and II satellites. Knudsen has also noted that the electric field will lift the layer with a vertical drift velocity given by $(\underline{E} \wedge \underline{B}/B^2)$ cos I, or $\underline{v} \times$ cos I, here I is the magnetic dip angle. During dayside hours the drift will act so as to raise the layer. This is in the opposite sense to the ion drift due to drag of the neutral air wind. A neutral wind of magnitude, u_n, $\sim$200-300 m s^{-1} is expected at cleft latitudes blowing towards the pole and hence the resultant ion drift $\sim(v - u_n)$ cos I. For ionisation crossing the cleft the result is to raise the layer $\sim$30 km. During magnetic storms the height by which the layer is raised can be increased by a factor $\sim$3-5. Hence, for this reason, theories involving neutral air winds to explain high latitude phenomena must be re-examined before any definite predictions as to their line effects can be made. Schematically, the effect of passage of flux tubes across the polar cleft is shown in Figure 1.14. Here, the redistribution of ionisation at high altitudes and the enhanced production at F-peak altitudes is seen in electron density profiles on entry to and exit from the cleft region.

In the polar cap the electron temperature drops with a time constant τ_t and the ionisation profile 'contracts' down the field line accordingly. At a convection velocity $\sim$.5-1 km s^{-1} the flux tubes are transported from $\sim$80°Λ on the dayside, across the cap, to 75°Λ on the nightside. The time taken allows the layer to decay with a rate $\tau_i \sim \beta^{-1}$ and by a factor 1/e.

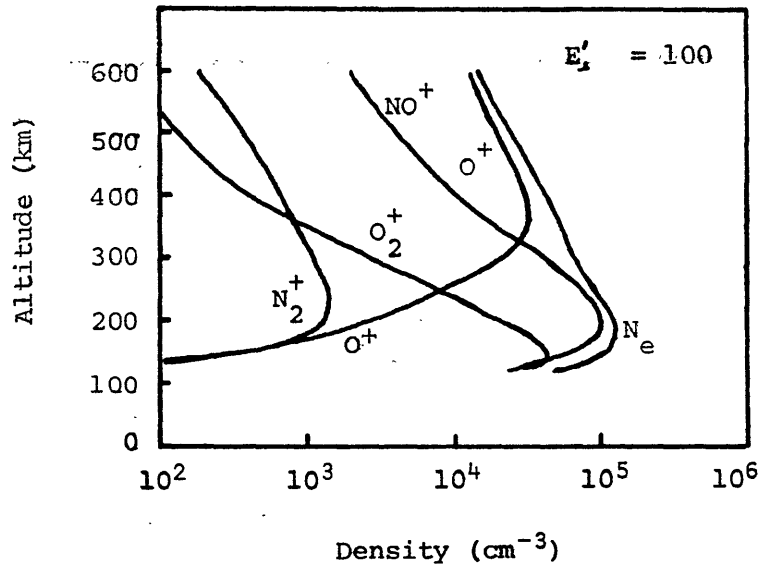


Figure 1.13 (From Schunk et al., 1975) Altitude profiles of ion and electron density calculated with an electric field dependent loss coefficient. Here a value appropriate to an ambient field of 100 mVm^{-1} is chosen and dayside production conditions are assumed.

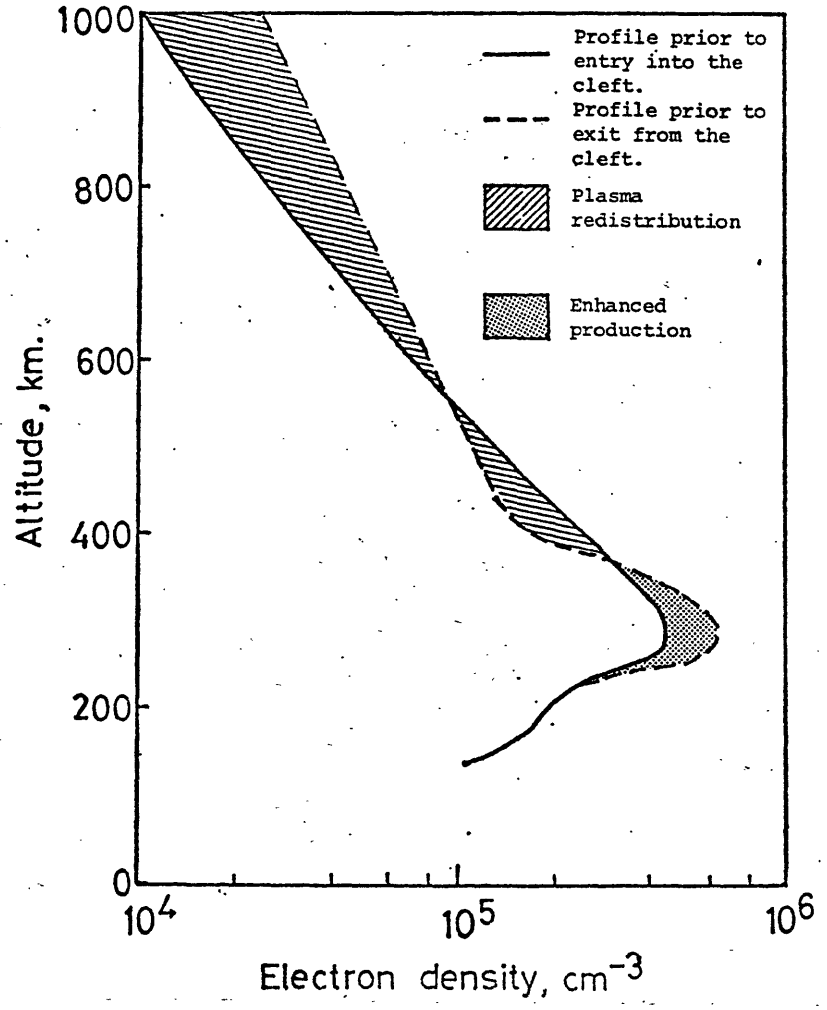


Figure 1.14 Schematic illustration of changes in the electron density profile on passage through the dayside cleft (after Knudsen, 1974).

On entry into the nightside precipitation zone the layer will experience similar behaviour to that in the dayside cleft region. However, due to the harder and more varied precipitating electron spectra (Appendix) and more complex convection patterns (Heppner, 1972) behaviour of the plasma is more complicated and less responsive to change. The main difference between dayside and nightside is that on the nightside the layer is lowered by the effect of the electric field drift.

Return flow is around the dawn side of the earth at $\sim$60°Λ. If there is no source of ionisation in this region then a decay in ionisation is expected, by a factor of 10 every 2-3 hours, through recombination and a nightside F-layer trough is formed.

Many investigations to observe these effects are proceeding e.g. Shepherd et al. (1976), but these are in their early stages and are not yet conclusive.

1.3 An Overall View

1.3.1 Polar ionospheric characteristics and processes

Gradually a combination of data from many techniques is leading to a clearer picture of the ionospheric layers in the polar regions. Observations have been carried out in three main altitudinal ranges, i.e. in the bottomside F-region (ground- and aircraft-based instruments), around the F-layer peak and just above (topside-sounding and in-situ measurements from satellites) and in the high-altitude topside at 1000 km and above (the Alouette, ISIS and in-situ experiments).

These measurements have revealed the ionosphere, at high latitudes, to be very different from that observed nearer the equator. It appears as a region of much fine structure in which many processes, not thought to be active in determining the mid-latitude distribution, are introduced largely through the influence of the magnetosphere. This separation of the ionosphere into two contrasting regions is best demonstrated by a

tabulation of typical values of the most important parameters encountered at points within the mid- and high-latitude regimes. Values given in this comparison (Table 1.3) are taken from data collected at altitudes near the F2 peak and just above.

The anomalous behaviour of the polar ionosphere is a result of processes largely arising from the high electric fields and low-energy particle input. Such parameters often neglected at low latitudes must be introduced into any high-latitude model. The necessary data is now available from fully-instrumented satellites such as the Atmospheric Explorer Series and from the versatile ground-based incoherent scatter radars.

Recently, a large body of data from low-energy particle detectors has become available and in this work particular attention is given to the role of low-energy electron fluxes in the particular problem of maintenence of the F-region in winter.

1.3.2 The electron density distribution

Two different pictures have begun to emerge. The spatially- and temporally-averaged data seem to suggest, quite strongly, a tongue structure, such as Figure 1.1, with a region of high density plasma extending from the dayside, across the polar cap, into the dark hemisphere. Closely-spaced data, along the track of satellites, indicate that the distribution consists of a ring-trough-cavity arrangement (Figure 1.5).

At very large sunspot number epochs there is some evidence that the electron density depletion in the polar cavity tends to disappear so that the trough-ring-cavity structure may be described as a trough-'tongue' structure. The in-filling of the cavity at high sunspot numbers may result from a number or combination of reasons such as enhanced convection of plasma or increased insolation. However, when such complications are relatively unimportant, as at night near solar minimum, it is clear that the magnetospherically-induced ionisation in the polar plasma takes the

basic form of a trough-ring-cavity structure as proposed by TA.

In subsequent sections a large scale survey of the topside polar plasma is reported in an attempt ot follow the evolution of the polar F-region over the solar cycle, and the distribution at epochs of high solar sunspot number is given particular attention.

Table 1.3

Typical mid- and high-latitude ionospheric characteristics

Parameter	Mid-latitude value	High-latitude value
Electron density (cm^{-3})	$10^5 - 10^6$	8×10^4 (Trough & cavity) $1.5\text{-}3 \times 10^5$ (Plasma ring)
Height of maximum h_mF2 (km)	260 (winter)	∿320 (winter)
O^+ density (cm^{-3})	2×10^4	1×10^4
H^+ density (cm^{-3})	5×10^3	2×10^2
Electron temperature T_e (°K)	1500	2 - 4000
Ion temperature T_i (°K)	1200	1500
Neutral temperature T_n (°K)	800	1000
Neutral constituents and densities $\times 10^{-7}$ (cm^{-3})	O 10.0 N_2 0.2	O 7.0 N_2 0.3
Plasma scale height (km)	200 (day)	150
Neutral scale height (km)	40	60
Electric field (mVm^{-1})	1 - 2	20 - 40

CHAPTER 2

A SURVEY OF THE POLAR PLASMA

2.1 Scope of the Data Survey

2.1.1 Objectives

Initial studies (Section 1.2) of topside-sounder satellite data have revealed many features of the plasma distribution in the polar region, notably the plasma ring-trough-cavity structure (Thomas and Andrews, 1970; Sato and Colin, 1969). These analyses were restricted to relatively small data bases concentrated at minimum epoch of the solar sunspot cycle.

However, topside sounder data are now available over the entire period from 1962 to 1970. This availability, together with routine processing programmes of 'raw' ionograms gives much denser observational coverage of the ionosphere at all epochs of the solar sunspot cycle. Thus large statistical surveys to investigate the functional variation of electron density distribution are now feasible.

Such surveys are important as a guide to the physical processes active in the polar upper atmosphere and as a test for theoretical models of this region, since these must attempt to reproduce the experimental observations. Thus, the aim of the present analysis is to determine the distribution of ionisation in the polar F-region and above and to specify both its long- and short-term variations. More particularly:

(i) Does the ring of enhanced plasma seen at solar minimum persist to times of high sunspot number?

(ii) How does the magnitude of the plasma density at the F_2 peak and at 1000 km vary over longer periods than that during the 1962/63 winter? Is there any long-term systematic variation as one proceeds from 1962 to 1968?

(iii) Is the ring of enhancement a complete ring or are there small sections missing at certain times as would occur if the enhanced region is in fact two superimposed 'horseshoes' of ionisation? The latter possibility has been suggested by magnetospheric models and by certain particle precipitation patterns.

(iv) How do magnetic activity and changes of universal time affect the 'ring'? Are the motions of the boundary correlated with changes in the interplanetary magnetic field? Are there any similar motions over the solar cycle when the interplanetary magnetic field is seen to vary in a cyclic way?

2.1.2 Data base

The basis of this analysis is data from the topside-sounding satellite Alouette I, though a small exploratory number of satellite passes from the Alouette II and ISIS satellites have also been examined.

Alouette I data have been reduced from the raw ionogram format (see Introduction) to many tabulated forms suitable for use in large data surveys. These include density and scale height data from ionograms monitored at Stanford (NASA, Thomas et al., 1967) and the so-called, Interpolated N(h) and ALOSYN (Alouette topside sounder synoptic) data (Canadian Department of Communications, Ottawa). Of these, only Alosyn fulfil the criteria of availability over the entire period 1962-1970.

Alosyn consists of selected ionospheric parameters obtained from Alouette I ionograms. An illustration of the parameters tabulated and their format is found in Phelps (1972). The relationship between the measured characteristic frequencies and the physically meaningful parameters electron density and altitude is given by

$$N = 1.24 \times 10^4 \quad (f_x^2 - f_x f_H) \qquad 2.1$$

where f_x is the reflected extraordinary ray frequency

f_H is the electron gyrofrequency at that altitude.

When f_x and f_H are measured in MHz, N is given in electrons cm^{-3}.

On substitution of $f_x = f_xF2$ (the extraordinary ray critical frequency) and $f_H = f_H(h_mF2)$, equation 2.1 yields the maximum concentration in the F-region, N_mF2. In the Alosyn tabulations, h_mF2 has been taken as 300 km in all cases, in order to calculate $f_H(h_mF2)$. The error resulting in N_mF2 by this assumption is generally $\lesssim$5%. This simplification is introduced to remove the need to derive the electron concentration profile with altitude from the observed virtual height-frequency traces recorded on ionograms. The resulting quantity, tabulated in Alosyn, is JF0F2, the ordinary wave penetration frequency of the F2 layer calculated from the measured f_xF2. Replacing the bracket in equation 2.1 by JF0F2 gives N_mF2.

Similarly the electron concentration at the satellite, N_S, ($\sim$1000 km above the earth) is derived by substituting $f_x = f_{xS}$ and $f_H = f_{HS}$ in equation 2.1, where

f_{xS} is the frequency at which the extraordinary ray has zero range

f_{HS} is the electron gyrofrequency at the vehicle.

2.1.3 Choice of coordinate system

One of the main influences of interest is that of the magnetosphere and it is natural to use a magnetospherically related coordinate system wherever possible. Therefore most of the data are required to be plotted as functions of generalised invariant latitude, this parameter being directly derived from McIlwain's L parameter of the magnetosphere. This coordinate system has had much success in ordering both magnetospheric and ionospheric data, and high latitude features seen on early topside-sounder observations displayed a tendency to follow L contours (Thomas et al., 1966).

The relationship between generalised invariant latitude at a point in space and the L coordinate is given by

$$\Lambda' = \cos^{-1} ([1 + h/R_E]/L)^{\frac{1}{2}} \qquad 2.2$$

where h = altitude of point above the earth

R_E = radius of the earth.

Of the various time elements used at high latitudes two time coordinates are selected; Local Time (LT) and Magnetic Local Time (MLT).

2.1.4 Data reduction and display

Because of the large number of satellite passes involved and the the fact that each pass contained on average 20 data points the survey is necessarily computer-based. For all passes at a geographic latitude greater than 40° values of N_mF_2 and N_S were computed and automatically displayed as a function of geographic latitude on plots such as Figure 2.1, showing typical passes at different epochs of the solar cycle. Each pass has been labelled by the universal time and the day of the year. Additional information has been extracted from the Alosyn tabulations and added to aid analysis. These are:

(1) The Kp index at the time of the pass

(2) The local time along the satellite orbit

(3) Values of solar zenith angle, χ, at the beginning and end of each pass.

Data in both the northern and southern hemispheres have been treated in this way though unavailability of southern hemisphere data, due to the lack of satellite telemetry stations, prevented any close examination of the electron density infrastructure in the Antarctic.

Conversion of any point on the satellite path from geographic coordinates to the chosen magnetospheric coordinate system was carried out in two ways:

(1) A computer routine was devised to give the generalised invariant latitude and magnetic local time of a point in space specified by its geographic latitude and longitude, altitude and universal time. This programme utilised the field model of Jensen and Cain (Cain et al., 1967).

(2) This method is illustrated by Figures 2.1 and 2.2. From a computer programme such as that described above, contours of generalised invariant latitude were generated, for altitudes of 300 and 1000 km, and plotted over a polar coordinate grid of geographic latitude and longitude. As shown by Figure 2.2 the scale of this grid was chosen so that when the electron density plots, represented by Figure 2.1, were superimposed on the polar grid the abscissae of these plots were tangential to the 80.5° geographic latitude contour. The abscissae of these plots thus closely approximated to the track of the Alouette I satellite. Variations of N_mF_2 and N_S as a function of geographic latitude and longitude, local time and invariant latitude could easily be determined from the appropriate polar coordinate grid. For ease of analysis a longitudinal reference mark, showing the direction of the 0° longitude line from the geographic pole, was computed from the Alosyn tabulations and added to the plots.

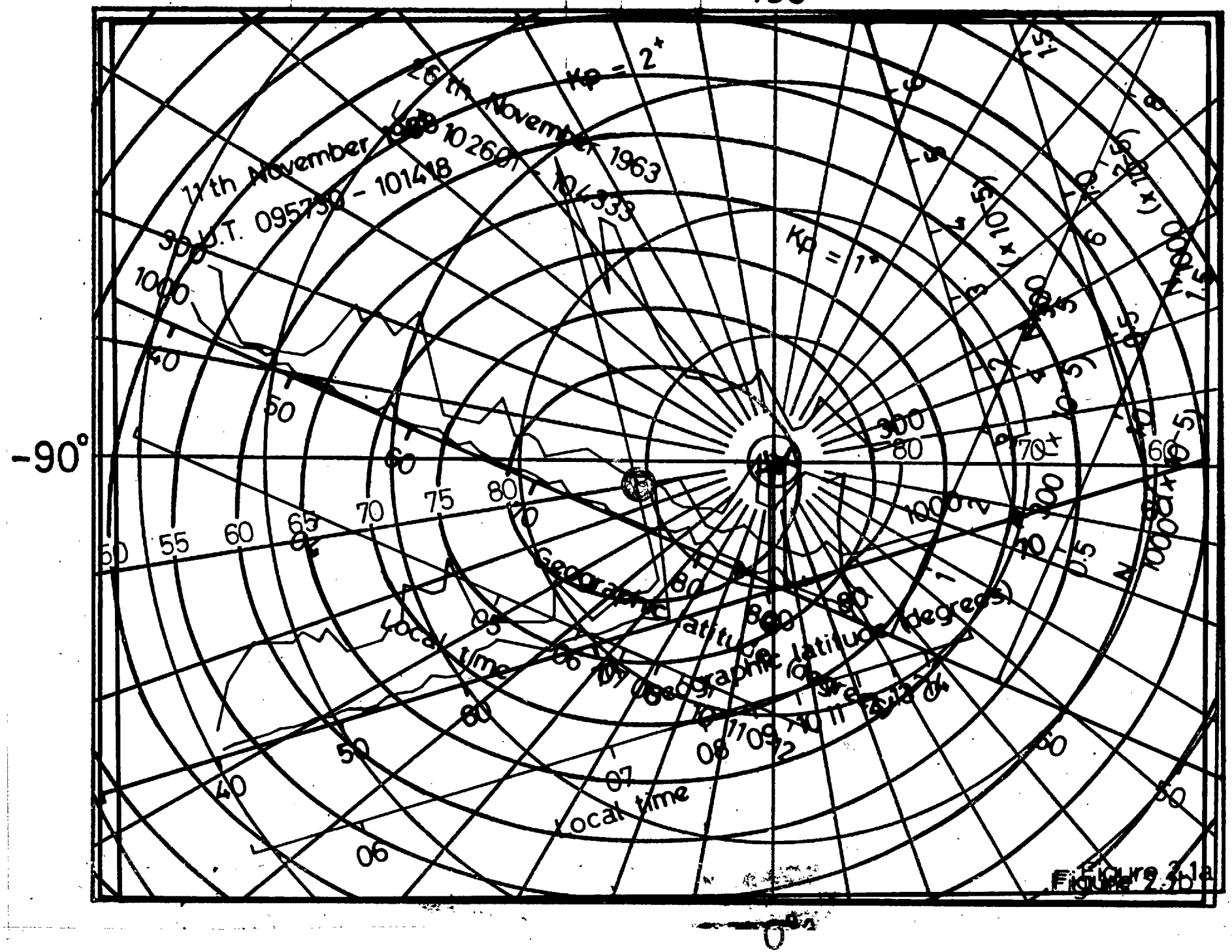
2.2 Analysis Procedure

Electron density plots from many months in the period from 1962 to 1970 have been generated from Alosyn data. These plots revealed the summer polar ionosphere to be largely a featureless region of high electron density dominated by solar insolation. During winter, however, when the solar influence is minimised, the ionosphere was of a very different nature. In agreement with previous work, the electron density distribution appeared very structured with large spatial gradients.

Figure 2.1a A typical computer plot of the variation of electron density at altitudes of 300 and 1000 km along the track of an Alouette I satellite pass during a period of low sunspot number. When the reference cross and line are laid over the geographic pole and the 0° longitude meridian on Figure 2.2 respectively, as shown, the abscissae approximates to the track of the satellite. The electron density at 300 and 1000 km as a function of geographic latitude and longitude, local time or invariant latitude can then be read off from the appropriate ordinate scale, N_{300} or N_{1000}. Details of local time along the satellite track, the universal time and Kp index during the time of the pass have been added from the Alosyn tabulations.

Figure 2.1b A typical computer plot of the latitudinal variation of density at 300 and 1000 km measured by the Alouette I satellite during a time of high sunspot number. Other details are as in Figure 2.1a.

Figure 2.2 Coordinate systems used in the analysis. The circles and lines represent contours of constant geographic latitude and longitude respectively. The open circle corresponds to the geographic pole. Overlaid on this grid are contours of constant generalised invariant latitude (Λ') at an altitude of 300 km calculated from the geomagnetic field model of Jensen and Cain (Cain et al., 1967). The solid circle shows the position of the L-pole, corresponding to $\Lambda' = 90°$.


180
-90°
90°
0
Kp = 2+
Kp = 1+
11th November 1963
10260
10333
0000 U.T. 095750 - 101418
Geographic latitude (degrees)
Local time
Figure 2.1a
Figure 2.1b

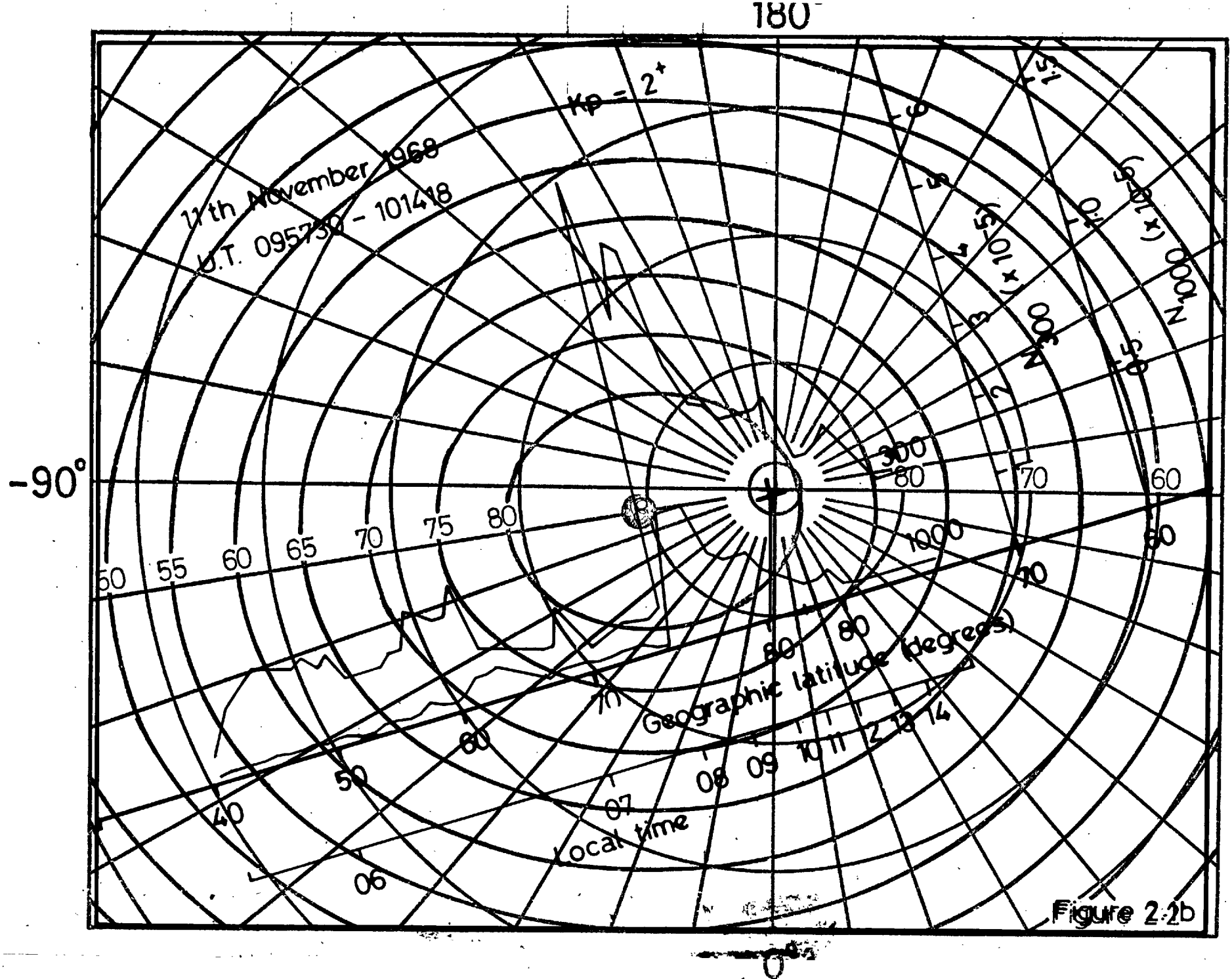

180
-90°
90°
0
Kp = 2+
11th November 1968
U.T. 095730 - 101418
Geographic latitude (degrees)
Local time
06
07
08
09
10
11
12
13
14
50
55
60
65
70
75
80
85
40
60
80
70
60
50
30
10
1000
300
N (x10⁵)
N (x10⁴)
N (x10³)
N (x10²)
N (x10)
N (x1)

Figure 2.2b

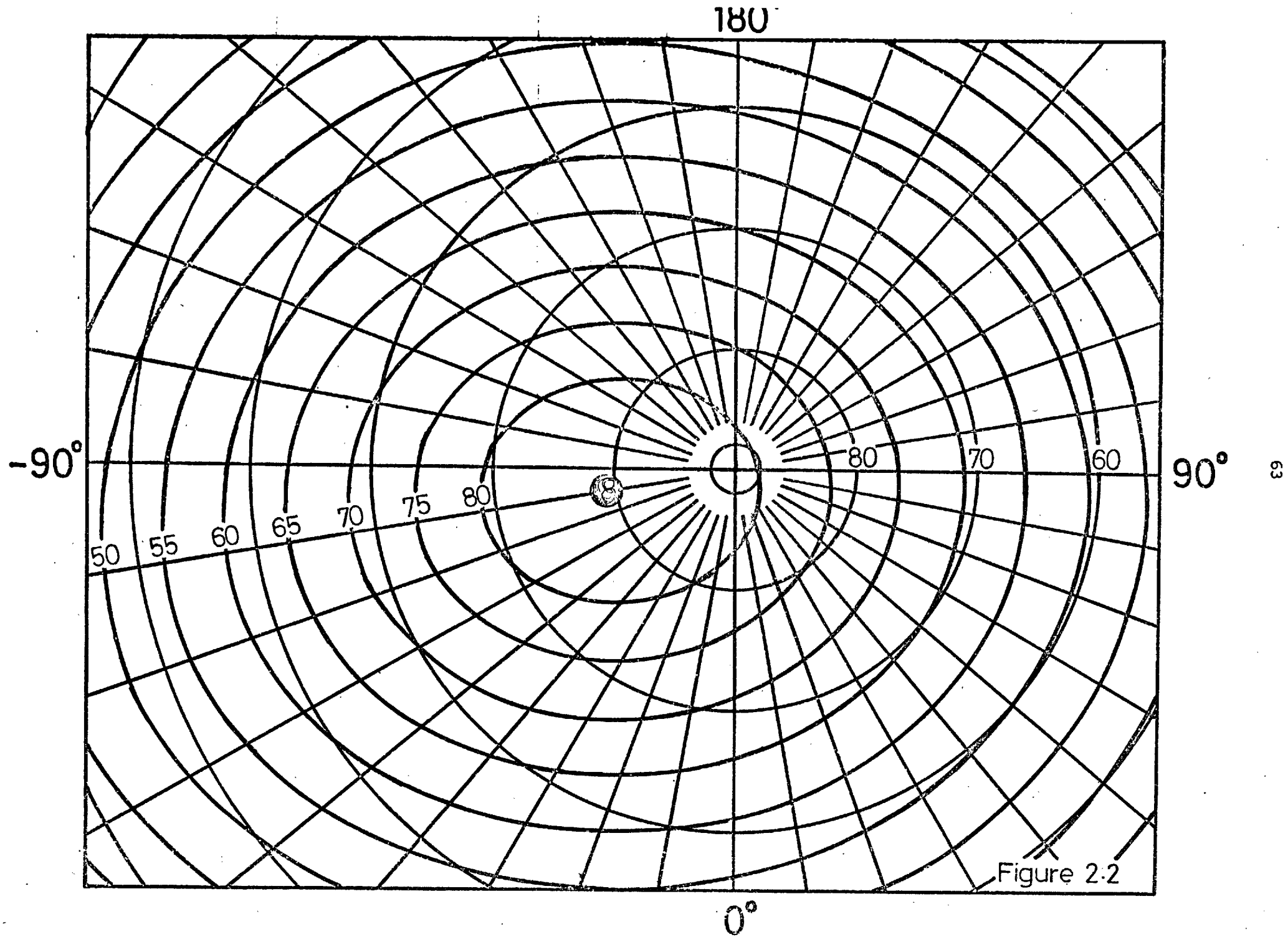


Figure 2·2

For these reasons sections of data obtained during the polar winter, when processes other than those due to the sun appear important, have been selected for further scrutiny. However, due to a lack of data at solar maximum some summer plots for this time were examined. Details of the number of passes examined in the months chosen are given in Table 2.1. The total number of passes involved is also shown.

The large number of passes used required the use of averaging. Figure 1.3 and the work of Hopkins (1974) have demonstrated that averaging values of electron density at a certain latitude over time, can lead to the masking of many interesting features. This consideration led to the present analysis in which persistent features of the distribution were recognised on each pass and then aspects of these features averaged. In particular, levels of electron density and the prominent enhancements of ionisation, as represented by the polar peak (Thomas et al., 1966), have been examined by the following proceedure.

1. A computer routine was devised to search for and select enhancements from the electron density data using the following criteria all of which must be satisfied simultaneously

   (a) Enhancements must occur at $\Lambda' \geq 60°$ for low values of Kp ($\leq 2^{o}$), and $\geq 50°$ for higher values of Kp.

   (b) The ratio of electron density at the maximum to that on either side of the enhancement maximum must be ≥ 2 except in the noon sector where a value of this ratio of 1.5 was chosen.

   (c) Values of logarithmic electron density gradients $(\Delta N_e/N_e)/\Delta\lambda$ at enhancement boundaries must be $\geq .1$ (degree of geographic latitude)$^{-1}$.

   (d) Each enhancement must be at least 5 data points in extent.

2. A visual inspection of the data was carried out in a search for possible enhancements. Passes which contained less than 20 values of

electron density were rejected.

Interpretation of high-latitude ionograms is difficult and large errors are often introduced. An error code is built into the Alosyn data and when this indicated that large errors were present in $\geq$ 50% of the electron density values in a plot, this too was rejected. The number of passes rejected for these reasons is also shown in Table 2.1.

3. On selection of an enhancement the following quantities were calculated and tabulated for further analysis

(a) The average value of electron density in the enhancement.

(b) The average value of electron density equatorward of the low latitude enhancement boundary.

(c) The average value of electron density poleward of the high latitude enhancement boundary.

(b) and (c) were assumed to be representative of densities in the trough and cavity respectively.

(d) The latitude and local times of the boundaries.

(e) The universal time and Kp of the satellite pass.

2.3 The Polar Plasma Distribution

2.3.1 The plasma ring

Each pass examined was divided into those which showed no evidence of electron density enhancements, passes indicating enhancements in positions predicted by the trough-ring-cavity distribution of Thomas and Andrews (1970) and data showing multiple enhancements. Other passes were ambiguous.

The results of this classification are shown in Table 2.1 in which numbers of passes falling into each category are presented.

These indicate that, over the entire period examined, $\sim$44% of all passes can be explained by the trough-ring-cavity structure and this is

YEAR	MONTH	Number of passes examined	Number of passes rejected	Number of passes showing ring structure	Number of passes showing no enhancements	Number of passes showing multiple peaks
1962	November	317	37	117	41	68
	December	284	61	112	41	26
1963	January	185	19	61	9	35
	February	51	2	25	5	13
	November	277	32	161	56	55
	December	240	24	144	40	32
1964	January	234	30	130	25	23
	November	226	33	134	25	32
	December	228	66	107	23	10
1965	January	224	39	124	29	20
	December	6	3	2	1	0
1967	January	132	23	39	22	22
	December	70	6	32	19	8
1968	May	74	2	25	38	4
	June	94	10	16	44	19
	November	89	24	30	14	11
1969	October	13	4	3	2	0
TOTALS		2744	415	1194	434	378

Table 2.1 Details of the data survey. Passes not included in this table exhibited ambiguous records.

the only repeatable and stable distribution seen. During summer months this percentage declines due to constant solar illumination of the polar regions.

It has been inferred from measurements of optical emissions during IGY (Sandford, 1970) and suggested by the electron density measurements of Pike (1971b) that levels of electron density near the geomagnetic pole may increase during the solar cycle thus obscuring the polar cavity of electron density at solar maximum. However, the figures in Table 2.1 show that the polar plasma distribution is dominated by the plasma ring arrangement at all stages of the solar cycle. If only winter data over periods representative of solar minimum (1963-4) and solar maximum (1967-8) are considered it is seen that the percentage number of passes indicating the presence of a plasma ring is 58% and 34% respectively. Hence the probability of observing the ring with Alouette I decreases from solar minimum to solar maximum but is still appreciable.

This initial result led to the adoption of the trough-ring-cavity structure as the model plasma distribution in the polar topside ionosphere. As the parameters controlling the ring are not well known further work concentrates on the long- and short-term variations to determine its controlling parameters.

2.3.2 Continuity of the plasma ring

Eather and Mende (1972), from investigations of high-latitude optical emissions, have suggested that the pattern of particle precipitation may consist of two 'horseshoes' placed around the geomagnetic pole and symmetric about magnetic noon and midnight respectively. In this interpretation discontinuities in auroral and low-energy particle fluxes appear around the 06 and 18 hr local time sectors with coincident absence of optical emissions. This has since been proved incorrect by the ISIS II Auroral Scanning Photometer (Lui et al., 1975) which has clearly demonstrated the continuity of the auroral oval.

However, it seems apparent from the very different properties of particle fluxes on the day and night-sides of the earth that there are two separate sources for such particles. This and certain theoretical magnetospheric models (Mishin et al., 1970) suggest that there may be some weakening of fluxes near the 06 and 18 local time sectors. A possible result of this situation could be that the plasma ring would, in fact, consist of two superimposed 'horseshoes' of ionisation.

This possibility has been examined in two ways:

(1) The continuity of the ring has been tested at maximum and minimum epochs of the solar cycle. This analysis involved dividing the satellite passes into hourly local time sectors. Then, considering each sector separately the ratio of the number of passes on which evidence of an enhancement was obtained to the total number of passes was computed.

(2) The local time variation of electron density in the ring has been studied. Values of density read off by the procedure of section 2.2, during quiet magnetic periods, have been classified into 24 hourly local time data sets, and averaged appropriately. Electron densities both at solar maximum and minimum have been treated in this way.

The results are presented in Figures 2.3(a) and (b) which show the variation of the percentage occurrence frequency and the average electron density at the F-layer peak with local time in the ring. The continuous and broken curves refer to solar minimum and solar maximum respectively.

Figure 2.3(a) shows that there are no discontinuities at the local times expected. The percentage occurrence of the enhancement region appears fairly constant throughout the local time day except for two sectors. Around noon the probability of detecting an enhancement increases. This is in agreement with previous observations of Nishida (1967) and Thomas and Andrews (1970). However, in the post noon sector there is a sharp decrease, the reason for which is not clear, followed by a recovery

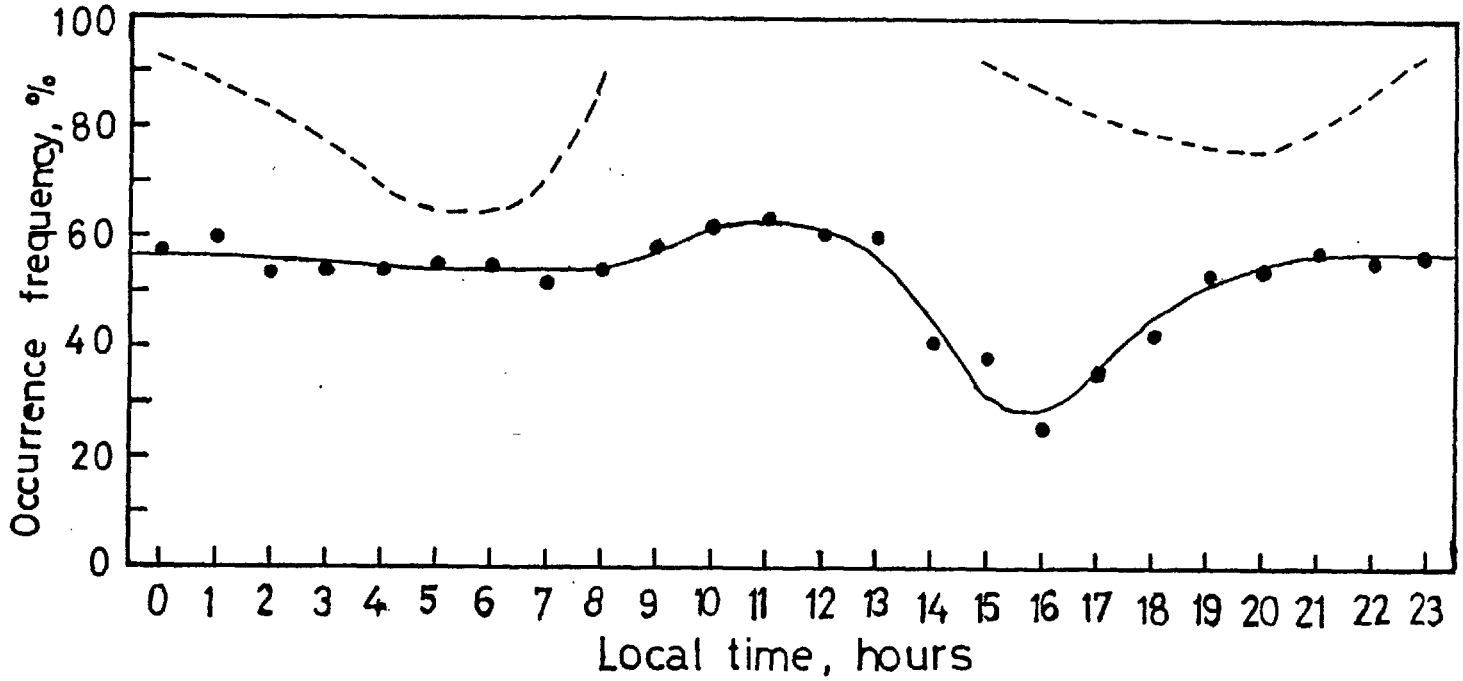


Figure 2.3a The variation of the percentage occurrence frequency of the enhancement with local time for minimum epoch (continuous curve) of the solar sunspot cycle and at all levels of magnetic activity. A smooth curve has been drawn through the points merely as a guide. A similar curve exists for conditions at high sunspot number. Also shown, (broken curves) are occurence frequencies of the visual aurora in the dawn and evening local time sectors for average levels of magnetic disturbance. (Taken from Feldstein and Starkov, 1967). The two observations do not show minima at the same local times.

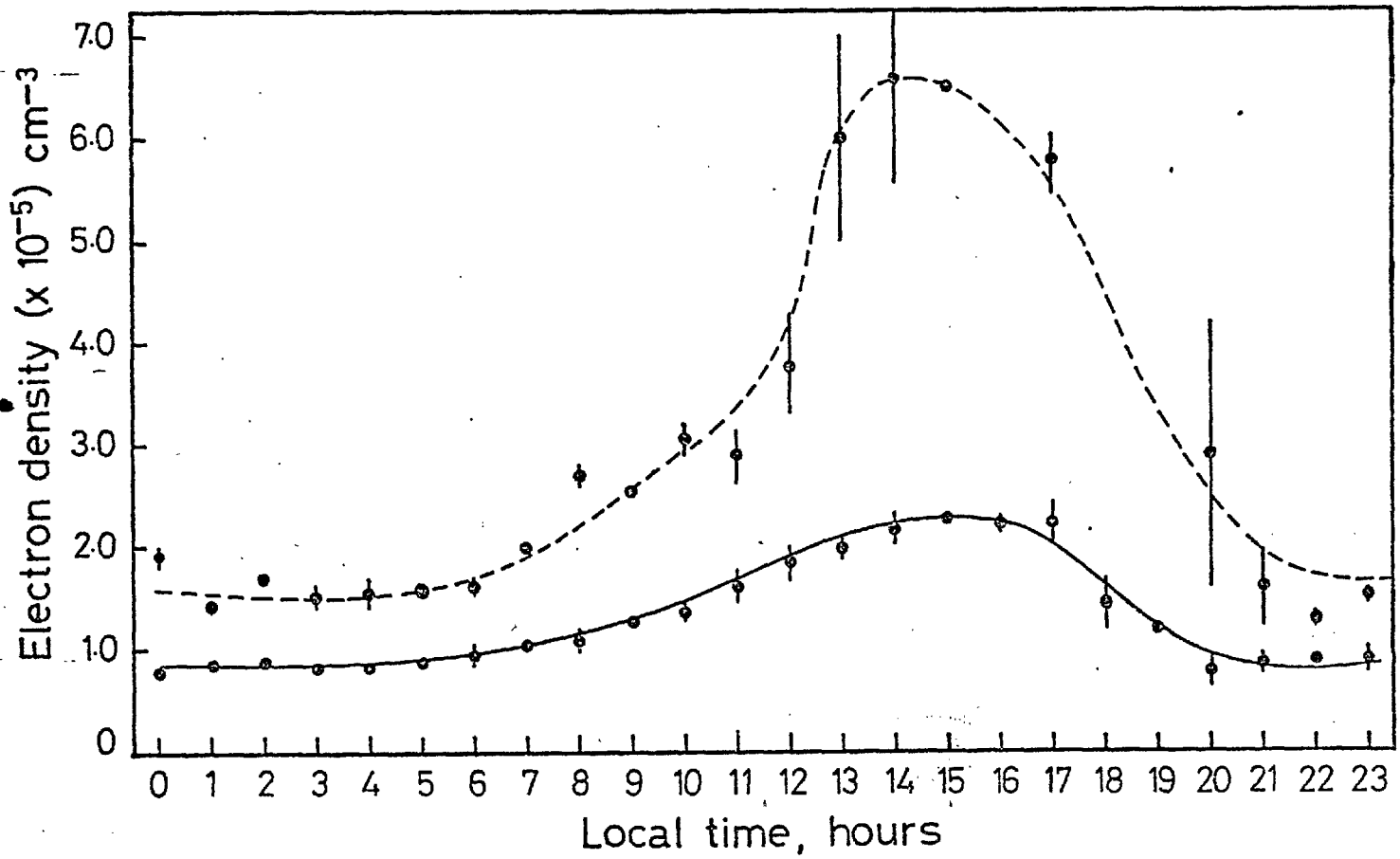


Figure 2.3b The variation of the average electron density at the F-layer peak, N_mF2, within an enhancement, for magnetically quiet periods for two cases:-
(1) the continuous curve - solar minimum; (2) the broken curve - solar maximum.
A smooth curve has been drawn through the points merely as a guide.

to previous values towards evening.

The local time behaviour of electron density in the ring at both epochs of the solar cycle (Figure 2.3(b)), is typical of solar control. Minimum densities reach their lowest values near local midnight while there is a prominent post-noon maximum. Again no discontinuities or electron density depletions are apparent near 06 hr local time to suggest break up of the ring into two 'horseshoes' of ionisation or weakening of the ionisation source. This result is, despite sparsity of data, repeated at 18 hr local time.

2.4 Modulations of Electron Densities

2.4.1 Solar cycle variations

It has long been known that electron densities at low- and mid-latitudes exhibit a strong modulation with the solar sunspot cycle (Rishbeth and Garriott, 1969). However, no analysis, at high latitudes, during winter conditions, has been attempted. It is thus not known whether the influence of the sun disappears completely at large solar zenith angles or if it is maintained by some process other than ultra-violet insolation.

A comprehensive investigation of the long-term changes of electron density has been carried out for winter conditions. Electron density data from the trough, cavity and ring regimes were obtained by the procedure described in section 2.2 for the months of November, December, and January at both the satellite altitude and the altitude of the F2 peak.

These data were first averaged irrespective of local time and Kp index. In cases where more than one month's data were available these were combined to produce a winter average electron density. The results of this were reduced to data tables.

To ensure that effects evident in such results were not due to high incidence of magnetic activity during some periods or because the satel-

lite sampled local times of high plasma density at a particular epoch further analyses were performed. The electron density values were first sub-divided into three data bins labelled by the ambient level of magnetic activity shown by the Kp index. The three categories and their assigned Kp ranges were:

LOW - $0 \leq Kp \leq 2^-$

MEDIUM - $2^o \leq Kp \leq 4^-$

HIGH - $4^o \leq Kp \leq$ maximum value

Winter averages for these three levels of magnetic activity were then obtained in a similar manner to that described above. Data samples from this investigation appears in Figure 2.8.

The local time sector of the polar ionosphere sampled by the satellite depended on the characteristics of orbit. In order that this should not bias results the data were classified as to their local time to form three sub-data sets. Again winter and monthly averages were computed. The sub-divisions of local time and their labels were:

MORNING - $00 < LT \leq 08$ hr

NOON - $08 < LT \leq 16$ hr

EVENING - $16 < LT \leq 24$ hr

The sunspot number, $\overline{R}$, used to illustrate the solar cycle variation was the smoothed observed Zurich Relative Sunspot Number. The period concerned, 1962-1969, was concurrent with the rising part of solar cycle No. 20 and $\overline{R}$ varied from a low of 9.06 (October 1964) to reach a maximum of 110.6 in November 1968. This maximum was far exceeded during previous cycles, particularly 1957-8 when $\overline{R}$ attained values as high as 205. To examine the variation of N_mF2 or N_S, the average values for a particular local time sector and level of Kp, were plotted as functions of the solar cycle epoch. The behaviour of $\overline{R}$ (Figure 2.4) is given for comparison.

The long-term variation of polar topside electron densities in the plasma ring is illustrated for specimen conditions in Figures 2.4, 2.5, 2.8 and 2.9. For comparison the behaviour of trough and cavity densities is shown in Figures 2.6 and 2.7. These show that there is a strong modulation of polar plasma densities at the F-layer peak and satellite altitudes irrespective of magnetic activity and local time selection criterion applied to data plotted. Densities in the cavity are somewhat less than those given by Phelps and Thomas (1974). This is due to the fact that in their definition, the polar cavity was found at all invariant latitudes >80°. Thus the sporadic enhancements of plasma density sometimes occurring within the polar cap will be included in their averages.

The magnitude of this effect has been investigated by a least-squares fit criterion regression analysis. The results of this are summarised in Table 2.2. Values of correlation coefficient obtained were mainly quite significant. In column 9 the relationship has been expressed in a similar form to that presented by Allen (1948) for mid-latitude ground-station data. In general, he and Wright (1962) have found

$$N_m F2 \propto (1 + .02 \bar{R}) \qquad 2.4$$

Comparison of this with Table 2.2 shows that the effect of the sunspot number on topside electron densities at high latitudes is similar to that at midlatitudes for regions of high density plasma but is much larger in the trough and cavity where densities are low.

Results shown in Figure 2.8 are typical of others in the survey concerning variations of polar densities with increasing magnetic activity. The behaviour is very varied and no clear trend is immediately apparent in either the ring, trough or cavity. However, the greater proportion of the Kp sub-divided into monthly averages showed an initial increase of electron density as Kp changed from low to moderate values. With a further increase to high magnetic activity the data show a tendency for

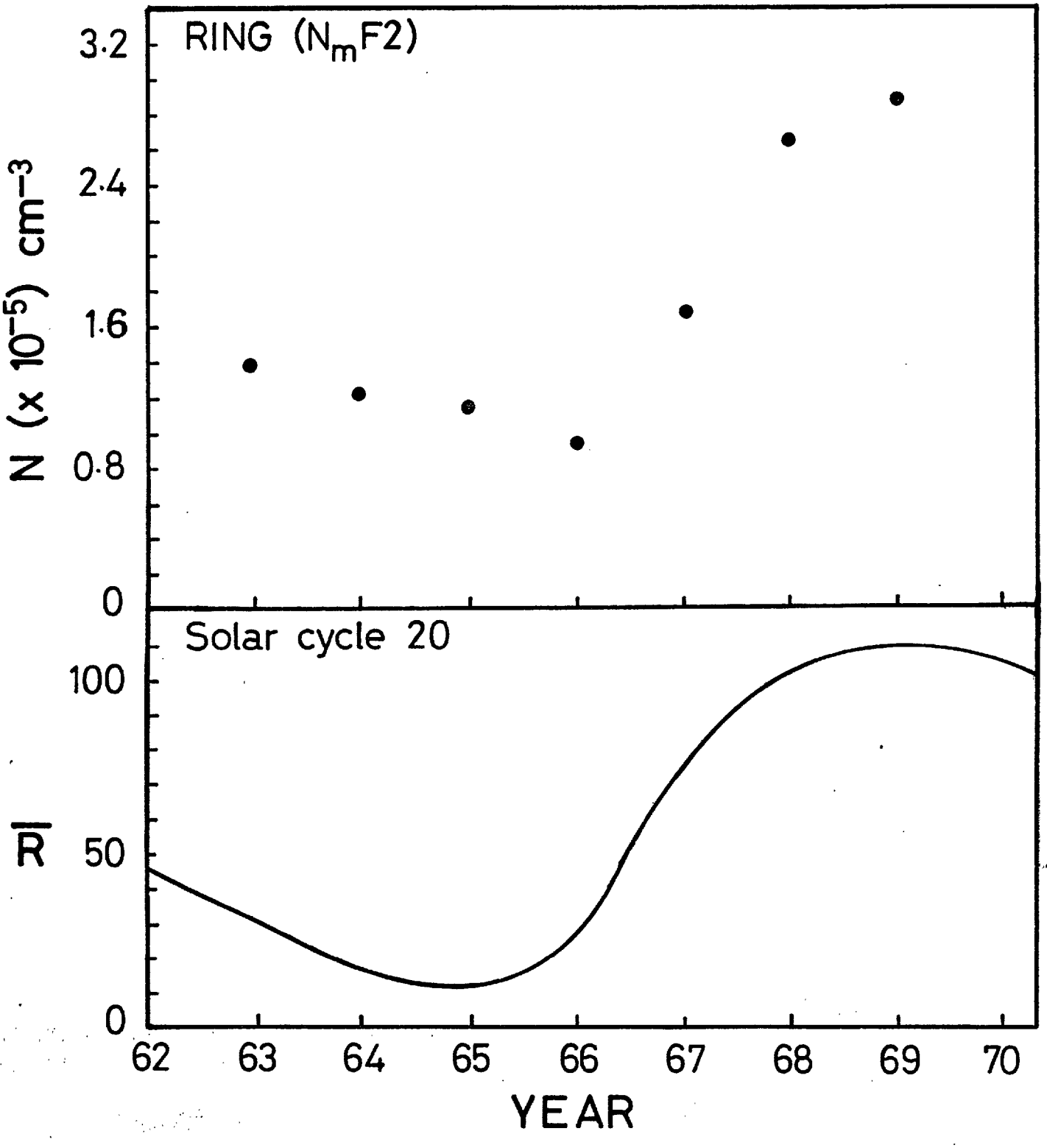


Figure 2.4 (Upper) Variation of N_mF2, the average electron density in all local time sectors and levels of magnetic activity, at the peak of the F2 layer, in the plasma ring, during the winters from 1962 to 1969. Error bars are too small to be significant on this scale.
(Lower) The modulation of the smoothed observed Zurich Relative Sunspot Number with time, largely during the rising part of solar cycle 20.

Density sampled	Kp	LT Sector	Altitude km	a_o $\times 10^{-4}$	a_1 $\times 10^{-3}$	n	r	N^{α}
Ring	All	All	300	7.2±1.7	2.0 ±.2	8	0.93	$(1 + .03\bar{R})$
Ring	All	All	1000	0.6 ±.2	0.2 ±.1	6	0.79	$(1 + .03\bar{R})$
Ring	All	Morning	300	5.4 ±.2	1.8 ±.2	8	0.90	$(1 + .04\bar{R})$
Ring	Low	All	300	5.3 ±.3	3.0 ±.3	8	0.93	$(1 + .07\bar{R})$
Trough	All	All	300	0.8 ±.1	1.0 ±.2	8	0.93	$(1 + .13\bar{R})$
Cavity	All	All	300	0.8 ±.1	1.0 ±.1	8	0.81	$(1 + .13\bar{R})$

where a_o and a_1 refer to the coefficients in the equation $N(\text{el cm}^{-3}) = a_o + a_1\bar{R}$

n is the number of points used in the regression analysis

r is the correlation coefficient

The significance of column 9 is explained in the text.

Table 2.2 Regression analysis of the function $N(\bar{R})$.

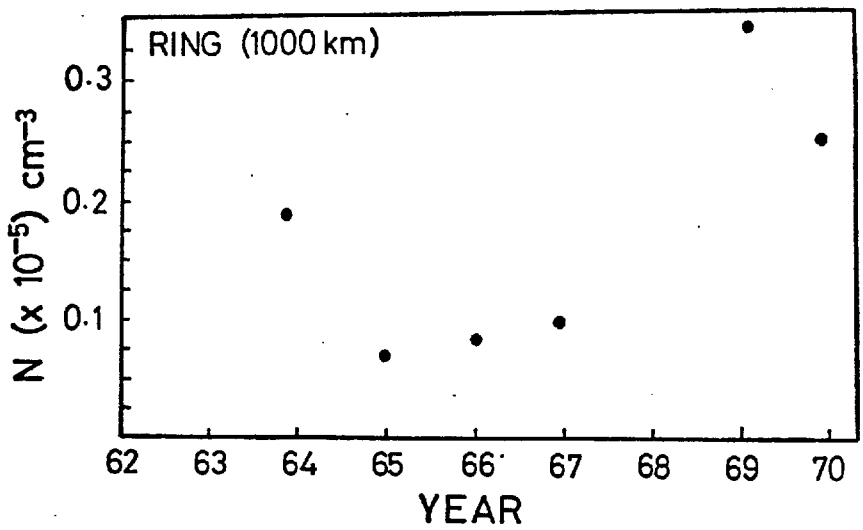


Figure 2.5 Variation of the average electron density at the satellite altitude (∿1000km), N_S, at all local times and levels of magnetic activity, for winters from 1963 to 1969.

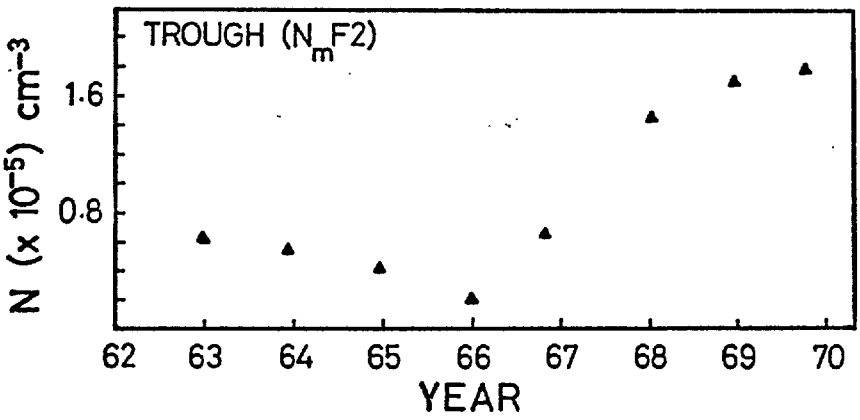


Figure 2.6 Variation of N_mF_2, the average electron density in all local time sectors and levels of magnetic activity at the peak of the F-layer, in the trough. Error bars are too small to be significant on this scale.

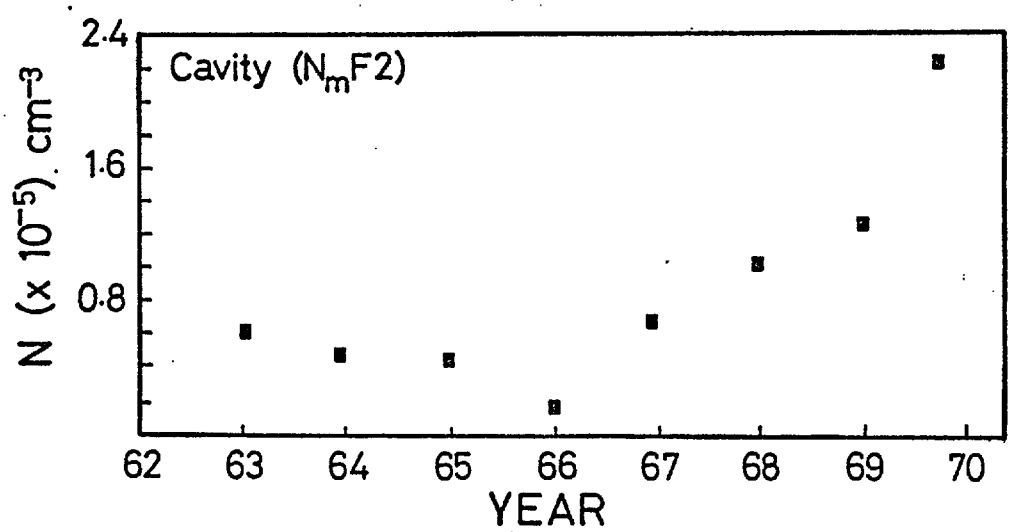


Figure 2.7 The variation of electron density averaged over all local times and levels of magnetic activity at the peak of the F-layer, N_mF2,

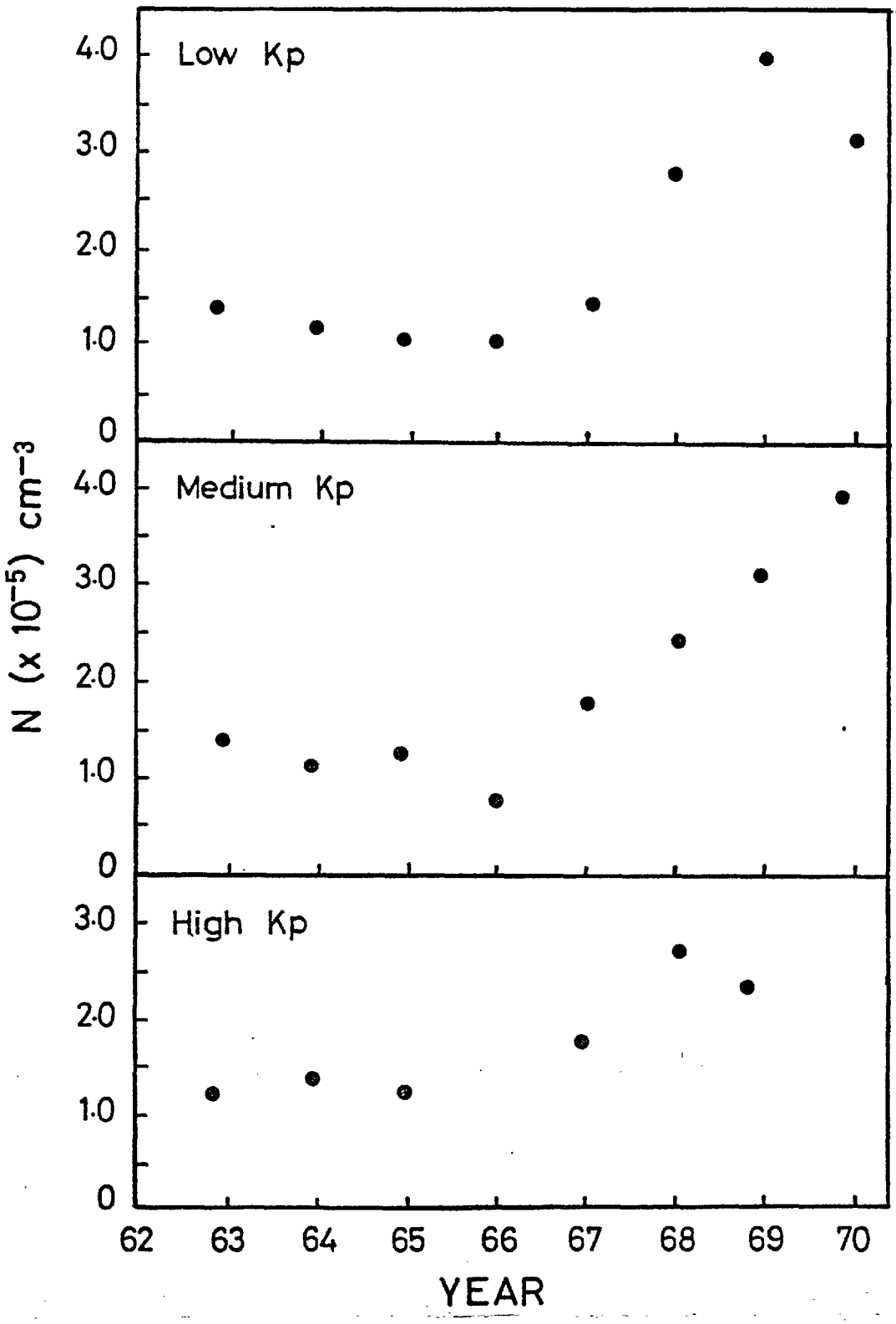


Figure 2.8 Electron density data in the ring. The three curves refer to the average electron density at the F-layer peak, N_mF2, at all local times but subdivided into the three Kp regimes as explained in the text.

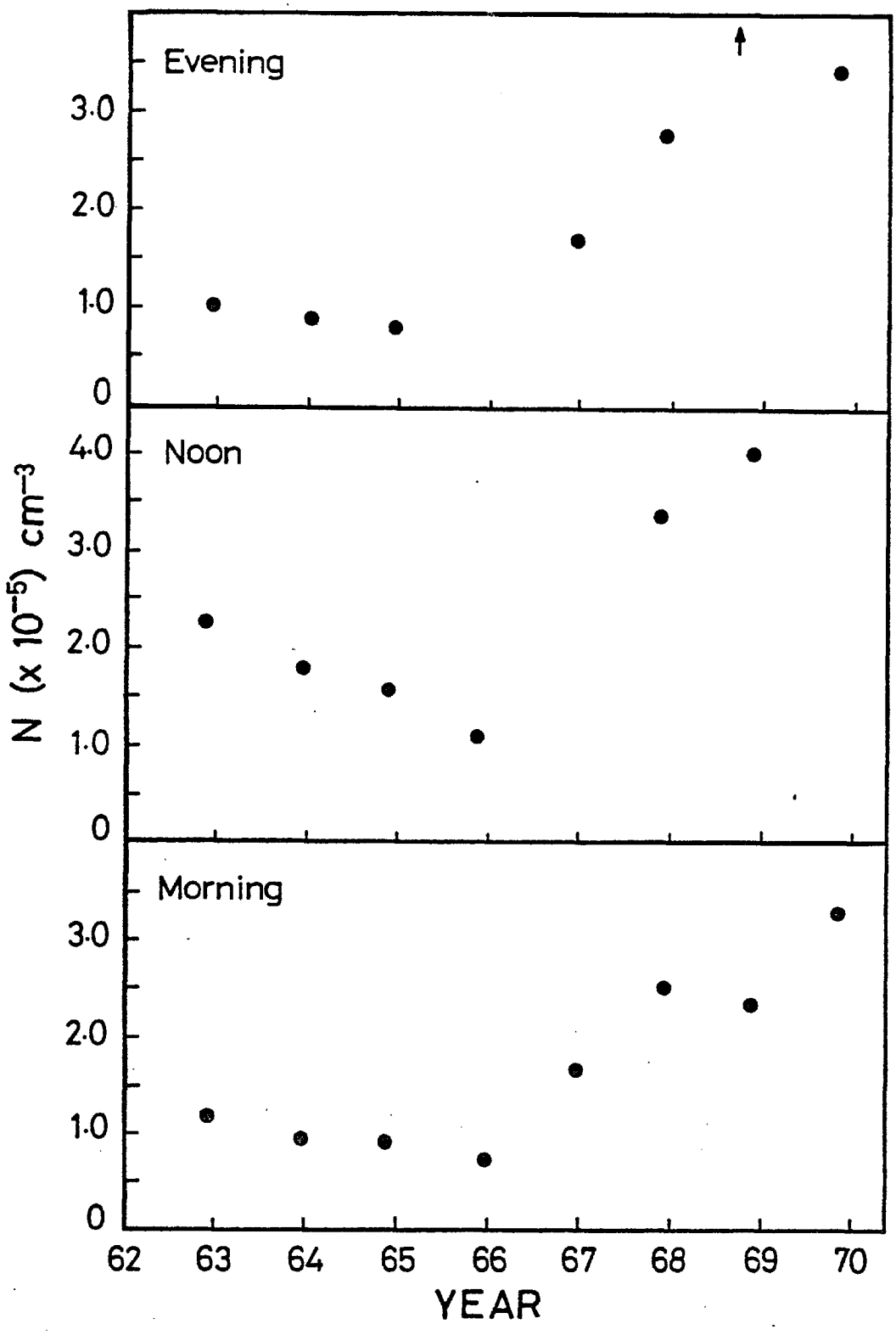


Figure 2.9 Variation of the average electron density at the F-layer peak, N_mF2, within the plasma ring for all Kp values, sub-divided into three local time sectors; morning, noon and evening.

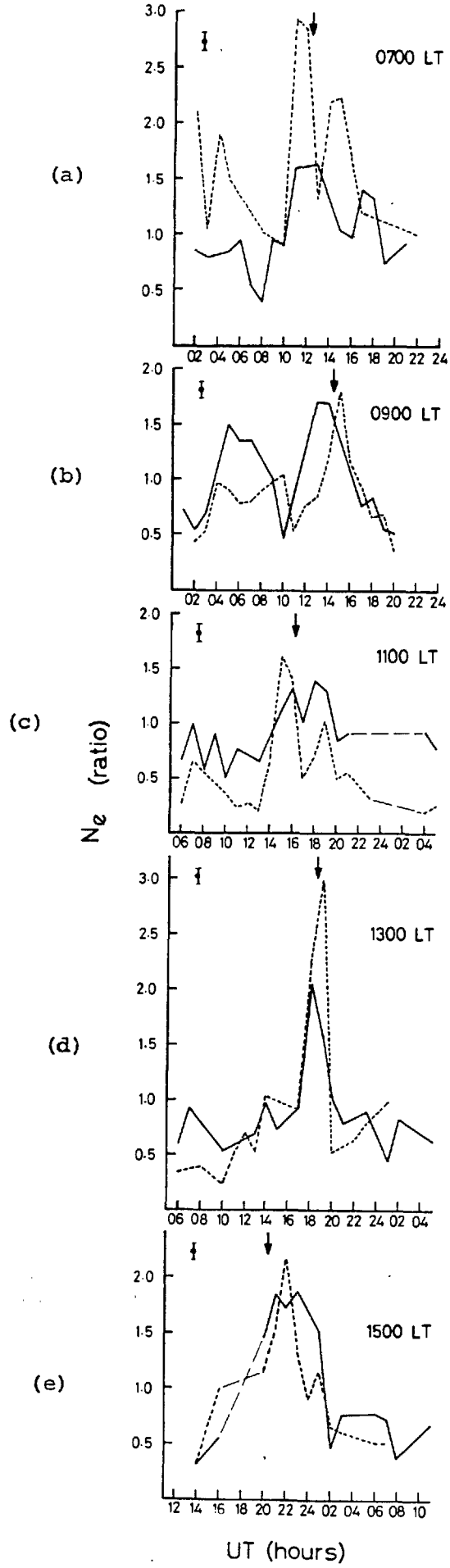


Figure 2.10

The universal time variation of electron density at various fixed dayside local times, for winter electron density data near maximum number of the solar sunspot cycle. The parameter plotted, $N_{e(ratio)}$, is the electron density normalised to the average value measured at that local time and over all universal times. The solid and broken lines refer to densities at 300 km and 1000 km respectively. The densities become very high at a universal time when the L-pole is in the local time sector concerned, denoted on the plots (left) by a small vertical arrow. A typical error bar is shown on each plot, others are omitted for clarity.

electron densities to decrease. Thus it seems that the response of the topside plasma to changes of magnetic activity is very complex.

2.4.2 Universal time effects on polar densities

Although universal time effects are a well-established feature of the high latitude plasma distribution very few satellite investigations of such behaviour have been attempted. One such analysis (Thomas et al., 1966; Thomas and Andrews, 1970) examined passes for the entire winter period of 1962-3 and plotted, as a function of the universal time, both

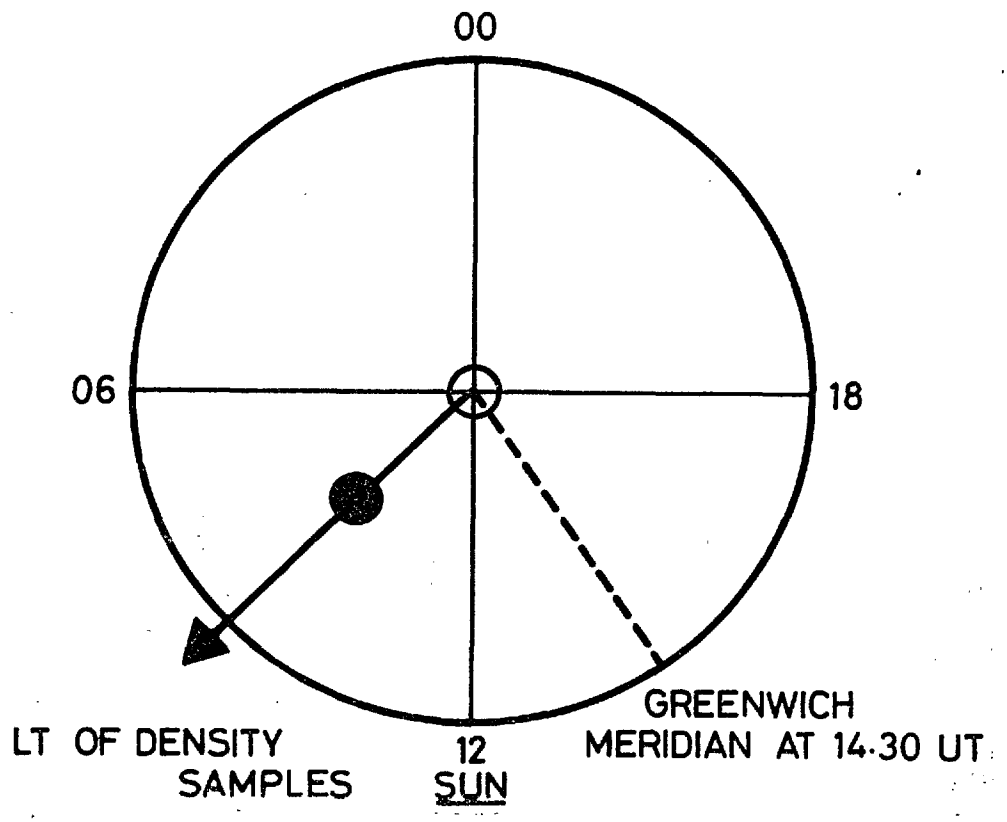


Figure 2.11 The L-pole, shown by a solid black circle, moves into the 0900 LT sector at 14.3 hrs.UT. As the L-pole rotates counter-clockwise with the earth it will move into later local time sectors.

the maximum density at 1000 km and the number of passes with a maximum exceeding a threshold of 1.5×10^5 cm^{-3} for each pass.

Both plots showed sharp peaks between 1700 and 2200 UT, corresponding to the time at which the eccentrically rotating magnetic pole, in the Northern hemisphere, was towards the sun.

However, as correctly pointed out by Thomas and Andrews (1970), this study did not separate out local time effects; thus the data set may be biassed due to passes occurring near 1800 UT sampling densities in the noon and afternoon local time sectors when photoionisation effects are expected to increase densities. It is the purpose of the following to present a comprehensive analysis of satellite observations of high latitude universal time effects. Both summer and winter months of Alosyn data have been selected from different epochs of the solar cycle. Periods were considered when the satellite orbit was such that the vehicle was travelling approximately perpendicular to the earth-sun line, and the solar zenith angle was kept almost constant over a pass. Values of critical frequencies were read at altitudes of 300 and 1000 km and local times of 07, 09, 11, 13, 15 and 17 hrs, for the dayside and 22, and 03 hrs on the nightside. Data were restricted to values of generalised invariant latitude $\sim 77.5° \pm 5.0°$ for dayside hours and $70° \pm 5.0°$ during local night to ensure that the frequencies were taken within the plasma ring feature (section 2.2). This excluded all but a few of the data obtained at 17 hrs. These frequencies were converted to electron densities and the results plotted for the winter of 1968-9 in Figure 2.10. In this diagram, each plot represents data for a different dayside local time. The average value of electron density was first normalised to the average electron density observed in the ring at the same local time and sunspot number and then plotted as a function of universal time. There is a clear peak of maximum densities for both altitudes at a universal time dependent on the local time of the sample. The small vertical arrows in Figure 2.10 denote the time when the magnetic L-pole (latitude 79°N

longitude 80°W) reaches the local time of the samples concerned. This is illustrated for the case of 0900 LT in Figure 2.11. At a universal time of 14.3 hrs, the local time at the L-pole is 0900. In all cases shown in Figure 2.10, at both altitudes, there appears to be a cyclic effect with a maximum density within two hours of the L-pole reaching the local time concerned. Before this densities are well below average and though high densities may persist after the passage of the L-pole eventually electron densities fall to low values. Data recorded at times during local night showed no obvious density maximum.

To examine the variation of this peak over the solar cycle an identical analysis was carried out on data from the winters of 1962-3, 1963-4 and 1964-5. Some of these results are presented in Figure 2.12. Here the appropriately normalised electron density, at the single local time, 1300 LT, is shown as a function of universal time for three successive winters. It is immediately apparent that the density maximum is present in every plot but is less well defined and of much smaller amplitude than that seen in 1968, a time of high sunspot number. In all of the 38 dayside cases examined at two altitudes, 6 possible local times and 4 different epochs of the solar cycle, 27 of these showed definite peaks at or around the time at which the L-pole coincides with the local time of the samples. Thus, on the dayside only there appears to be a 'wave' of higher density plasma rotating in unison with the L-pole. The modulation appears to be confined to winter as curves similar to Figures 2.10-2.12 for summer months showed little variation of density with universal time.

The early satellite results showing a maximum of Arctic densities near magnetic noon have been associated with the well known ground-based observations of 'UT control' (Maehlum, 1968; Pike, 1970). In the Antarctic this takes the form of a maximum in f_oF2 at all stations at 0600 UT. Duncan (1962) also reported a similar but weaker variation in the north

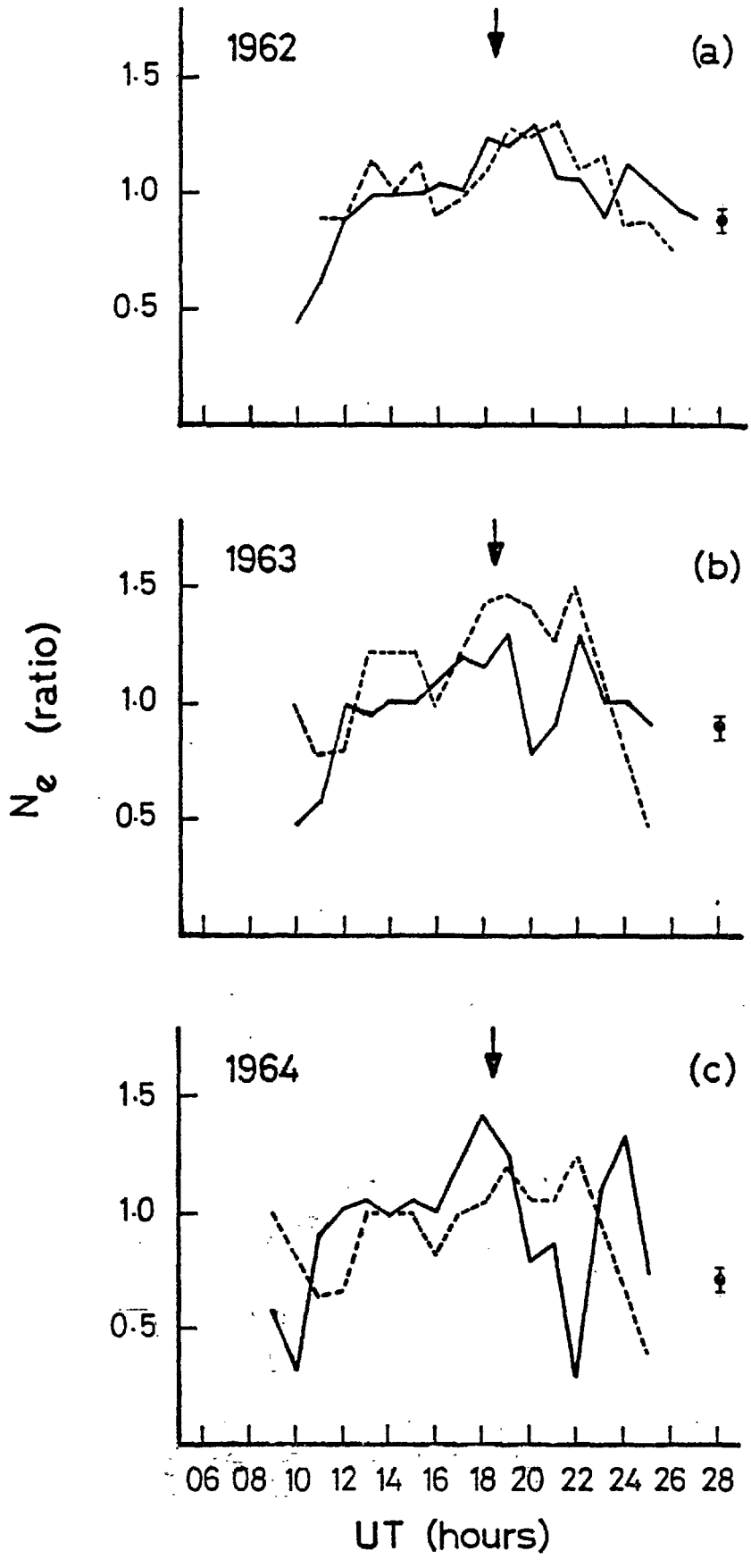


Figure 2.12 The universal time variation of electron density at a single (1300 LT) local time at various stages of the solar sunspot cycle. Error bars, a sample of which is shown on each plot, are omitted for clarity. A comparison of these results with Figure 2.10(a) indicates that the UT peak has a distinct solar cycle variation.

polar F-region near 2000 UT. Challinor (1970) has since re-examined Arctic station data and has found this not to be the case. The time of occurrence of the maximum in the f_oF2 diurnal variation is different at each station and is at times predicted by an explanation of the effect involving the neutral thermospheric wind system (King et al., 1968) These winds produce drifts of ionisation along magnetic field lines and these drifts have a vertical component. At any particular station, as the position of the magnetic pole changes so does the angle between the neutral wind and the magnetic field. The maximum upward drift occurs at a time when the wind is directed along the magnetic field and away from the magnetic pole. Vertical drifts produced in this way can act to raise the layer into a region of lower loss rate with a consequent increase in critical frequency. This mechanism, however, does not seem able to account for the Alouette I observations since it should be expected to be equally operative on the nightside and in summer but this is not the case.

Other possible interpretations of this effect should be considered. The universal time marked by the arrow in Figure 2.10 corresponds to the highest magnetic latitudes (since the L-pole is closest to the point of observation at that time). The evidence of this figure alone is consistent with simply a single enhancement of density over the pole itself, at least on the dayside of the earth. This eccentrically placed enhancement would then cause the progressive delay in the universal time of the maximum. Such an explanation is, however, inconsistent with the form of the density records along vehicle tracks, which show a definite spatial structure (Figure 2.1). Assuming the presence of this structure (i.e. a ring of enhanced density), and that the position of the pattern is magnetically controlled, density maxima as in Figure 2.10 could arise geometrically. The observations at a fixed local time means that they are made at approximately the same geographic latitude, while the geomagnetic latitude (as

a result of the geomagnetic axis eccentricity) fluctuates cyclically throughout the UT day. Density maxima might then arise as the observation point (in magnetic co-ordinates) moves cyclically across this pattern. However, data have been taken from a restricted invariant latitude range to minimise this effect. Again the effect should be seen on the nightside.

The UT maxima are, in fact, very similar to that reported in neutral densities by Reber and Hedin (1974). They have analysed OGO 6 data from both hemispheres for equinox and summer solstice conditions in 1969. At invariant latitudes above 65°, during magnetically quiet periods, persistent density enhancements in neutral N_2 and O and coincident depletions of He have been observed, which appear to co-rotate with the L-pole. Figure 2.13 illustrates this result in an identical format to that used for Alouette I electron density measurements. Here N_2 densities sampled at approximately 450 km, 1300 LT and 75° Λ' but normalised to the average N_2 density, at this invariant latitude and local time, are plotted against universal time. The result is a curve very similar, in slope and amplitude, to those of Figure 2.11, indicating possible connection between the neutral and charged particle UT peaks.

Reber and Hedin (ibid) maintain that the observed neutral particle effects are consistent with a high-latitude heat input controlled by the position of the geomagnetic pole. This mechanism seems sufficient to raise the neutral particle temperature by ∿50°K.

If a similar heating process is responsible for the UT electron density effect, an idea of the temperature enhancement necessary to produce the modulation can be easily obtained. Assuming an exponential layer consisting of O^+ ions in diffusive equilibrium, then the ratio of post- to pre-heating electron densities at 1000 km is given by:

$$\frac{N_e'(1000)}{N_e(1000)} = \frac{N_e'(300)}{N_e(300)} \times \frac{\exp(-700/H_p')}{\exp(-700/H_p)} \qquad 2.5$$

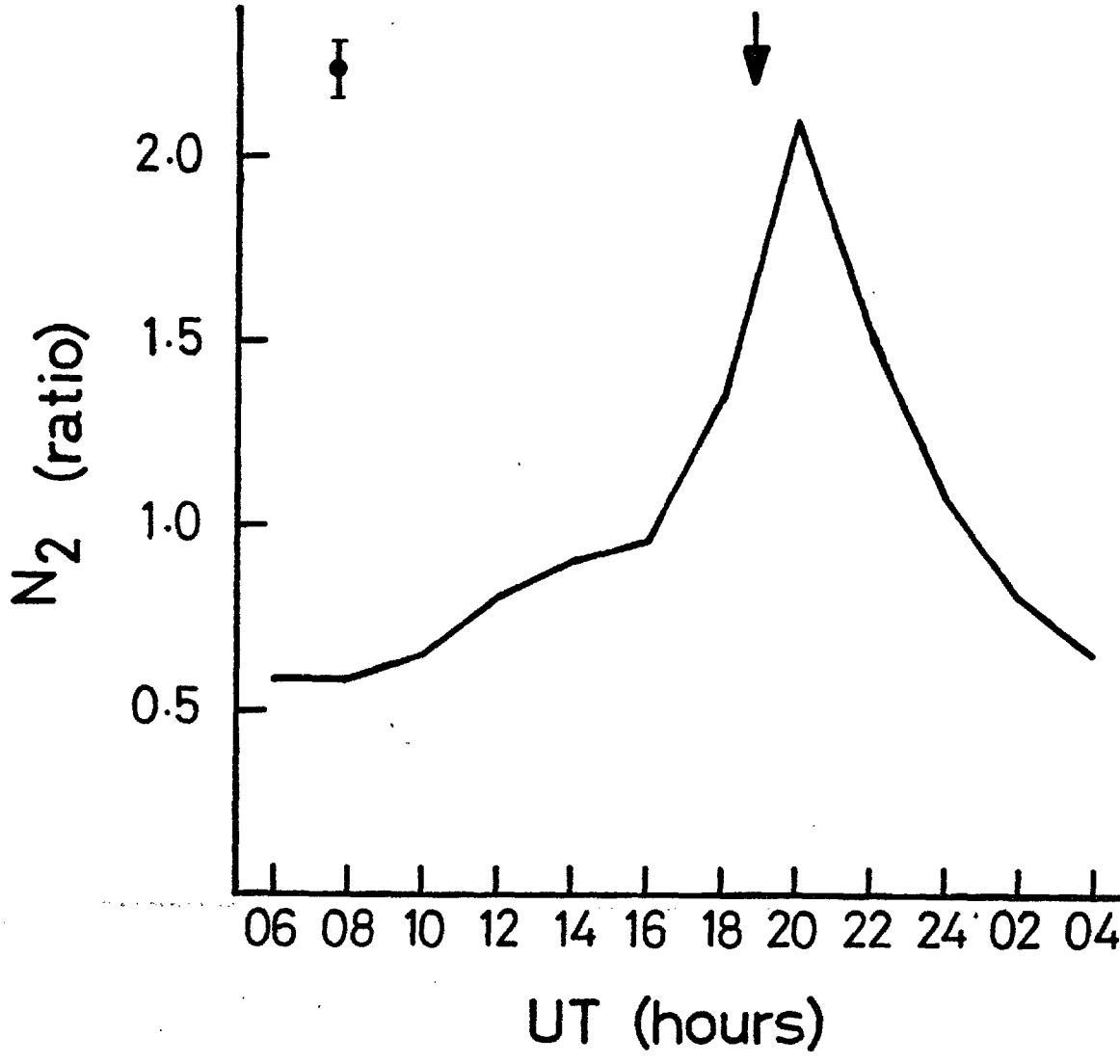


Figure 2.13 OGO 6 sunspot maximum results on the universal time modulation of neutral N_2 densities at 450 km for solstice conditions in 1968. The data have been presented in identical format to Figure 2.10. The quantity $N_{2(ratio)}$, represents the N_2 density, normalised to the average density over all universal times, at ∿75° Λ' and 1300 LT and is plotted against universal time. Error bars, of which a single example is shown, are omitted for clarity. The curve is very similar to that shown in Figure 2.10(d), with very high particle densities near the universal time at which the L-pole enters the 1300 local time sector, denoted above by a small vertical arrow.

where $H_p = \frac{k\,T_p}{m(O^+)g}$ with $T_p = T_e + T_i$ and the dashed quantities refer to the heated gas condition. For an unperturbed plasma temperature, T_p, of 3000°K and the average values of the electron density ratios, for solar maximum conditions shown in Figure 2.10, indicate that an enhancement of T_p, ∿200°K, is sufficient to cause the observed modulation. There is thus a possible close correspondence of the charged and neutral particle temperature enhancements in the UT maxima and this could again indicate a possible connection between them.

No mechanism able to account for these observations is yet known. Two known energy input processes, important at high latitudes, are heating effects of precipitating particles and enhanced joule heating due to the electric fields of the auroral electrojet. Some enhancement of one of these inputs caused by a certain orientation geomagnetic field or an interaction between the charged and neutral particle distributions seems the most likely area of investigation.

2.5 Movement of Enhancement Boundaries

2.5.1 Preamble

The variability of the polar F-region is reflected in the fluctuations of the latitudinal position of the steep electron density gradients which form the boundaries of the ring. However, knowledge of these boundaries is important in identifying the region of the magnetosphere with which they are connected by the magnetic field lines.

The Alouette data have thus been used in a statistical survey to specify the changes of boundary position as a function of local time, magnetic activity, universal time, and epoch of the solar cycle. From consideration of related observations (e.g. those on low-energy particles), the amplitude of the variation associated with each of the above parameters was assessed. Individual effects were then investigated in order of decreasing amplitude keeping all other variables constant within a small

range. To ensure availability of large data bases for determination of small amplitude effects pronounced variations were removed by application of corrections specified earlier in the analysis.

The data used were the geographic and generalised invariant latitudes of the foot of the electron density gradients forming the poleward and equatorward boundaries of the enhancements. The local and universal times, and Kp index associated with satellite sightings of these features were also noted.

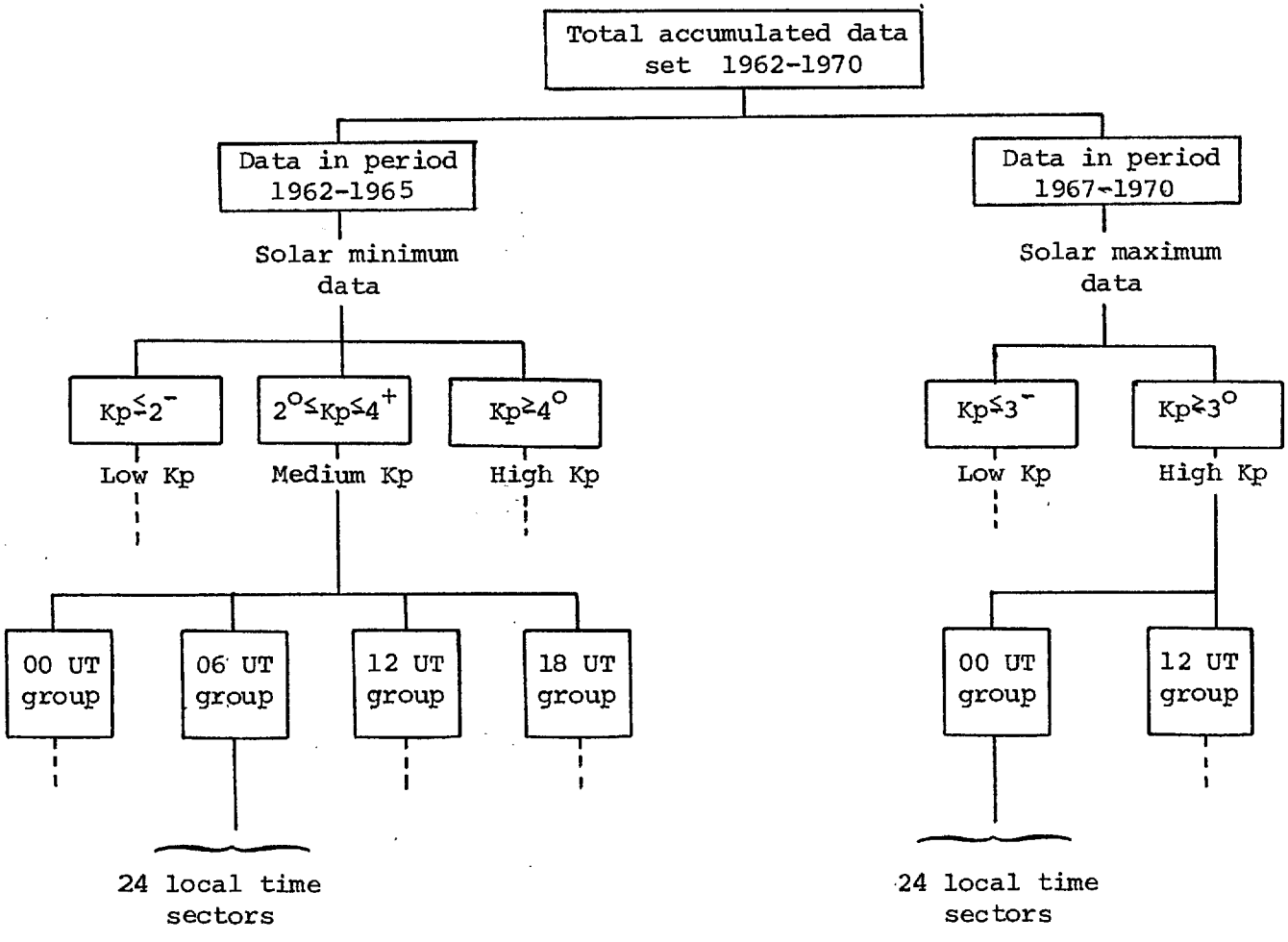
It is important to note that the term boundary as used here does not imply a contour of constant electron density but specifies the position of a steep gradient or density discontinuity. Thus on the day- and night-sides, levels of density bordering the boundary may vary considerably.

2.5.2 Local time variations

Many high latitude phenomena exhibit large local time asymmetries. In order to facilitate comparison with such observations an analysis has been performed to determine the local time behaviour of the electron density enhancement boundaries.

The modulations due to changes of magnetic activity, universal time and solar cycle were minimised by classifying the data into groups labelled by Kp index, UT and sunspot number. The latitudes of the poleward and equatorward boundaries were treated separately by the scheme outlined in Figure 2.14. Thus data from the solar minimum set could fall into a possible 288 data bins. All values appearing in each bin were then averaged to give a mean latitudinal value and standard deviation for the enhancement boundary at solar minimum, for that local time, Kp group and universal time zone. Sparse data coverage at maximum epoch of the solar sunspot cycle did not allow as close an examination of the ring boundaries and, as shown in Figure 2.14,a maximum of only 96 data bins were used to give realistic values for the mean and standard deviation of each sample.

The results of such an analysis are best summarised as smoothed plots



1. Remove solar cycle variations. Two periods sub-divided according to ambient sunspot number.

2. Remove changes due to magnetic activity.

3. Remove universal time variations. Solar minimum - 4 universal time groups centred on 00, 06, 12 and 18 UT each with a width of 6 hours. Solar maximum - 2 UT groups centred on 00 and 12 UT each with a width of 12 hours.

4. Final local time modulations. 24 hourly local time zones.

Figure 2.14

of ring boundaries on a grid of (invariant) geographic latitude - (magnetic) local time. There are 12 possible plots for the solar minimum data set and 4 for the solar maximum period each representative of a different level of magnetic activity and UT zone. If, within each Kp and UT regime, modulations due to magnetic activity and universal time are assumed small, then each plot can be regarded as an instantaneous view of the electron density distribution over the north polar region. However, since data were collected over extended periods, as is necessary to obtain a complete local time coverage from a slowly precessing satellite orbit, there is an inbuilt seasonal variation which is difficult to assess.

A selection of these plots is presented in Figures 2.15-2.17. In all cases contours representing the smoothed boundaries of the ring are presented. Some samples of the points from which such contours were derived are shown. Solid circles are used for equatorward edge points and crosses denote poleward boundary data. The coordinate system used here is generalised invariant latitude and magnetic local time, since this was found to organise the data in a better manner. This is indicative of magnetospheric control.

Similar boundary contours, drawn in generalised invariant latitude and magnetic local time, derived from electron density data at the satellite altitude, ∿1000 km, were in close agreement with those obtained near the F-layer peak. This would point to the structure being strongly field-aligned at both the high- and low-latitude edges. This extends Taylor's (1973) earlier finding for the low latitude wall. The 1000 km contours are omitted from the invariant latitude - magnetic local time grids for added clarity.

Figure 2.15 shows plots obtained at solar maximum epoch for two different Kp regimes and two ranges of universal time. Similar plots, derived from data collected at periods of low sunspot number, are shown in Figure 2.16. Here a complete range of UT zones and Kp ranges is presented. Examination of single plots shows a marked local time asymmetry in

Figure 2.15 The smoothed extent of the plasma ring near the F-layer peak, shown as shaded regions, for conditions of high sunspot number and the two levels of magnetic activity and the UT zones marked. The grids are of generalised invariant latitude and magnetic local time. The light and heavy lines denote, respectively, the plasma enhancement region as seen by Ariel 3 at 450 km, and the FLIZ phenomena of the bottomside F-region.

Figure 2.16 The smoothed extent of the plasma ring as derived from observations at periods of low sunspot number. A complete coverage of UT and magnetic activity plots is included. The upper set are for the three Kp regimes considered, at a constant UT of 12 hr. The three lower plots and the 12UT, Low Kp diagram from the upper set, form a family of UT diagrams for constant (low) Kp. The 18 UT low Kp plot is shown with a typical set of points from which the smoothed extent was derived. Points and crosses denote low and high latitude data respectively. Other details are found in the caption of Figure 2.16.

Figure 2.17 Diurnally-averaged (all UT's) plots of the extent of the plasma ring near the F-layer peak, for low/medium Kp conditions and at two stages of the solar sunspot cycle. The ring boundaries are shown as dot-dashed contours and can be compared with observations of other high-latitude parameters. Here the electric field reversal contour (solid line) and the 75% occurrence auroral oval (lightly shaded region) are given. Reference to Figure A8 (Appendix) gives the correlation between the region of low-energy electron fluxes and the area of plasma enhancements.

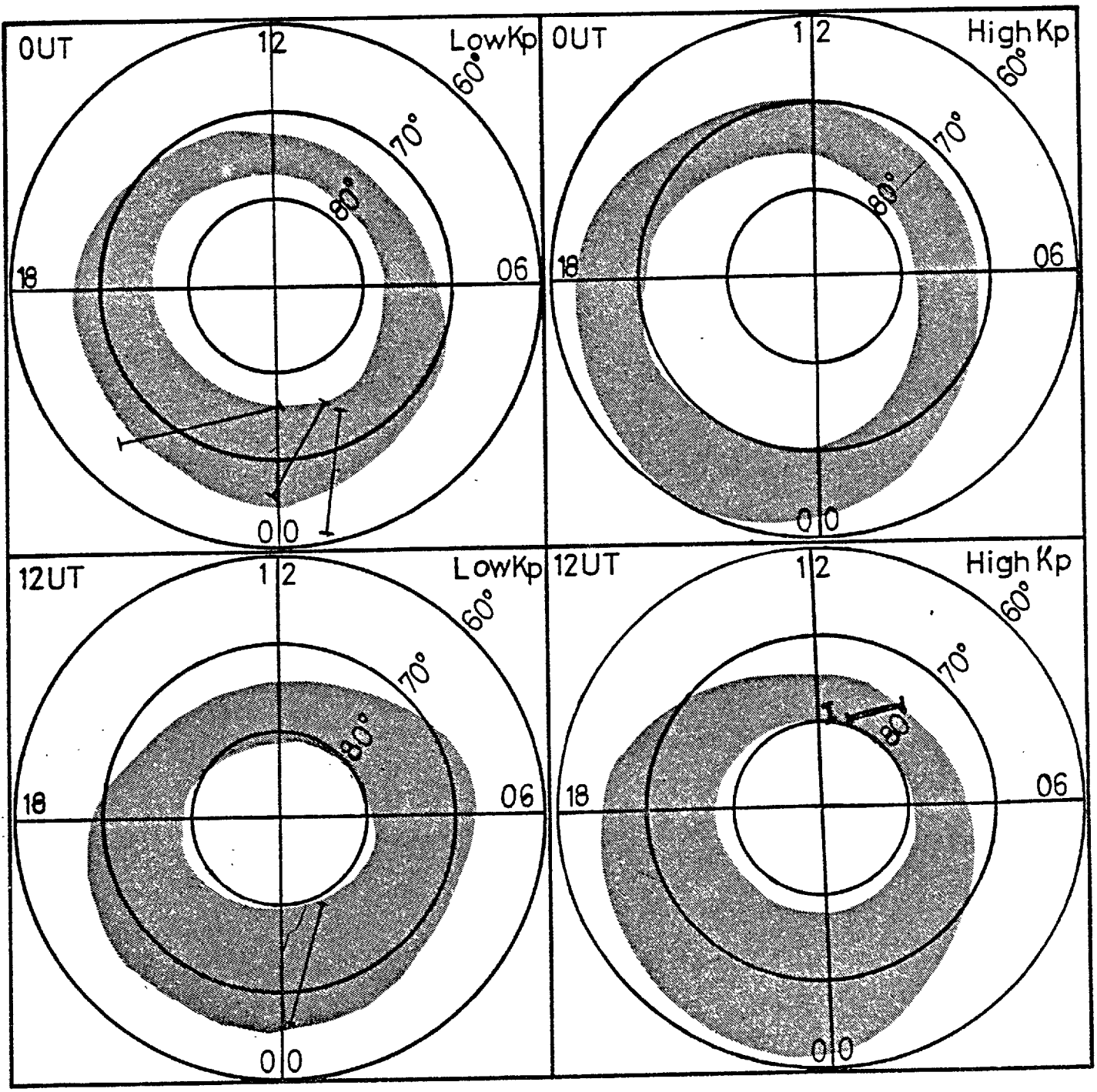
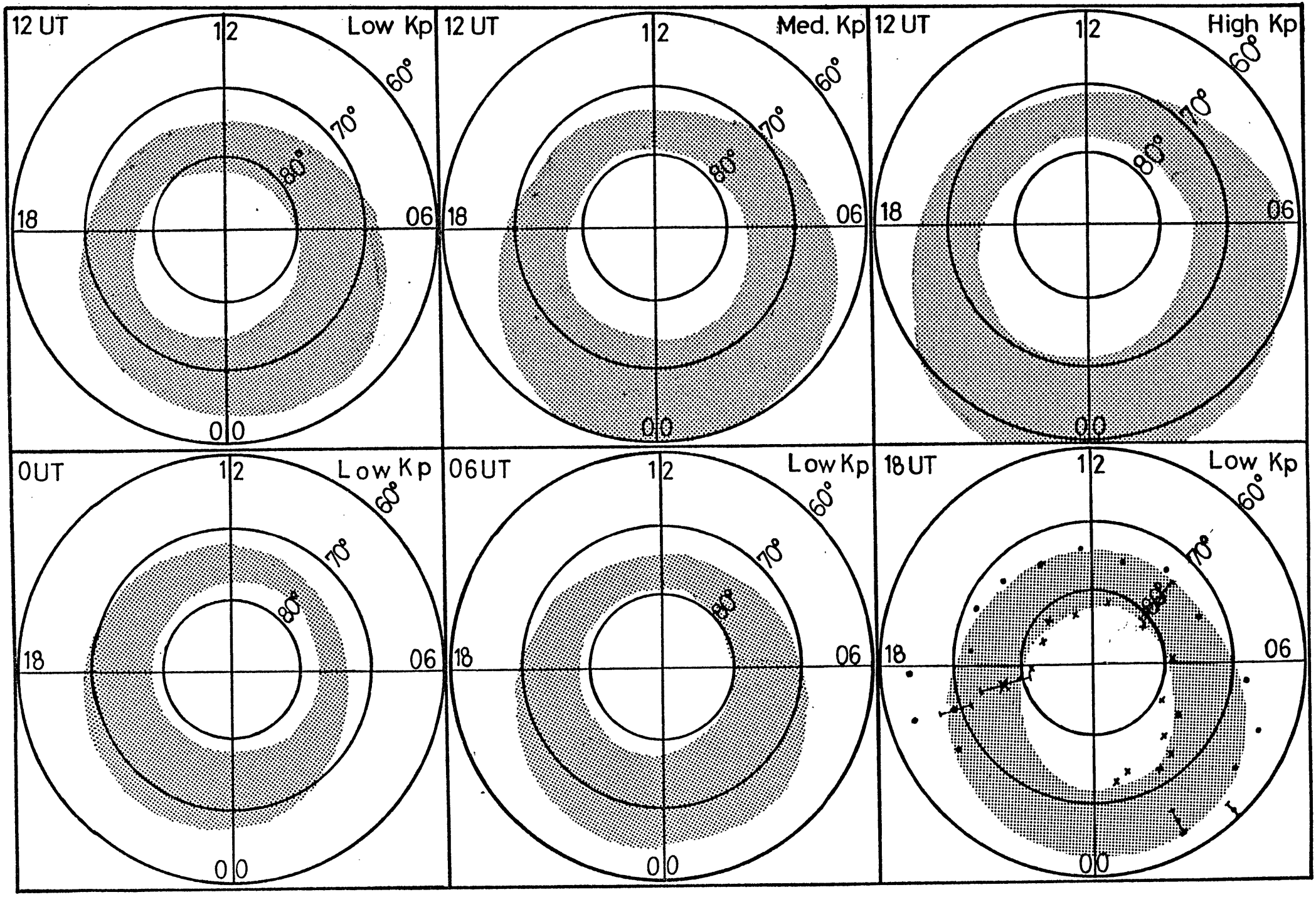

0UT
12
LowKp
60°
70°
80°
18
06
00
0UT
12
HighKp
60°
70°
80°
18
06
00
12UT
12
LowKp
60°
70°
80°
18
06
00
12UT
12
HighKp
60°
70°
80°
18
06
00

Figure 2.15

Figure 2.16


12 UT
Low Kp
12 UT
Med. Kp
12 UT
High Kp
0UT
Low Kp
06UT
Low Kp
18UT
Low Kp
12
60°
70°
80°
18
06
00

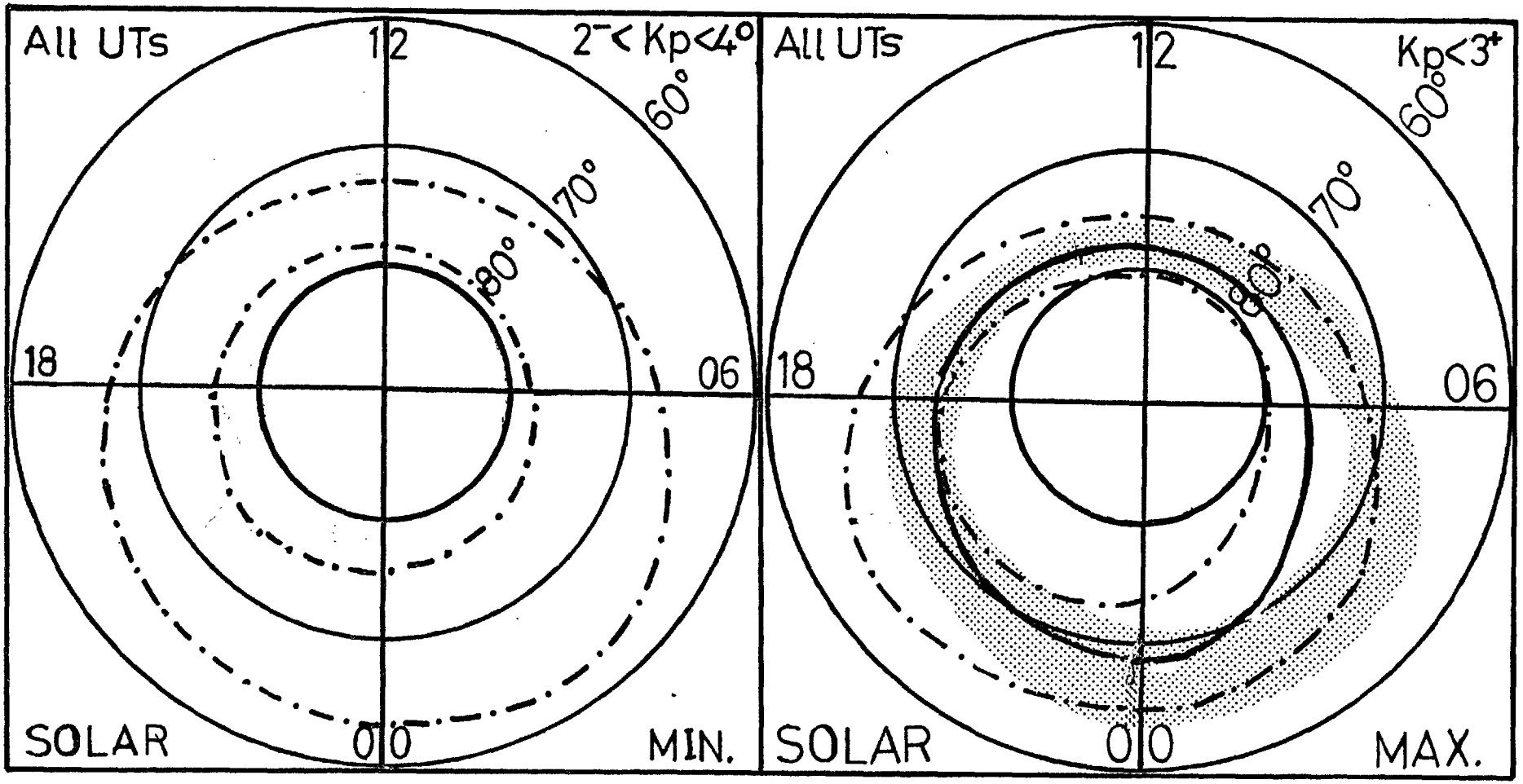

All UTs
12
2⁻ < Kp < 4°
60°
70°
80°
18
06
SOLAR
00
MIN.
All UTs
12
Kp < 3⁺
60°
70°
80°
18
06
SOLAR
00
MAX.

Figure 2.17

that ring boundaries on the dayside are found at much higher invariant latitudes than those found during night hours. This again is representative of magnetospheric control since such an asymmetry mirrors the general magnetospheric configuration with field lines compressed on the dayside and extending into the geomagnetic tail in the anti-sunward direction. On average a large local time variation of the equatorward ring boundary is apparent, amounting to ∿10° throughout the local time day. The local time response of the poleward boundary is more varied but is much less than the equatorward, the nightside location being approximately 5° equatorward of its daytime position. The result is a region of enhanced density much wider on the nightside (∿10-15°) than on the dayside (∿5°).

Also shown, on the appropriate grid, are the extent and position of various other high latitude observations thought to be related to the plasma enhancement region. These include the region of enhanced electron density, at 450 km, seen in individual passes of the Ariel 3 satellite over the north polar region in winter 1967, by Hopkins (1974). Considering the averaging involved and the different altitude at which these data were acquired the extent of the enhanced plasma region as seen by Alouette I and Ariel 3 shows remarkably good agreement.

In addition the relationship with the F-layer irregularity zone (FLIZ) of Pike (1971b), thought to be the bottomside extension of the plasma ring, can be assessed. Some of Pike's data have been rejected because they were obtained during summer. However, the locations of the FLIZ and the ring are in good agreement for winter sunspot maximum conditions, indicating that the plasma ring in the topside F-region maps into the bottomside F-layer without great modification.

To enable comparison of the ring to be made with the location of other satellite observations, diurnal average (i.e. over all universal times) boundary positions were computed. These are shown, for two extremes of sunspot number and moderate magnetic activity, in Figure 2.17.

A correlation can be made with the electric field reversal (identical with the boundary between sunward and anti-sunward convection regimes), as derived by Gurnett (1973), from INJUN V results, or the auroral oval of Akasofu (1968), which is superposed on one coordinate grid. Reference to Figure A8 (Appendix) allows consideration of the spatial and temporal coincidence of the ring and regions of precipitating fluxes. The following conclusions can be drawn from this study:

1. The electric field reversal shows close agreement with the poleward boundary especially in the pre-midnight sector. This is in agreement with Rostoker et al. (1976). In the post-midnight zone and during dayside hours the electric field reversal is located a few degrees equatorward of the ring poleward boundary.
2. The auroral oval and particle measurements show excellent agreement with the extent of the plasma ring. In nearly all cases the auroral oval lies within the ring boundaries showing the close relationship between optical and F-region plasma density data. The spatial and temporal concurrence of particle and electron density data is discussed further in Chapter 3 but a cursory examination reveals similar local time asymmetries.

An idea of response of boundary location due to changes of magnetic activity and universal time can be obtained from a study of plots of the type contained in Figure 2.16. By following a series of patterns at constant universal time through increasing Kp regimes it is clear that both boundaries move equatorward; an identical response to the soft precipitation region. Similarly UT behaviour can be inferred. However, these contours have been obtained by an averaging process and more accurate analysis is necessary to reveal the effect of these two parameters on the ring boundaries.

2.5.3 Response to changes in magnetic activity

It is now well-established that the equatorward boundary of the top-side plasma ring moves to lower latitudes in response to increasing geomagnetic activity (Andrews, 1969). Pike (1972b) has observed similar relationship for the FLIZ on the dayside. However, no data on the poleward boundary has yet been presented. Both Andrews and Pike found its location to be so variable that their small data sets were unable to reveal statistically significant trends. However, some systematic changes have been noted in the present large survey and it is now possible to give a comprehensive account of the location of the ring under differing magnetic disturbance conditions.

Only winter data from 1962, 1963 and 1964 were used and these were first sorted into two-hourly local time bins to remove LT variations. For all local time groups linear and polynomial regression equations were computed using a least-squares fit criterion for both equatorward and poleward boundaries. Regression lines were fitted to data sets further classified into three-hourly universal time zones.

Sample computational results are summarised in Figure 2.18. In all cases no improvement of fit to the points was obtained by the use of a polynomial relation. In Figure 2.18 the value of the coefficient a_1 in the linear regression equation $\Lambda' = a_o - a_1 \times Kp$ for both the poleward and equatorward ring boundary is plotted as a function of local time. In most cases the points plotted have been calculated from over 100 observations and the correlations for both poleward and equatorward boundaries were very significant.

The results indicate that in all local time sectors both enhancement boundaries move to lower latitudes with increasing Kp but, as is clear from Figure 2.18, there is a different response on the day- and nightsides of the earth. On the dayside the plasma ring appears to contract slightly since the poleward boundary moves farther towards the equator with increasing Kp than does the low-latitude boundary. However, during night-time

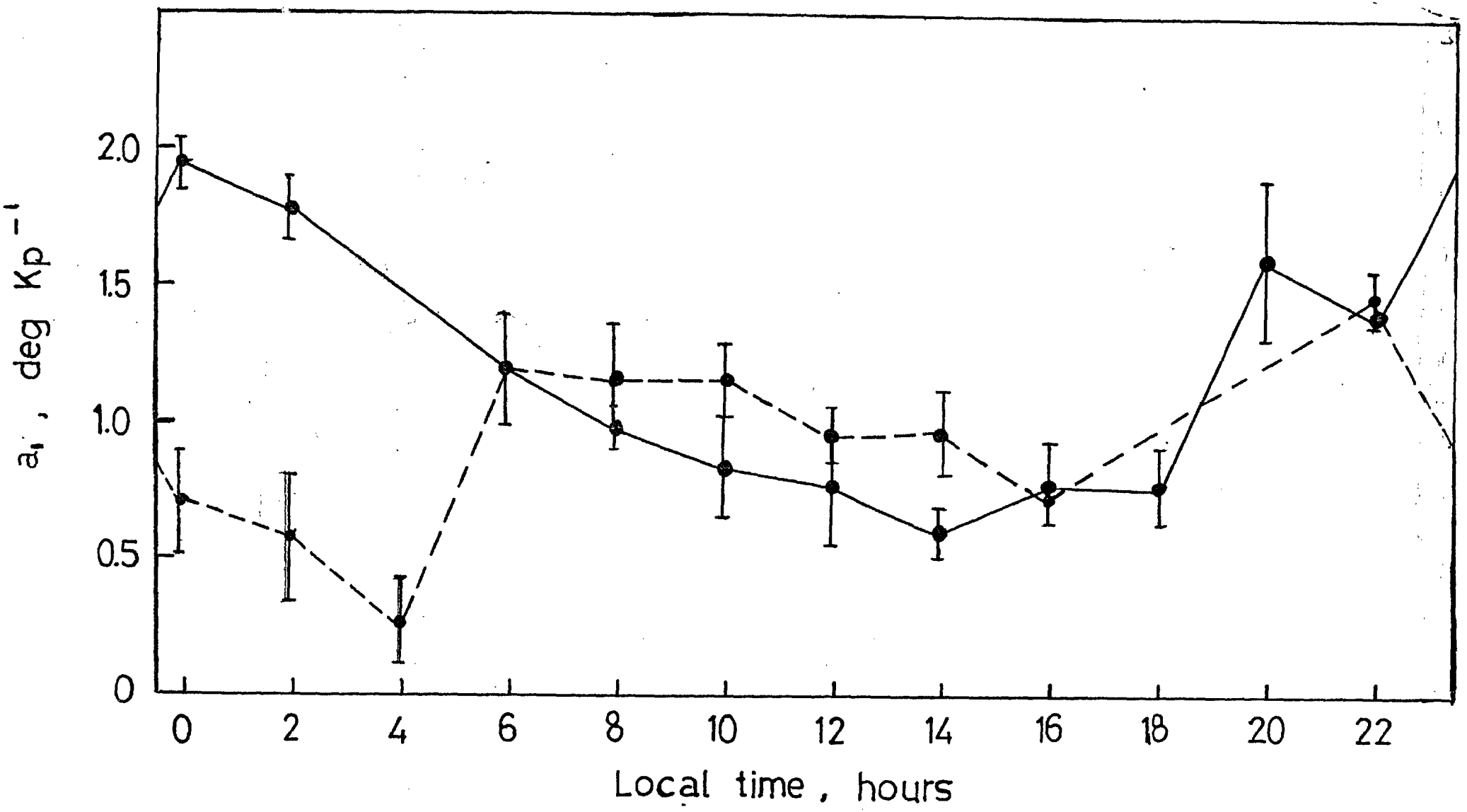


Figure 2.18 The response of the plasma ring near the F-layer peak and at low sunspot number to changes of magnetic activity. The satellite observations of the latitudinal position of the low- and high- latitude boundaries of the plasma ring have been subject to a regression analysis, and the results of this study are shown here. Plotted as ordinate is the coefficient, a_1, in the equation $\Lambda' = a_0 - a_1 \times Kp$, for the ring boundary latitude, which determines the level of response to changes of Kp index. The abscissa is the local time of the sample. A continuous line joins equatorward boundary points while a broken line is used to join data concerned with the high-latitude ring edge.

the situation is reversed, and there is large latitudinal expansion of the region of enhanced plasma.

A possible explanation of this effect is the connection of the Kp index with the strength of the auroral electrojet. It is well known that low-energy electron precipitation regions, such as the cleft on the dayside and the nightside high-latitude zone, move equatorward with increasing Kp index (Burch, 1969; Hoffman, 1972). If, as suggested in Chapter 3, the density enhancements are due to the collisional ionisation effects of these particles, then the plasma ring would follow the precipitation regions to lower latitudes. This occurs on the dayside where the poleward or high-latitude boundary appears more responsive to magnetospheric control. An analogous weak effect is observed on the nightside but, as noted above, this is accompanied by an expansion of the ring not observed during the day. This expansion is probably caused by the auroral electrojet which, during times of high Kp, is enhanced and dissipates large amounts of energy in the nightside ionosphere and neutral atmosphere. This energy is in the form of joule heating and thus one would expect an increase of Kp to be accompanied by a horizontal and vertical expansion of the charged and neutral density distributions.

2.5.4 Universal time (UT) modulations

TA's early model of the topside polar plasma distribution (Section 1.1.3) comprised a ring of enhanced electron density eccentrically located about the geographic pole. To a first order this was considered to be stationary with respect to the sun and the earth rotated beneath it. UT or longitudinal variations were introduced only as small unspecified perturbations on the position of the ring to explain the transition between double- and single-peaked patterns of electron density observed in satellite pass records. This section will be concerned with an examination of these perturbations.

Thomas et al. (1966) first noted that ionisation troughs, at a par-

ticular local time, were constrained to move along a contour of generalised invariant latitude as the earth rotated, beneath them. This and the evidence of the constant UT plots, showing that, approximately, the ring similarly aligns along L-shells, indicate a possible meridional oscillation of the boundaries as the magnetic pole moves eccentrically about the earth's rotational axis.

This effect would be most pronounced in geographic coordinates and consequently the plots in Figure 2.16 have been redrawn on a grid of geographic latitude and local solar time. The sample curves in Figure 2.19 were derived from these new diagrams and show, at the local times marked, the latitudinal modulation of the equatorward and poleward edges of the ring with universal time. Despite the averaging involved the effect is repeatable and very clear. It consists of a meridional oscillation of amplitude as large as 20° for the low-latitude edge and 10-15° for the poleward boundary. The phase of this oscillation is strongly controlled by the position of the L-pole. Thus in all cases investigated the most equatorward position of the boundaries occurred at the universal time at which the L-pole reached the local time of the observations. The most poleward position of these boundaries was found $\sim$12 hours later when the L-pole was on the opposite side of the geographic pole.

This oscillation can be pictured as a 'wobble' of the plasma ring from its fixed position with respect to the earth-sun line, in unison with the motion of the L-pole.

It is clear from the constant UT plots that the ring boundaries do not follow the motion of the L-pole exactly since at a given local time they are found at different invariant latitudes at different times in the UT day. This could be due to many reasons, only one of which will be treated here in any detail.

Hoffman (1972) and Burch (1973) have examined the position of the equatorward boundary of the soft electron precipitation zone in invariant latitude, as a function of the angle between the magnetic axis

and the earth-sun line, β. After elimination of substorm effects Burch has found a 4° shift in the position of the dayside soft precipitation in response to the annual variation of tilt angles. This change occurs such that for winter and summer conditions the invariant latitude of the boundary shifts by 2-3° and +1° respectively from its equinox location. The maximum universal time dependence was consequently observed at the winter solstice and amounted to only 2.5°.

The identical relationship has been investigated for the plasma ring. For two-hourly time sectors the invariant latitudes of the high- and low-latitude boundaries of the ring were noted along with the ambient Kp, universal time and date of observation. Data with Kp > 2^{o} were then discarded to avoid inclusion of substorm effects. The remaining boundaries were normalised to a Kp of 1^{0} by use of the relations derived in Section 2.5.2. The universal time and day of observation were then used to calculate the dipole tilt angle, β. The method used is illustrated in Figure 2.20. This shows the situation at magnetic noon at the winter solstice. It is assumed that throughout the UT day, therefore, β fluctuates by 22° as the geomagnetic axis moves about the rotation axis and there is an annual variation of 47° as the rotation axis oscillates about the ecliptic plane. The diurnal and annual modulations were represented by

$$\delta D = 11 \sin\left(\frac{\pi}{12}\{UT + .67\}\right) \qquad 2.6$$

$$\delta A = 23.5 \cos\left(\frac{2\pi}{365} t\right) \qquad 2.7$$

where UT is the universal time in hours and t is the time in days measured from the winter solstice. δA and δD are measured in degrees.

The expression for β, in degrees, is therefore

$$\beta = 90 + \delta D + \delta A \qquad 2.8$$

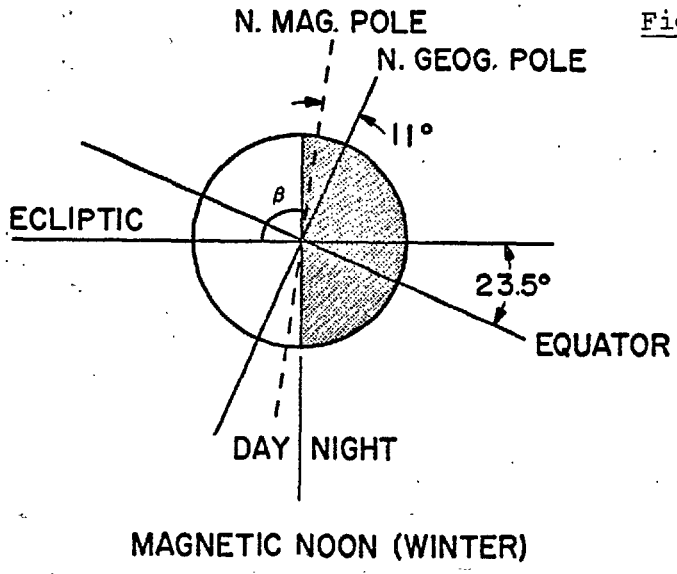


Figure 2.20 The tilt angle of the geomagnetic axis with respect to the earth-sun line, β. The diurnal variation is due to the eccentric rotation of the magnetic pole about the geomagnetic axis. A further modulation is introduced by the annual oscillation of the equator about the ecliptic.

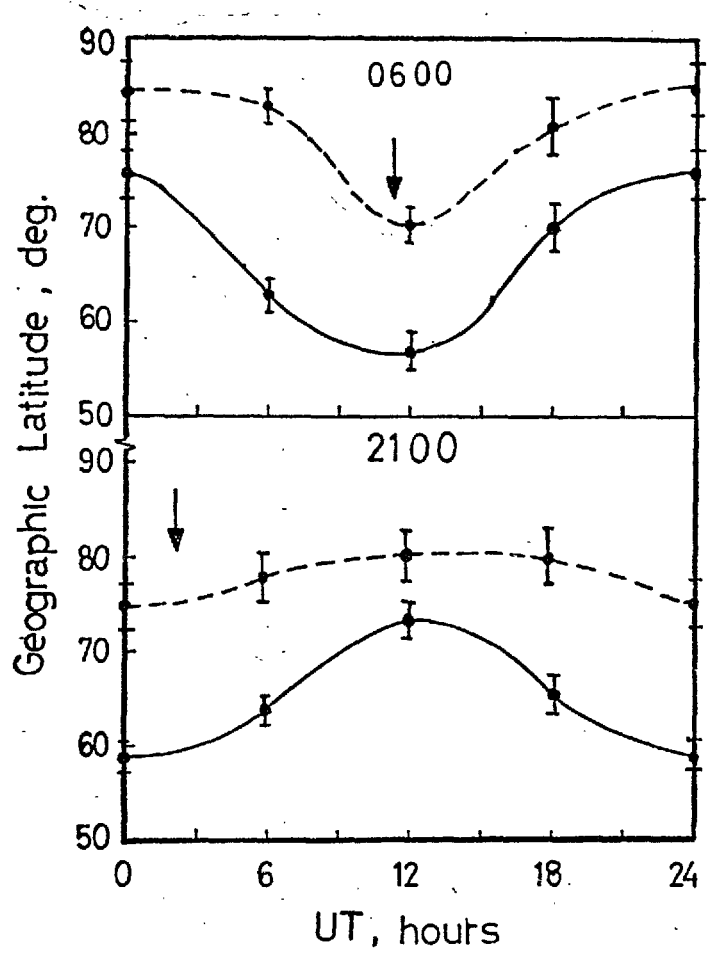


Figure 2.19 The meridional oscillation of the position of the plasma ring boundaries in geographic latitude with universal time. The upper and lower diagrams refer to the two different local times; 0600 and 2100. Points were derived from smoothed ring boundaries plotted on a grid of geographic latitude and local solar time. The size of error bars was obtained from the degree of smoothing applied at that local time and universal time. Continuous and broken guidelines have been drawn through the points referring to the low- and high-latitude boundaries respectively and the arrow marks the universal time at which the local time at the L-pole coincides with that of the sample.

The results of a statistical analysis of data close to the winter solstice showed that the ring behaved in a similar way to the soft precipitation zone. An increase of β was accompanied by an equatorward motion of the ring and the diurnal variation of the position of the ring was found to be ∿2°.

Despite the close agreement of this result with that of Burch (1973), for cleft electrons, the correlations obtained were not very significant and other possible explanations of the modulation of the invariant latitude ring boundaries with UT must be sought. Andrews (1969) has argued that contours of invariant latitude are calculated from diurnally-averaged field models. The effects of field distortions are not considered. Small UT effects, dependent on the field, are therefore likely to be insignificant when compared to these short term fluctuations.

2.5.5 Long-term modulations

The results of section 2.4.1 concerning the level of electron densities over the solar sunspot cycle prompt the question: Are there similar modulations of the boundary positions of the plasma ring? This has been tested over the period 1962-69 at altitudes near the peak of the F-layer. As for previous analyses the data were first organised into two-hourly local time sectors. Using the results of sections 2.5.3 and 2.5.4 the latitudinal positions of the ring walls were then corrected for Kp and dipole tilt angle (β) variations and solstice average locations obtained. These averaged, corrected values were plotted as a function of year to assess the solar cycle influence.

The points exhibited large scatter but of the 5 local time sectors considered all the poleward and 4 equatorward boundaries were found to move to slightly lower latitudes from 1962 to 1969. Two of the most regular examples are presented as Figure 2.21. Reference to Figure 2.5 indicates that the modulations appear closely anti-correlated with the smoothed Zurich Relative Sunspot Number, $\bar{R}$. Thus in addition to solar

cycle control of electron densities, there is also a small solar cycle effect on the position of the high-latitude electron density gradients.

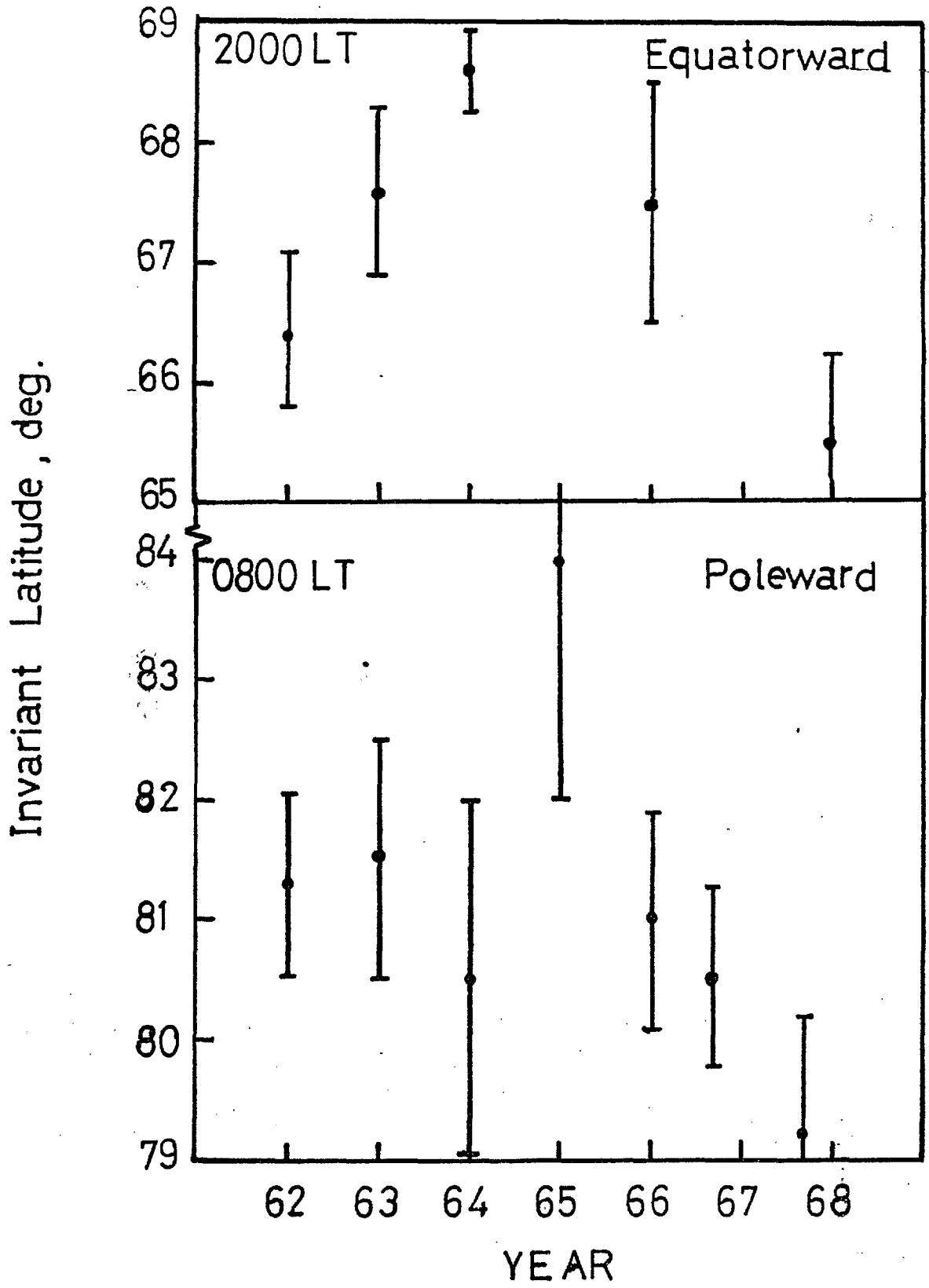


Figure 2.21 The long-term variation of the plasma ring boundaries. The most regular examples of the high- and low-latitude boundaries are shown for the local time sectors given.

2.6 Discussion

2.6.1 The Solar-Terrestrial relationship

The foregoing survey has established that there are substantial solar cycle variations of the polar topside ionosphere. It has been shown that plasma ring electron densities increase in unison with the smoothed Zurich Relative Sunspot Number by a factor comparable with that at midlatitudes while electron densities in other high latitude regions show a much larger variation. In addition, the boundaries of the plasma ring show systematic changes and the 'UT peak' of densities increases from solar minimum to maximum. Other geophysical phenomena which exhibit analogous fluctuations include the frequency of the visual aurora (Fälthammer, 1972) and the level of geomagnetic activity. The process responsible for these changes, as with many other aspects of the Solar-Terrestrial relationship, is not understood.

At midlatitudes the rise in electron density is presumably due to increased production of ionisation by higher solar photon fluxes. Eclipse observation of ultra-violet radiations and electron density indicate that this assumption is justified. That solar ionising emissions show this trend is now well established. Hall et al.(1969) and Ivanov-Kholodnyi et al. (1972) have reported measurements from various wave bands of the solar spectrum. In the wavelength range 1310-270 Å, which contains most radiations responsible for the formation of the F2 layer, the integrated flux changes by a factor of $\sim$1.4-1.5 for a change in the 10.7 cm solar radio flux (a good indicator of solar sunspot activity, Bauer , 1973) of 2.

However, at high latitudes, especially during winter, insolation from the sun at F-region altitudes is very weak and - would not be expected to have the large effects observed. There are also no known ways in which the ultra-violet radiations could affect plasma boundaries, or cause modulations of the UT peak and the possibility of other mechanisms

by which the solar cycle affects the high-latitude terrestrial ionosphere must be considered.

It is to be shown (Chapter 3) that fluxes of low-energy particles form an important source of ionisation in the polar region, especially in the plasma ring. Although such modulations have not yet been observed, it is possible that changes of low-energy flux properties are responsible for solar cycle effects on electron densities. Sufficient data coverage must now be available to allow an analysis to search for these changes. However, the high temporal and spatial variability of such fluxes would make isolation of solar cycle effects difficult. Ultimately the source of these particles is the solar wind, which has direct entry to low altitudes on the dayside through the cleft region (Heikkila and Winningham, 1971). Particles incident on the nightside atmosphere are thought to have come from the solar wind by way of the plasma sheet where they are subject to some acceleration process. Thus fluxes of low-energy particles could be modified by changes of the source, viz the solar wind, or parameters specifying this acceleration. The magnetic field geometry could also be a participant in this modulation by allowing entry of a greater number of particles to low altitudes or by enhancement of magnetospheric-solar wind interaction.

2.6.2 Solar cycle variations of the solar wind

Since the solar wind with its magnetic field is the principal link between the Sun and the geomagnetic field, on a simple view we might expect solar cycle variations of solar wind properties analogous to those found in solar activity and in geophysical parameters.

After Wolfe's (1972) unsuccessful search for a change in solar wind conditions over the solar cycle, Neugebauer (1975), pointed out the need to consider parameters averaged over periods of six months to remove solar latitude variations. With such corrections, an analysis of data from many spacecraft has been repeated to search for changes in the solar

wind.

Various investigations using many different solar wind velocity parameters, such as the yearly average or the six-monthly average of the fractional time the velocity was greater than 500 km s^{-1}, have failed to find any correlation with solar cycle. In fact, fluctuations appear near random. Intrilligator (1974), however, has found a correlation between solar-wind stream structure and solar activity. This is shown in Figure 2.22 in which the number of high-velocity streams is shown for each year of the solar cycle. On a simple interpretation, streams may be thought as consisting of sudden enhancements of solar-wind flow velocity and are usually accompanied by an increase in solar-wind density. So by modification of the interaction between the solar wind and the magnetosphere through changes in the magnetic field or particle fluxes, the stream structure could have important effects on geophysical processes.

Solar-wind density and composition could have a geophysical influence. Diodato et al. (1974) have clearly demonstrated a 50% modulation of proton density anti-correlated with solar activity. Also Neugebauer (ibid) has reported a strong solar cycle dependence of six monthly averages of the ratio of number densities of helium to hydrogen in the solar wind. Unfortunately there are no satisfactory theories to suggest how these properties affect geophysical behaviour.

2.6.3 The interplanetary magnetic field and the geophysical response

Figure 2.23 shows the variation of the interplanetary magnetic field over the solar cycle. Here the mean value, modal* value of field magnitude, and the fraction of time the field was greater than 10 γ, all show solar cycle variations. Schatten (1971) has indicated that the size of this modulation in the mean and modal values is as large as 50% and arises from the greater frequency of occurrence of higher field magnitudes.

*The modal value is the most probable measurement of the magnetic field.

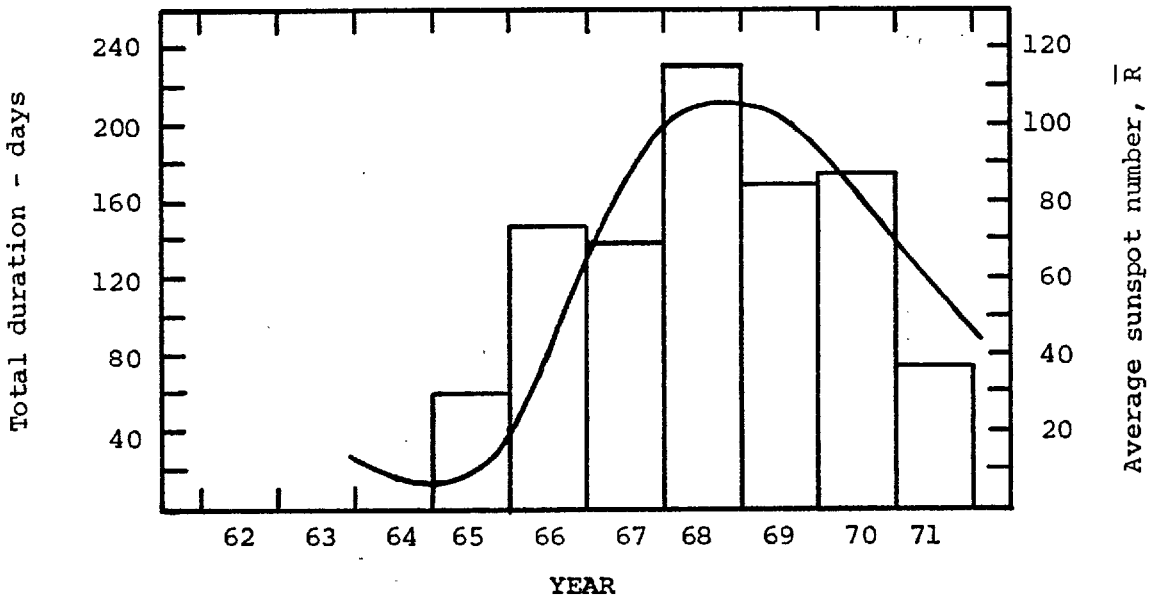


Figure 2.22 The number of days high speed streams were observed in the solar wind for each year in the period from 1965 to 1969. The data are taken from Intrilligator (1974) and were obtained from spacecraft observations. Also shown is the smoothed Relative Zurich Sunspot Number, $\overline{R}$, for the same period.

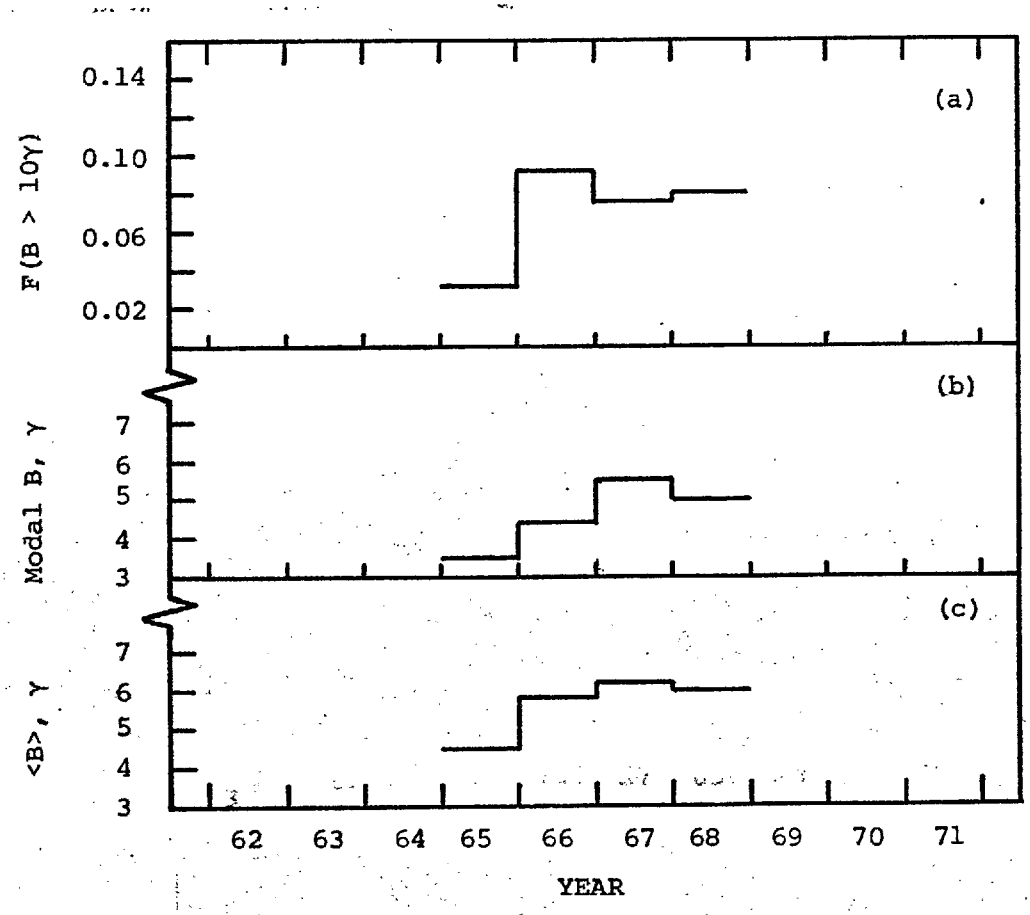


Figure 2.23 The variation of the fractional time the interplanetary field was greater than 10γ (a), the modal magnitude (b) and the mean value of the interplanetary magnetic field (c) from 1964 to 1969. The solid lines are annual averages taken from Schatten (1971).
($1\gamma = 10^{-5}$ G)

However, similar magnetic field measurements during 1968-1972, the falling part of the cycle, showed no significant variation of interplanetary magnetic field magnitude. (Hedgecock, private communication, 1977). The only obvious trend over this period was a decrease in the occurrence frequency of high magnetic field values, probably related to the number of observed high speed streams.

Since the interplanetary magnetic field is 'frozen-in' to the solar wind, the field lines move with the plasma and, ideally, take an Archimedean spiral configuration due to the combined solar rotation and radial solar wind expansion velocities. The field can be directed toward or away from the sun and is usually ordered into sectors in which the average field direction is the same for several consecutive days. There is then an abrupt transition and, for the ensuing period, the field is in the opposite sense. The field, divided thus into four sectors, exhibits a solar latitude dependence. During solar cycle 20 this was observed as a transition from predominantly solar-pointing to anti-solar-directed fields, as the earth passed across the ecliptic from northern to southern heliolatitudes. Recent observations suggest that this latitude variation was not visible near solar maximum, (after 1967), but it reappeared in 1971 with opposite phase.

The relationship between magnetospheric phenomena and the state of the interplanetary magnetic field has been intensively studied over the last decade. Much of this work has been concerned with the response of the earth's plasma envelope to changes in direction of the individual magnetic field components. Probably due to the difficult nature of the problem, no correlation between interplanetary field strength and magnetospheric parameters has been attempted.

Burch (1973) has reviewed present knowledge regarding the north-south or Z-component (in the solar-magnetospheric coordinate system) of the field. Whilst this has a positive or northward component conditions in the magnetosphere seem quiet. However, a transition to a south-

pointing field appears to trigger the growth of the magnetospheric substorm phenomena and an increase in the number of geomagnetic field lines reconnecting to the interplanetary field. One immediate effect is an enhancement of energy transfer from the solar wind to the magnetospheric plasma, which results in an intensification of plasma convection and electric fields in the polar cap. At low altitudes the soft dayside precipitation zone, associated with the cleft and the bottomside F-layer irregularity zone, move sharply equatorward (Pike, 1972b).

A reversal of the azimuthal or Y-component of the interplanetary magnetic field occurs as the earth crosses a sector boundary between regions of oppositely-directed field.

Known terrestrial response to these transitions has been summarised by Svalgaard(1973). Satellite observations have revealed that the azimuthal component has a distinct influence on the distribution of polar cap convection and electric fields. Thus when the Y-component is positive eastward, the convection pattern is shifted towards dawn where maximum convection velocities are found. The reverse occurs for a positive westward azimuthal component. In summary, it would seem that the north-south component of the interplanetary magnetic field determines the level of convection over the polar cap, while the east-west component controls the asymmetry of the convection pattern with respect to the noon-midnight meridian.

In the polar topside ionosphere Titheridge (1974) and Shepherd et al. (1976) have tentatively identified sector-structure-dependent behaviour. Examining the heating effects of low-energy particles entering the cleft, Titheridge found that the position and width of the high-temperature region depended on whether the earth was in an 'away' or a 'toward' sector. Similarly, Shepherd et al., after analysis of optical and particle data from five ISIS 2 winter passes through the cleft region, concluded that differing characteristics of precipitating low-energy electrons and ambient thermal density distribution were due to transitions between sectors.

Thus there is some evidence to suggest that ionospheric parameters have some dependence on the state of the interplanetary magnetic field.

2.6.4 Conclusions

Through a large scale analysis of topside-sounding satellite electron density data it has proved possible to follow the development of the charged particle distribution over the rising portion of the solar cycle. The work of Thomas and Andrews (1969) has been extended and the trough-ring-cavity arrangement of the topside plasma has been shown to persist and dominate at all sunspot numbers. The survey has also provided new information on long- and short-term modulations of this infra-structure.

The close agreement between the spatial and temporal extent of the plasma ring at 300 and 1000 km show that this feature is strongly field-aligned. It appears as a continuous band of enhanced plasma, eccentrically placed around the geomagnetic pole. Both the influence of solar produced photoionisation and low-energy electrons are evident. The low-energy electron precipitation zone and the ring show close agreement in invariant latitude and local time, indicating a possible connection between them. There is also a correlation of the decrease of observational frequency of the plasma enhancement in the post-noon local time sector with the distinct hardening of the cleft electron energy spectrum in the afternoon hours (McDiarmid et al, 1975). However, the local time variation of densities is similar to that seen at mid-latitude ground stations, taking the form of a steady increase during local day to a post-noon maximum. Thus there is evidence for both photo- and particle-ionisation processes in the plasma enhancement region.

Further evidence of particle involvement is found in the behaviour of the ring boundaries with magnetic activity and dipole tilt angle. On the dayside the ring and soft-electron precipitation zone simultaneously move equatorward with increasing Kp With dipole tilt, a similar observation to that for particle fluxes (Burch, 1972) is reported for the ring

region. This takes the form of a weak dependence of the position of the enhancement region on the angle between the geomagnetic axis and the earth-sun line. Nighttime behaviour is more complex since here substorm effects are more pronounced and the ionosphere acts as a sink for energy stored in the magnetosphere. Again the response of the low-energy particles and the ring (Winningham et al., 1975) to increase of magnetic activity are similar. The large expansion of the ring is not reflected by the electron precipitation zones, however, and this could be part of the ionospheric effect of dissipation of substorm energy.

Thomas and Andrews' (ibid) original model viewed the high-latitude electron distribution as a ring of high density plasma, surrounding the geomagnetic pole, and stationary with respect to the earth-sun line. The earth, rotating beneath this pattern, had no effect. This must now be modified to include the precessional motion of the structure revealed by universal time variations of ring boundaries. As the magnetic pole rotates eccentrically about the geographic pole, a ring boundary oscillates along a line of constant local time, with an amplitude of $\sim$10°.. The boundary reaches its lowest latitude when the magnetic pole enters the local time sector concerned (Figure 2.11). Overall, this motion takes the form of a 'wobble' or precession of the ring about the geographic pole.

Electron densities show an analogous dependence on the rotating magnetic pole. Sampling data within the ring at a constant local time and as a function of universal time, average electron densities are shown to be enhanced when the local time of the L-pole and the sample coincide. This modulation cannot be explained by any geometric effect and is unlike any universal-time-controlled neutral air wind phenomena observed at high-latitude ground stations (King et al., 1967). After analysis of OGO 6 neutral composition data, Reber and Hedin (1974) have interpreted identical polar enhancements of O and N_2 concentrations as a source of heating, controlled by the geomagnetic pole. Here, it is found that similar magnitudes of

temperature enhancement can account for the observed changes of electron density, and, although processes involved are unclear, this remains a plausible explanation.

The extended period of observation has allowed an investigation of the response of the polar topside ionosphere to the solar cycle. Briefly two main effects have been identified:

1. Trough, ring and cavity densities at all altitudes, local times and levels of magnetic activity increased in unison with the sunspot number.
2. The ring shifted to lower latitudes with increasing phase of the solar cycle implying a slight expansion of the polar cavity.

The decrease in observational frequency of the ring over the period 1962-1968 was due to these modulations. Although all densities increased, there was proportionately a larger effect in the trough and cavity. Thus the difference between densities in the ring and those in the trough and cavity decreased over the solar cycle and enhancements at solar maximum would be less likely to fulfil the selection criteria of section 2.2.

The cause of the solar cycle changes is not known. Similar processes to those at mid-latitudes, where a smaller functional dependence of electron density on sunspot number has been reported, seem inadequate. Also temperature enhancements at solar maximum, which could cause an increase of densities, have not been observed (Miller, 1970). Other aspects of the solar-terrestrial relationship have been examined in an attempt to isolate the source of these modulations. Of solar wind parameters investigated, the interplanetary magnetic field (IMF), which has been identified as the source of polar cap magnetic field and ionospheric fluctuations, shows no clear trend with solar cycle.

It is now well-established that with increasing sunspot number energy transfer from the solar wind is increased, resulting in enhanced energy density, tail field and convection velocity. The impor-

tance of convection is suggested by the large response (compared with that of the ring) of the trough and cavity, where it is considered to be the dominant process (Schunk et al., 1976). Thus, on a simple view, with the swifter convection velocities at solar maximum, field tubes, constrained to convect across the polar cavity, would spend a shorter time in regions where ionisation production is small. Thus there would be consequent smaller loss of ionisation and increased densities in the trough and cavity. Expansion of the polar cavity would support this argument for intensified convection. Density variation in the ring present no problem since they are of the same order as those observed at mid-latitudes.

2.6.5 Future work

Rostoker et al. (1976) have recently suggested that the topside plasma distribution can be used to monitor the level of convection within the polar cap and hence the strength of the solar-terrestrial interaction. They have correlated the position of the poleward edge of the ring, in the morning and evening local time sectors, with the boundary of the auroral electrojet. (Coincident with the transition from sunward to anti-sunward convection regimes.) Results presented here suggest that their proposal may be justified during evening hours, but on the dayside the boundary of the electrojet is found at consistently lower latitudes than that of the plasma ring. Thus although care must be taken, the tentative suggestion, that conditions in the topside F-region can be used to infer magnetospheric conditions and the strength of the solar-terrestrial interaction, still remains.

The present survey has been used to specify the variation of the topside ionosphere with the sunspot number - a coarse measure of interplanetary conditions. An extension of this analysis, to correlate the fluctuations of ionospheric parameters with individual solar wind properties (e.g. individual components of the interplanetary magnetic field) and

magnetospheric measurements, would form a useful part of an investigation into the ability of the ionosphere to respond to changes of interplanetary and magnetospheric conditions. It would also represent a rigorous test of ionospheric effects of IMF polarity recently seen by Titheridge and Shepherd et al.

CHAPTER 3

EFFECTS OF PRECIPITATING PARTICLES

3.1 Introduction

The ionosphere at high latitudes often exhibits very different characteristics from that found in mid- and low-latitude regions (Thomas and Sader, 1964). On ionograms, this transition is often seen to occur in the region of the low latitude wall of the trough, the ionospheric extension of the plasmapause (Andrews, 1969). At high latitudes the ionosphere seems much more irregular and disturbed, and, in addition, very large electron densities and spatial electron density gradients are often detected, in both summer and winter, day and night (Chapter 2).

These large values of electron densities are difficult to explain in terms of the processes responsible for creating the mid-latitude ionosphere and therefore a very important problem arises as to the source of the ionisation in the polar domain. Figure 3.1 illustrates the problem. The plot labelled 1 shows the maximum electron density at the F2 layer (N_mF2) as observed by the topside-sounder, Alouette I, for a pass over the north polar region in December 1967. The electron density is plotted as a function of solar zenith angle, $\chi°$, and on standard theory would be expected to decrease with increasing values of χ. However, an enhancement, typical of the plasma ring phenomenon, (Chapter 2), is clearly seen and appreciable electron densities ($\sim 10^5$ electrons cm^{-3}) are present even at larger zenith angles. Curve 2 (figure 3.1) shows how simple Chapman theory fails to account for these observed densities. Using the modified Chapman function and recently derived numerical approximations to it (Smith and Smith, 1972), the electron density associated with the Chapman production mechanism was calculated under the following conditions (taken from Bauer, 1973):

1. A neutral atmosphere of oxygen atoms only at a constant temperature of 1000°K

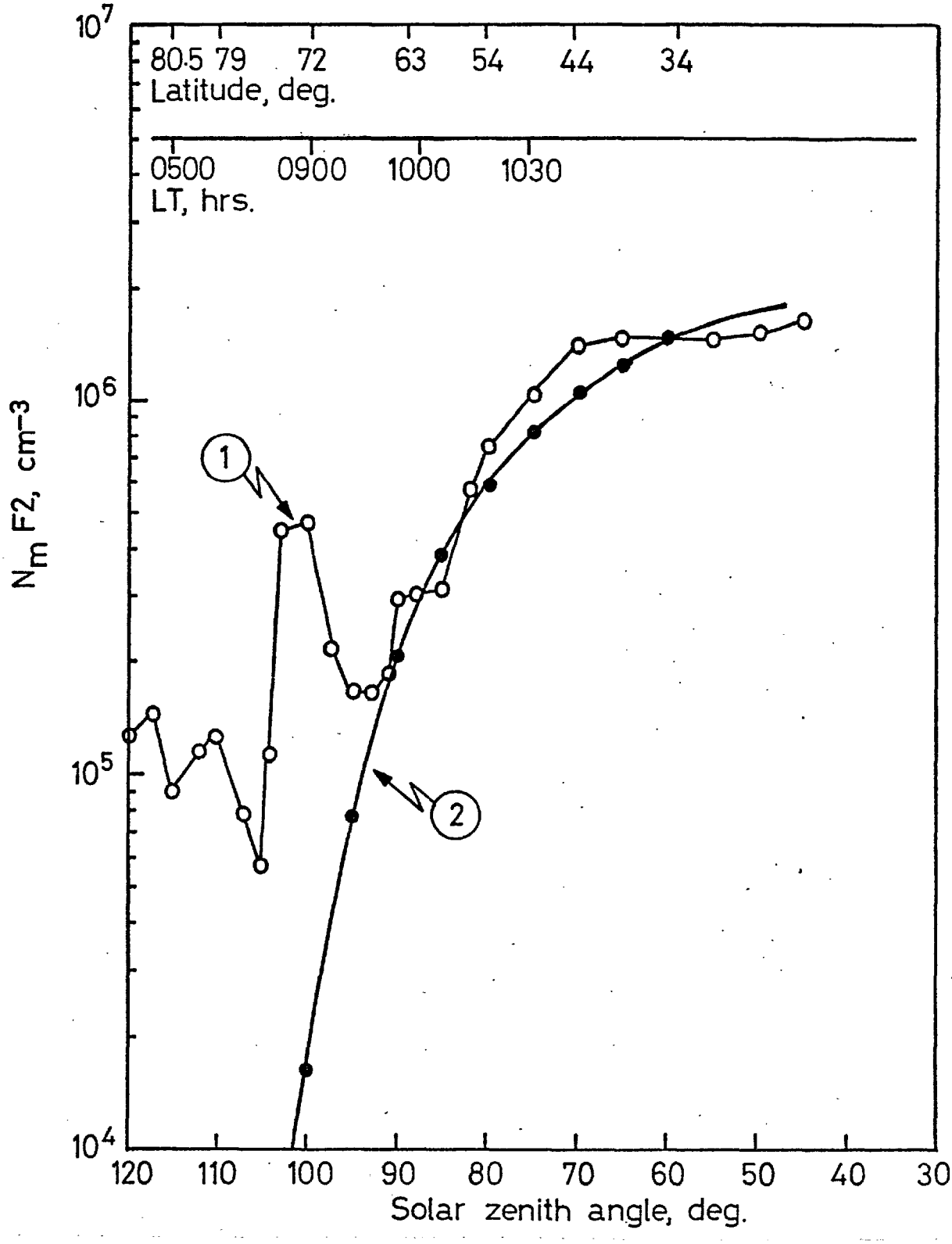


Figure 3.1 A demonstration of how simple Chapman theory fails to account for observed polar F-region densities in winter. 1 shows the electron density at the F-layer peak, as measured by the satellite Alouette I on 27th December 1967, 1608-1623 UT. The electron density is shown as a function of solar zenith angle and the geographic latitude and local time of the satellite are shown at the top of the figure. 2 shows the electron density at the F-layer peak expected from modified Chapman theory. Details of this calculation are found in the text. While there is good agreement at zenith angles $\lesssim 80°$, appreciable electron densities persist at large zenith angles where Chapman's modified theory predicts no production of ionisation.

2. A solar photon flux outside the atmosphere of 9.0×10^{9} cm^{-2}

3. An absorption cross section of 5.0×10^{-18} cm^{2}

4. A linear loss mechanism of the type βN. ($\beta \sim 2\times10^{-12} cm^{-3} s^{-1}$).

The two curves agree well for zenith angles <90° but above 90° there exist appreciable densities not predicted by simple theory. Hence, there appear to be active processes other than insolation maintaining levels of ionisation over the polar region. Since the lifetime of ionisation in the F-region is approximately a few hours, these processes must be continuous and of a kind which can account both,for the magnitudes,and the distribution pattern of the observed densities. It is intended in the sections which follow to examine quantitatively the extent to which the production processes associated with precipitating low-energy particles are able to account for the values of electron density, and the structure of the spatial electron density distributions observed and summarised in Chapter 2. An assessment of the importance of this process will be given. Here the term 'precipitated particles' refers to those electrons and protons which are found to be entering the atmosphere from great altitudes. They are to a first approximation tied to magnetic field lines and as well as motion parallel to the magnetic field,they are acted on by forces which cause them to spiral about the field line and drift in a direction perpendicular to it.

It should be noted that the enhanced electron densities observed at high latitudes may be caused, or contributed to, by a number of other possible mechnisms such as:

1. Horizontal movements of ionisation from low latitudes (Knudsen, 1974)

2. Vertical lifting of the ionosphere (Knudsen, 1974)

3. Enhanced production processes (Sato, 1973; Kamiyama, 1966)

4. Depleted loss processes (King et al., 1967).

Among enhanced production processes are the ionisation effects of so-called precipitated particles. Other processes cannot be excluded,

but there appears to be little direct experimental evidence that they can be of primary importance in providing the large observed densities. Whilst they remain plausible contributory sources of plasma they will not be considered in great detail in this work. However, it is possible that effects such as vertical and horizontal drifts may considerably modify the ambient density distribution and be responsible for the unusual diurnal variation of high-latitude h_mF2.

3.2 Precipitating Particle Fluxes

It has long been known that particles, chiefly protons and electrons, of energies greater than a few keV are found at high latitudes. Such particles have been identified as the cause of auroral emissions through processes of collisional excitation of the neutral atmosphere. However, recently, with the improvement of experimental techniques large populations of protons and electrons with lower energies of hundreds of eV have been detected. The vast majority of such measurements have been made by satellites though important contributions have arisen out of ground- and aircraft-based studies. Since a knowledge of the energy spatial and temporal distributions of such particles is now acknowledged as a prerequisite of any high latitude F-region study, such measurements have been collected and summarised in the Appendix.

To reduce particle measurements to a form suitable for use in theoretical calculations of F-region behaviour various simplifications and approximations have been made. For reasons discussed in the Appendix only electrons are considered and the spatial distribution is represented by four spectral types. These spectra and their characteristics are given in Table A3 (Appendix).

It was found possible to make analytic approximations to these spectra expressing them in a Maxwellian ($C_1E \exp -E/\alpha$) and power law ($C_2E^{-\delta}$)

form. These approximations, which are within the accuracy of the measuring instruments, are illustrated in Figure 3.2. Table 3.1 shows values of the constants C_1, C_2, α and δ used in these approximations.

For each spectrum the total energy of the particles available, E_{TOT}, is calculated by:

$$E_{TOT} = \int_{E_{min}}^{E_{max}} E\, J(E)\, dE \qquad 3.1$$

where E_{max} is the high-energy spectral cut-off

E_{min} is the low-energy spectral cut-off.

This important parameter, which specifies the total energy input into the atmosphere is listed for each spectrum in Table 3.1.

3.3 Calculation of Production Rates

3.3.1 Energies from .1 - 10 keV

One of the most widely-used approaches, to the calculation of production rates, involves the use of the Monte Carlo method and takes account of straggling effects (e.g. Maeda, 1965). This is a very lengthy computational technique so here a much simpler method is used which gives an approximate answer, yet represents a great simplification. In addition calculations involve so many other unknowns, such as the polar neutral atmosphere, that the use of a very involved, and accurate technique seems of little practical value.

The method employed is due to Rees (1963, 1964, 1969) and is based on the earlier experimental work of Grün (1959). Grün approached the problem by injecting mono-energetic electron beams, of energies varying from 5 - 54 keV, into air. The luminosity produced was examined and isophotes of emission presented. In this way, all orders of scattering were included and a choice of scattering cross-sections avoided. Using these luminosity profiles, and assuming that the luminosity produced was

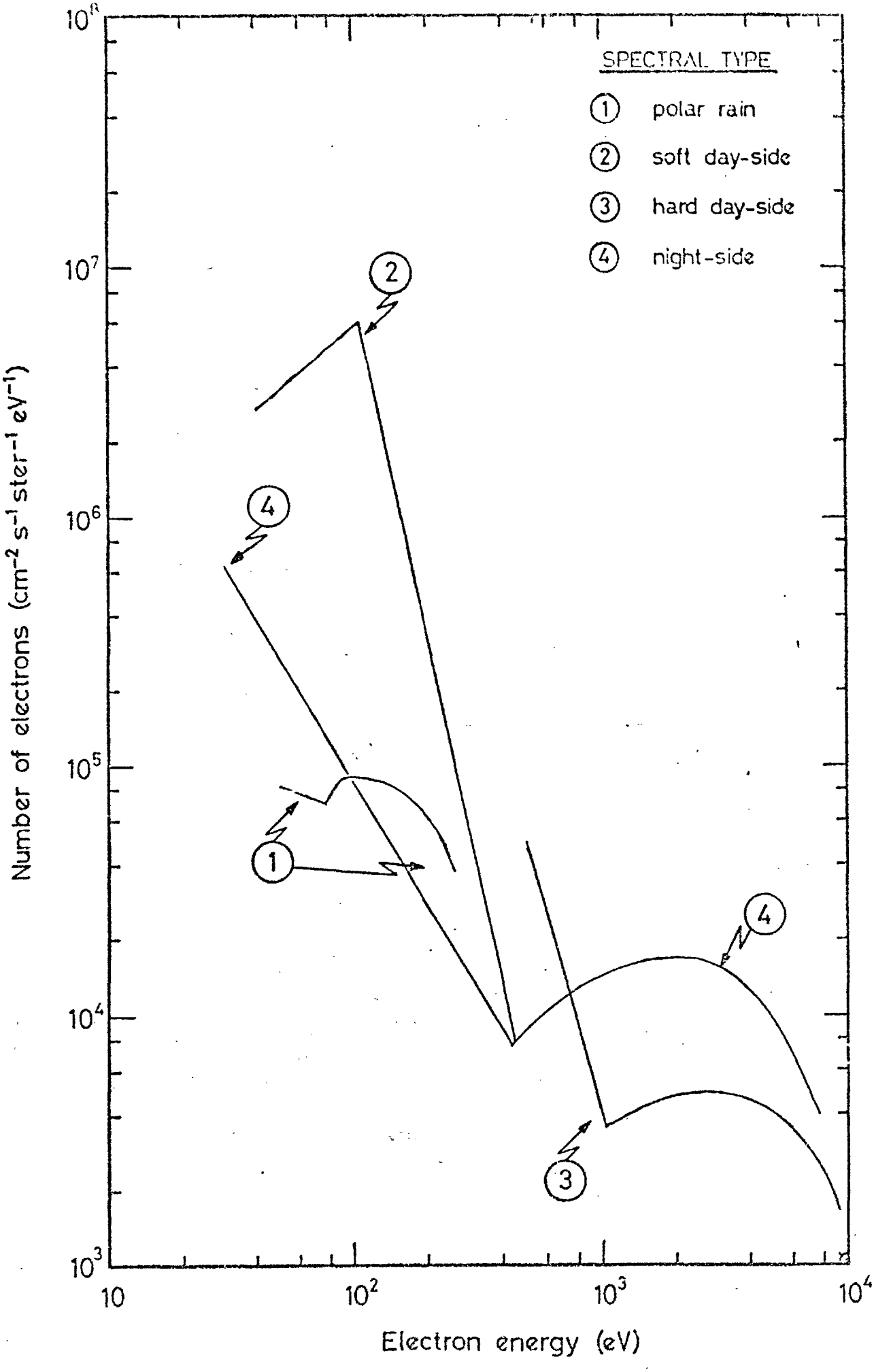


Figure 3.2 Analytic approximations to the typical low-energy electron fluxes shown in Figure A9 (Appendix). Details of parameters specifying these curves are given in Table 3.1.

Table 3.1

Differential energy spectra of precipitated particle fluxes in the polar ionosphere (electrons cm^{-2} s^{-1} $ster^{-1}$ keV^{-1})

	Power law	Maxwellian form	Transition* energy (keV)	Energy range (keV)	E_{TOT} (ergs $cm^{-2}s^{-1}$)
Polar rain	$2.90\ 10^{2}\ E^{-.37}$	$2.65\ 10^{9}\ E \exp(-E/0.1)$	0.08	.05 - .25	0.01
Soft dayside	$3.75\ 10^{10}\ E^{.833}$ $2.61\ 10^{5}\ E^{-4.55}$	Best approximation in energy range was two power law spectra	0.15	.05 - .5	0.28
Hard dayside	$4.00\ 10^{6}\ E^{-3.67}$	$4.71\ 10^{5}\ E \exp(-E/3.0)$	1.0	.5 - 10	0.64
Nightside	$8.36\ 10^{6}\ E^{-1.68}$	$4.20\ 10^{8}\ E \exp(-E/1.1)$	0.40	.05 - 10	2.35

* The transition energy defines the energy ranges where the observed flux is best represented by the power law or Maxwellian form. In the case of the soft dayside the transition is between two power law type spectra.

proportional to the energy deposited, Rees (1963) performed a numerical integration of the angular distribution of Grün's isophotal contours and was able to calculate a normalised, energy dissipation distribution function, $\lambda(z/R)$, which gives the fraction of energy deposited at a depth z in the atmosphere for a particle of total range R. This was possible for an arbitrary pitch angle distribution of incident electrons and the functions for the pitch angle distributions indicated are shown in Figure 3.3. Over the energy range 5 - 54 keV, λ was found to be independent of energy, and although Maeda questioned this at low energies, Rees has extended the lower limit of energy to .4 keV.

The range data used are those collected by Gledhill (1973). Gledhill has gathered together the experimental and theoretical results on the range-energy relationship for electrons in the energy range .1 - 600 keV. He found many workers had used different definitions for range. He standardised the data by converting them to practical range, R_p, as defined by Berger et al. (1970). Figure 3.4 shows Gledhill's curve of practical range R_p (in gm cm^{-2}) versus energy in keV. These results fall off on the low range side of the graph and show a tendency to deviate more at low energies, indicating larger ranges than expected. Most experimental data necessarily use a monodirectional beam and Gledhill used the results of Berger et al. (1970) to convert these results to isotropic (over downward hemisphere) pitch angle distributions. Thus using

$$\frac{R_{p(IDH)}}{R_{p(MD)}} = 0.926$$

Gledhill has given the range-energy relationship as

$$\log_{10} R_{p(IDH)} = -5.133 + 1.358x + .215x^2 - 0.43x^3 \qquad 3.2$$

Here $x = \log_{10} E$ and the units of R_p are gm cm^{-2} for an energy, E, in the range .1 - 100 keV.

The electrons precipitate into two model atmospheres, the static diffusion model of Jacchia (1971), for an exospheric temperature of 1200°K,

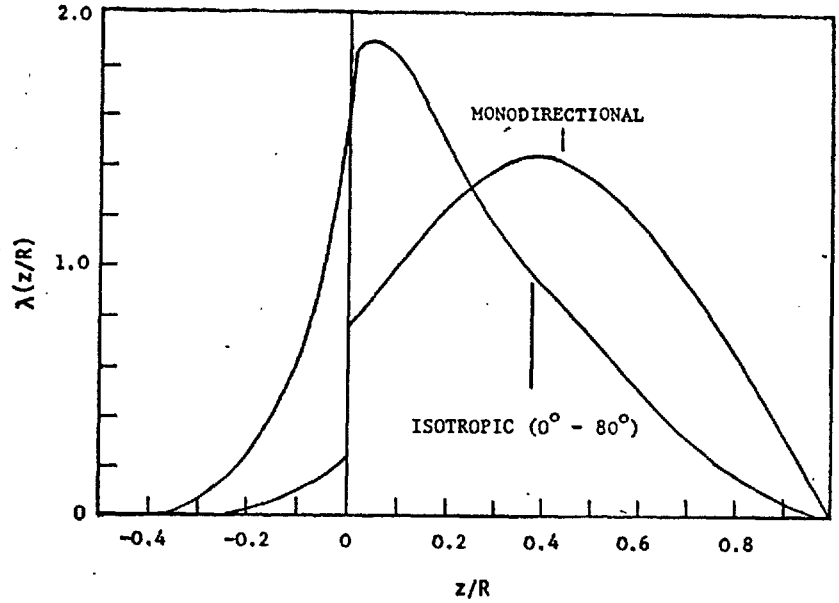


Figure 3.3 The energy dissipation distribution function, $\lambda(z/R)$, as a function of the fractional penetration distance, z/R, for two pitch angle distributions of electrons. (From Rees, 1964).

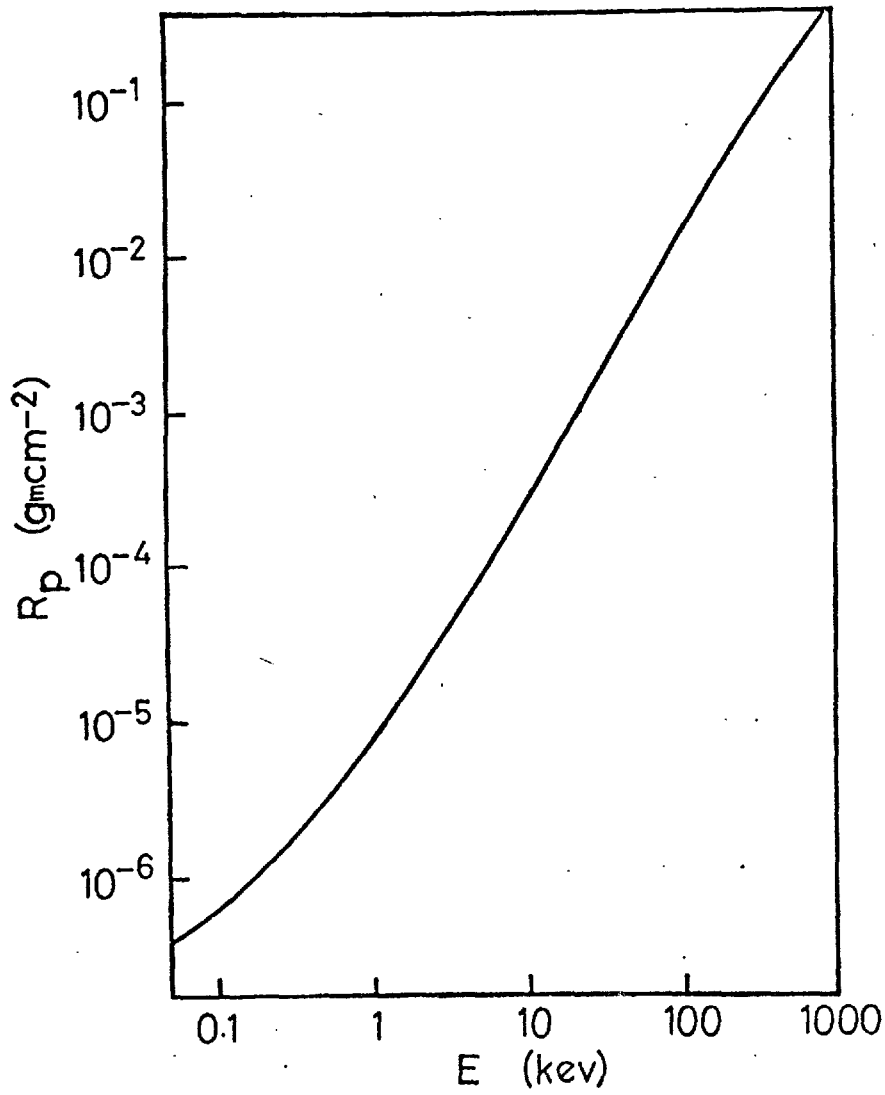


Figure 3.4 The practical range, R_p, (as defined by Berger et al., 1970) in units of atmospheric depth, as a function of electron energy. The curve is from Gledhill (1973) drawn by him as a best fit through the experimental data for which references are given in Gledhill.

and the CIRA 1965 atmosphere. Unfortunately atmospheric models are constructed for midlatitudes and it is not known if they are closely representative of the polar atmosphere. It is known, however, that temperatures in the polar regions are greater than those at midlatitudes and this parameter was adjusted accordingly by comparing model densities and composition with results obtained over the Antarctic region by the satellite OGO 6 (Hedin and Reber, 1972). Atmospheric parameters and characteristics are shown in Table 3.2.

Defining depth as length times density for a homogeneous absorber, then for the atmosphere at an altitude h, the depth is given by:

$$z = \int_h^\infty \rho dh = \frac{1}{g}\int \rho g \, dh = \frac{p}{g} \qquad 3.3$$

where p is the pressure

From Rees (1963) the energy deposited between depths z and z+dz from electrons of initial energy E is given by

$$\delta E = \lambda(z/R)\ J\ (E,\alpha)\ \frac{E}{R}\, dz\ dE\ d\Omega\ eV\ cm^{-2}\ s^{-1} \qquad 3.4$$

where $d\Omega$ is the solid angle from which the particles are incident

If J is not isotropic and is a function of pitch angle α then it should be noted that R now refers to practical range, shown as R_p in equation 3.2 and $\lambda(z/R)$ is the appropriate energy dissipation function for the pitch angle distribution chosen. In equation 3.4 $J(E,\alpha)dEd\alpha$ is the flux of electrons in the range of energies E to E+dE and pitch angles α to $\alpha+d\alpha$. Hence from the definition of z

$$dz = \rho(h)\ dh$$

and the energy deposited per unit volume is given by δE where

$$\delta E = J\ (E,\alpha)\ \frac{E}{R}\ \lambda(z/R)\,\rho(z) dE\ d\Omega\ eV\ cm^{-3}\ s^{-1} \qquad 3.5$$

Altitude km	$\log_{10} \rho$ gm cm^{-3}	Temperature °K	$\log_{10} z$ gm cm^{-2}
90	-8.461	183.0	-2.7300
96	-8.942	187.1	-3.1070
100	-9.264	197.1	-3.4821
106	-9.729	227.4	-3.8824
110	-10.015	259.0	-4.1137
120	-10.618	373.4	-4.5450
130	-11.061	512.7	-4.8342
140	-11.386	642.7	-5.0517
150	-11.638	750.3	-5.2272
160	-11.844	834.7	-5.3773
180	-12.175	952.2	-5.6345
200	-12.447	1027.5	-5.8564
220	-12.684	1078.3	-6.0563
250	-12.998	1126.7	-6.3273
300	-13.449	1166.7	-6.7291
360	-13.916	1185.4	-7.1572
400	-14.198	1191.0	-7.4240
500	-14.841	1196.7	-8.0290
600	-15.427	1198.5	-8.5746
700	-15.966	1199.2	-9.0424
800	-16.447	1199.5	-9.4110
1000	-17.152	1199.8	-9.8748

Altitude km	$N(N_2)$	$N(O_2)$	N(O)	N(A)	N(He)
90	13.7498	13.1724	11.6094	11.8276	8.9685
96	13.2691	12.6623	11.8914	11.3469	8.4878
100	12.9472	12.3118	11.8132	11.0250	8.1659
106	12.4825	11.7986	11.5658	10.5339	7.7591
110	12.1969	11.4804	11.3784	10.1506	7.6914
120	11.5899	10.8096	10.9636	9.3527	7.5288
130	11.1378	10.3128	10.6464	8.7668	7.3985
140	10.8012	9.9422	10.4120	8.3285	7.3036
150	10.5372	9.6503	10.2324	7.9807	7.2338
160	10.3186	9.4071	10.0877	7.6886	7.1805
180	9.9578	9.0032	9.8571	7.1985	7.1017
200	9.6519	8.6585	9.6683	6.7764	7.0422
220	9.3759	8.3461	9.5016	6.3917	6.9927
250	8.9943	7.9131	9.2755	5.8557	6.9291
300	8.4057	7.2428	8.9328	5.0228	6.8378
360	7.7380	6.4811	8.5485	4.0735	6.7391
400	7.3059	5.9879	8.3009	3.4583	6.6764
500	6.2560	4.7888	7.7003	1.9619	6.5254
600	5.2404	3.6289	7.1200	0.5139	6.3800
700	4.2550	2.5033	6.5571		6.2390
800	3.2976	1.4097	6.0102		6.1022
1000	1.4615		4.9616		5.8398

Table 3.2 An example of atmospheric parameters used in the computations. These were taken from Jacchia (1971) and are appropriate to an exospheric temperature of 1200°K. Number densities are logarithms to the base 10.

For primary energies in the range considered, 35eV is necessary for the production of one ion-electron pair. As the major neutral constituents of the upper atmosphere are O and N_2, the ions formed are O^+ and N_2^+. The latter rapidly recombines so we are concerned with the production of oxygen ions only. Making the assumption that within a volume the production of an ion species is proportional to its contribution to the total mass density, the production rate of ions of the ith species due to all incident energies and for precipitating pitch angles i.e. $|\alpha| \leq \pi/2$ is

$$Q_i(z) = \frac{1}{35} \int_{\Omega} \int_{E} J(E,\alpha) \frac{E}{R} \lambda(z/R) \rho(z) \frac{\rho_i(z)}{\rho(z)} \, dE \, d\Omega \qquad 3.6$$

Since the polar region only is under consideration it is justifiable to neglect effects of the magnetic field. Rees (1964) has shown that its effects for an extended source are negligible, since the field lines are almost vertical. Magnetic reflection has been neglected by assuming that any untrapped particles at 1000 km will not suffer any noticable effect due to a converging magnetic field.

3.3.2 Energies below .1 keV

It is clear from spectral measurements that densities at F-region altitudes may be very dependent on the large population of electrons in the spectra at $\lesssim$100 eV. Although measurements of spectra at these energies are very reliable the part they play in the creation of ionisation is uncertain. Above, the method of Rees (1963), initially used in the energy range .4 - 10 keV, has been extended down to even lower energies (viz. .1 keV). This can be justified due to the uncertainties of other parameters and the fact that polar conditions are so variable that an order of magnitude calculation only is feasible. However, criticisms of this method at such low energies are based on the fact that 35 eV is lost from the primary electron in the creation of each electron-ion pair. Such quantal steps in primary electron energies seem well-established

where the primary electron energy is much larger than that necessary for the creation of each electron-ion pair, but at low energies - a few hundred electron-volts - this approximation appears likely to become progressively less applicable (Banks et al., 1974).

Recently, theoretical approaches to energy deposition and thermalisation of soft electrons have made great progress (Banks et al., 1974; Mantas and Walker, 1976), largely through attention to photoelectron problems. Mantas (1975) used the Boltzmann equation, to determine heating, ionisation and excitation of the neutral gas, for isotropic monoenergetic electron beams of 25, 50 and 100 eV. Figure 3.5 shows Mantas' calculation of the ionisation due to monoenergetic beams of unit flux at the three energies. Here, using an adaption to Mantas's technique, (Mantas, private communication, 1976) an estimation of production rates is obtained for experimentally measured intensities of electrons, at energies $\leq$ 100 eV (typical of the dayside soft region), in order to determine the relative importance of the contribution to the total ionisation of these electrons.

From Figure 3.5 it is possible to interpolate at various altitudes to find the ionisation rate due to unit incident fluxes of energies intermediate to 25 and 100 eV. For each specific altitude this gives a measure of the way energy is deposited as a function, $q_h(E)$, of energy between 25 and 100 eV. Such a function is shown in Figure 3.6 for an altitude of 900 km. The smooth curve drawn through the true points at 25, 50 and 100 eV has no physical basis and is purely for convenience. For measured spectra the total production at each altitude for electrons $25 \leq E \leq 100$ in electrons cm^{-3} s^{-1} can be written simply as

$$q_h = \int_{25}^{100} q_h(E)\ J(E)\ dE \qquad 3.7$$

In this way a very approximate altitude production profile can be synthesised.

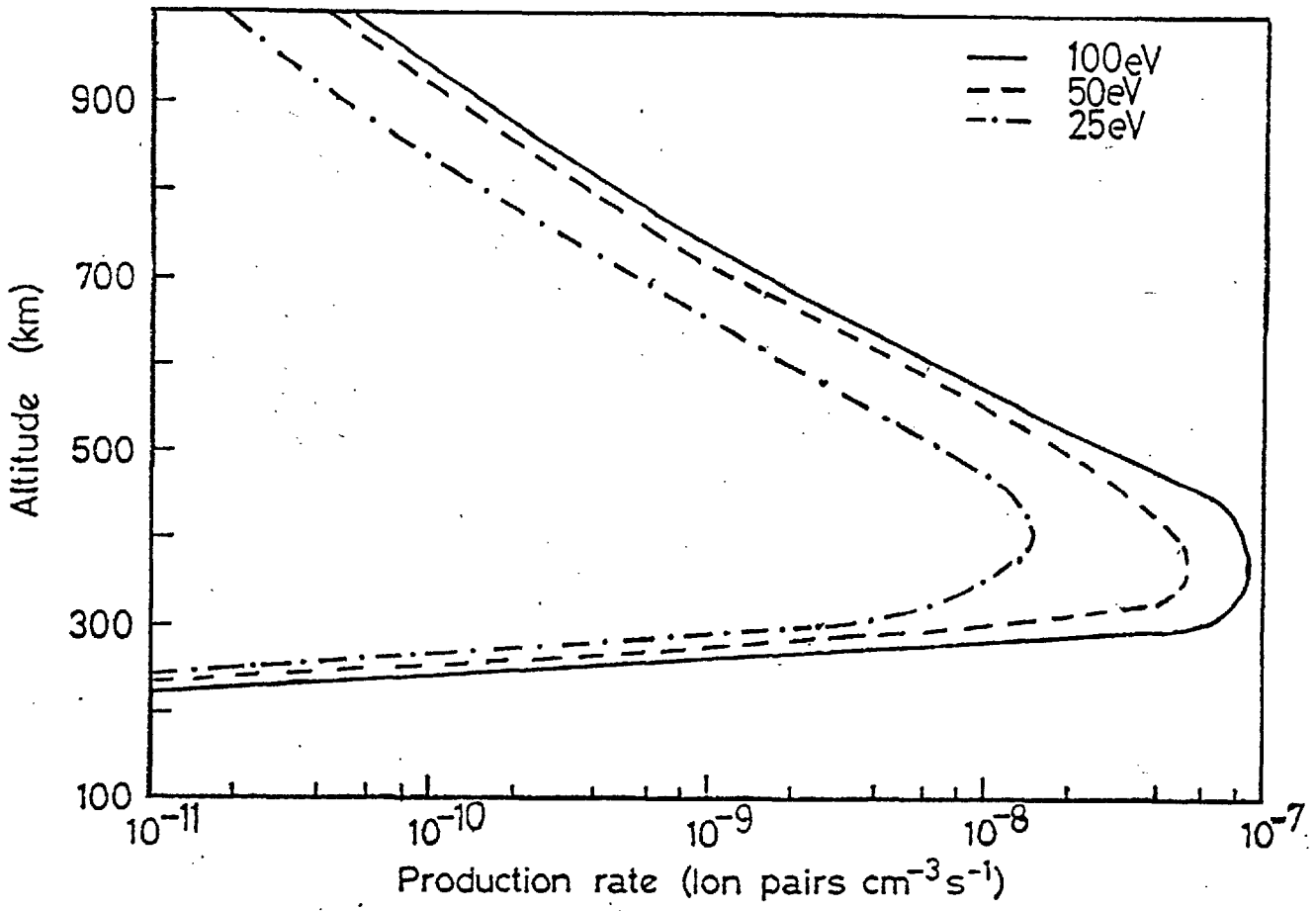


Figure 3.5 Production rate profiles for O^+ ions for a unit flux beam with the three indicated energies. Other ion production rates have been omitted from the original diagram since they are orders of magnitude less than O^+ production except at very low altitudes. (From Mantas, 1976)

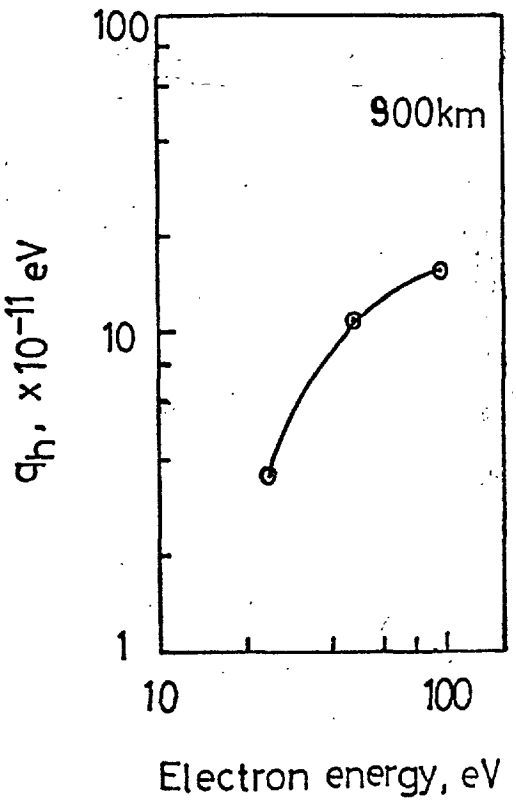


Figure 3.6 The energy deposited in unit time and unit volume at an altitude of 900 km for a unit energy flux by electrons between 25 and 100 eV. The points have been obtained from Mantas' calculations by interpolation and are used to give an estimate of the effect of the electrons <100 eV as an ionisation source.

3.4 Derivation of Electron Density Profiles

To determine the electron density profiles which would result from the calculated production rates it is necessary to solve the continuity equation for electrons in the polar ionosphere. We take into account only the production, loss and ambipolar diffusion terms. Ionisation of neutral particles by collisions with energetic electrons is assumed to be the only production mechanism and effects such as neutral winds and atmospheric heating are neglected. The calculation is simplified by assuming diffusive equilibrium to exist at 1000 km (although this is open to doubt over the polar regions, especially over the polar cap). This restriction was later removed and the effect of an upward and downward flux of particles at 1000 km was investigated.

To obtain an expression for the diffusion flux, the coupled equations of motion of ions and electrons in a vertical magnetic field at a level z are taken (Thomas et al., 1966):

$$m_i \nu_{in} v_i = m_i g - eE - \frac{k}{n_i} \frac{d}{dz} (n_i T_i) \qquad 3.8$$

$$m_e \nu_{en} v_e = m_i g + eE - \frac{k}{n_e} \frac{d}{dz} (n_e T_e) \qquad 3.9$$

where e is the modulus of the electronic charge,

E is the electric field due to the charge separation in the ionosphere,

z is the atmospheric depth measured positive "downwards",

and all other symbols have their usual meaning. These equations are valid at heights near and above the F-region peak where collisional terms are much greater than inertial ion and electron acceleration terms. However, despite the fact that the polar ionosphere is considered to be of a single ion species (O^+) and collision terms are not large, at great altitudes, where the mean free path becomes large, the full momentum equations may be required (Banks and Holzer, 1968). Sato (1973) in fact

neglects the use of the full momentum equations even at 2500 km which may not be a valid assumption.

Thus with charge neutrality

$$n_i = n_e = N \quad \text{charged particle density}$$

$$v_i = v_e = v \quad \text{diffusion velocity}$$

and $$m_i \nu_{in} >> m_e \nu_{en}$$

the ambipolar diffusion flux can be written approximately as

$$F = Nv = \frac{Hg}{\nu_{in}} \left(\frac{N}{H} \left(1 - \frac{dH}{dz}\right) - \frac{dN}{dz} \right) \qquad 3.10$$

where the plasma scale height H is a function of altitude given by

$$H(z) = \frac{K(T_e + T_i)}{m_i g} .$$

Hence

$$\frac{dN}{dz} = \frac{N}{H} \left(1 - \frac{dH}{dz}\right) \qquad \text{for } Nv = 0$$

or

$$\frac{dN}{dz} = \frac{N}{H} \left(1 - \frac{dH}{dz}\right) - \frac{F}{D} \qquad \text{for } Nv \neq 0$$

and Hg/ν_{in} is the ambipolar diffusion coefficient D for a vertical magnetic field and F is the diffusion flux.

Taking the familiar continuity equation for steady state conditions in the absence of horizontal movement at a level z;

$$Q(z) - L(z) - \frac{d}{dz} (Nv) = 0 \qquad 3.11$$

where Q(z) and L(z) are production and loss rates per unit volume at a

level z, equations 3.10 and 3.11 combine to give a second-order differential equation which can be solved by a fourth order Runge-Kutta technique from 1000 km (where diffusive equilibrium was initially assumed, Nv = 0) to a level where diffusion is again small at an altitude of 180 km. The altitude where diffusion is again small is not precisely known and several different values, between 150 and 200 km were used. The second constraint placed on the solution was that between the integration limits the integrated production and loss rates must be equal.

An expression for the altitude-dependent scale height was found from experimentally measured polar values and approximated as a linearly increasing function, rising from 80 km at 150 km, altitude, to 200 km at 1000 km. (This corresponds to $T_i + T_e = 3780°K$ if $g = 980$ cm s^{-2}.) Thus, remembering that z is positive downward,

$$H_z = H_{1000} - bz \qquad 3.12$$

where b = 12/85.

The value of the diffusion coefficient used was that of Dalgarno (1964) and Yonezawa (1966), namely 2×10^{-19}/neutral density ($cm^2\ s^{-1}$). The height dependence comes in the exponential fall-off of neutral density which dominates other effects such as those of temperature.

The ionospheric chemistry involved in the loss processes at the heights concerned and for polar latitudes is not known so only three reactions were considered. The important processes were taken to be dissociative recombination with the rate determining interchange processes:

$$O^+ + N_2 \longrightarrow NO^+ + N \qquad \text{rate } k_1$$

$$O^+ + O_2 \longrightarrow O_2^+ + O \qquad \text{rate } k_2$$

$$N_2^+ + O \longrightarrow NO^+ + N \qquad \text{rate } k_3$$

A comprehensive review of laboratory measurement of ionospheric reaction rates has been given by Ferguson (1975) who summarised attempts to measure rate constants at temperatures approaching those encountered in the ionosphere. Lindinger et al. (1974) have measured rate constants at temperatures up to 900°K but this still falls short of the desired 2000 - 2500°K range. The difficulty is, to some extent, being surmounted by experiments in which reactant ion kinetic energy can be varied from thermal energy into the eV range corresponding to temperatures of several thousand degrees. It is not yet clear if knowing the rate constant as a function of ion kinetic energy is equivalent to knowing it as a function of temperature. However, it must be a large step forward since a prominent, if not dominant, effect of temperature is to increase the collision energy of the reactants. Making this assumption gives values, at 2000°K, for the rate constants shown in Table 3.3.

Table 3.3

k_1	6×10^{-13} (cm^3 s^{-1})
k_2	1×10^{-11} "
k_3	7.9×10^{-12} "

(from Ferguson, 1974)

Hence:

$$L = k_1[N_2][O^+] + k_2[O_2][O^+] + k_3[O][N_2^+]$$

With oxygen the dominant ion at polar latitudes and molecular oxygen densities well below those of nitrogen the above expression reduces to the first term of the right hand side only, so that the rate of removal of oxygen ions (assuming the dissociative recombination processes go

quickly enough to ensure that $[O^+] = N$, the total charged particle density) becomes

$$L = k_1 \times (\text{molecular density}) \times N. \qquad 3.13$$

Thus the discrepancy between laboratory data and that inferred from ionospheric measurements appears to be resolved by assuming that the temperature and energy dependence of rate constants are identical. Yonezawa (1966) found the rate constant k_1 in 3.13 to be 6.5×10^{-13} while k_1 from table 3.3 gives 6×10^{-13} giving remarkably good agreement.

3.5 Computational Results

Using equation 3.6, altitude profiles of ion production rates have been derived, and used as input data for the calculation of the resulting electron densities from equations 3.10 and 3.11. A variety of atmospheric conditions, incident energies and pitch angle distributions have been considered so that the most important factors governing the magnitude and form of a corpuscularly-produced polar ionosphere could be isolated.

3.5.1 Production rates

The main results of production rate calculations are presented in Figure 3.7 and Table 3.4. The former shows altitude production rate profiles of oxygen ions for various electron spectra, in the energy range .1 - 10 keV, incident on a Jacchia (1971) model atmosphere. Also a similar curve, marked Mantas (<100 eV), is shown to indicate production rates due to fluxes of electrons of energies $25 \leq E \leq 100$ eV, in the soft dayside or cleft spectrum. These have been calculated from the approximation to Mantas and Walker's (1976) method. The latter, in tabular form, summarises the maximum values and altitudes of production found in Figure 3.7.

Several points are immediately apparent regarding profiles such as

those of Figure 3.7:-

1. Production rates. Over the range of spectra, corresponding to different regions of the polar ionosphere, there is a substantial variation in the height and magnitude of the production maxima. This appears to be due to the distribution, in number and energy, of electrons in the incident flux. Thus the soft day- and nightside spectra give similar peak production rates, but these are found at very different altitudes because of contrasting number fluxes at a few hundred eV and a few keV. Electrons with characteristic energies of around a few hundred eV are found to deposit a larger proportion of their energy at F-region altitudes between 250 and 400 km. Those with energies $\geq$ 1 keV, however, have greatest effect at altitudes $\lesssim$200 km.

The values of peak O^+ production rate and the altitude at which this occurs, shown in Table 3.4, should be compared with those given by Rishbeth and Garriott (1969), for photoionisation in a similar neutral atmosphere by an overhead sun. They have shown that for typical mid-latitude conditions photoionisation can provide a peak production rate $\sim$800 ion pairs cm^{-3} s^{-1} at $\sim$170 km altitude. Such a figure would indicate that the fluxes in the polar cavity and the hard-dayside region are incapable of supporting appreciable electron densities. However, precipitation in the nightside ring gives an ionisation production peak magnitude and height in close agreement with the solar-produced layer and would, by analogy, be expected to support similar densities. The case of the cleft region, where the soft-dayside flux is found, is very interesting. Peak production is of similar order to that given by solar effects, but it is formed at a much higher altitude. In this region chemical loss is less effective and thus higher densities than those at mid-latitudes might be predicted. However, production rates are not an observed ionospheric parameter and conversion to electron densities is necessary to allow comparison with experiment.

2. Integration limits Initial investigation of the effects of extending the energy integration range by reduction of the lower limit to 75 and 50 eV showed that only production by the soft dayside electrons was appreciably increased. To assess the contribution of electrons <100 eV to total production in the soft dayside region, where the spectrum has the largest content of these electrons, the approximation to the calculation of Mantas and Walker (1976) was applied. This showed that energies of <100 eV could be safely neglected without modification of the conclusions, since production rates (for E < 100 eV) were a factor $\sim 10^2$ down on those above 100 eV. It is clearly indicated by the curve marked Mantas in Figure 3.7.

As previously mentioned, artificial cut-offs were applied to the lower and upper bounds of the fluxes of Figure 3.2 by the energy limits of sensor sensitivity. Extrapolation of the curves through these limits and computation of the resulting production rates, allowed an assessment of the importance of such cut-offs to be made. For all spectra, except those in the hard-dayside region, the modified profiles exhibited unimportant production enhancements of ~2-5%. Only the hard-dayside flux, when extrapolated from its lower limit, gave substantial increases in production, even to levels above those predicted for the soft dayside flux. However, there is very little evidence for such an extension of the energy range for this flux. In fact, data from the ISIS I SPS, for transits of the dayside high-latitude region, show order-of-magnitude decreases of fluxes of electrons with energies approximately a few hundred eV and less, on passing from the soft to hard-dayside zones.

3. The neutral atmosphere Although the curves in Figure 3.7 were computed for a Jacchia (1971) model neutral atmosphere at 1200°K similar calculations were performed for models of different composition and temperature. The shape and level of the altitudinal production profile proved to be very dependent on the distribution and concentration of neutral constituents. Although individual constituents do not

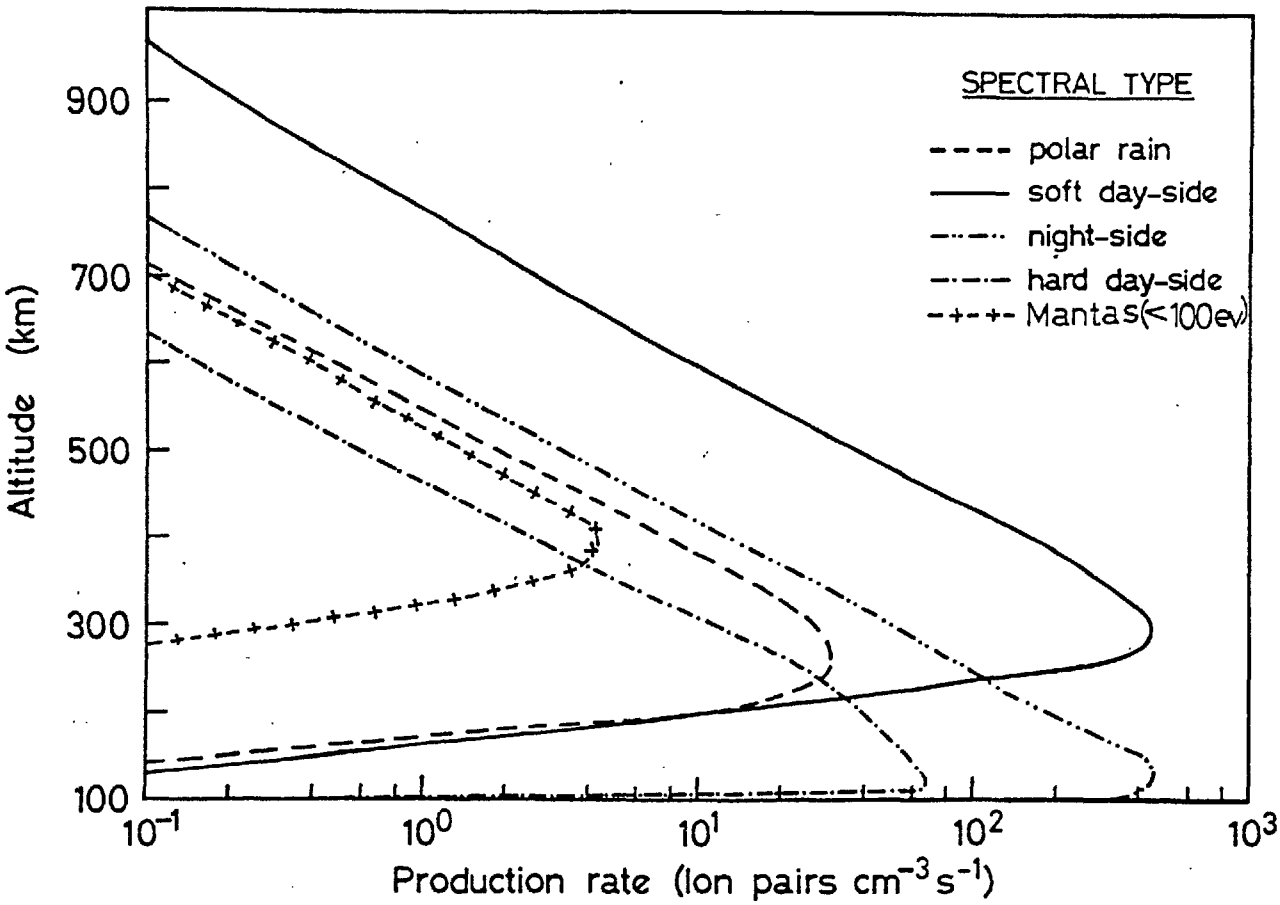


Figure 3.7 Altitude profiles of oxygen ion production rates computed for the spectra noted in the key. All except the curve marked Mantas (<100 eV) have been computed for atmospheric conditions noted in Table 3.2 and integrated over the energy range 0.1 - 10 keV. This Mantas curve is the contribution to production rates of the soft dayside flux component with energies $25 \leq E \leq 100$ eV and has been calculated from an adaption of the results obtained by Mantas and Walker (1976).

Table 3.4

Maximum O^+ production rates

Region	Production rate (ion pairs $cm^{-3} s^{-1}$)	Altitude (km)
Polar rain	$2.9 \pm 0.2 \times 10^1$	280 ±10
Soft dayside	$4.0 \pm 0.1 \times 10^2$	300 ±15
Hard dayside	$6.8 \pm 0.3 \times 10^1$	110 ±10
Nightside	$2.2 \pm 0.1 \times 10^2$	120 ±15

The errors are introduced by uncertainties in reading from the curves of Figure 3.7.

exhibit the same ease of ionisation, the separate ion production rate profiles closely follow the altitude profile of the corresponding species neutral number density. Thus in the F-region, and above, where the dominant neutral is the oxygen atom, computations have shown that inclusion of production of all ion species has a minor effect on total ion density. In addition, N_2^+ ions, produced by ionisation of the N_2 molecule, are quickly lost through the recombination and cannot make an appreciable contribution to the stable electron density distribution. However, at altitudes above 1000 km, where hydrogen and helium become more abundant, and below 200 km, where O_2 and N_2 are important, ionisation of these neutrals must be considered. Thus at the very low energies ($\sim$.1 keV) at which electrons deposit their energy in the F-region,the magnitude of the neutral temperature can, through its influence on the neutral oxygen distribution, play a considerable part in fixing the height of the production maximum.

Further investigations of the temperature effects have shown that consequent redistribution of neutral constituents influences all aspects of the production rate profile.

It is known that temperatures are often enhanced in the high-latitude precipitation regions and also largely govern the distribution and composition of the neutral atmosphere. The neutral atmosphere, in response to an increase of temperature, undergoes a thermal expansion with the result that at a fixed altitude the total density and the proportion of heavier species are increased.

Temperature effects have been examined in two ways. Figure 3.8 shows one of these investigations. Here the variation of the height of the production maximum with electron energy is given, for identical model atmospheres, with differing exospheric temperatures. At high electron energies, approximately a few keV, the temperature variation has negligible effect since electrons penetrate to low altitudes where the atmosphere is dense and has small thermal response. However, at energies

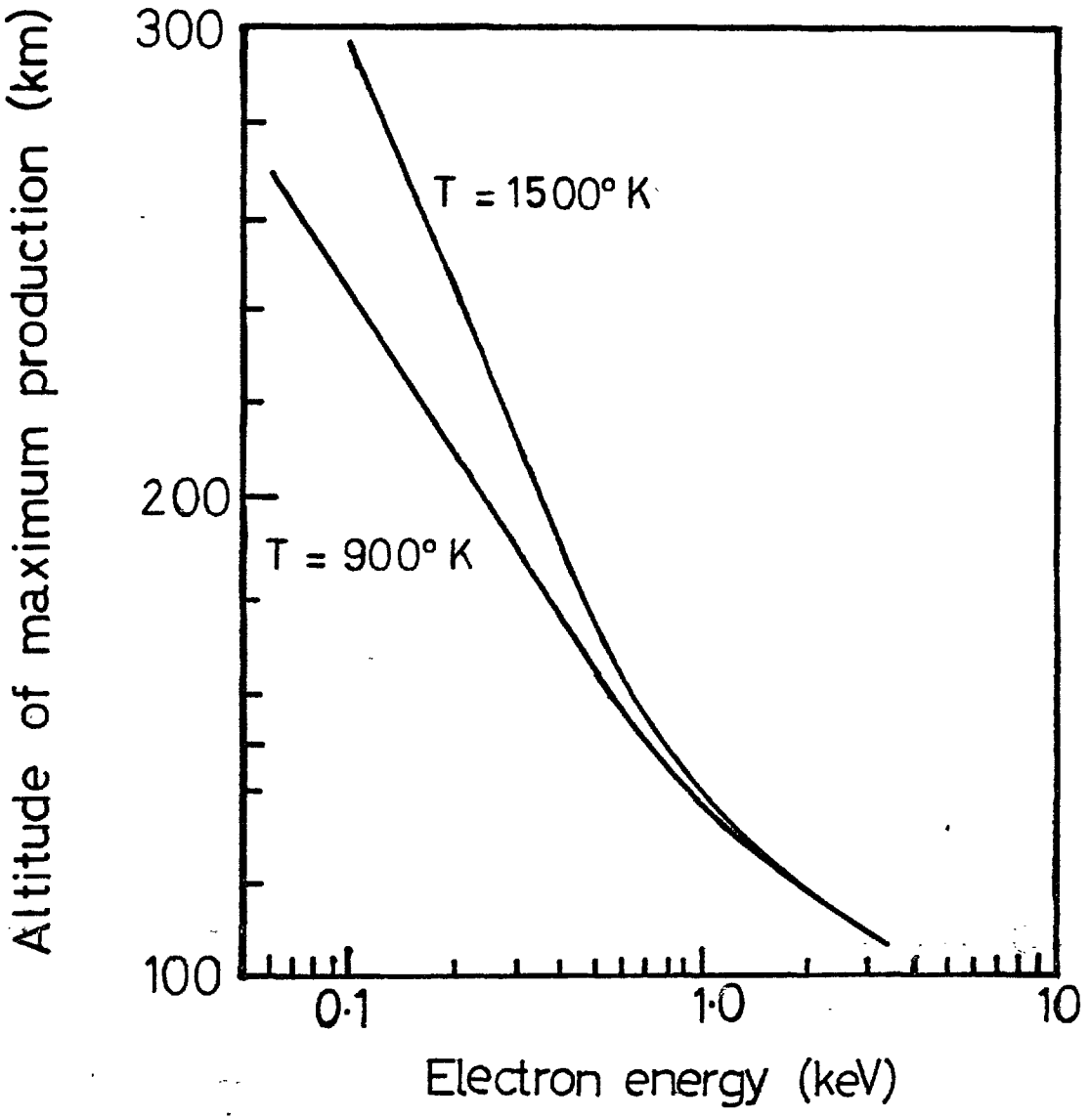


Figure 3.8 The dependence of production of oxygen ions on exospheric temperature. The curves show the altitude of maximum production of oxygen ions as a function of energy for two extreme temperatures. It is apparent that the effect of temperature becomes increasingly important at progressively lower energies.

$\lesssim$1 keV temperature effects become noticeable and increase with decreasing energy.

3.5.2 Electron density profiles

Using the production rate profiles, given in Figure 3.7 as input data, the corresponding altitudinal variations of electron densities have been computed from equations 3.10 and 3.11 and are presented in Figure 3.9. Values of peak electron density and the altitude, read from those profiles are set out in Table 3.5. The results show that in the soft dayside, nightside (roughly coincident with the ring) and, to a lesser extent the hard dayside region, soft electron fluxes can support considerable levels of electron density. The structure of the incident electron spectrum plays an important role in deciding the electron density profile. This is evident in the soft day- and night-side zones. Production maxima resulting from these two spectra are very similar (Figure 3.7) but the heights of peak production are very different. The nightside spectrum peaks at very low altitudes where chemical loss is large. Hence, the maximum densities expected in the nightside region are a factor of 2 down on those on the dayside, whose spectrum gives peak production $\sim$300 km where recombination is small and diffusion is the most important process. To some extent, therefore, the 'softness' of the spectrum can determine the magnitude and altitude of the peak in electron density.

As expected, from the low intensity of the spectrum, the steady electron densities predicted for the polar cap are small, being orders of magnitude less than those in the intense precipitation regions. However, computations of production due to sporadic fluxes (polar squall, polar shower) noted by Winningham and Heikkila (1974) indicate the possibility that such events could support values of electron density similar to those of the soft dayside region. The variable nature of these observations makes conversion of these production rates to steady densities very

difficult.

Further calculations were performed to investigate the effects of a 'polar wind' type flux of ions and variations in the value of the loss rate due to high latitude electric fields and temperatures.

Figure 3.10 shows the results of relaxing the constraint of diffusive equilibrium placed on earlier integrations. The region used is the soft dayside since this appears to be the region where polar wind type flows are most often observed (Shepherd et al., 1976). Two cases are presented corresponding to upward and downward fluxes, through the 1000 km level, of 5×10^8 ions $cm^{-2}s^{-1}$. The results, in accordance with Banks and Holzer (1969) and Banks et al. (1975), show, that for the major ion, there is very little modification of the diffusive equilibrium distribution. An upward flux slightly decreases densities at all altitudes above 200 km and has little effect on the height of the peak. The reverse occurs for a downgoing ion flux of identical magnitude.

The way in which the loss coefficient, β, is modified by fluctuations in electric field magnitude and temperature, and resultant changes in the auroral ionosphere have been outlined in section 1.2.4. Of these parameters, electric field is the dominant influence, for one of magnitude 50 mVm^{-1} often observed at the polar cap boundary, and an extreme temperature enhancement of 820°K, produce an identical response in β.

Investigations have been carried out, with a constant neutral temperature of 1000°K, for values of β from 8.8×10^{-12} to 1.6×10^{-11} cm^{-3} s^{-1}, corresponding to field strengths of 0 and 100 mVm^{-1} respectively. Typically, with β increasing over this range, maximum oxygen ion density decreased by a factor of 4 and the layer was raised by $\sim$50 km.

Recently, Schunk et al. (1976) have extended such investigations far beyond the scope of the present work by deriving ion density profiles for all important F- and E-region constituents with varying ambient electric fields. Their calculations show that such fields can cause important reductions in F-region electron densities.

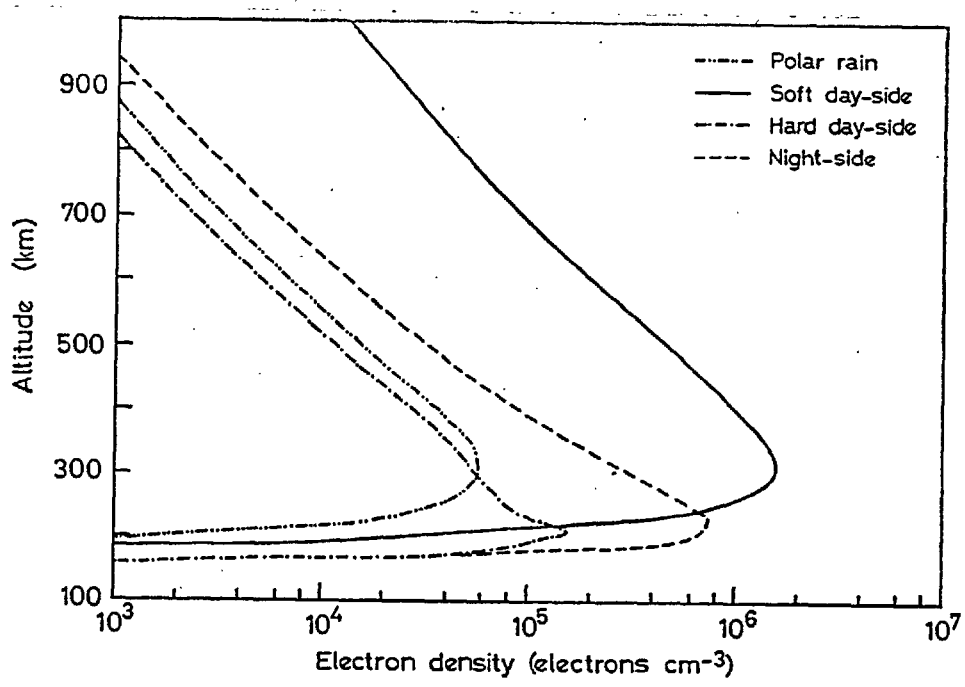


Figure 3.9 The altitudinal variation of electron density computed by the procedure outlined in Section 3.4 for the production rates presented as Figure 3.7.

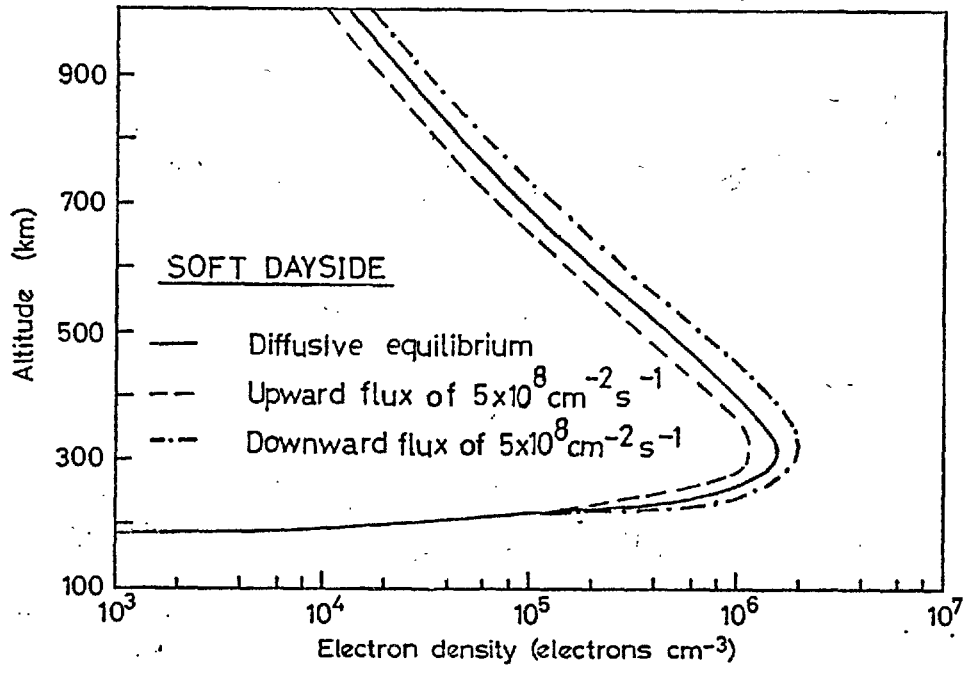


Figure 3.10 Consideration of 'polar wind' type flows of ions on the topside electron density profile. Here the diffusive equilibrium constraint on the solution has been relaxed in the soft dayside region to allow both upward and downward fluxes of ions through the upper boundary at the 1000 km level. As is apparent the effects are small. The scale height of the plasma is modified at high altitudes, being different for each curve. However, the variation is too small to be noticeable on such a scale.

Table 3.5

Maximum calculated electron densities

Region	N_{max} (electrons cm^{-3})	h_mF2 (km)
Polar cap	$5.70 \pm 0.5 \times 10^4$	310 ±15
Soft dayside	$1.55 \pm 0.2 \times 10^6$	310 ±10
Hard dayside	$1.52 \pm 0.2 \times 10^5$	212 ±10
Nightside	$7.40 \pm 0.4 \times 10^5$	235 ±10

The errors are estimated from consideration of the integration step length and uncertainties introduced by reading from the curves of Figure 3.8.

3.6 Discussion

3.6.1 The production rate calculation

As an electron penetrates into the atmosphere it experiences collisions which cause it to be scattered in two ways:

1. Elastic or Coulomb scattering from atoms and nuclei at high energies, and orbital electrons lower down the energy scale. The effect of such scattering is to deflect the particle through wide angles but leave the particle energy unchanged.
2. Inelastic scattering with orbital electrons. The incident beam is degraded in energy but is not deviated from its original trajectory by this process.

As a result of these collisions, the primary flux of electrons, as it penetrates the atmosphere, has a constantly changing energy and pitch angle distribution. In addition, ionisation and excitation of neutral

constituents can occur to produce additional or secondary ionisation which are also subject to collisional interactions.

To describe this behaviour models of increasing sophistication have been derived. At energies $\lesssim$500 eV the problem is vastly simplified by the introduction of a continuous slowing-down approximation. For the keV energies there is great uncertainty concerning specific processes involving both inner and outer shell electrons. However, the fractional incident energy lost per collision is small, and the discreteness of losses becomes unimportant in following the history of an incoming electron. Thus the problem is reduced to that of calculating the energy deposition as a function of altitude and assuming that one the creation of one ionisation pair (ion-electron) requires 35 eV. Kamiyama (1974) has used this approximation as have Berger et al. (1970, 1974) in their Monte Carlo computations for incident electrons $\gtrsim$2 keV

At lower energies ($\lesssim$400 eV), electron transport must be considered since interactions involving outer shell electrons and the discrete processes of energy loss and scattering become increasingly important. This energy range has received little attention since ionisation and excitation cross sections at energies of approximately a few hundred eV are not well known. However, recently, Banks et al. (1974) and Mantas and Walker (1976) have approached a solution from the analogous treatment of photoelectron transport. In the domain of discrete effects these aforementioned treatments use two-stream transport equations to describe elastic and inelastic scattering through the atmosphere.

The approach used in section 3.3 of this thesis utilises experimental data on the pentration of electrons into air, reported by Grün (1959) for energies 4-5keV, and extended down to 2 keV by Cohn and Caledoneia (1969). The use of experimental results avoids selection of cross-section and choice of scattering processes modelled, since all physical effects considered in theoretical investigations must be present. Rees (1969) has processed Grün's data to derive the fraction of incident energy deposited in an

absorber by electrons, as a function of its fractional penetration. He found this to be almost independent of energy in the range 4 - 5 keV and extended its applicability down to 400 eV. At energies $\geq$500 eV Banks et al. (1974) have found excellent agreement between ionisation predicted by their theoretical model and that calculated by Rees, using a continuous slowing down approximation and experimental data. This work further extends Rees' computations down to 100 eV, to take some account of the large population of electrons recently observed with energies of several hundred eV. Although large angle scattering and electron-electron interactions become increasingly important it is not expected that there are extreme differences between microscopic processes at energies $\sim$2 keV and hundreds of eV.

The shortcomings of treating discrete processes as continuous are recognised but are not considered to be so substantial as to modify the conclusions of the study. The work merely seeks orders of magnitude of electron density made possible by collisional ionisation. No information on optical excitation or evolution of the flux with altitude is attempted. Due to the variability and lack of knowledge on the polar neutral atmosphere and the derived dependence of calculated production rates on this parameter, no production computation can expect to reproduce measured densities with $\gtrsim$20% accuracy.

Uncertainties introduced by the extension of continuous losses to lower energies would need to be large to be the major source of error In general, production rates, obtained from computations involving discrete processes form broader peaks, at slightly higher altitudes and magnitudes than those derived using the continuous loss approximation. Furthermore, it is thought that adoption of theoretical models with discrete energy losses would not modify the conclusions of this study regarding the ability of soft electron fluxes to account for the observed electron density distribution.

3.6.2 Comparison of theoretical calculation with observed electron density distributions

It is appropriate at this point to determine how closely an ionosphere, created solely by the collisional effects of low-energy electrons, and the observed distribution, agree. Comparison can be made in the following three areas:-

1. Positions, in magnetic latitude and time, of the plasma ring and precipitation region boundaries.
2. Measured and calculated peak electron densities ($N_m F2$).
3. Measured and calculated peak altitudes ($h_m F2$).

Such comparisons should be carried out with topside sounder data collected near solar maximum (1967-9) since this corresponds to the epoch of low-energy particle measurements.

1. Spatial correlation The precipitation zones and the experimentally determined plasma infrastructure can be compared by examining Figures A8 (Appendix) and 2.17. The patterns are very similar and, to a first order, it would appear that the soft day- and night-side flux zones can be closely identified with the day- and night-side sections of the plasma ring. This agreement is extended to the polar cavity which co-rotates with the region where steady polar rain fluxes are found. The position of the hard day-side region is more confused, but possibly may coincide with the equatorward portion of the plasma ring.

Some 'smearing out' of the particle-produced ionisation will occur and close correspondence of the patterns is not to be expected. However, agreement between the distributions is good, especially with regard to the day-night asymmetry, evident in both diagrams.

2. Electron densities Since the soft electron precipitation and electron density distributions appear closely related, this identification can be extended to compare maximum possible electron densities computed using the appropriate form of spectrum, and electron densities at the F2 peak measured by topside sounding satellites during 1968-9. (Taken from

Chapter 2.) This comparison is set down in tabular form as Table 3.6.

Table 3.6

Comparison of observed (1968-69) and calculated densities

Region	Calculated	Observed
Polar cavity	5.70×10^4	1.20×10^5
Day-side 'ring'	1.55×10^6	4.00×10^5
Night-side 'ring'	7.40×10^5	4.20×10^5

Several conclusions are apparent from these figures.

1. Electron densities predicted for the day- and nightside plasma ring regions exceed observed densities by a factor of $\sim$4 and $\sim$2 respectively. Thus there is order of magnitude agreement but the discrepancies require further investigation.
2. The calculated cavity densities fall below those measured, by a factor of $\sim$2. Again there is order of magnitude agreement.

Further examination of the disagreement between calculated and observed densities reveals that the predicted excess noted in (1) above is not disturbing. One assumption, implicit in the computation, is that all the energy in the incident beam is used to create ionisation. Theoretical and experimental considerations (e.g. Mantas and Walker, 1976) suggest that a substantial fraction of the incident energy appears as thermal and optical excitation of the neutral and charged particle populations. The exact partition of the incoming energy into its various forms (e.g. optical emission, heat) is not known and it is probable that it is a complex function of the initial electron energy. Mantas and Walker's (ibid) computations indicate that at energies $\lesssim$100 eV only $\sim$25% of the incident spectral energy is lost in the creation of additional ionisation. Such considerations would reduce all calculated densities,

but, because of the functional dependence of energy partition on the incident electron energy, an assessment of the magnitude of the reduction is difficult. Clearly, however, it is possible that dayside ring densities may be reduced closer to the observed values while those calculated for the nightside ring region may fall below measured densities. Furthermore, the deficit between predicted and computed cavity densities, considered further in section 3.6.3,will be increased. The variability and complex nature of the polar regions means that the computational technique is a very simplified picture and many processes active in the polar regions are not considered. Among these are joule and particle heating effects. Such processes will act both to increase the loss rate through the temperature effect and enhance production through increasing the ionised neutral species density. Large local variations are expected to result from these perturbations but over wide spatial and temporal extents it is thought that the calculation can give a good idea of the levels of electron density expected.

From this and considerations regarding energy partition the agreement between observed and predicted levels of electron density in the ring is satisfactory but there appears a clear discrepancy for the region of the polar cavity.

3. F-layer peak altitudes One further parameter which permits comparison between theory and experiment is h_mF2, the altitude at which the maximum electron density occurs. h_mF2, however, is not a freely-available parameter. Topside ionograms rarely allow its unambiguous identification, especially in polar regions where they often contain blurred or multiple traces. In the case of ground stations, information on the height of the layer is generally presented in the form of monthly median values of the M(3000)F2 quantity and not direct measurements of h_mF2. At high latitudes the interpretation of M(3000)F2 is not clear but it has been used to infer h_mF2 values at high southern latitudes by King et al. (1971) from

$$h_mF2\ (km) = \frac{1490}{M(3000)F2} - 176. \qquad 3.14$$

Using this relation the average diurnal variation of h_mF2, at several Arctic stations, for the months of December and January from 1966 to 1969, has been examined to assess the accuracy of the calculated values. Subject to the availability of data for this epoch, the stations were selected so that data coverage of the polar cavity, day and night-side ring regions and the trough was provided. Details of the stations used in this analysis are provided in Table 3.7.

Tromso and Kiruna, stations lying almost entirely outside the plasma ring structure, as defined in Chapter 2, exhibited an h_mF2 variation typical of those found at mid-latitudes and described by Rishbeth and Garriot (1969). In each case the F-layer appeared at altitudes above 350 km during local night but fell to a deep minimum of $\lesssim$280 km near noon.

In contrast, the three stations at higher latitudes showed little significant statistical variation of h_mF2 over the local day. The details are summarised at Table 3.8. Resolute Bay, at an invariant latitude of 83°, lies within the polar cap at all times and, for such conditions, the calculation predicts that h_mF2 should be at a constant height of 310 km. Within the limits of experimental accuracy the constant height of the layer is observed but it is in general found $\sim$20 km higher than is predicted by theory. This descrepancy is not serious, however, since vertical drifts induced by electric fields and the neutral air wind system can act to raise the layer. The magnitude of modifications to the layer produced by these processes have been investigated by Knudsen (1974) and King et al. (1967), and they are quite capable of accounting for 20 km variations in the height of the layer.

Though the plasma ring is subject to large variations, consideration of its position and the invariant latitudes of the stations at Heiss and Dixon Islands suggest that, during local night, they are positioned beneath

Table 3.7

Station	Geographic Latitude (°N)	Geographic Longitude	Invariant Latitude	Polar regions acessible to station
Resolute Bay	75	95°W	83	Polar cavity
Heiss Island	81	58°E	75	Day- & night-side ring
Dixon Island	73	82°E	67	Trough & ring (possible)
Tromso	70	19°E	66	Trough - day & night
Kiruna	68	27°E	64	Trough - day & night

Table 3.8

Station	$h_m F2$ (Day), km	$h_m F2$ (Night), km
Resolute Bay	329 ±3	332 ±5
Heiss Island	331 ±4	337 ±8
Dixon Island	334 ±4	340 ±6

the nightside section of the plasma ring. In local daytime, the position is more complex and, although Heiss Island is expected to be beneath the cleft or dayside ring region, Dixon Island is probably positioned outside the ring region.

Computations indicate that in the day- and night-side sections of the plasma ring, - the soft and nightside precipitation zones, the maximum electron density should be located at 310 and 235 km respectively (Table 3.5). Furthermore, the measured h_mF2 for a station passing beneath these regions would exhibit a large diurnal modulation. Table 3.8 shows that at Heiss and Dixon Islands this is clearly not observed. If the aforementioned neutral air winds and electric fields influence the daytime layer then agreement in the dayside region is satisfactory,with predicted and measured vlaues of h_mF2 falling at 310 and $\sim$335 km respectively. However, there appears to be a serious descrepancy on the nightside where the calculated values find the layer at much lower altitudes than are observed. Layer modifications of the type mentioned above are not capable of producing the large changes necessary to reconcile theory and experiment. One possible explanation lies in the form of spectrum chosen for the nightside zone computations. In this region the spectrum is known to be harder than that on the dayside and much more varied. The differential energy spectrum is thought to consist of a 'soft' logarithmic part with the harder electrons of keV energies forming a Maxwellian distribution. The position in energy and magnitude of this Maxwellian section of the spectrum has a great influence on the production maximum in the nightside region. It is possible that in seeking the maximum possible electron density produced by such a spectrum the importance of the Maxwellian peak may have been overemphasised thus forcing the altitude of the production and, consequently, the electron density maximum to lower altitudes.

3.6.3 Sources of polar cavity ionisation

It is possible that the following processes, either singly, or in combination, play some part in the removal of the polar cavity electron density deficit:

1. Grazing incidence of solar ultra-violet radiation.
2. Sporadic intense fluxes of low-energy particles.
3. Movement of plasma from high-production regions.

In winter, at high latitudes, although the atmosphere at F-region heights is not illuminated by the sun, this is not the case at higher altitudes. However, the solar zenith angle, χ, is large, often $\gtrsim 100°$, and the situation is referred to as one of *grazing solar incidence*. Kamiyama (1966) has determined possible production from this source. He found that the peak of ion production, formed at very high altitudes, $\sim$800 km, could provide maximum electron densities $\sim 8\times10^4$ cm^{-3} for $\chi \lesssim 103°$. For larger zenith angles the densities produced quickly became insignificant. Thus for zenith angles <103°, grazing solar photoionisation is capable of supporting observed cavity densities. At $\chi > 103°$, however, the contribution is slight. Furthermore the plasma ring cannot be explained by this mechanism.

As outlined in the Appendix, in addition to the steady 'polar rain'

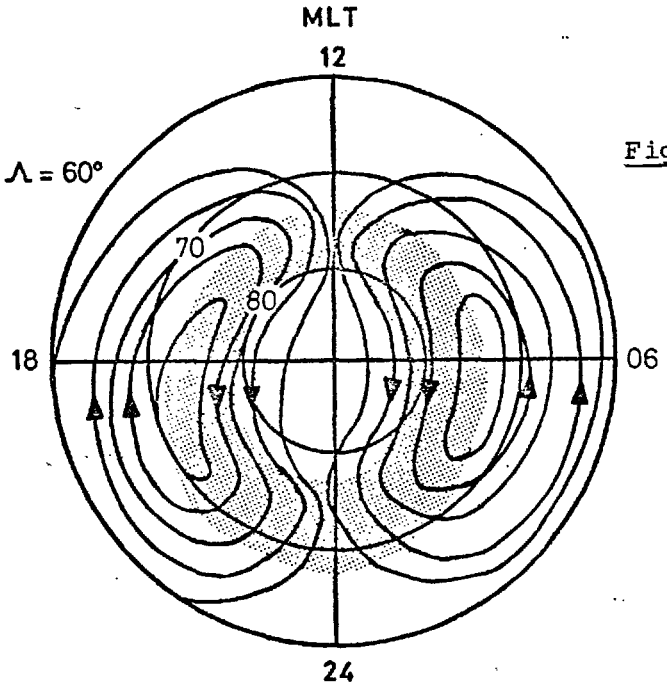


Figure 3.11 A schematic view of the possible convection of plasma from regions of high production to the polar cavity. The convection pattern is the normal two-cell magnetospheric form with modified convection paths so that field tubes spend an appreciable period in the precipitation regions. The shaded area is the high production region.

discussed previously, Winningham and Heikkila have reported two other types of electron flux. These they labelled *'polar showers'* and *'polar squall'*. Due to their temporal characteristics they have not been considered in great detail in this work, but they are often very intense. Trial computations, with Winningham and Heikkila's spectra, indicate their possible importance as a source of ionisation, for calculated production rates are comparable with those found in the soft dayside or cleft region. Unfortunately, no data concerning the occurrence frequency of such fluxes are available and an assessment of their contribution to polar cap densities is not possible.

It has been recently suggested (Knudsen, 1974) that convection of the F-layer plasma over the polar cap in response to the magnetospheric electric field plays a profound role in the behaviour of the layer. Since the exact convection pattern remains unknown, the effects are uncertain but it is probable that as magnetic field tubes convect anti-sunward, through the cleft, ionisation is added by the soft electron flux. Field tubes continue to move into the polar cap carrying this extra ionisation with them. Thus some production from the dayside soft region may be responsible for the level of polar cap electron density. Schematically, this is shown in Figure 3.11. Typical paths of field tubes, convecting through the cleft and across the cap from day- to nightside are shown. The pattern given by Knudsen has been modified so that field tubes spend a greater period of time in the soft dayside high production zone. Such a pattern ensures that appreciable ionisation can be transferred to the polar cavity. Electron densities can be enhanced by a few $\times 10^5$ electrons cm^{-3} in a flux tube as it passes through the cleft. This ionisation will of course experience chemical loss in the cavity but from an initial estimate it seems that this may be an adequate source.

3.6.4 Conclusions and future work

Previous calculations of ionisation effects of low-energy

electrons in the ionosphere have either used model spectra,not representative of those observed, e.g. Sato (1973), or have concentrated on other problems, such as the evolution of the spectrum with altitude(Banks et al. 1974). In previous sections,using recently acquired data regarding the overall structure o low-energy particle precipitation, the levels and distribution of electron density, produced solely by collisional effects of these fluxes, have been given. It has been shown that the features of the plasma ring are well explained by such processes but in the polar cavity, near the geomagnetic pole, further production sources must be active. An assessment of the importance of some possible sources indicates that plasma convection may be the most effective. Assumptions and the major errors inherent in the calculation do not appreciably modify these conclusions.

An attempt has been made to isolate parameters important in determining production rates and electron densities. The neutral atmosphere, especially its temperature,is a major influence on ion production. Thus the approach to levels of ionisation,through a continuous energy deposition function,seems a justifiable approximation,given the uncertainties in the neutral atmosphere. Before any further calculation to reproduce measured densities,using the theoretical models of Banks et al. and Mantas and Walker (1976) is made, a more definite specification of the neutral distribution is required.

Recent and future satellite experiments make further investigation of electron penetration phenomena feasible. One new approach, different from the present analysis, in which an attempt is made to reproduce average densities from average fluxes, is to examine simultaneous electron density and particle precipitation data from satellites such as ISIS II. A direct comparison,between electron densities produced and the electron spectrum,could then be obtained. However, this does not overcome the lack of information on the neutral atmosphere,but this will become better known through measurements of the Atmospheric Explorer series of satellites.

In the laboratory, further investigation is needed. Gledhill (1973) has drawn attention to the sparsity of experimental data on the range of electrons in air at energies ∿1 keV. Indeed, no work has been reported since that of Cohn and Caledonia in 1969. The experimental difficulties concerned with the measurement of particle energies of approximately a few hundred eV have recently been overcome (Storey, private communication, 1977). Extension of the work on the range-energy relation of electrons, described by Gledhill, now seems possible. This, in turn, would give additional input to the theoretical models concerning the interaction of electrons and charged and neutral particle atmospheres at very low energies, where the greatest uncertainty as to the microscopic processes lies.

References

Akasofu, S.I., Polar and Magnetospheric Substorms, D. Reidel, Dordrecht, 1968.

Allen, C.W., J.G.R. 53, 433, 1948.

Andrews, M.K., Ph.D. Thesis, University of London, 1969.

Andrews, M.K. and J.O. Thomas, Nature 221, 223, 1969.

Angerami, A.A. and J.O. Thomas, J.G.R. 69, 4537, 1964.

Axford, W.I. and C.O. Hines, Can. J. Phys. 39, 1443, 1961.

Banks, P.M. and J.R. Doupnik, Planet. Sp. Sci. 22, 79, 1974.

Banks, P.M. and T.E. Holzer, J.G.R. 73, 6846, 1968.

Hanks, P.M. and T.E. Holzer, J.G.R. 74, 6317, 1969.

Banks, P.M. and G. Kockarts, Aeronomy, Academic Press, New York, 1973.

Banks, P.M., Chappell, C.R. and A.F. Nagy, J.G.R. 79, 1459, 1974.

Banks, P.M., Schunk, R.W. and W.J. Raitt, Geophys. Res. Lett. 1, 239, 1974.

Banks, P.M., Schunk, R.W. and W.J. Raitt, Planet. Sp. Sci. 23, 1103, 1975.

Bauer, S.J., Physics of Planetary Ionospheres, Springer-Verlag, Berlin, 1973.

Bellchambers, W.H., Barclay,L.W. and W.R. Piggott, The Royal Society IGY Expedition to Halley Bay 1955-9, 179, Roy. Soc., London, 1962.

Berger, M.J.,Seltzer, S.M. and K. Maeda, J.A.T.P. 32, 1015, 1970.

Berger, M.J., Seltzer, S.M. and K. Maeda, J.A.T.P. 36, 591, 1973.

Bowman, G.G., Planet. Sp. Sci. 17, 777, 1969.

Buchau, J., Pittenger,E.W. and A.H. Sizoo, AFCRL Environmental Res. Papers No. 320, 1970.

Burch, J.L., In Polar Ionosphere and Magnetospheric Processes, Gordon and Breach, New York, 1969.

Burch, J.L.,J.G.R. 77, 6696, 1972.

Burch, J.L., Goddard Space Flight Center Maryland, Report X-646-73-390, 1973.

Cain, J.C., Hendricks, S.J., Langel, R.A. and W.V. Hudson, Goddary Space Flight Center Maryland, Report X-612-67-173, 1967.

Challinor, R.A., J.A.T.P. 32, 1959, 1971.

CIRA, COSPAR International Reference Atmosphere, North Holland, Amsterdam, 1965.

Cohn, A. and G. Caledonia, Bull. Am. Phys. Soc. 14, 523, 1969.

Dalgarno. A. J.A.T.P. 26, 939, 1964.

Deshpande, M.R., Nature 244, 109, 1973.

Dieminger, W., Ster. finn. Akad. Wiss. 199, 1959.

Diodato, L., Moreno, G., Signorini, C. and K.W. Ogilvie, J.G.R. 79, 5095, 1974.

Duncan, R.A., J.G.R. 67, 1823, 1962.

Duncan, R.A., J.A.T.P. 31, 1003, 1969.

Eather, R.H. and S.B. Mende, In Ionosphere-Magnetosphere Interactions, Universitetsforlaget, Oslo, 1972.

Fälthammer, C.-G., In Cosmical Geophysics, Universitetsforlaget, Oslo, 1972.

Feldstein, Y.I. and G.V. Starkov, Planet. Sp. Sci. 15, 209, 1967.

Ferguson, E., Rev. Geophys. Sp. Phys. 12, 703, 1974.

Gledhill, J.A., J. Phys. A: Math; Nucl. Gen. Vol. 6, 1420, 1973.

Grebowsky, J.M., Chen, A.J. and H.A. Taylor, Goddard Space Flight Center Maryland, Preprint 1974.

Grün, A.E., Zeit. fur Naturforschung 12a, 89, 1959.

Gurnett, D.A., In Earth's Magnetospheric Processes, D. Reidel, Dordrecht, Holland, 1973.

Hagg. E.L., Can. J. Phys. 45, 27, 1967.

Hall, L.A., Higgins, J.E., Chagnon, C.W. and H.E. Hinterhegger, J.G.R. 74, 4181, 1969.

Hays, P.B., Jones, R.A. and M.H. Rees, Planet. Sp. Sci. 21, 559, 1973.

Hedin, A.E., Mayr, H.G., Reber, C.A., Spencer, N.W. and C.R. Carignan, J.G.R. 79, 215, 1974.

Hedin, A.E. and C.A. Reber, J.G.R. 77, 2871, 1972.

Heikkila, W.J. and J.D. Winningham, J.G.R. 76, 883, 1971.

Heppner, J.P., In Critical Problems of Magnetospheric Physics, ed. E.R. Dyer, National Academy of Sciences Washington, 1972.

Hill, G.E., J. Atmos. Sci. 20, 492, 1963.

Hoffmann, R.A., In Magnetosphere-Ionosphere Interactions, Universitetsforlaget, Oslo, 1972.

Hopkins, H.D., Ph.D. Thesis, University of Birmingham, 1974.

Intrilligator, D.S., Astrophys. J. 188, L23, 1974

Ivanov-Kholodnyi, G.S., Nikolski, G.M., Nauka, Moscow, 1969. (Eng. Trans., NASA TTF-654, 1972).

Jacchia, L., Smithsonian Contrib. Astro. 8, 215, 1965.

Jacchia, L., Smithsonian Sp. Astro. Rep. No. 332, 1971.

Kamiyama, H., Sci. Rep. of Tohuku Univ., Series 5: Geophys. 18, 1, 1966.

Kamiyama, H., Rep. Ion. Space Res. Japan 20, 171, 1974.

King, J.w., Eccles, D. and H. Kohl, J.A.T.P. 33, 1067, 1971.

King, J.W., Kohl, H. and R. Pratt, J.A.T.P. 29, 1529, 1967.

Kohl, H. and J.W. King, J.A.T.P. 29, 1045, 1967.

Knecht, R.W., J.G.R. 64, 1243, 1959.

Knudsen, W.C., J.G.R. 79, 1046, 1974.

Lindinger, W., Fehsenfield, F.C., Schmeltekopf, A.L. and E. Ferguson, J.G.R. 79, 1974.

Liszka, L., J.A.T.P. 27, 1243, 1967.

Lui, A.T.Y., Anger, C.D. and S.I. Akasofu, J.G.R. 80, 3603, 1975.

Maeda, K., J.A.T.P. 27, 259, 1965.

Maeda, K., J.A.T.P. 38, 197, 1976.

Mantas, G.P., Planet. Sp. Sci. 24, 409 , 1976.

Mantas, G.P. and J.G. Walker, Planet. Sp. Sci. 23, 355, 1975.

Maehlum, B.N., J.G.R. 73, 3459, 1968.

McDiarmid, I.B., Burrows, J.R. and E.E. Budzinski, Preprint, 1975.

McIlwain, C., J.G.R. 65, 2727, 1960.

Miller, N.J., J.G.R. 75, 7175, 1970.

Miller, N.J., J.G.R. 79, 3795, 1974.

Mishin, V.M., Saifudinova, T.I. and I.A. Zhulin, J.G.R. 75, 797, 1970.

Muldrew, D.B., J.G.R. 70, 2635, 1965.

Neugebauer, M., Space Sci. Rev. 17, 221, 1975.

Nishida, A., J.G.R. 72, 6051, 1967.

Nishida, A. and T. Obayashi, In Critical Problems of Magnetospheric Physics, Ed. E.R. Dyer, National Academy of Science, Washington, 1972.

Oguti, T. and K. Marubashi, Rep. Ion. Sp. Res. Jap. 20, 96, 1966.

Patton, D.E., Peterson, V.L., Stonehocker, G.H. and J.W. Wright, Geomag. and Aeron., Ant. Res. Ser. 4, A.G.U., 1965.

Penndorf,R., Geomag. and Aeron., Ant. Res. Ser. 4, A.G.U., 1965.

Phelps, A.D.R., Int. Rep. Imp. Coll., 1972.

Phelps, A.D.R. and J.O. Thomas, Int. Rep. Imp. Coll., 1974.

Pike, C.P., J.G.R. 75, 4871, 1970.

Pike, C.P., J.G.R. 76, 6875, 1971a.

Pike, C.P., J.G.R. 76, 7745, 1971b.

Pike, C.P., In Air Force Cambridge Res. Labs. Rep., No. AFCRL-72-0305, 1972a.

Pike, C.P., J.G.R. 77, 6911, 1972b.

Raitt, W.J., Schunk, R.W. and P.M. Banks, Planet. Sp. Sci. 23, 1103, 1975.

Reber, C.A. and H. Hedin, J.G.R. 79, 2457, 1974.

Rees, M.H., Planet. Sp. Sci. 11, 1209, 1963.

Rees, M.H., Planet. Sp. Sci. 12, 722, 1964.

Rees, M.H., Sp. Sci. Rev. 10, 413, 1969.

Rishbeth, H., J.A.T.P. 34, 1, 1972.

Rishbeth, H., J.A.T.P. 39, 111, 1977.

Rishbeth, H. and O.K. Garriot, Introduction to Ionospheric Physics, Academic Press, New York, 1969.

Rostoker, G., Sharma, R.P. and M.P. Hron, Planet. Sp. Sci. 24, 1081, 1976.

Sandford, B., In The Polar Ionosphere and Magnetopsheric Processes, Gordon and Breach, New York, 1969.

Sato, T., Rep. Ion. Sp. Res. Jap. 13, 91, 1959.

Sato., T., Preprint, Tokyo University, 1973.

Sato, T. and L. Colin, J.G.R. 74, 2193, 1969.

Sato, T. and G.F. Rourke, J.G.R. 69, 4591, 1964.

Schatten, K.H., Rev. Geophys. Sp. Phys. 9, 773, 1971.

Schunk, R.W., Raitt, W.J. and P.M. Banks, J.G.R. 80, 3121, 1975.

Schunk, R.W., and J.G. Walker, Planet. Sp. Sci. 20, 581, 1972.

Schunk, R.W., Banks, P.M. and W.J. Raitt, J.G.R. 81, 3271, 1976.

Sharp, R.P., J.G.R. 71, 1345, 1966.

Shepherd, G.G., Whitteker, J.H., Winningham, J.D., Hoffmann, J.H., Maier, E.J., Brace, L.H., Burrows, J.R. and L.L. Cogger, J.G.R. 81, 6092, 1976.

Smith, F.L. III and C. Smith, J.G.R. 77, 3592, 1972.

Stoffregen, W., In Magnetosphere-Ionosphere Interactions, Universitets-forlaget, Oslo, 1972.

Svalgaard, L., SUIPR Report 521, 1973.

Taylor, H.A., Planet. Sp. Sci. 20, 1593, 1972.

Taylor, H.A. and W.J. Walsh, J.G.R. 77, 6716, 1972.

Taylor, G.N., J.A.T.P. 35, 647, 1973.

Thomas, J.O. and M.K. Andrews, J.G.R. 73, 7407, 1968.

Thomas, J.O. and M.K. Andrews, Planet. Sp. Sci. 17, 433, 1969.

Thomas, J.O. and M.K. Andrews, In Polar Ionosphere and Magnetospheric Processes, Gordon and Breach, 1970.

Thomas, J.O. and A.Y. Sader, J.G.R. 69 4561, 1964.

Thomas, J.O., Rycroft, M., Colin, L. and K.L. Chan, The topside ionosphere: I. The analysis of Alouette I topside ionograms II. Experimental results from Alouette I, In Electron Density Profiles in Exosphere and Ionosphere, North Holland, Amsterdam, 1966.

Thomas, J.O., Rycroft, M. and L. Colin, Electron Densities and Scale Heights in the Ionosphere, NASA SP 3026, 1967.

Titheridge, J.E., Tech. Rep. 74/1, Radio Res. Centre, Univ. Auckland, August 1974.

Wagner, R., Snyder, A.L. and S.I. Akasofu, Planet. Sp. Sci. 21, 1911, 1973.

Willmore, A., J.A.T.P. 36, 2255, 1974.

Winningham, J.D. and W.J. Heikkila, J.G.R. 79, 949, 1974.

Winningham, J.D., Yasukara, F., Akasofu, S.I. and W.J. Heikkila, J.G.R. 80, 3148, 1975.

Whitteker, J.H., J.G.R. 81, 1279, 1976.

Whitteker, J.H., Brace, L.H., Maler, E.J., Burrows, J.R., Dodson, W.H. and J.D. Winningham, Planet. Sp. Sci. 24, 25, 1976.

Wolfe, J.H., In Solar Wind 3, NASA SP-308, 1972.

Wright, J.W., Nature 194, 462, 1962.

Yonezawa, T., Sp. Sci. REv. 5, 3, 1966.

APPENDIX

LOW-ENERGY PARTICLE PRECIPITATION - A SURVEY

CONTENTS

ABSTRACT

At high invariant latitudes, $\geq 60°$, large fluxes of low-energy ($\sim$0.1 -10 keV) electrons and protons have been noted entering the upper atmosphere. Measurements of the characteristics of these particles are reviewed including those obtained from both 'direct' satellite observations and 'indirect' methods, such as analysis of optical emissions. An 'average' picture of such precipitating fluxes is derived.

1. Introduction

An influx of particles at high latitudes has long been postulated as the cause of the various auroral emissions. Early measurements to determine the characteristic properties of these particles were made by flying particle detectors on board rockets into visible auroral forms. These observations revealed that the particles responsible for such emissions were electrons and protons of energies in the range of a few - 10 keV. The advent of satellites enabled more detailed information to be obtained and the improvement in detection techniques allowed observations to be made at progressively lower energies, so that measurements are now possible over the entire magnetospheric particle energy range from thermal energies $\sim$ 1 eV to $\sim$50 keV. These data indicated that at very low energies, typically a few hundred eV, there were considerable fluxes of particles. Calculations (Rees, 1969) have shown that such particles deposit much of their energy at F-region altitudes and thus any study of the high latitude F-region ionosphere demands a knowledge of the properties of such particles so that their influence can be clearly specified. In addition it has been postulated (Hoffman, 1972) that these low-energy particles are responsible for many other interesting phenomena both above and in the ionosphere, such as the generation of VLF hiss, magnetic purturbations, and various optical emissions.

The following sections contain a review of the most important of these observations which are pertinent to F-region physics. Section 2 details the so-called 'direct' techniques which are largely satellite-based. A few of the most successful 'indirect' techniques are included in Section 3. Finally in Section 4, an average, in so far as it is possible in dealing with such variable processes, picture of quiet-time incident particle fluxes is obtained. The Appendix collects many of the terms and summarises units common in reports of such experiments.

2. Direct Measurements of Precipitating Particle Fluxes

2.1 Introduction

Now the most commonly used method of investigating the structure of high latitude particle precipitation is that of satellite-borne particle spectrometers. From rather crude instruments with high (a few keV) energy thresholds these detectors have advanced considerably and are now capable of detecting particles with energies from a few electron-volts to approximately fifty kilo-electron-volts. They are now carried as a matter of routine on most satellites investigating the ionospheric/magnetospheric plasma, and detector and satellite specifications of the most successful of these are found in Table A1. In fact detectors, such as the LEPEDEA (low energy proton and electron differential energy analyser) and the SPS (soft particle spectrometer), have become so widely used that it is instructive to describe the principles of operation of one such example - the Soft Particle Spectrometer flown on ISIS I and II.

2.2 The ISIS Soft Particle Spectrometer (SPS)

This instrument (Heikkila et al., 1970), shown schematically in Figure 1a, was designed to count electrons and protons with energies from 10 - 10000 eV. The instrument consists of two collimating apertures, placed perpendicularly as in Figure 1b, and these provide both an axial and radial beam positioned on the the satellite as shown. Energy selection is controlled by electrostatic deflection and the energy range is sampled in 19 steps (including 0 eV). Each step is held for 1 second, during which time electrons and protons in the two perpendicular directions are counted for 0.5 seconds. An alternative mode allows a fast energy sweep to provide quicker spectrum coverage.

Pitch angle information can be obtained in two ways. One is to compare simultaneous data from both the radial and axial beams but better pitch angle information results can be obtained from the spin of the satellite allowing the radial beam to scan through the sky.

Counting is accomplished by means of electron multipliers which for the detection of 1 primary particle provide a voltage pulse suitable for amplifying and counting.

The basic advantage of the SPS is that its radial beam geometric factor of 1.2 10^{-3} cm^2 ster is considerably larger than that of comparable instruments. This coupled with a wide energy passband ($\Delta E/E \simeq .8$) gives a very large count rate and this sensitivity allows measurements of very weak particle fluxes to be made.

2.3 Early Measurements

The first experiment to measure electron fluxes (Evans et al., 1964) detected integral fluxes of electrons at energies $\geq$80 eV and $\geq$20 keV respectively. Such detectors flown by Sharp et al. (1969) and others on board the satellites Aurora I (Burch, 1968, 1969), and Ogo 4 (Hoffman 1969, 1972) revealed several distinct regions of particle precipitation.

Around the magnetic pole, at $\gtrsim$80°, a region, which has become known as the polar cavity (Maehlum, 1968) was discovered. Here, (except for occasional bursts), there were no detectable fluxes, and these observations placed an upper limit on the energy influx for this region of .001 ergs cm^{-2} $ster^{-1}$ s^{-1} (Burch, 1969).

Surrounding the cavity, on the dayside, early data suggested that these were two adjacent but distinct zones of precipitation, each a few degrees in latitudinal (north-south) extent. The poleward zone, very intense and with great structure, extended from $\Lambda \sim 80°$ to $\Lambda \sim 75°$ and has been termed the 'burst' (Hoffman, 1969) or 'soft' (Burch, 1968; Eather, 1969) zone. Electrons in this region were of very low energies, typically approximately a few hundreds of eV and deposited $\sim$.16 ergs cm^{-2} s^{-1} of ($.1 \leq E \leq 1.3$ keV) energy into the atmosphere. Only pitch angle information was available at $\sim$2.3 keV energy (Hoffman, 1969). This indicated that fluxes in the burst region tended toward anisotropy.

However, measurements at this energy may not have been typical of the lower energy regime of the spectrum.

The low latitude region was of very different character. Here fluxes were steady and weak but were of markedly higher energies - typically approximately a few keV. Burch (1968) designated this region, which stretched from $\Lambda \gtrsim 69°$ to $\Lambda \gtrsim 75°$ the 'hard' zone which corresponds to Hoffman's (1969) 'band' region. However, unlike the soft zone this precipitation did not extend through all dayside hours but often terminated around local noon. Of these experiments only that of Sharp and Johnson (1968) was sensitive enough to protons and this revealed the possibility of two proton precipitation zones coincident with electron precipitation. An illustration of the early data and differing spectral characteristics observed in these differing precipitation regions is shown in Figure 2. Here, Aurora 1 observations for a north polar pass are reproduced together with 6-point energy spectra sampled at designated points along the satellite track.

During local night behaviour appeared more complicated. Again precipitation was detected equatorward of the polar cavity but fluxes were of higher average energy than on the dayside. Sharp et al. (1969) and Hoffman (1969) found it difficult to delineate hard and soft zones as on the dayside. Instead they outlined a single precipitation domain at invariant latitudes $\Lambda \gtrsim 65°$ to $\Lambda \gtrsim 70°$ in which the average energy increased with decreasing latitudes. Thus the proportion of low energy electrons present in the flux decreases with decreasing latitude. Burch's data, however, again allowed separation into high (soft) and low (hard) latitude zones analogous to those on the dayside. He measured the precipitated energy flux to be $\sim.008$ ergs cm^{-2} s^{-1} in the high latitude region. Hoffman and Burch (1973) discovered that measured spectra were approximately isotropic over the downward hemisphere. However, Hoffman has reported very intense bursts of field-aligned precipitation. These have since been related to the

inverted 'V' events of Frank and Ackerson (1971, 1972) (Hoffman and Berko, 1972). They are temporally very variable and of limited latitudinal extent (typically $\sim 1 - 1\frac{1}{2}°$ invariant latitude, Λ). In this work they have not been considered in great detail since their spatial extent is too small to have any appreciable effect on the global distribution of electron density above $\Lambda = 60°$

2.4 Results of Frank and Co-Workers

Frank and his co-workers have derived much detailed information on the distribution of particle fluxes and their characteristics through investigations with the LEPEDEA (low energy proton and electron differential energy analyser) instrument mounted on a variety of satellites, notably INJUN V and Ariel IV.

They have concluded that with gross simplification the over-all character of charged particle distributions at high latitudes can be described in terms of two major zones, one located poleward and the other located equatorward of the trapping boundary for energetic electron ($E > 4.5$ keV) intensities. Equatorward of this boundary at invariant latitudes $\sim 70 - 76°$ (dayside) and $64 - 70°$ (nightside) there is a population of electrons with typical energies in the 1 - 3 keV range. These fluxes are broad, diffuse and steady and are more intense ($\sim 10^4$ electrons cm^{-2} s^{-1} $ster^{-1}$) during local midnight and early morning. Intensities seem to decrease as the average electron energy increases. In the early evening sector this zone contracts being typically 1 - 2° of invariant latitude in width. Poleward of this region appears a region of soft precipitation. Here data show that fluxes of electrons with energies ~ 200 eV form a background for the so-called very intense ($10^9 - 10^{10}$ electrons cm^{-2} s^{-1} $ster^{-1}$) 'inverted-V' events. These form narrow bands of precipitation that are characterised by increasing average energy to a maximum (approximately a few keV) with a subsequent decrease in average energy as the satellite passes through the band. These events, located just poleward of the

trapping boundary, increase in occurrence frequency, width (20-30 km), intensity and average energy from local noon to a maximum near local midnight. Typical energy deposited by such an event is $\sim$10 ergs cm^{-2} s^{-1} $ster^{-1}$. In addition, protons were also detected by the LEPEDEA including soft protons similar to those detected by ISIS I in the high latitude region and often more energetic protons at low latitudes, especially in the evening local time sector.

An Ariel 4 mounted LEPEDEA has also revealed the very complex nature of the pitch angle structure of these particle fluxes (Craven and Frank, 1975a, b; Frank et al., 1975). The data show that for all local time sectors fluxes equatorward of the trapping boundary were approximately isotropic outside the loss cone except toward the equatorward boundary where the average energy begins to increase and the pitch angle distribution begins to peak at 90°. Poleward of the trapping boundary the pitch angle distribution encountered depended on the local time sector sampled. In the post-midnight sector, all fluxes measured exhibited approximate isotropy. However, during the local day, local evening and around midnight there often occurred sporadic field-aligned intensities rising out of an isotropic background. Field-aligned intensities reached a maximum at energies approximately a few keV and at energies greater than this fluxes became entirely isotropic. At energies $\lesssim$ a few hundred eV, however, there were strongly field-aligned intensities. 'Inverted-V' events displayed similar complexity with isotropic pitch angle distributions in the centre bounded by regions showing strongly field-aligned intensities.

2.5 The results of Heikkila and co-workers

Due to its sensitivity (section 2.2), Heikkila and his co-workers using the SPS, were able to obtain significant counting rates over the polar cavity (section 2.3) where previous detectors had registered counting rates below threshold levels. Winningham and Heikkila (1974)

have divided the fluxes observed in this region into three types, according to their temporal and spatial properties. Very soft (energies $\sim$100 eV) and weak electron fluxes ($\sim 10^5$ cm^{-2} ster^{-1} eV^{-1} s^{-1}) were present at all times and over the entire polar cap region. These so-called 'POLAR RAIN' fluxes (Winningham and Heikkila, 1974) were isotropic, except for a single loss cone when the instrument looked down the magnetic field line, and carried an energy flux $\sim 10^{-3}$-10^{-2} ergs cm^{-2} s^{-1}.

Further fluxes of limited spatial extent (a few tens - 300 km) were measured sporadically in the polar cap and named 'POLAR SHOWERS'. These precipitating fluxes were more intense than 'polar rain' and electrons of energies $\sim$100 - 200 eV, assuming an isotopic pitch angle distribution carried an energy flux from 10^{-2} - 10 ergs cm^{-2} s^{-1}. The 'POLAR SHOWERS' spectra appeared to change in character as the satellite passed from dayside into darkness, becoming harder and of greater spatial extent. The final precipitation type noted in the polar cavity was present only during times of enhanced geomagnetic activity. This was an intense flux of electrons with energies of a few keV. Winningham and Heikkila's (1974) data showed that this region of fluxes, which they designated "POLAR SQUALL", extended for $\sim$5°Λ poleward of the soft zone, and carried a total energy flux of $\sim 10^{-1}$ ergs cm^{-2} ster^{-1} s^{-1}.

ISIS I also confirmed the existence of the soft zone on the dayside. During dayside hours from 0800 - 1000 MLT from invariant latitudes 75°-80°, Heikkila and Winningham (1971) detected a very soft but intense flux of electrons interspersed with higher energy bursts. As is seen in Figure 3 the peak in the spectrum occurs near 100 - 200 eV and at energies below this peak there was often a minimum at $\sim$80 eV followed by a steep increase to very high flux values at even lower energies. Heikkila and Winningham (1971) attributed these observations to the direct entry of solar wind plasma down the magnetic cleft or cusp formed by the neutral lines in the geomagnetic field. Through this mechanism the solar wind was seen to

deposit an energy ∿2 ergs cm^{-2} $ster^{-1}$ s^{-1}. In addition, a coincident proton flux of average energy ∿300 eV was also detected but both number and energy flux levels were typically two orders of magnitude below those associated with electrons. As far as Heikkila and Winningham's (1971) data were able to show, these fluxes appeared isotropic over the downward hemisphere. Equatorward of this region observations of a hard zone similar to that specified by Burch (1968) were reported, where fluxes, at least at auroral latitudes were isotropic over the downward hemisphere (Heikkila, 1972).

ISIS I measurements have also been reported in the 2100-0300 MLT sector (Winningham et al., 1975). Although this investigation was largely concerned with geomagnetic substorms a quiet-time picture was presented. From $\Lambda \sim 75°$ to $\Lambda \sim 70°$ there appeared a region of cleft-like precipitation (average energy and number fluxes very similar to those in the cleft). However, the fluxes on the nightside were an order of magnitude less intense and possessed a trapped pitch angle distribution (i.e. a peak at $\alpha_p = 90°$). Equatorward of this region, at invariant latitudes 70°-66°, the spectrum hardened appreciably to an average energy ∿800 eV and then softened to ∿30eV. Here the pitch angle distribution appeared isotropic for energies <1 keV and trapped for greater energies. At lower latitudes a weak trapped flux of electrons with energies >1 keV was apparent and was probably the low energy component of the Van Allen Radiation Belts. Proton fluxes of somewhat higher energies were present in both of the first two regions described above. Typical examples of these flux types observed with the SPS are shown in Figures 3 and 4.

3. Indirect Observations

3.1 Introduction

Hartz and Brice (1967) first demonstrated that by global mapping of particle related phenomena it was possible to indicate patterns of electron and proton precipitation. They utilised various effects including auroral absorption, VLF emissions and optical emissions observed from the ground and deduced that on both day and nightside there were basic types of particle flux separated in latitude.

3.2 Distribution of particle fluxes

Work later than Hartz and Brice (1967) has developed studies of (Eather, 1969; Eather and Mende, 1972) particle excited optical emissions both from aircraft and the ground and it is now possible, using theoretical calculations on particle excitation, to derive a great deal of information on the particle fluxes from the study of the latitudinal distribution and intensity of certain spectral lines. The most important Lines are the 4861Å H emission, N_2^+ emissions at 4278Å and 3914Å and the O^+ oxygen lines at 5517Å and 6300Å. Table A.2 indicates how these emissions can be used to derive the precipitating particle species, total energy flux and average energy of the flux.

A useful summary of such measurements given by Eather and Mende (1972), indicated satisfactory agreement with satellite observations. From a great deal of data, they determined the average properties of precipitating particles and their global distribution. Dayside electron patterns showed a demarcation into 'hard' (a few keV) and 'soft' (100 - 200 eV) zones with protons ∿300 eV present in the high latitude 'soft' electron zone. The probable average energy fluxes are ∿.1 ergs cm^{-2} s^{-1} $ster^{-1}$ for the 'hard' and 'soft' electrons and ∿.05 ergs cm^{-2} s^{-1} $ster^{-1}$ for protons. The lower latitude proton zone proposed by Sharp and Johnson (1969) was not indicated by photometry.

Optical data obtained during local night were interpreted in terms of a broad region of low-intensity electron precipitation extending from the polar cavity to the trough region. The average energy of these electrons increased with decreasing latitude from 500 eV to 1.5 keV. Superimposed on these fluxes at lower latitudes there were high intensity fluxes of energies approximately a few keV. Thus the position appeared more complicated than on the dayside. The soft fluxes apparently extended right across the precipitation region with high energy electrons present only at lower latitudes. Protons of approximately a few keV energy were indicated by Hβ emissions equatorward of the 'hard' electron zone during local evening moving poleward until in the post-midnight sector the proton and 'hard' electron zones were concurrent.

Precipitation patterns for electrons and protons in polar regions derived from such measurements are shown in Figure 5. Question marks indicate zones of uncertainty and the Q= 1 (quiet-time) auroral oval is superposed for reference. Different source regions were proposed for day and nightside fluxes and hence the precipitation pattern suggested consisted of two 'horse shoes' with discontinuities at 0600 and 1800 MLT.

However, recent data from the ISIS II Auroral Scanning Photometer at wavelengths of 4278 Å and 5577 Å (Liu et al., 1975) have shown that the diffuse and discrete auroral emissions form clear and continuous adjacent bands approximately along the 70° invariant latitude contour. Collaborative observations have been reported by Buchau et al. (1972), who followed the discrete and diffuse aurora for complete cicumpolar aircraft flights. Hence, since optical emissions are continuous it appears that the gaps of Figure 5 near 0600 and 1800 MLT must be filled in to form continuous precipitation regions.

3.3 Solar cycle variations

Eather and Mende's (1972) data indicated no measureable fluxes in the polar cavity and placed an upper limit on the precipitated energy

flux of $\lesssim$.001 ergs cm^{-2} s^{-1} $ster^{-1}$. Sandford (1969) has suggested, however, that this may not be the case at times of high solar sunspot number. From a collection of optical data for a time period extending over the International Geophysical Year (1957-8) to a solar minimum epoch at 1962-3, Sandford has shown the occurrence of very large intensity and latitudinal variations. Of especial interest were the very high intensities of the 6300 Å and 3914 Å emissions over the polar cavity *inferred* from IGY data. Typical data are reproduced in Figure 6 and, despite the lack of data at high latitudes, show the possibility of considerable levels of emission and hence particle fluxes in the polar cavity at high sunspot number. This effect has not since been observed, however, and was possibly due to the abnormally high sunspot number of that period.

4. An Overall View

There still remain unresolved inconsistencies in data concerning the influx of low-energy particles not least among these being observations in the soft dayside region or cusp. Here Frank and Ackerson (1971) see a zone of 1° latitudinal width with field-aligned fluxes. Heikkila and Winningham (1971), however, note a region of isotropic electron and proton fluxes 5° in latitudinal extent. Such observations and the very varied nature of patterns and structure of precipitating particle fluxes make any statistical analysis of related phenomena or calculations involving such parameters very difficult. To make such problems tractable an averaged picture of this mass of observations, involving gross simplification, has been attempted.

The first of these is to neglect protons. Recent observations from the Atmospheric Explorer-C satellite at energies $\geq$200 eV (Torr et al., 1976), summarised in Figure 7, show that the total high latitude energy flux, carried by protons, between 250 and 300 km, is often two orders of magnitude less than that carried by electrons. On average it is found that protons carry only 2-4% of the energy in the total energy flux. In addition, protons, on entering the neutral atmosphere, often undergo charge exchange reactions such as:

$$H^+_{fast} + H_{thermal} \longrightarrow H_{fast} + H^+_{thermal}$$

$$H^+_{fast} + O_{thermal} \longrightarrow O_{fast} + H^+_{thermal}$$

and thus possess short lifetimes. Hence they are an unimportant factor in most particle effects which originate lower in the atmosphere ($\lesssim$1000 km).

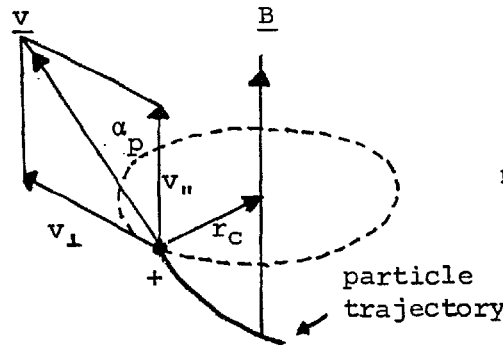
For the purposes of further analysis a second simplification of dividing the polar regions into the four precipitation zones, shown in Figure 8, is made. The characteristics of electron fluxes found in these regions are tabulated in Table A.3. Typical spectra for the regions are shown in Figure 9. That for polar rain was taken from Winningham and

Heikkila (1974) whilst the dayside and cusp and nightside spectra were averaged from those presented by Winningham et al. (1973). Where curves become discontinuous fluxes drop below sensitivity levels. The dayside hard spectrum was derived from Hoffman (1972). Here an artificial cut-off is applied since the lower energy limit shown was the lowest energy at which Hoffman's detector was operable.

In this account only events of a temporally 'steady' nature have been included. Fluxes such as polar showers and inverted 'V'-events have been excluded. However, spectra have been published for such events and hence computations as to their effects can be made. However, simplification necessitated their removal from the pattern presented here and it is considered unlikely that they contribute significantly to the average polar F-region electron density.

Units and Definition of Particle Fluxes

For a charged particle moving in a magnetic field there can be three types of periodicity; cyclotron motion, bounce motion and drift motion. To a first approximation in a homogeneous field the precipitating particles can be regarded as entering the atmosphere spiralling down field lines. They are then characterised by their mass, velocity or energy and pitch angle. The radius of this spiral is given by the particle gyro-radius.



Spiral motion of a positive particle around a magnetic field line

The pitch angle α_p is the angle between the velocity vector and the magnetic field.

$$\sin \alpha_p = v_{\perp}/v \qquad \text{A1}$$

Precipitating particles

A term often referred to in particle measurements is the loss cone. Those charged particles which are lost into the atmosphere, as opposed to those which are reflected back into the magnetosphere by the convergent geomagnetic field at decreasing altitudes, are said to populate a loss cone. The loss cone is defined at any location with a magnetic field B as any pitch angle $\alpha_p < \alpha$ where

$$\sin^2\alpha = B/B_M \qquad \text{A2}$$

where B_M is the maximum field likely to be encountered. Particles with pitch angles less than α are called precipitating particles. For auroral

precipitating particles the equatorial loss cone is ∿3° in angular extent.

The count rate of a particle detector, R, is given by

$$R = G \int_{E_{min}}^{E_{max}} \eta\,(E)\; J(E)\; dE$$

where G is the geometric factor of the detector

$\eta(E)$ is the efficiency of the detector or the number of counts for each incident particle

E_{max} and E_{min} are the upper and lower bounds of the energy passband.

$J(E)$ is the differential directional flux.

The differential directional flux $J(E,\alpha_p)$, is the number of particles in a unit energy interval and unit solid angle crossing unit area perpendicular to the velocity direction in unit time. Hence if dn particles impinge on area ds, in time dt and solid angle $d\Omega$

$$dn = J(E,\alpha_p) ds\; d\Omega\; dE\; dt$$

Units are $[cm^2\ s\ ster\ eV]^{-1}$. This is the quantity usually specified when a particle energy spectrum is presented.

The integral particle flux, F, is the total number of particles incident on a sphere of unit cross-section in unit time from all directions between energies E_{min} and E_{max}.

$$F = \int_{E_{min}}^{E_{max}} \int_{4\pi} J(E,\alpha_p) d\Omega\; dE \qquad \text{A3}$$

Units are $[cm^2\ s]^{-1}$

The integral energy flux, T, is the total energy of all particles between E_{max} and E_{min} incident on unit area in unit time. Hence

$$T = \int_{E_{min}}^{E_{max}} \int_{4\pi} E\, J(E,\alpha_p) d\Omega\, dE \qquad A4$$

Units are ergs cm^{-2} s^{-1}.

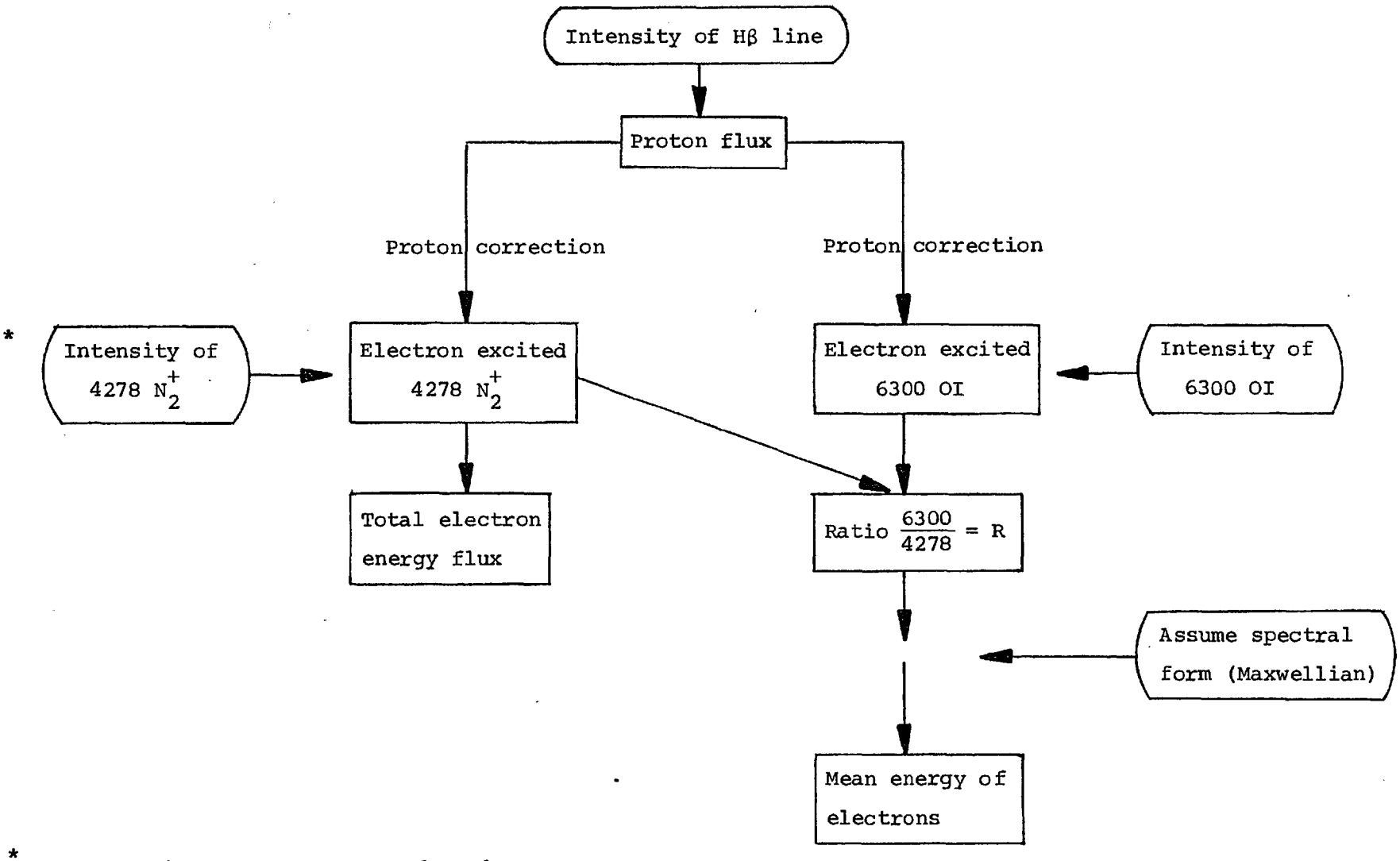
Hard and soft electrons The terms 'hard' and 'soft' are generally subjective classifications of the particle energy. In the regime of F-region physics, electrons of energy $\gtrsim$1 keV are said to be hard, while those below this energy are designated soft. However, this is not a strict rule and in some literature the transition between soft and hard particles may be as high as a few keV. An example of spectral division into hard and soft components is given in Figure A4.

Thus the spectral descriptions soft- and hard-dayside, etc. become self-explanatory, with the former adjective referring to the energy content and the latter the local time sector in which it is found.

Table A.1

Satellites carrying low-energy particle detectors								
		1963 - 42A and 1965 - 90A	AURORA 1	OGO 4	INJUN 5	ISIS 1	ATMOSPHERIC EXPLORER C	ATMOSPHERIC EXPLORER D
Launch date			29.6.67	28.7.67	8.8.68	30.1.69	16.12.73	March 1975
Perigee (km)		∿300	3808	412	677	570	150 (variable through thrust capability)	
Apogee (km)		∿300	3928	908	2528	3500	4000 (variable through thrust capability)	
Inclination (°)		89.9 (94 for 1965 - 90A)	89.92	86	81	88.5	68.4	98
Type of detector		Plastic scintillator and metal foil	← Electrostatic selection of particle energy and species →					
Energies and particles	Electrons	Total energy in electrons of $E \geq 2$, 10, 21 keV $.08 \leq E \leq 1.6$ keV	3 channels 46 - 280 eV 540 - 2400 eV 6200-25000 eV	4 channels with peak energies at 700 eV 2300 eV 7300 eV 2400 eV	40 - 12000 eV	10 - 12000 eV (22 bands)	.2 - 25 keV in 16 bands. Fixed energy detector at 5 keV (AEC + AED). Additional fixed energies, .2, .38, .72, 1.4, 1.9, 2.6, 6.9, 18.0 keV (AED only).	
	Protons	Total energy in protons of $E \geq 4000$ eV			50 - 1500 eV		.2 - 25 keV (in 16 bands)	
Resolution (ΔE/E)		-	-	16%	25%	40%	30%	30%
Pitch angle information (α_p°)		Pitch angles sampled at 0, 60	All pitch angles accepted <30	Pitch angle ranges sampled centred on 0, 30, 60, 90	Pitch angle ranges sampled centred on 0, 90	Pitch angles sampled in range 0 - 180 obtained by satellite spin	$0 \leq \alpha_p \leq 180$ when spinning. α_p = 45 (only) when oriented.	As AEC when spinning. When oriented 2(45, 60) and 4(±7, 35,45,60) pitch angles sampled
Additional Experiments		None	Various photometers	Ion mass spectrometer	DC electric field sensor	Topside sounder; Scanning photometers; Langmuir probes; Ion probe.	Various spectrometers (ion, neutral, uv, photoelectron); Langmuir probes; Photometers.	

TABLE A.2



* Here the figures refer to wavelengths measured in ångström units.

STRUCTURE OF HIGH LATITUDE ELECTRON PRECIPITATION				
	Polar cavity	Soft dayside	Hard dayside	Nightside
Latitudinal extent ($\Lambda°$)	$\gtrsim$81	∿77 - 81 (afternoon) ∿75 - 82 (dawn)	∿68 - 75 (dawn) ∿70 - 74 (noon)	66 - 75
Local time extent	All LT's	0600 - 1800	0600 - 1400	1800 - 0600
Average particle energy (keV)	∿0.10	∿0.20	∿3 - 5	Decreases with increasing latitude from ∿5 - .5
Integrated energy content (ergs $cm^{-2}ster^{-1}s^{-1}$)	0.003	0.1 - 0.2	0.1	Average ∿.15
Pitch angle distribution	Isotropic	Isotropic?	Isotropic	Variable
Temporal behaviour	Steady	Very variable	Steady, diffuse	Variable

Table A.3

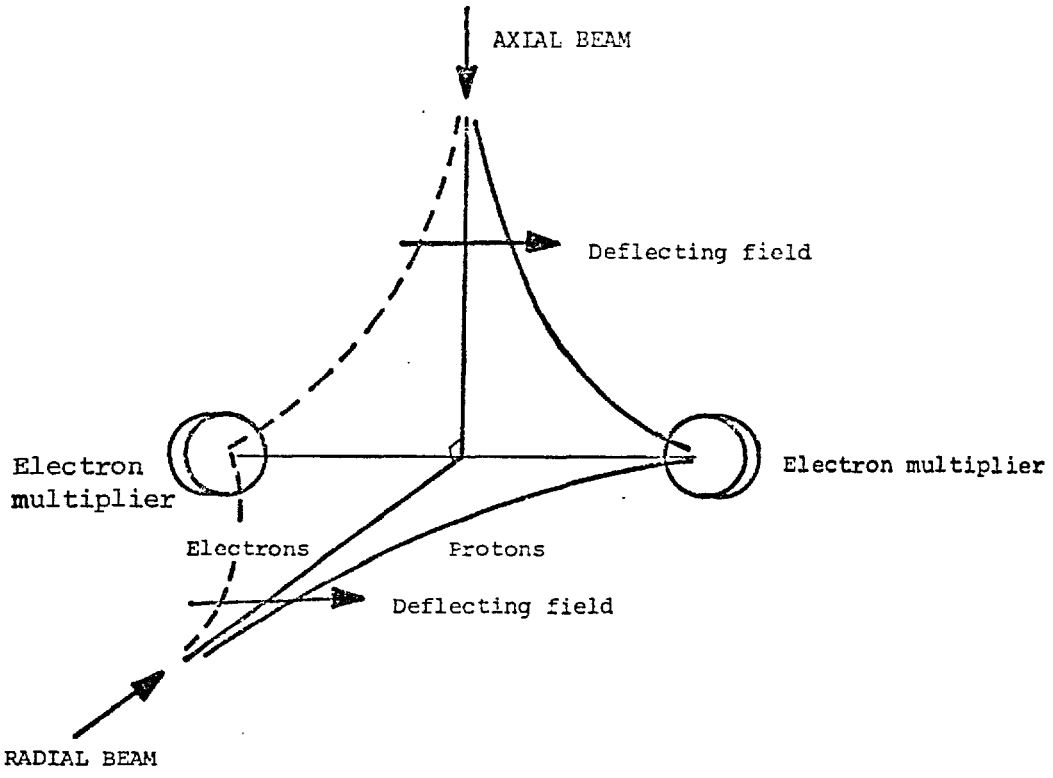


Figure A1(a) A schematic diagram of the soft particle spectrometer (SPS) carried on ISIS I. Electrons and protons from two mutually perpendicular directions are electrostatically separated and guided to particle detectors where they are counted. The broken and solid lines represent idealised trajectories for electrons and protons respectively.

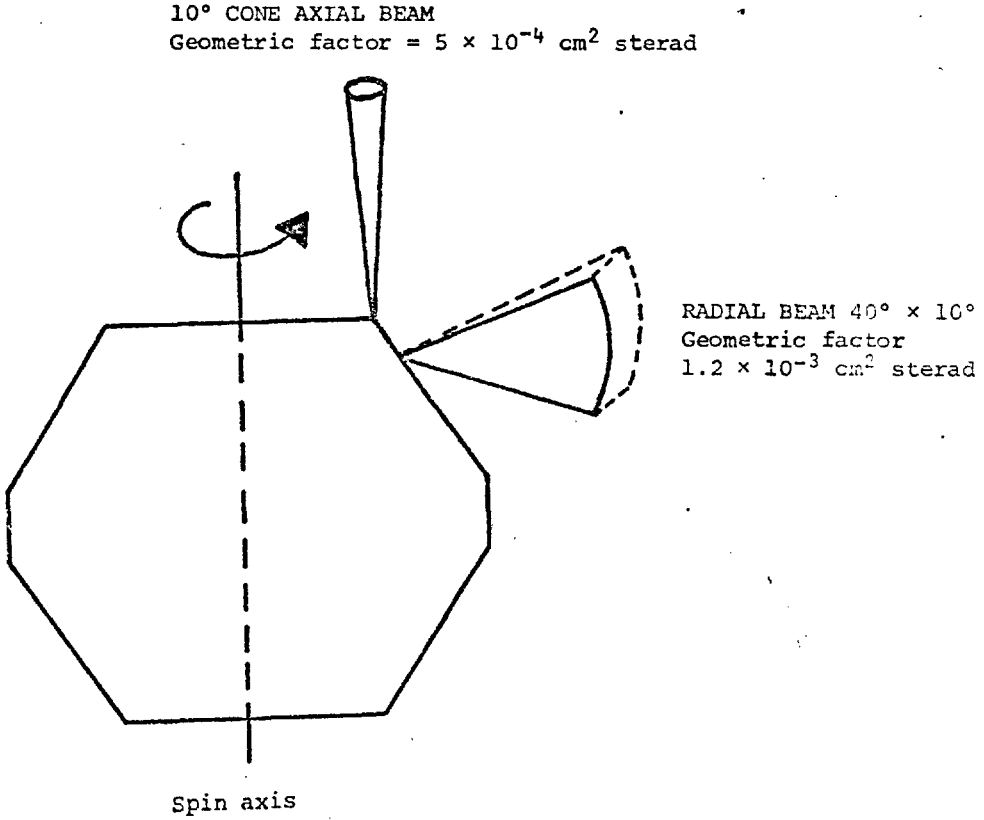


Figure A1 (b) The positioning and viewing angles of the particle entry apertures on ISIS I. The positioning of the radial beam and the satellite spin combine to sample a selection of pitch angles while the axial beam points in a constant direction.

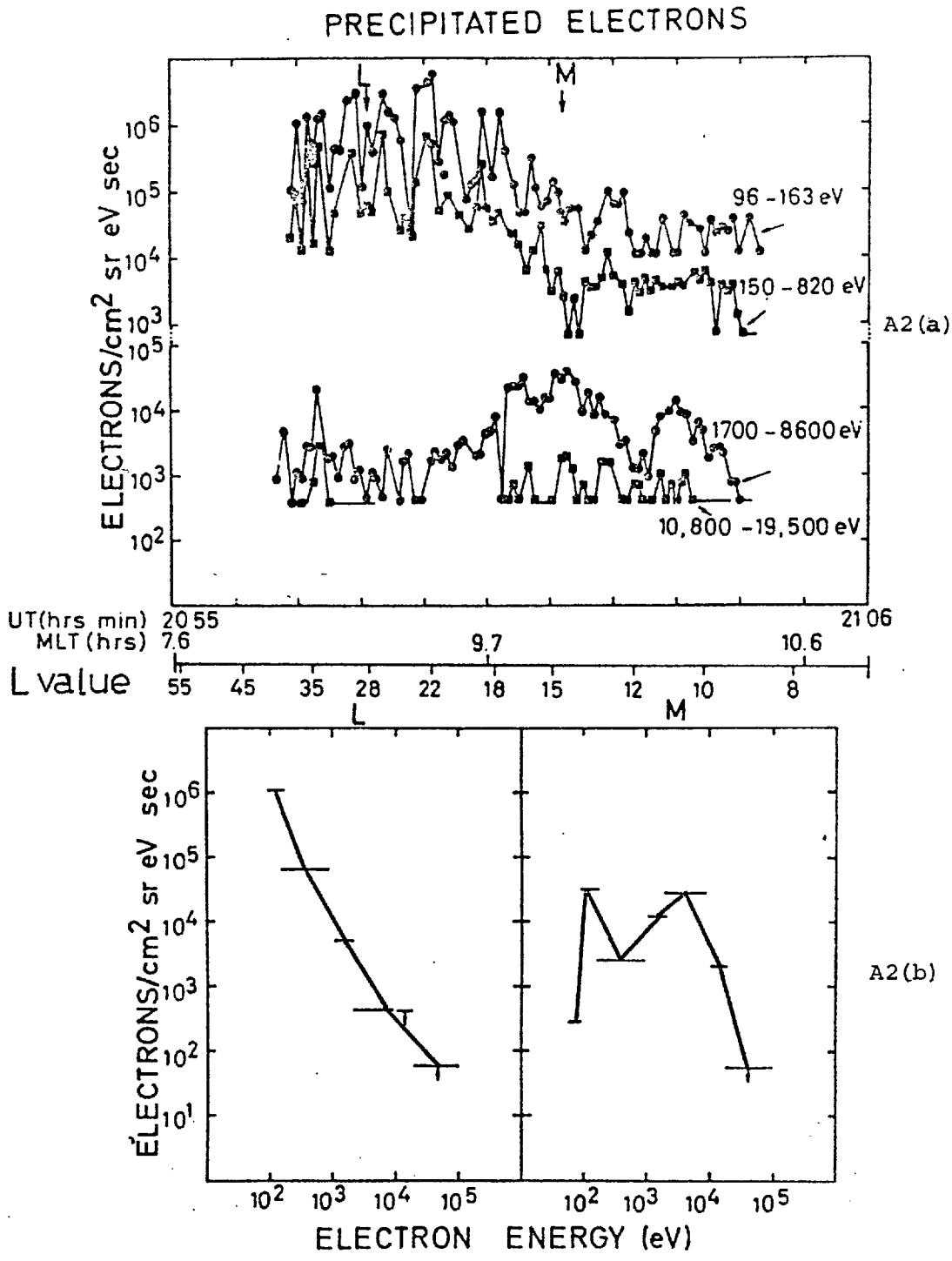
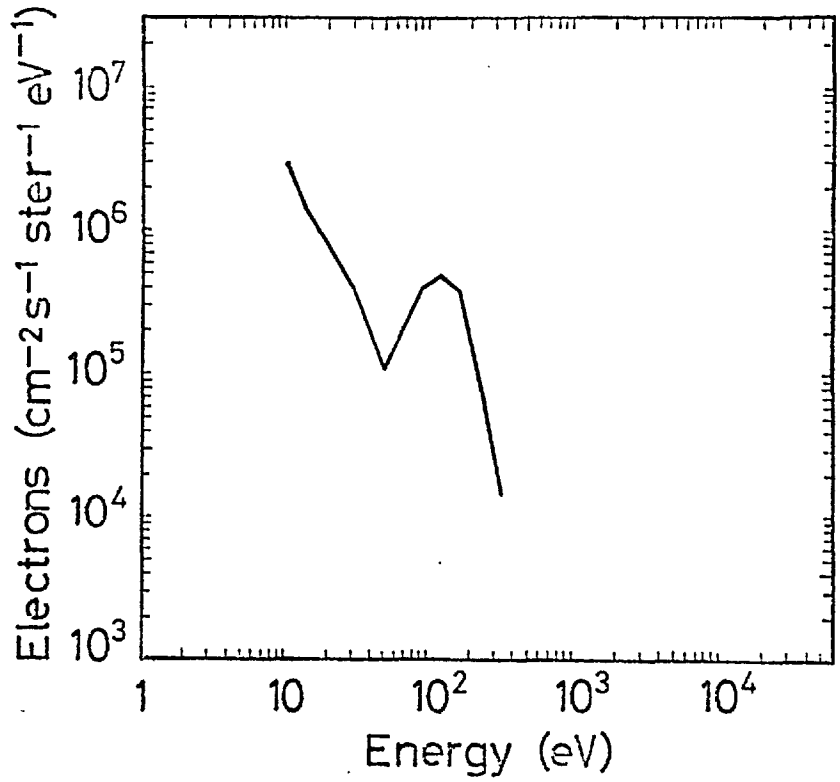


Figure A2 (After Burch, 1969). Typical measured and derived data from an early satellite particle detector on board Aurora 1, for a day-side pass, 159, on 18th July 1967, during quiet magnetic conditions (Kp = 1).

(a) Simultaneous measurements of electron fluxes in the labelled energy bands as a function of L-coordinate, MLT and UT. The 'soft' zone is evident for $L \lesssim 40$, with high structured fluxes in the 96-163 eV and 150-820 eV energy bands. A clear transition to a 'hard' zone occurs near $L \sim 18$ when fluxes of particles with energies $\lesssim 1$ keV decrease and an accompanying steady increase in the number of particles with higher energies is seen.

(b) Measured spectra in the 'soft' and 'hard' zones showing their different characteristics. The spectra designated L and M were constructed from measureemnts obtained at the points similarly indicated in the upper diagram. As discussed above these were

Figure A3 A typical electron spectrum observed in the cleft or soft dayside region by the ISIS 1 soft particle spectrometer for orbit 1603 on June 22nd 1969. Data were taken from Winningham et al. (1973).



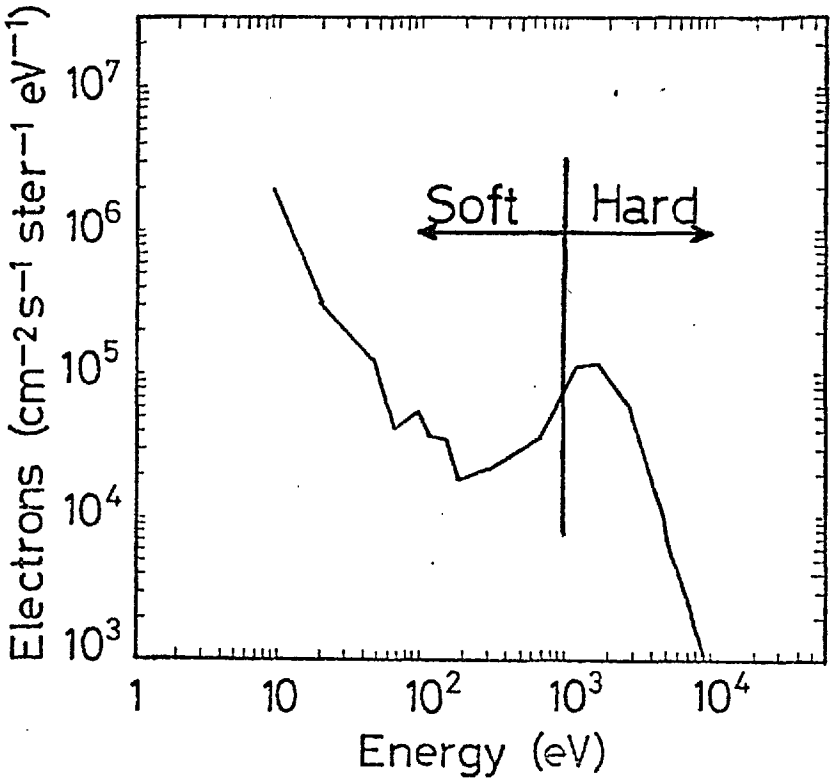


Figure A4 A typical hard nightside electron spectrum observed by ISIS 1 on the same orbit detailed in the caption of Figure A3. The essential difference between the soft day and hard night spectra is immediately apparent, the energy peak shifting from energies of approximately a few 100 eV to the keV range, from day to night. The spectrum is roughly divided into soft and hard components.

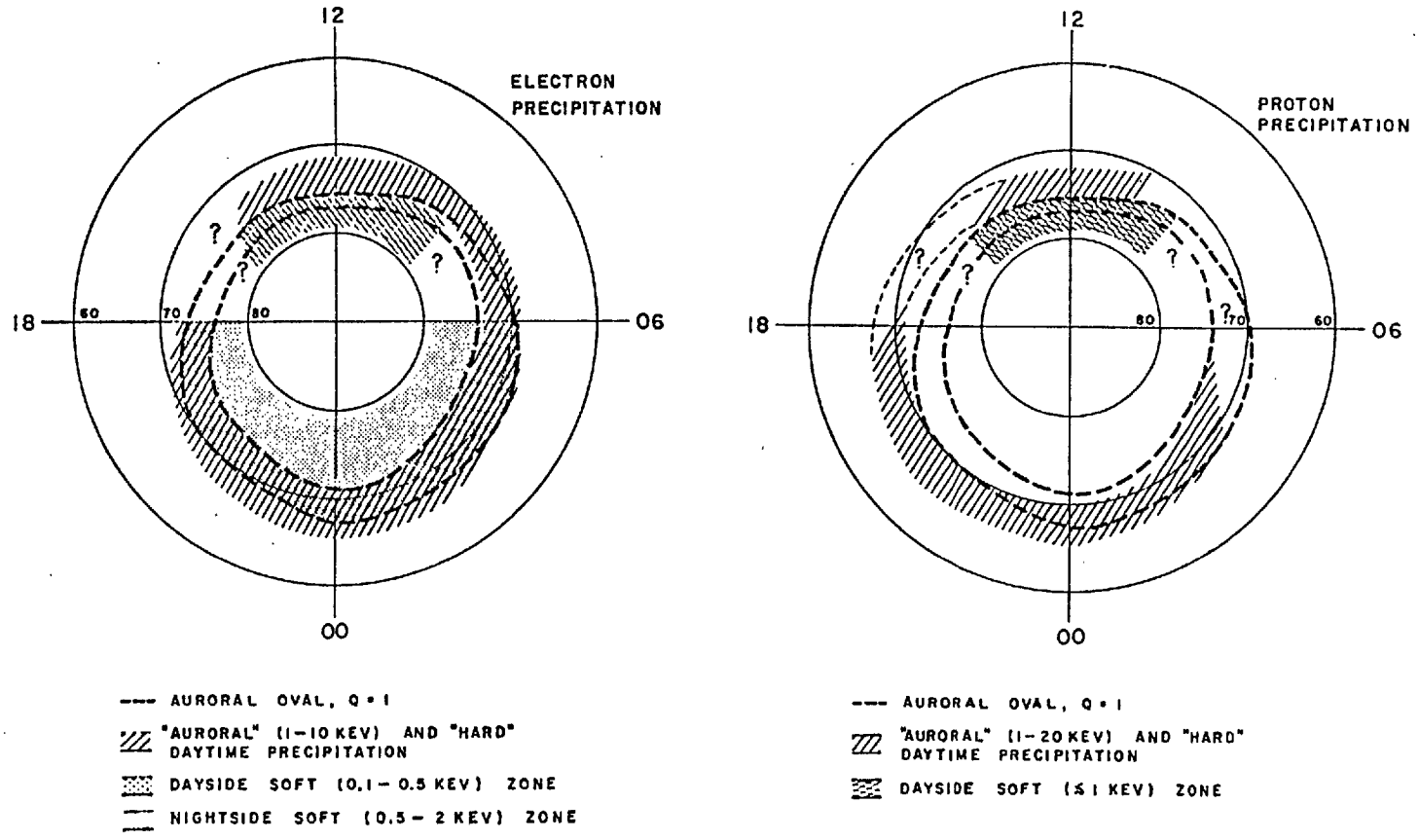


Figure A5 (From Eather and Mende, 1972). Electron and proton precipitation patterns obtained from observations of optical emissions, shown on a grid of invariant latitude and magnetic local time. The shading is explained in the key. The auroral oval (Q = 1) is superposed for comparison and the question marks refer to areas where evidence for precipitation is uncertain.

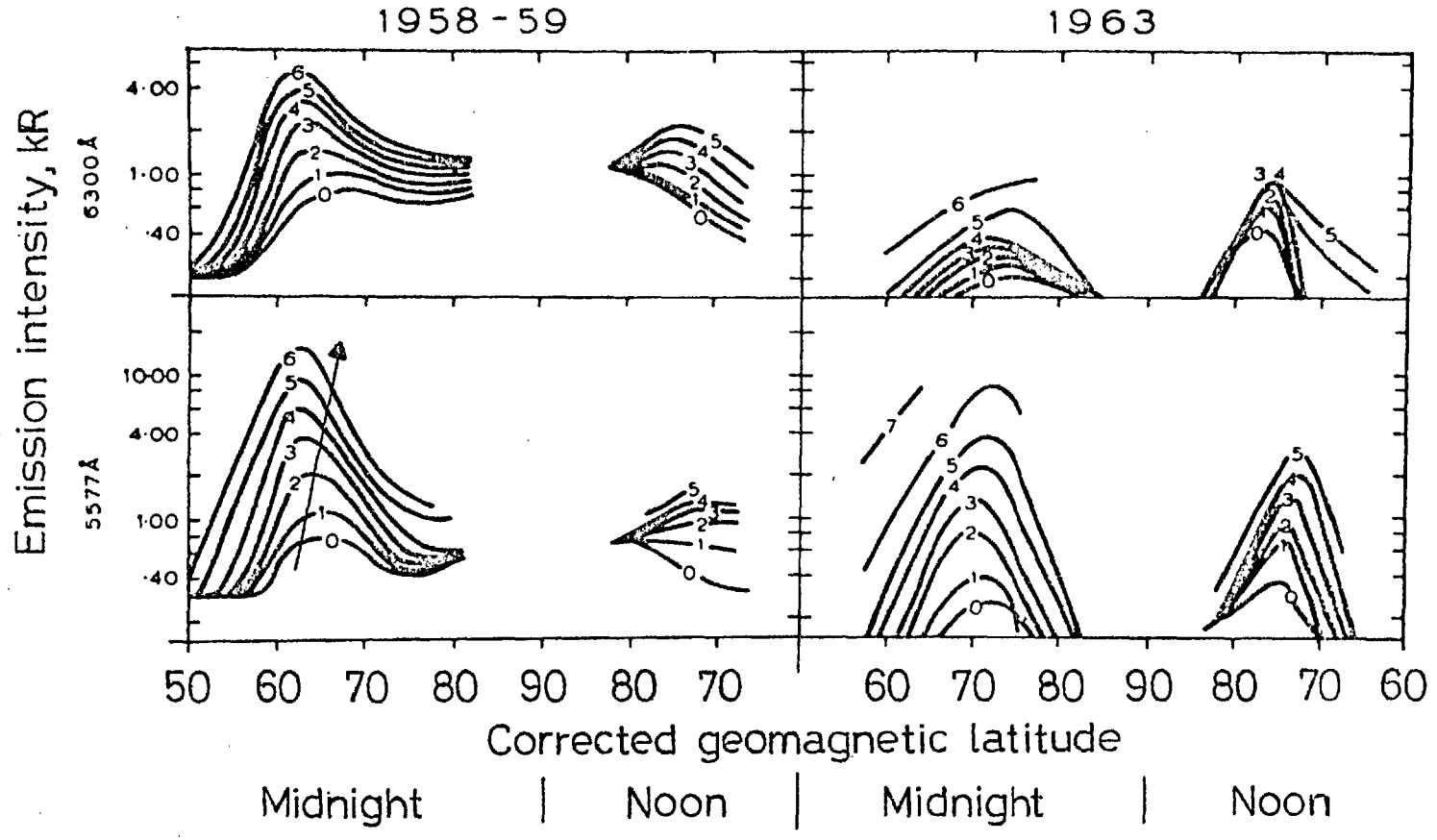


Figure A6 (After Sandford,1969). A comparison of the latitude profiles of the median intensities (in kilorayleighs) of 5577 and 6300 Å emissions observed at periods of high (1958-9) and low (1963) sunspot number along the noon-midnight corrected geomagnetic time meridian. The numbered curves refer to values of local K-index with the arrow indicating the direction of increasing K. The greatest difference between these two sets of curves appears at latitudes >80° corrected geomagnetic latitude where emission in 1963 fell sharply to low levels while at solar maximum relatively high values seem to be maintained. Thus although there are no data at higher latitudes Sandford (ibid) inferred from these results that high emission levels are present over the polar cap near maximum epoch of the sunspot number cycle.

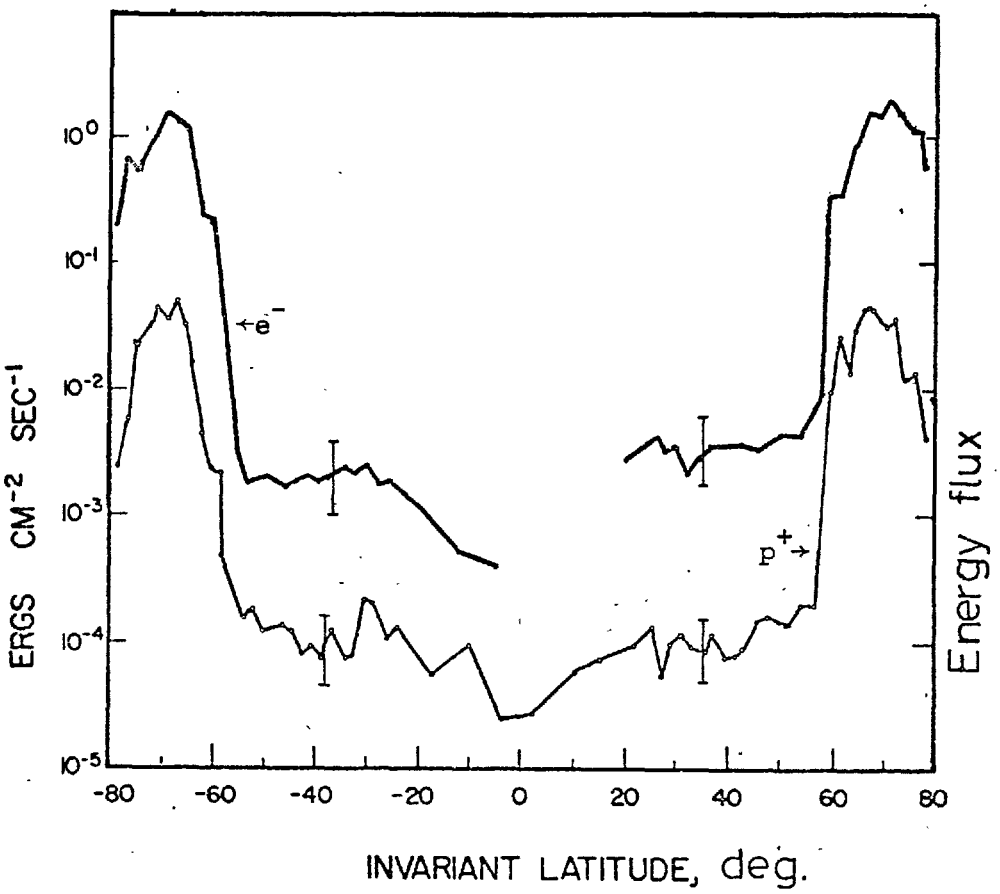


Figure A7 (After Torr et al., 1976). The total energy influx between altitudes of 250 and 300 km, for protons (thin curve) and electrons (heavy curve), for nightside hours, as a function of invariant latitude. The figure demonstrates the relative importance of particles as a source of energy at high and low latitudes and that electrons are the largest source of this energy.

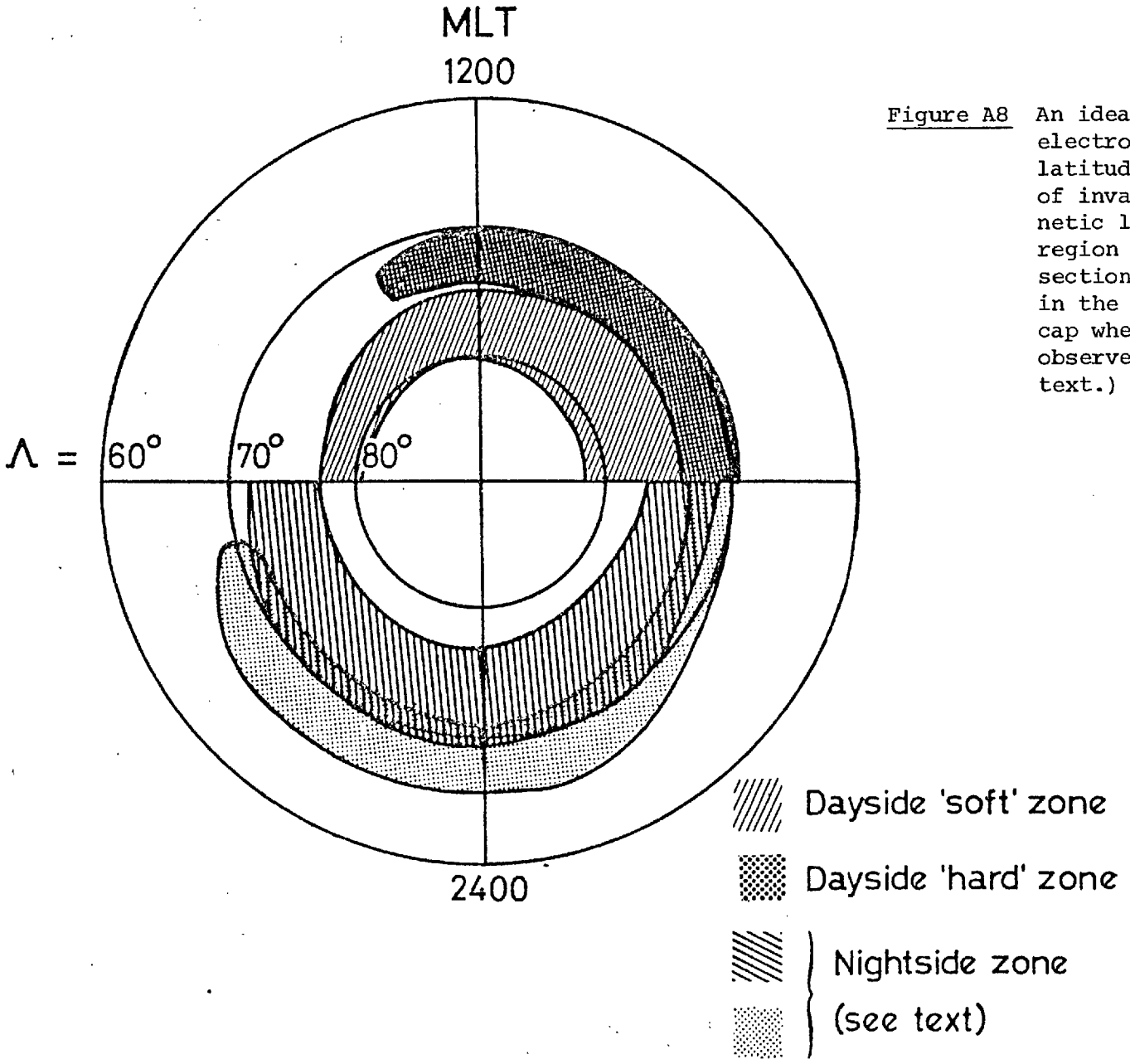


Figure A8 An idealised map of steady electron precipitation at high latitudes superposed on a grid of invariant latitude and magnetic local time. The polar region has been divided into 5 sections comprising those shown in the key and also the polar cap where various flux types are observed. (Explained in the text.)

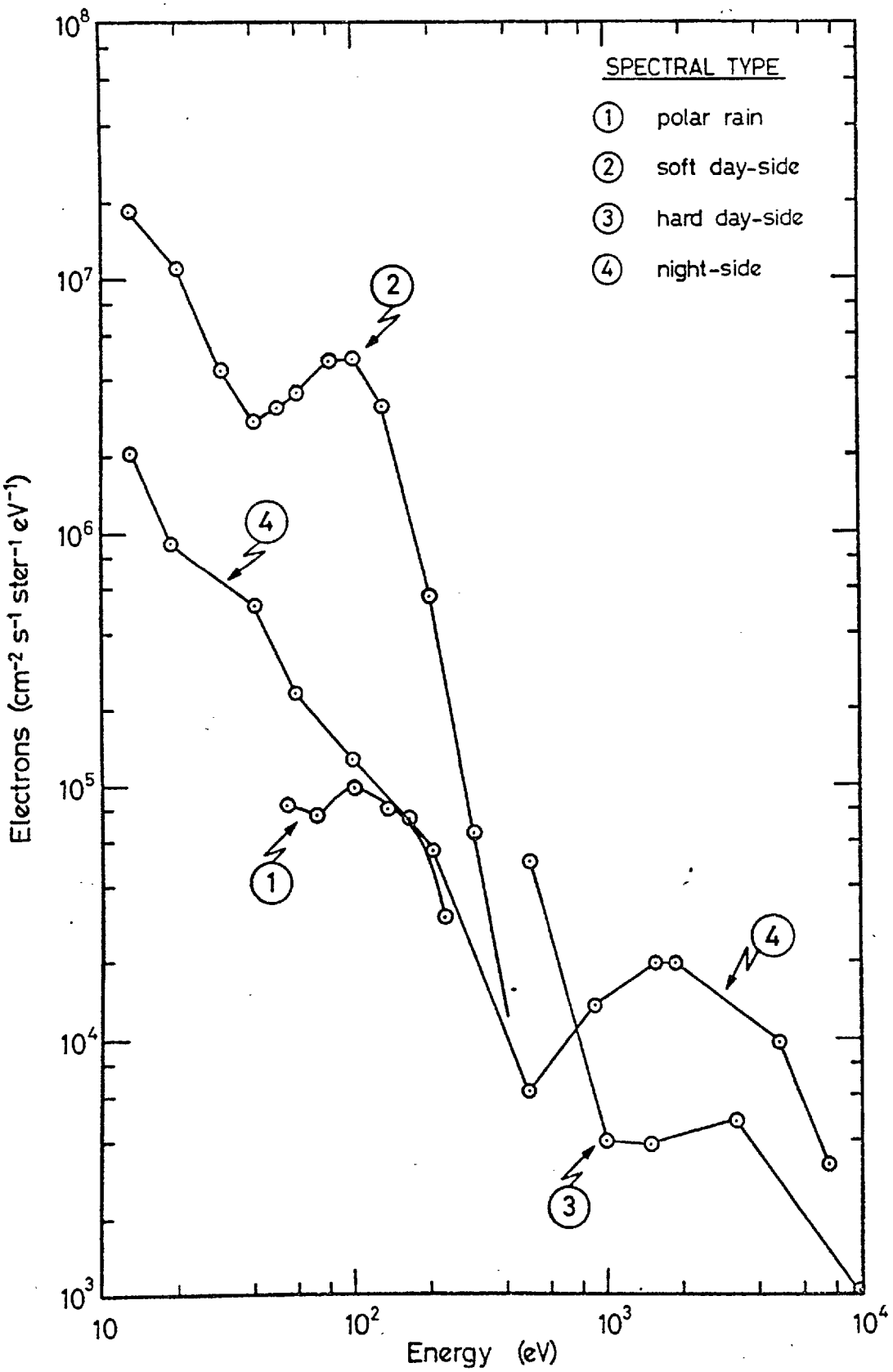


Figure A9 Typical electron spectra observed in the regions shown in Figure A8. The nomenclature is explained in the key In many cases, particularly spectrum 3, artificial cut-offs are applied by the energy passband of the particle detector. However, few observations of low-energy particles have been reported in the 'hard' dayside regime.

References

Buchau, J., G.J. Gassmann, C.P. Pike, R.A. Wagner and J.A. Whalen, Ann. Geophys. 28, 443, 1972.

Burch, J.L., J.G.R. 73, 3585, 1968.

Burch, J.L., in The Polar Ionosphere and Magnetospheric Processes, ed. G. Skovli, Gordon & Breach, New York, 1969.

Craven, J.D. and L.A. Frank, Proc. Roy. Soc. Lond. A343, 167, 1975a.

Craven, J.D. and L.A. Frank, University of Iowa, Dept. of Physics & Astro., Preprint No. 75-23, 1975b.

Eather, R.H., J.G.R. 74, 153, 1969.

Eather, R.H. and S.B. Mende, in Magnetosphere-Ionosphere Interactions, ed. K. Folkestad, Universitetsforlaget, Oslo, 1972.

Frank, L.A. and K.L. Ackerson, J.G.R. 76, 3612, 1971.

Frank, L.A. and K.L. Ackerson, J.G.R. 77, 4116, 1972.

Frank, L.A., N.A. Saflekos and K.L. Ackerson, University of Iowa, Dept. of Physics and Astro., Preprint No. 75-17, 1975.

Hartz, T.R. and N.N. Brice, Planet. Space Sci. 15, 301, 1967.

Heikkila, W.J., Space Res. 12, Akademic Verlag, Berlin, 1972.

Heikkila, W.J. and J.D. Winningham, J.G.R. 76, 883, 1971.

Heikkila, W.J., J.B. Smith, J. Tarstrup and J.D. Winningham, Rev. Sci. Instrum. 41, 1393, 1970.

Hoffmann,R.A., in The Polar Ionosphere and Magnetospheric Processes, ed. G. Skovli, Gordon and Breach, New York, 1969.

Hoffmann, R.A., in Magnetosphere-Ionosphere Interactions, ed. K. Folkestad, Universitetsforlaget, Oslo, 1972.

Hoffmann, R.A. and F.W. Berko, J.G.R. 76, 2967, 1971.

Hoffmann, R.A. and J.L. Burch, J.G.R. 78, 2867, 1973.

Lui, A.T.Y., C.D. Anger and S.I. Akasofu, J.G.R. 80, 3603, 1975.

Maehlum, B.N., J.G.R. 73, 3459, 1968.

Rees, M.H., Space Sci. Rev. 10, 413, 1969.

Sandford, B., in The Polar Ionosphere and Magnetospheric Processes, ed. G. Skovli, Gordon and Breach, New York, 1969.

Sharp, R.D. and R.G. Johnson, J.G.R. 73, 969, 1968.

Sharp, R.D., D.L. Carr and R.G. Johnson, J.G.R. 74, 4618, 1969.

Torr, D.G., M.R. Torr, R.A. Hoffmann and J.C.G. Walker, J.G.R. Letts. 3, 305, 1976.

Winningham, J.D. and W.J. Heikkila, J.G.R. 79, 949, 1974.

Winningham, J.D., S.I, Akasofu, F. Yasuhara and W.J. Heikkila, J.G.R. 78, 6579, 1973.

Winningham, J.D., F. Yasuhara, S.I. Akasofu and W.J. Heikkila, J.G.R. 80, 3148, 1975.

PART II

EXPERIMENTS WITH PLASMA WAVES

Nomenclature (Part II)

$\underline{B}$, $\underline{B}_o$	magnetic induction = $\mu_0\underline{H}$
c	speed of light in vacuo
$\underline{D}$	Electric displacement
e	Electronic charge (positive)
$\underline{E}$, $\underline{E}_o$	Electric field
f, f_o	wave frequency or particle distribution function
f_H	electron gyrofrequency, $\omega_H/2\pi$
f_T	upper hybrid frequency
G	conductance
$\underline{H}$	magnetic field
$\underline{k}$	wavevector
K	Boltzmann constant
m	electron mass
n	an integer, 1, 2, 3, ...
N	electron density
r	the electron gyroradius, $1/\omega_H\sqrt{(KT/m)}$
$\underline{r}$	position vector
S	susceptance
t	time
T	absolute electron temperature
$\underline{v}$	particle velocity
$\underline{v}_g$	group velocity
v_T	thermal speed of an electron, $\sqrt{(KT/m)}$
Y	admittance, $1/Z = G + iS$
Z	impedance
$\underline{\underline{\varepsilon}}$	tensor permittivity
$\varepsilon_\perp$	perpendicular component of $\underline{\varepsilon}$, $\varepsilon_\perp(k,\omega) \equiv \varepsilon(k_\perp,\omega)$
ε_o	permittivity of free space
λ	wavelength, $2\pi/k$

λ_D	Debye length, $= (\varepsilon_o KT/Ne^2)^{\frac{1}{2}}$
Λ	$k_I^2 r^2$
μ_o	permeability of free space
ν	collision frequency between electrons and neutral particles
ϕ	phase angle or electrostatic potential
ω, ω_o	$2\pi f$ or $2\pi f_o$
ω_H	electron angular gyrofrequency, $\lvert eB_o/m \rvert$
ω_p	plasma frequency $2\pi f_N$
ω_T	$2\pi f_T$

CHAPTER 4

THEORY AND EXPERIMENT IN PLASMA WAVES

4.1 Introduction

Plasma waves occur naturally in a wide variety of situations ranging from astrophysical plasmas to the earth's upper atmosphere. They can also be excited artificially in the laboratory by antennae immersed in the ambient plasma, and, as the so-called 'resonance spikes' on topside-sounder records illustrated, in the ionospheric plasma. Our subsequent understanding of these waves has resulted largely from the research arising from the latter observation.

The topside-sounder resonances indicate that at certain frequencies signals are received continuously, and initial work centred on identification of these frequencies. Analyses showed that the frequencies depend on parameters local to the satellite and are the local electron plasma frequency, f_N, the local upper hybrid frequency, f_T, and the electron gyroharmonics, nf_H, $(n = 1, 2, \ldots)$. Physically, an explanation of the resonances requires a class of undamped wave at the identified discrete frequencies, whose group velocity matches that of the satellite, so that a propagated wave packet remains at the antenna.

The presence of resonances at the frequencies f_N and f_T suggest that electrostatic oscillations of electrons, in directions parallel and perpendicular to the static magnetic field, might be involved. In this process, the electrons oscillate against the restoring force of the stationary ions. However, this type of cold plasma mechanism is unable to predict waves at the electron gyroharmonic frequencies.

When electron thermal motions are considered, a plasma in an applied magnetic field is capable of sustaining a great variety of wave motions.

Many of these wave motions are severely damped, and Landau (1946) showed that electrostatic oscillations, at the plasma frequency would, in the absence of a magnetic field, be changed into slowly damped propagating waves. Later, Bernstein (1958), using a similar analysis, was able to show that for a hot magneto-plasma and, in the particular case of propagation strictly perpendicular to the applied magnetic field, damping disappears and pure real solutions of the dispersion relation exist at harmonics of the electron gyrofrequency. It is now acknowledged that these slow, propagating, undamped waves - the Bernstein modes, are responsible for topside-sounder resonance observations at the electron gyroharmonics.

Furthermore, numerical simulations of ray trajectories near f_N and f_T by McAfee (1968, 1969), and recent analytical ray tracing of electron oscillations near nf_H (Bitoun, 1974), have shown that waves of the type thought to be responsible for satellite resonances can be reflected obliquely from an electron density gradient. Buckley (1974) has demonstrated the principle of a plasma wave radar in showing that, in certain circumstances, at the resonance frequencies, a swept-frequency transmitter in a satellite can excite two waves of slightly different frequency, which can be received back at the satellite. The resulting interference pattern can be interpreted in terms of the local plasma parameters; electron density and temperature, and the possibility of using such waves as a space diagnostic is suggested.

The Bernstein modes have become the subject of extensive laboratory investigation. Initially, these experiments were primarily designed to verify the dispersion relation derived by Bernstein (ibid), (Crawford, 1970). Recently, they have been extended to include all aspects of their propagation characteristics, including their application as a plasma diagnostic tool. Two techniques have been developed for experimental investigation of hot plasma properties and the Bernstein waves

excited by applied external alternating voltages:

1. Direct measurement of excited Bernstein mode wavelengths.
2. Examination of the frequency variation of hot plasma admittance.

Studies of the type 1, (Andrews, 1969; Clinckemaille, 1970) have indicated that cyclotron harmonic waves possess distinct advantages as a plasma diagnostic, giving accurate measurements of electron density and temperature. They have very small wavelengths which allow probing of small-scale electron density irregularities and, at distances greater than several Debye lengths from the antenna, measurements appear to be unaffected by the form of exciting probe and the surrounding ion sheath. This is important since the sheath is a little-understood region where current theory requires much development. It is the region where the alternating voltages couple to the plasma through perturbations to the particle trajectories and non-linear processes are thought to be active.

Experiments concerned with plasma admittance (Andrews, ibid; Mantei, 1967) give great insight into plasma behaviour. They can also be used for the inference of plasma properties, but may suffer the disadvantage of dependence on the exciting probe and its surrounding sheath.

The subsequent sections will be concerned with an extension of previous measurements, with special reference to two areas where previous work has been sparse. The first of these is an examination of propagation above the upper hybrid frequency, f_T, where theory suggests that two Bernstein modes are excited. A further study centres on the effect of the antenna dimensions on the measured plasma admittance and possibilities for the use of such methods in the determination of sheath size. The implication of these experiments is discussed, but first a brief theoretical background will be supplied.

4.2 Theoretical background

The theory of wave propagation in an infinite magnetoplasma has recently received extensive attention, see for example, Dougherty, (1974); Thomas, (1975); and references therein. Thus only the most relevant aspects will be treated here.

4.2.1. Maxwell equations for the plasma. There are two possible approaches to the theoretical treatment of a plasma. One can regard it as a region of free space with permittivity ε_o, and an internal assembly of free charges and currents, ρ_{int} and j_{int}. If a driven antenna is immersed in the plasma the driving charge and current density form an external source, specified by ρ_{ext} and j_{ext}. The first and last Maxwell equations become:

$$\begin{aligned} \nabla.\underline{D} &= \nabla.\underline{\underline{\varepsilon}}_o.\underline{E} = \rho_{int} + \rho_{ext} \\ \nabla\wedge\underline{H} &= \underline{j}_{ext} + \underline{j}_{int} + \varepsilon_o\dot{\underline{E}} \end{aligned} \qquad 4.1$$

Alternatively the plasma may be considered as a dielectric with a tensor permittivity, $\underline{\underline{\varepsilon}}$. 4.1 is then written:

$$\begin{aligned} \nabla.\underline{D} &= \nabla.\underline{\underline{\varepsilon}}.\underline{E} = \rho_{ext} \\ \nabla\wedge\underline{H} &= \underline{j}_{ext} + \underline{\underline{\varepsilon}}.\dot{\underline{E}} \end{aligned} \qquad 4.2$$

If all quantities are assumed to vary as $\exp i(\omega t - \underline{k}.\underline{r})$ then the operators ∇ and $\partial/\partial t$ take the form $-i\underline{k}$ and $i\omega$ respectively. Maxwell's equations can then be combined to give the wave equation, with sources, and applicable only after the initial transients have decayed.

$$\underline{k}\wedge(\underline{k}\wedge\underline{E}) + \omega^2 \mu_o \underline{\underline{\varepsilon}}.\underline{E} = i\omega \mu_o \underline{j}_{ext}. \qquad 4.3$$

4.2.2. The electrostatic approximation. This widely-used approximation greatly simplifies the analysis of wave propagation, and it is worthwhile to examine its physical meaning. Waves which are electrostatic in character possess a negligible magnetic wave field, implying that the wave vector, $\underline{k}$, is very nearly parallel to the wave electric field,

$\underline{E}$. Hence as $\underline{k} \wedge \underline{E} = 0$, the wave electric field is derivable from a scalar potential, ϕ, i.e.

$$\underline{E} = -\nabla\phi = i\underline{k}\phi \qquad 4.4$$

An equivalent statement of the electrostatic assumption is given by Stix (1962). The electric field components perpendicular ($\underline{E}_\perp$), and parallel ($\underline{E}_\parallel$), to $\underline{k}$ are related by the full wave equation (4.3). This can be written in the form:

$$(\varepsilon_o n^2 - \underline{\underline{\varepsilon}})\,\underline{E}_\perp = \underline{\underline{\varepsilon}}.\underline{E}_\parallel$$

where $n = \frac{ck}{\omega}$, the refractive index. For $\underline{E}_\perp \to 0$, the condition for pure longitudinal or electrostatic waves, n must be very large.

Hence $\frac{ck}{\omega} \gg 1$ or $\omega/k \ll c$

The nature of electrostatic waves is clear from this relation, showing them to be slow waves, with phase velocities comparable with the electron thermal speed, and of short wavelengths (the mm and cm range).

The electrostatic approximation greatly simplifies the wave equation 4.3, which reduces to :

$$\omega\underline{\underline{\varepsilon}}.\underline{E} = i\underline{j}_{ext} \qquad 4.5$$

4.2.3. Transform techniques. It has proved convenient, when using equations of the type represented by the Maxwell and the Boltzmann equations, to use transform analysis. In this procedure quantities which are functions of $\underline{r}$ and t become functions of $\underline{k}$ and ω by the application of time and space Fourier transforms. For the electric field then

$$\underline{E}(\underline{k},\omega) = \int_{-\infty}^{+\infty} d\underline{r} \int_{0}^{\infty} dt\, \underline{E}(\underline{r},t)e^{-i(\omega t - \underline{k}.\underline{r})} \qquad 4.6$$

As an example the transformed permittivity can be defined by the relation

$$\underline{D}(\underline{k},\omega) = \underline{\underline{\varepsilon}}(\underline{k},\omega).\underline{E}(\underline{k},\omega). \qquad 4.7$$

The use of these methods introduces two assumptions. These are that the plasma must be infinite, or at least of very large dimensions compared with the wavelength of the waves, and that the wave fields must be small, so that linear theory can be applied.

Equations are consequently obtained for the wave fields in terms of $\underline{k}$ and ω. If comparison with experiment is desired, the inverse transform must be applied.

$$\underline{E}(\underline{r},t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{+\infty} d\underline{k} \int_C d\omega \underline{E}(\underline{k},\omega) e^{i(\omega t - \underline{k}.\underline{r})} \qquad 4.8$$

The evaluation of 4.8 introduces many difficulties involving the form of C, the contour of integration.

4.2.4. The dispersion relation. If an external driving term is introduced, its current-charge relationship is given by the continuity equation;

$$\omega \rho_{ext} = \underline{k}.\underline{j}_{ext} \qquad 4.9$$

then an expression for the electric field can be derived by substitution of 4.9 into the wave equation.

$$\underline{E}(\underline{k},\omega) = \frac{i\underline{k}\,\rho_{ext}}{\underline{k}.\varepsilon(\underline{k},\omega).\underline{k}} \qquad 4.10$$

The condition for propagating waves is that non-zero values of $\underline{E}$ must exist in the absence of an external source. This implies that the dispersion equation relating ω and $\underline{k}$ is given by:

$$\underline{k}.\underline{\underline{\varepsilon}}(\underline{k},\omega).\underline{k} = 0 \qquad 4.11$$

By chosing an appropriate orientation of co-ordinate axes with respect to the magnetic field, 4.11 can be simplified further for propagation perpendicular to $\underline{B}_o$, the static applied field, to give:

$$k^2\varepsilon = 0 \quad \text{or} \quad \underline{\underline{\varepsilon}}(\underline{k},\omega) = 0 \qquad 4.12$$

The transformed permittivity is a highly significant quantity since its zeros determine the spectrum of waves which propagate. Much

effort has been directed to the determination of the form of $\underline{\varepsilon}_{\perp}(k,\omega)$[#] in different plasmas. Two special cases, cold and hot plasmas, are considered below.

[#] N.B. In subsequent work the subscript $\perp$ will be omitted and it is to be understood that $\varepsilon(k,\omega)$ is substituted for $\varepsilon_{\perp}(\underline{k},\omega)$ and that the wave vector k refers to a direction $\underline{k}_{\perp}$, perpendicular to the field $\underline{B}_o$.

4.3 The plasma permittivity

4.3.1. Cold plasma considerations. An expression for the cold plasma permittivity can be obtained using the magnetoionic theory (Clemmow and Dougherty, 1969). In this approximation charged particle motion arises only from Lorentz forces acting on the plasma and the permittivity is a function of ω only. For propagation perpendicular to the applied magnetic field, the permittivity, normalised to the free space value, is given by:

$$\frac{\varepsilon_c}{\varepsilon_o} = 1 - \frac{\omega_p^2}{\omega^2 - \omega_H^2} \qquad 4.13$$

The form of this expression is shown in Figure 4.1, in which the modulus of $\varepsilon_c/\varepsilon_o$ is plotted as a function of ω_H/ω. The salient features of this curve are the null at the upper hybrid frequency and the infinity at $\omega_H/\omega = 1$.

4.3.2. The hot plasma permittivity. Thermal velocities are introduced via the particle distribution function, $f(\underline{r},\underline{v},t)$. The wave is considered as a first order perturbation on the particle trajectories, and this interaction is described as a first order correction to the distribution function. The wave fields, which must simultaneously satisfy Maxwell's equations for the plasma, and the distribution

function, are linked through Boltzmann's equation. This implies that only collisions modify f along a trajectory in phase space. If the collision frequency is neglected, the Liouville equation can be used

$$\frac{\partial f}{\partial t} + \underline{v}.\frac{\partial f}{\partial \underline{r}} - \frac{e}{m}(\underline{E} + \underline{v}\wedge\underline{B}).\frac{\partial f}{\partial \underline{v}} = 0 \qquad 4.14$$

Assuming that f can be expressed in the form $f = f_o + f_1$ where f_1, the first order perturbation, varies as $\exp i(\omega t - \underline{k}.\underline{r})$, 4.14 can be solved for f_1 by linearising to include first order terms and using spatial and temporal transform techniques. Two approaches may be adopted. By the use of the full equations all possible waves can be derived or, as here, the wave magnetic field can be set to zero and only longitudinal modes with $\underline{k} \| \underline{E}$ derived. The first order perturbation to the electric field and distribution function are related through current. From Maxwell's third equation.

$$\underline{j}_{int} = -e\int f_1 \underline{v}\, d^3\underline{v} = -(\varepsilon_o - \underline{\underline{\varepsilon}}).\omega\underline{E}_1 \qquad 4.15$$

Here $d^3\underline{v}$ is an integration over velocity space.

If the initial distribution function, f_o, has a Maxwellian form

i.e. $f_o = N\left\{\frac{m}{2\pi KT}\right\}^{3/2} \exp -\frac{mv^2}{2KT}$

then the equation for f_1 can be integrated to give

$$1 + k^2a^2 = \frac{s}{\omega_H}\int_0^\infty dy \exp\{-\frac{sy}{\omega_H} - \lambda(1 - \cos y) - \frac{1}{2}\mu y^2\} \qquad 4.16$$

where $s = i\omega$ = Laplace transformed frequency

$\lambda = k^2r^2 \sin^2\theta$

$\mu = k^2r^2 \cos^2\theta$

$y = \phi' - \phi$ = variable of integration

This integral is defined for all s if $\mu > 0$, or if s has a small real component. Bernstein (ibid) showed that when $\mu = 0$, for propagation perpendicular the magnetic field $\underline{B}_o$, s is purely imaginary and Landau damping is absent. As the wave vector swings away from the purely

perpendicular direction then damping increases extremely rapidly, although it is very wavelength dependent. With $\mu = 0$ the dispersion relation is

$$\frac{\varepsilon_\perp(k,\omega)}{\varepsilon_o} = 1 - \frac{\omega_p^2}{\omega_H^2} \sum_{n=1}^{\infty} \frac{2\exp(-\Lambda)\ I_n(\Lambda)}{\Lambda((\omega/n\omega_H)^2 - 1)} = 0 \qquad 4.17$$

with $\Lambda = (k^2r^2/\omega_H^2) \times (KT/m) = (k_\perp r)^2$, r is the electron gyroradius and I_n is a Bessel function of imaginary argument. Poles are produced in this expression when ω/ω_H takes integer values.

The wavelengths which propagate are limited, to a certain extent, by the electrostatic approximation and the Debye length. Although infinitely short wavelengths are predicted, in theory one would not expect wavelengths less then the Debye length, since the wave fields would then be subject to influences from single particle trajectories and not collective plasma effects. Hence $\frac{2\pi}{k}$ must be $>\lambda_D$. The electrostatic assumption breaks down at very long wavelengths and this can be expressed as $\frac{\omega}{\omega_H kr} < \frac{c}{v_T}$, where v_T is the electron thermal speed. Since $\frac{c}{v_T} \sim 10^3$, the expression 4.16 is invalid for $\omega/\omega_H \geq 10^3 kr$ (termed the light line) and in this region of ω/ω_H electro-magnetic effects must be considered.

The dispersion relation (4.17) has been the subject of much discussion (Buckley, 1970; Tataronis and Crawford, 1970), and it has been derived for many forms of initial distribution function. Those for a Maxwellian plasma are summarised diagrammatically in Figure 4.2. Such curves take two forms depending on whether $\omega_o <$ or $> \omega_T$. In the former case there is a continuous transmission band and any real frequency can excite two wave numbers, $\pm k_1$. For $\omega_o > \omega_T$, propagation is only possible if ω_o falls into a passband situated at, and just above, the electron gyroharmonic frequencies. The dispersion curves permit any frequency ω_o, falling in a passband, to excite four wavenumbers, $\pm k_1$ and $\pm k_2$. Between the passbands there are frequency domains where propagation is not possible due to strong Landau damping. The reason for the

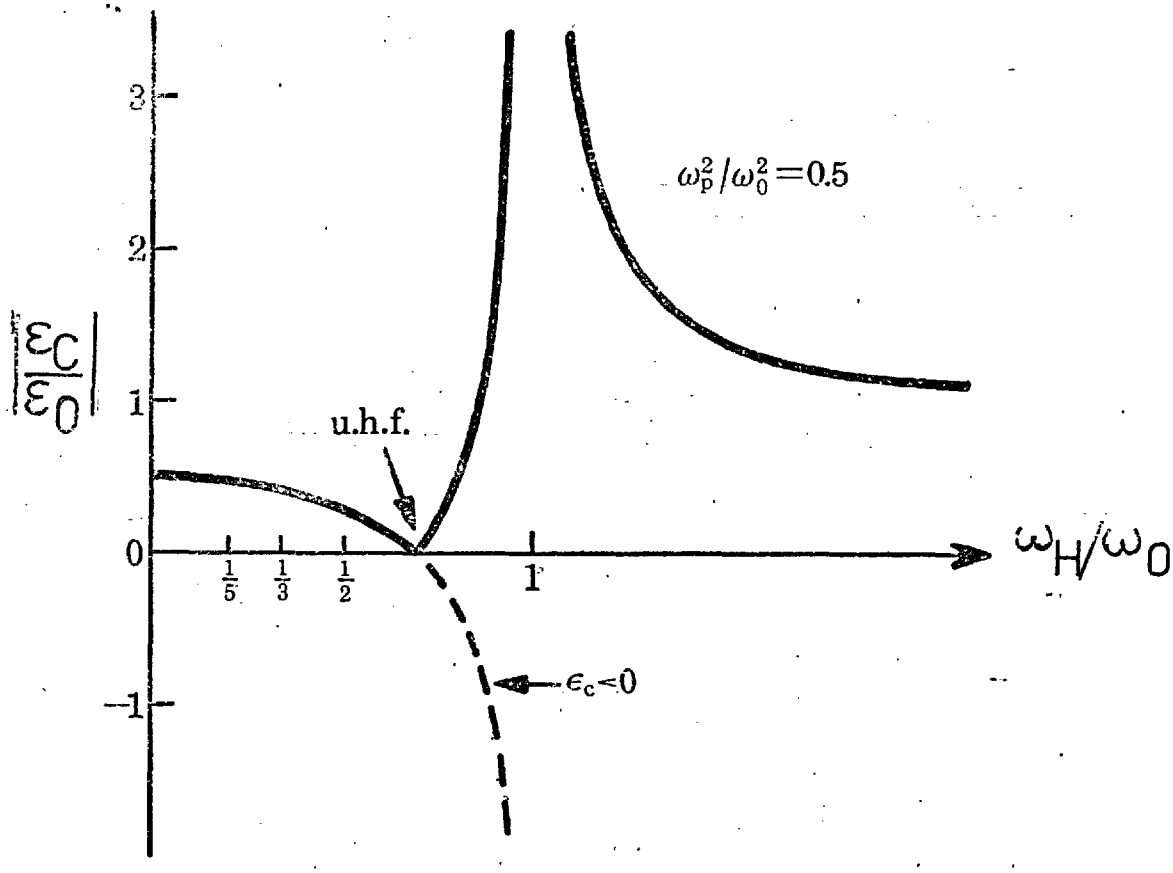


Figure 4.1 The modulus of the normalised cold plasma permittivity, as a function of ω_H/ω_o, given by equation 4.13. This particular situation corresponds to conditions such that $\omega_p^2/\omega_o^2 = 0.5$. The admittance null at the upper hybrid frequency is clearly evident. (From Thomas 1975)

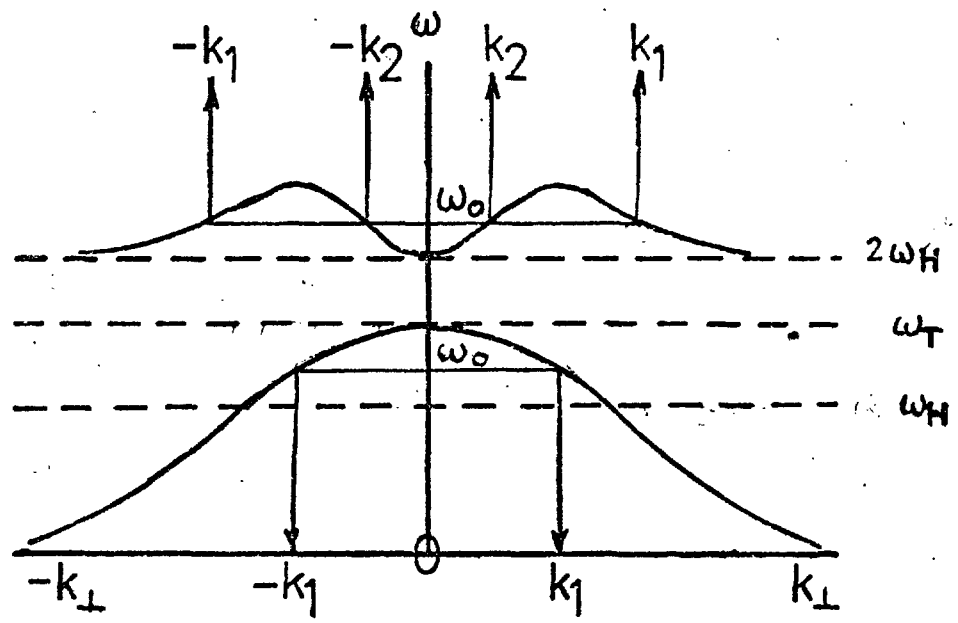


Figure 4.2 A schematic representation of the Bernstein mode dispersion relations for the situations where $\omega_o < \omega_H$ and $\omega_o > \omega_H$ (lower and upper curves respectively). These show how, depending on the value of ω_o, one or two propagation modes are possible and that transmission at wave numbers $+k_1$ and $-k_2$ is not physically acceptable, as they correspond to negative group velocities.

disappearance of damping at the cyclotron harmonics is as follows. The particles gyrate about the field lines and return to the same position every cyclotron period. Thus as the motion repeats itself, there is no phase-mixing as in parallel propagation and the resultant Landau damping is zero.

Of the wave numbers excited, only those with positive group velocities are physically meaningful, since only they can carry energy away from the source. The sign of the group velocity, $d\omega/dk$, is given by the slope of the dispersion curves (Figure 4.2). Hence, the propagating wave numbers must have positive gradients and only $-k_1$ and $+ k_2$ satisfy this criterion.

For specific solutions of the dispersion equation numerical techniques must be employed. Examples of computed solutions, due to Andrews (1969), are given in Figure 4.3. The two sets of curves give the relationship between the normalised variables, f/f_H and kr. There is also a negative branch, along the kr axis, not shown here. The two families of curves correspond to the two possible ways of sweeping the parameter, f/f_H, used in experiments. In each case either fN/f or fN/f_H is kept constant while the magnetic field or the input frequency is varied.

4.4. Theoretical aspects of experimental situations.

In practice the linear theory described above must be modified so that it is applicable to actual physical situations. Experiments are normally carried out by immersing antennae in lightly ionised plasmas and applying small radio frequency voltages, specified by ρ_{ext} and j_{ext}. The procedure adopted is then to measure the spatial or frequency variation of some amenable parameter, such as the electric field or the plasma admittance. The fine structure of the electric field and admittance is governed by

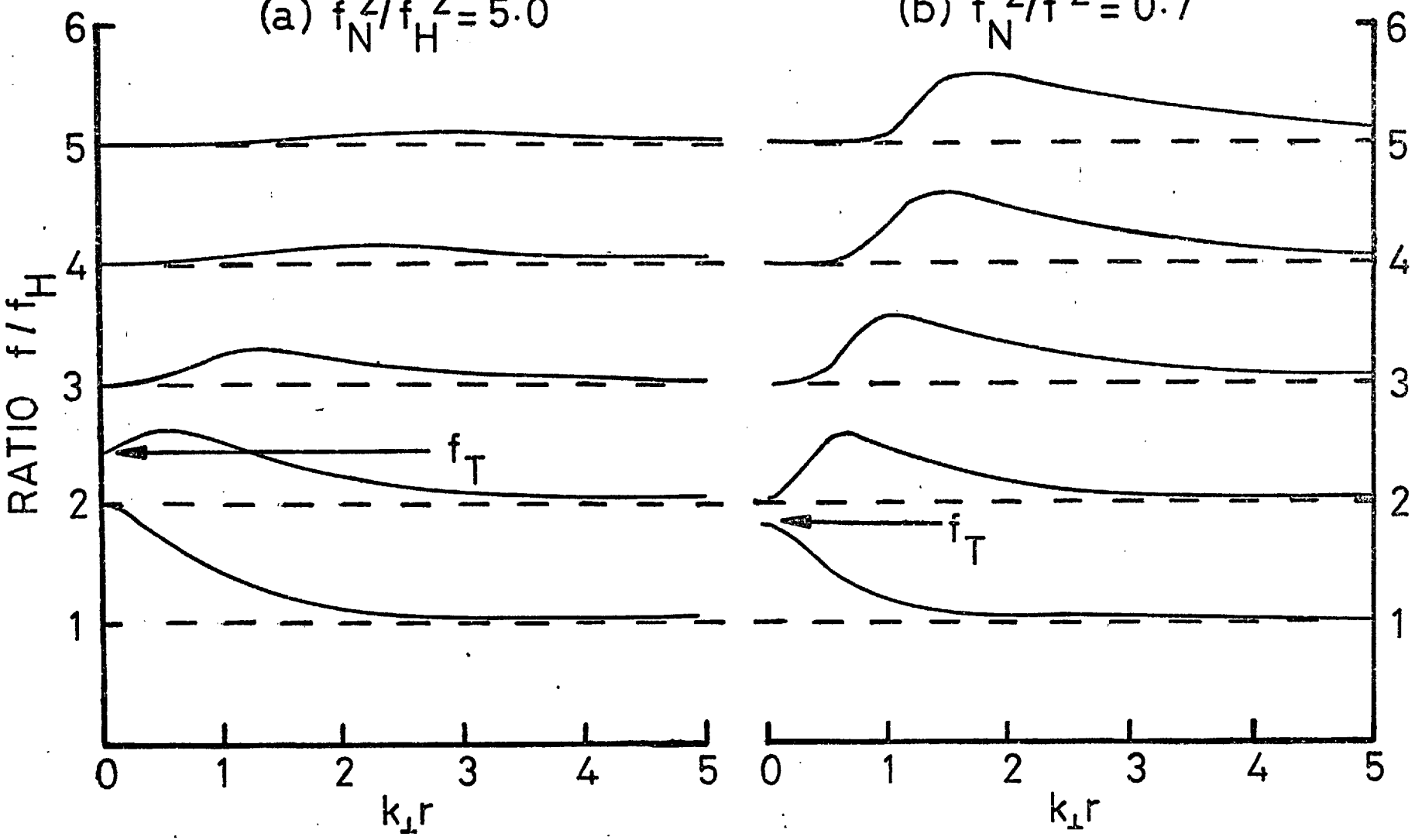


Figure 4.3 Numerical solutions of the dispersion relation (4.11) for perpendicular propagation of cyclotron waves. The contours represent values of ω and k for which $\varepsilon(k,\omega) = 0$. In each experimental situation there is continuous propagation below the upper hybrid and transmission in bands above this frequency.

the form of antenna used. Two particular probe geometries have been treated in detail. These are the cases of parallel infinite grids (Mantei, 1967, Buckley, 1970) and parallel cylindrical wires (Mantei, ibid). This linear theory only applies at distances greater than several Debye lengths from the antenna, for close to the probe the sheath region is found. The theory can also be improved by the consideration of momentum transfer between electrons and neutral constituents.

4.4.1. The zeros of $\varepsilon(k,\omega)$ and a collision term. To make the formulation physically more realistic account can be taken of the effect of momentum transfer between electrons and neutral particles. This is done by adding a collision term to the Boltzmann equation, thus

$$\left.\frac{\partial f}{\partial t}\right|_{\text{collisions}} = -\nu f_1$$

where ν is the electron-neutral collision frequency for a lightly ionised plasma. Crawford & Harp (1964), have shown that the resulting change in the dispersion relation (4.17) is to replace ω by $(\omega - i\nu)$ and ω_p^2 by $(1 - \frac{i\nu}{\omega})\omega_p^2$.

The effect of collisions can be examined by considering an infinite plane grid placed at $x = 0$, parallel to the magnetic field $\underline{B}_o$. To this an alternating charge of frequency ω_o is applied, giving

$$\rho_{ext} = e^{i\omega_o t}\,\delta(x)$$

where $\delta(x)$ is the Dirac delta function. The resultant temporal electric field perpendicular to the grid is

$$\underline{E}(k,t) = \underline{k}\,\frac{1}{2\pi k^2}\int_c \frac{e^{i\omega t}\,d\omega}{\varepsilon(\omega - i\nu,k)\,(\omega - \omega_o)}$$

In the absence of collisions the roots of $\varepsilon(k,\omega) = 0$, are for real k, real. However, the zeros of $\varepsilon(\omega - i\nu, k)$ lie at $\omega + i\nu$ and the electric field, due to real wave number (Thomas, 1975), is

$$\underline{E}(k,t) = i\underline{k}\,\frac{e^{i\omega_o t}}{k^2\varepsilon(\omega_o - i\nu,k)} + F(e^{-\nu t}) \qquad 4.18$$

The second term in this expression represents a series of damped terms arising from the complex zeros of ε. They all are severely damped because of the time dependence $e^{-\nu t}$, and consequently, for large values of t, the plasma responds only to the driving frequency. In this case only the first term need be considered and the spatial electric field is given by the inverse transform of 4.18.

$$E(x,t) = \frac{e^{i\omega_o t}}{2\pi} \int_{-\infty}^{+\infty} \frac{ik\, e^{-ikx}\, dk}{k^2 \varepsilon(\omega_o - i\nu, k)} \qquad 4.19$$

If $\nu = 0$ then the roots of $\varepsilon(\omega,k)$ fall on the real k axis, where strictly speaking the integrand is not defined. To perform the integration the integrand is analytically continued on to the real axis, by deforming the integration path around the roots of $\varepsilon(k,\omega)$ as shown in Figure 4.4. In this case the field is then the sum of the principal value of the integral evaluated over the real k axis (the cold plasma term since it is independent of k), and the residues evaluated at the singularities (the field due to the Bernstein modes).

Collisions shift the poles away from the real k axis so that a real root found at (ω_o, k_o) is found at $(\omega_o - i\nu,\ k_o - \frac{i\nu}{v_g})$. Purely real solutions no longer exist and collisional damping effects are introduced. The new positions of the poles are shown in Figure 4.4. The direction in which they move from the real k axis is dependent on the sign of the group velocity, obtained from the slope of the dispersion curves. In this case the integration can be carried out along the whole real axis and, for $x>o$, the contour must be closed in the lower half k-plane. This results in contributions to the field from the modes $-k_1$ and $+k_2$ only, i.e. those with positive group velocity.

4.4.2. The electric field in the plasma

To examine the structure of the electric field, between parallel grids, Buckley (ibid), has expanded 4.19 for vanishingly small ν. He has

given, for unit charge per unit area on each grid

$$E(x,t) = -e^{i\omega_o t}\left(\frac{1}{2\varepsilon(\omega_o)}\Big|_{k=0} + \frac{e^{+ik_1 x}}{-k_1 \frac{\partial\varepsilon}{\partial k}\Big|_{-k_1}} + \frac{e^{-ik_2 x}}{+k_2 \frac{\partial\varepsilon}{\partial k}\Big|_{+k_2}} + 2\mathrm{Re}\sum_{n=3}^{\infty}\frac{e^{-ik_n x}}{k_n \frac{\partial\varepsilon}{\partial k}\Big|_{k_n}}\right) \qquad 4.20$$

The terms in this equation can be considered separately.

1. The first term arises from the pole at k = 0 and is the contribution due to the capacitively-transmitted field with infinite wavelength in the electrostatic approximation. It can be shown that the denominator reduces to the cold plasma perpendicular permittivity given by 4.13.

2. The terms involving $-k_1$ and k_2 represent the hot plasma effects and are the fields due to the undamped Bernstein mode waves. Expressions of the type $\left(k\frac{\partial\varepsilon}{\partial k}\right)^{-1}$ are the corresponding amplitudes, and the exponential factors indicate that, if the probe separation is varied, oscillations of the electric field are found. In the stopbands, between the harmonics of the electron gyrofrequency, both k modes are non-existent for propagation perpendicular to the ambient magnetic field. At frequencies below the upper hybrid only that mode at $-k_1$ is excited. In general, these terms are expected to be of small amplitude and consequently they act as corrections to the cold plasma term (1).

3. The final term expresses the contribution of the spatially damped modes due to the complex zeros of ε , which occur when ω_o lies in a stopband. At small distances from the probe these modes are present and are thought to be partially responsible for effects in the sheath region.

At distances greater than several Debye lengths from the probe, the resultant electric field between the grids can be considered as a superposition of the capacitively-coupled cold plasma field and, depending on the frequency,one or two propagating Bernstein modes.

4.4.3. The plasma admittance/impedance. The plasma admittance or impedance can be derived from consideration of the electric field across the plasma and the current supplied to the antenna, which is given by the time derivative of ρ_{ext}. Hence, for an antenna current per unit area I_o, and a separation $2x_o$, the unit area impedance, Z, and admittance, Y are given by

$$\frac{1}{Z} = Y = \frac{I_o}{\int_0^{2x_o} E(x,\omega)\, dx}$$

1. Infinite parallel grids. Mantei and Buckley have used separate approaches, differing only in detail, to determine the electric field and plasma admittance variation across a static ambient field, between two infinite planar parallel grids. If the separation of these grids is $2x_o$, the driving charge density is given by

$$\rho_{ext}(x,\omega) = \rho_o e^{i\omega t} \{\delta(x - x_o) + \delta(x + x_o)\} \qquad 4.21$$

where δ is the Dirac delta function. This expression is substituted into equation 4.10 for the electric field, and the inverse transform applied to give

$$E(x,\omega) = \frac{\rho_o e^{i\omega t}}{\pi} \int_{-\infty}^{+\infty} \frac{\sin(k\, x_o)\, e^{-ikx}}{k\, \varepsilon(k,\omega)}\, dk \qquad 4.22$$

Integrating the electric field over grid separation as in 4.22 provides the voltage between the grids, and this, divided by the current per unit area, I_o, provides the impedance per unit area, $Z(\omega)$. This can be treated in several ways - either splitting the impedance into real and imaginary parts, or inverting to give the admittance, $Y(\omega)$. Mantei (1967) has shown that $Z(\omega)$ normalised to the vacuum impedance, Z_o (= $1/\omega C_o$ where $C_o = \varepsilon_o/2x_o$) is

$$\frac{Z(\omega)}{1/\omega C_o} = \frac{1}{Y} = \frac{-i\varepsilon_o}{\pi x_o} \int_{-\infty}^{+\infty} \frac{\sin^2(kx_o)}{k\varepsilon(k,\omega)}\, dk \qquad 4.23$$

For grids placed at x = o and x = d, Buckley has obtained an expression for the normalised impedance similar to that given

for the electric field, 4.20, and presented it as separate real and imaginary components.

Thus for grids placed at x = 0 and x = d, Z = R + iX where

$$\frac{R}{1/\omega C_o} = \frac{2}{d}\left[\frac{1-\cos k_1 d}{-k_1^2\,\partial\varepsilon/\partial k\big|_{-k_1}} + \frac{1-\cos k_2 d}{k_2^2\,\partial\varepsilon/\partial k\big|_{k_2}}\right] \qquad 4.24$$

$$\frac{X}{1/\omega C_o} = \frac{\varepsilon_o}{\varepsilon_c} + \frac{2}{d}\left[\frac{\sin k_1 d}{-k_1^2\,\partial\varepsilon/\partial k\big|_{-k_1}} + \frac{\sin k_2 d}{k_2^2\,\partial\varepsilon/\partial k\big|_{k_2}}\right]$$

The hot plasma admittance can be expected to approximately follow the cold plasma behaviour. Mantei (1967) has shown, by numerical solution of 4.23, that the admittance minimises at ω_T, and it is clear from 4.24 that as $\omega \to \omega_T$, X and $R \to \infty$, in agreement with cold plasma theory. At the gyrofrequency cold plasma theory predicts an infinite admittance. Here, the dispersion curves (Figure 4.3) show that the wave number becomes infinite. Again 4.24 show that at $\omega = \omega_H$, R and X = 0. However, these formulae were obtained with a vanishing collision frequency. The inclusion of collisions in a numerical calculation of the plasma admittance shows that only the imaginary part reaches a peak (the conductance does not maximise) at the gyrofrequency.

Thermal effects are most prominent near the cyclotron harmonics. The major behaviour is governed by the denominators of 4.23 and 4.24, and the roots of the dispersion relation $\varepsilon(k,\omega) = 0$. In the passbands the real part of the admittance is due solely to the Bernstein modes. Within the passbands above the gyroharmonics $\omega = n\omega_H$, $n>2$, both real and imaginary parts of the admittance reach a maximum.

There are also subsidiary variations superposed on the principal admittance maxima and minima. These are due to the oscillatory

component present in the form of the $\sin^2(kx_o)$ term in 4.23, and the sine and cosine factors of 4.24. Both the real and imaginary components of the admittance and impedance contain these terms, which result in subsidiary admittance maxima corresponding to minima of $\sin^2 kx_o$ (4.23). These occur when the electric field, integrated between the plates (i.e. voltage), reaches a minimum. This condition is fulfilled when the value of the Bernstein mode wavelength is such that an even number of half-wavelengths can be accommodated between the grids. Conversely, when an odd number of half-wavelengths fit between the grids, subsidiary minima appear in the admittance. Thus between admittance peaks or troughs, ω_H/ω changes to allow one more or less cyclotron wavelength to fill the space across the grids.

If in the expressions 4.24 $d=\infty$ is allowed, which is equivalent to the cold plasma limit (Debye length and Larmor radius = 0), then $\omega C_o R \to 0$ and $\omega C_o X \to \frac{1}{\varepsilon_c}$. The admittance of the plasma in this approximation is proportional to $|\varepsilon_c/\varepsilon_o|$. Thus, the general form of the plasma admittance follows this cold plasma expression, and hot plasma effects merely serve to increase the admittance in the passbands, close to the gyroharmonics.

2. Infinite parallel wires. The majority of experiments are performed with parallel wires as exciting antennae and Mantei (ibid) has used a similar derivation to that given above for grids, to obtain expressions for the admittance between two infinite cylindrical wire probes, immersed in a hot plasma, aligned along the ambient magnetic field. For wires of radius a and separation $2x_o$, Mantei has shown

$$\frac{1}{Y(\omega)} = Z(\omega) = -\frac{1}{\omega C_o} \frac{i\varepsilon_o}{\ln(2x_o/a)} \int_0^\infty \frac{J_o(ka) - J_o(2ka)}{k\,\varepsilon(k,\omega)}\,dk \qquad 4.25$$

where C_o is the vacuum capacitance between parallel conducting cylinders and J_o is a Bessel function.

Qualitatively, this result is similar to that obtained previously for planar grids. There is a peak at each gyroharmonic and a null at the upper hybrid frequency. In each passband the subsidiary maxima and minima described previously, occur, due to the oscillatory behaviour of the numerator in 4.25. However, these are accentuated relative to the grid situation because of the lower vacuum capacitance of the parallel wires.

Mantei has inverted the expression given above and has evaluated the integral, first making it tractable by introducing a non-zero collision frequency, to obtain numerical solutions for the relative parallel wire admittance. His results, replotted by Andrews (1969) are shown in Figure 4.5 together with the corresponding dispersion curves. These curves can be compared with Figure 4.1 which show the cold plasma permittivity, as a function of ω/ω_H.

They are essentially similar with a maximum and a null in admittance, respectively at the gyro- and upper hybrid frequency. However, in the hot plasma case enhanced admittances appear at harmonics of the gyrofrequency as well as subsidiary oscillatory components in the Bernstein mode passbands. The hot plasma effects thus appear as perturbations superposed on the cold plasma background admittance.

4.5. Previous laboratory observations

There is now a considerable literature concerned with observations of effects at frequencies related to the electron gyrofrequency. Reviews

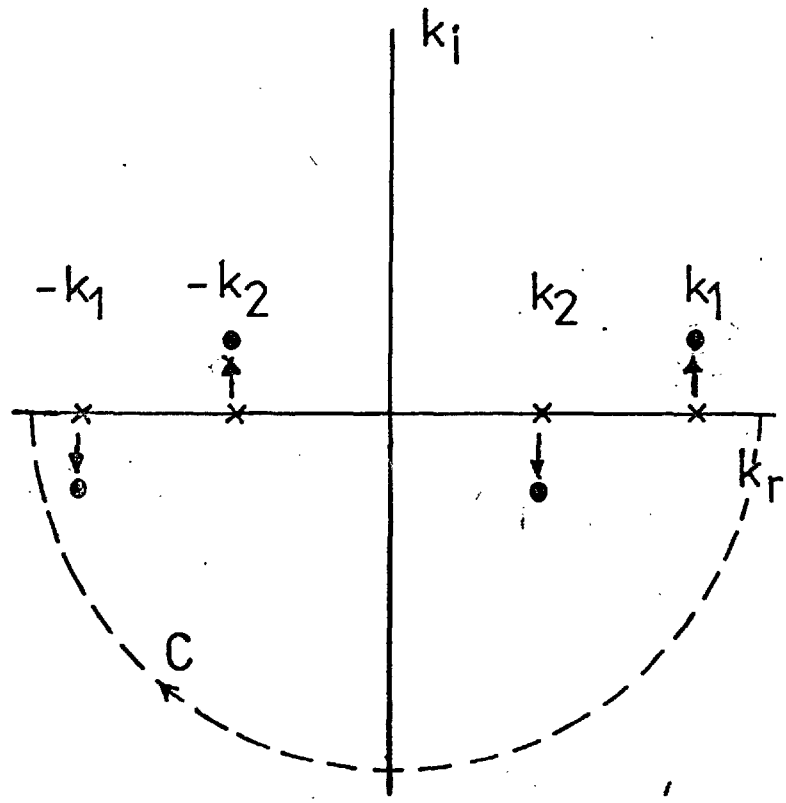


Figure 4.4 The contour of integration used to evaluate the integral 4.19. With the introduction of collisions, the undamped poles (crosses) on the real axis move, in the directions of the arrows, to the points indicated. The integration contour then becomes the real k axis and is closed in the lower half k-plane (shown by arrowheads). Only those poles with positive group velocities ($-k_1$ and $+k_2$) contribute to the electric field.

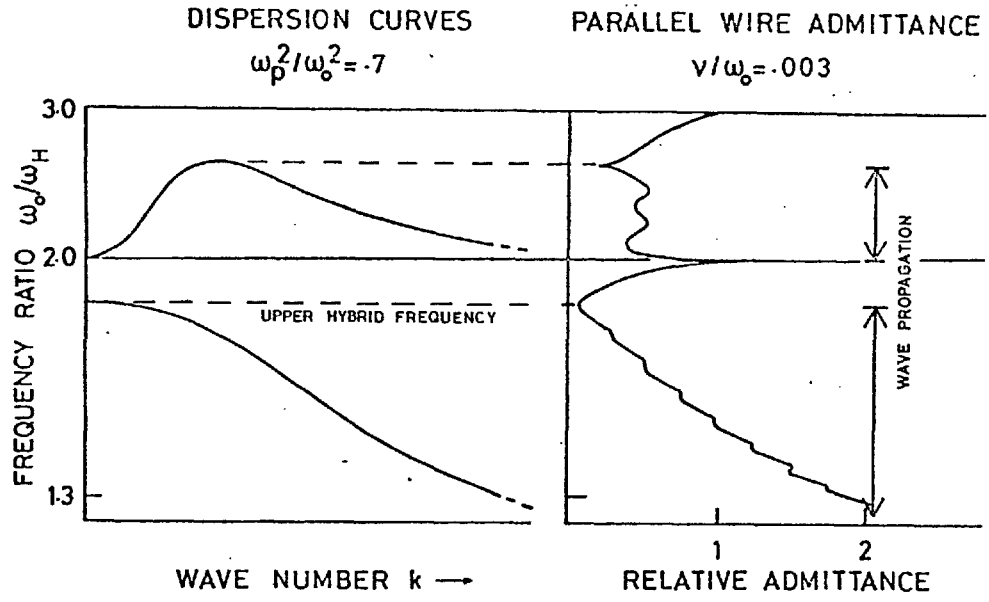


Figure 4.5 The relative admittance, for parallel infinite wires calculated numerically from 4.25, for a normalised collision frequency $\nu/\omega_o = .003$. The curves approximately follow the cold plasma admittance with thermal effects at the gyroharmonics and subsidiary periodic ripples in between. The way in which they arise is explained in the text. On the left the corresponding dispersion curves are given for comparison (from Andrews, 1969).

of early work have been given by Crawford (1970) and recent experiments have been summarised by Thomas (1975).

The first indications that cyclotron harmonic wave phenomena were observable in the laboratory came from the reports of noise emissions from a plasma perpendicular to the magnetic field. Wharton (1960) reported that peaks in the emission occurred at the electron gyroharmonics. Later a complimentary experiment examined the absorption of microwave energy in a plasma. Buchsbaum and Hasegawa (1964) noted that the frequency response exhibited complex variations. Strong absorption peaks were noted at the cyclotron harmonics and additional oscillatory fine structure. The modulations were attributed to standing longitudinal oscillations in the plasma, and indirectly Buchsbaum and Hasegawa (ibid) were able to verify the dispersion characteristics of cyclotron harmonic waves over a limited frequency range. After these initial results investigations proceeded along two separate lines - measurement of plasma admittance and determination of Bernstein mode wavelengths.

1. Plasma admittances. The earliest attempts to purposely examine the Bernstein modes were made by Crawford et al (1964). They, for the first time, immersed parallel antennae in the plasma along the direction of the static magnetic field. An alternating radio-frequency voltage was applied to one antenna and the signal transmitted through the plasma, in a direction perpendicular to the magnetic field, was received with the second probe. The dispersion parameter, ω/ω_H, was then varied by sweeping ω_H, via the magnetic field, from low through to high values. The received signal, proportional to the plasma admittance, was found to vary in a characteristic manner. Peaks of received signal, and hence admittance, were observed at harmonics of the electron gyrofrequency and an admittance null appeared at the upper hybrid frequency. In addition to the resonance peaks there was a subsidiary fine structure

of minor peaks and troughs between successive harmonics.

Subsequently Mantei (1967) and Thomas et al (1970) rigorously investigated the form of such curves. An example is shown as Figure 4.6, in which the main features described above are clearly visible. Curves of this type show a close resemblance to the theoretical predictions of Mantei (1967). Such curves can be used to estimate electron density by identification of the upper hybrid frequency. The periodic subsidiary admittance ripples have been used to verify the dispersion relation. Harp (1965), verified that in each wave passband, for an antenna spacing d, successive admittance maxima correspond to the wave number, $k = 2\pi/d$. Thus, Mantei, using such measurements of wave number from many corresponding experimental records of (ω_H/ω), built up large numbers of data for constant electron density conditions. These, when plotted alongside theoretical dispersion curves, gave excellent agreement.

In practice such techniques of investigating plasma behaviour possess several disadvantages. The electron density can often change appreciably as ω_H is swept. Thus some technique is required, in which the magnetic field is kept constant. Furthermore, recent measurements, Andrews (1969), suggest that the form of the admittance records may be affected by the geometry of the probes and their surrounding sheaths. At present, theories of sheath behaviour are not well-developed and it is desirable to avoid such departures from linear theory.

2. Experiments with moving probes. To overcome problems such as variation of plasma density, Thomas et al (1970) devised an experiment in which the magnetic field and input frequency were kept constant and the distance between the transmitting and receiving antennae varied. Subsequently experiments of a similar type have been performed by Clinckemaille (1970), Leuterer (1972) and Christopoulos and Christiansen (1974).

The variation of electric field between the antenna has been discussed previously (equation 4.20). It is made up of two or three undamped components, and the resultant field is the superposition of the capacitively-coupled cold signal, and one or two propagating Bernstein modes.

If the wavelength of the capacitive signal is taken to be infinite then this signal will remain unchanged with distance. Then, at distances greater than several Debye lengths from the probe, the field takes the form

$$E(x,t) = E_o\{\alpha e^{i\omega t} + \beta e^{i(\omega t - k_1 x)} + e^{i(\omega t - k_2 x)}\}$$

where α and β represent the respective amplitudes. If the exciting frequency $\omega < \omega_T$ then only one Bernstein mode is present and the potential relative to the electrode is given by

$$A(x) = -E_o\left\{\alpha x + \frac{1}{k}\sin kx - \frac{i}{k}(1 - \cos kx)\right\} e^{i\omega t} \qquad 4.27$$

The amplitude is consequently given by

$$A^2(x) = E_o^2\left\{\alpha^2 x^2 + \frac{2}{k^2} + \frac{2}{k}(\alpha^2 x^2 + \frac{1}{k^2})^{\frac{1}{2}} \sin(kx - \phi)\right\} \qquad 4.28$$

where $\phi = \tan^{-1}\frac{1}{k\alpha x} = \frac{\text{(cyclotron voltage amplitude)}}{\text{(cold amplitude)}}$

On moving one probe, therefore, if ϕ is assumed small (i.e. Bernstein mode amplitudes are very much less than cold plasma signal amplitudes) the received signal will be periodically modulated by the sin kx term and maxima and minima appear in the signal. Voltage maxima are separated by a distance δx where

$$\delta x = 2\pi/k = \lambda$$

In principle, the result is a direct determination of Bernstein mode wavelength.

In this way dispersion characteristics were obtained by Thomas et al. (1970) and Clinckemaille (1970). In practice, the probe separation is continuously

varied at different operating frequencies, for fixed discharge current and magnetic field. Many pairs of ω - k measurements allow experimentally-determined dispersion curves to be synthesised and Andrews (1969) has shown that they are in good agreement with theoretical computations.

The variation of received signal, as a function of probe separation, for a given ratio of applied frequency and gyrofrequency, is shown in Figure 4.7. The oscillatory behaviour, caused by interference between the capacitive and Bernstein mode signals, is prominent. Decay of the signal amplitude with increasing probe separation can also be observed.

Both Thomas et al (1970) and Clinkemaille (1970) have proposed that these electron oscillations may be used as a plasma diagnostic. This technique in a sense, involves the inverse procedure to that described above. Assumptions must first be made, regarding the velocity distribution of the plasma. Computed dispersion curves, such as those of Figure 4.3 are then derived for a wide variety of electron densities and temperatures. These parameters enter 4.17 through ω_p and the electron gyroradius, $r = \left(\frac{kT_e}{m\omega_H^2}\right)^{\frac{1}{2}}$. At this stage there are two alternative approaches. Thomas et. al. have used the transmission null at the upper hybrid frequency to fix the electron density (usually accurate to a few per cent). Then for a single input frequency, ω, fixing the parameter $\frac{\omega}{\omega_H}$, the corresponding wave number, k, is found. Using r as an adjustable parameter, and the measured k value, it is possible to predict the range of variation of kr. This can then be matched to the theoretically derived dispersion curves and using a best fit approximation r, and hence T_e, can be specified with sufficient accuracy.

Clinckemaille, adopted the alternative technique of deriving plasma

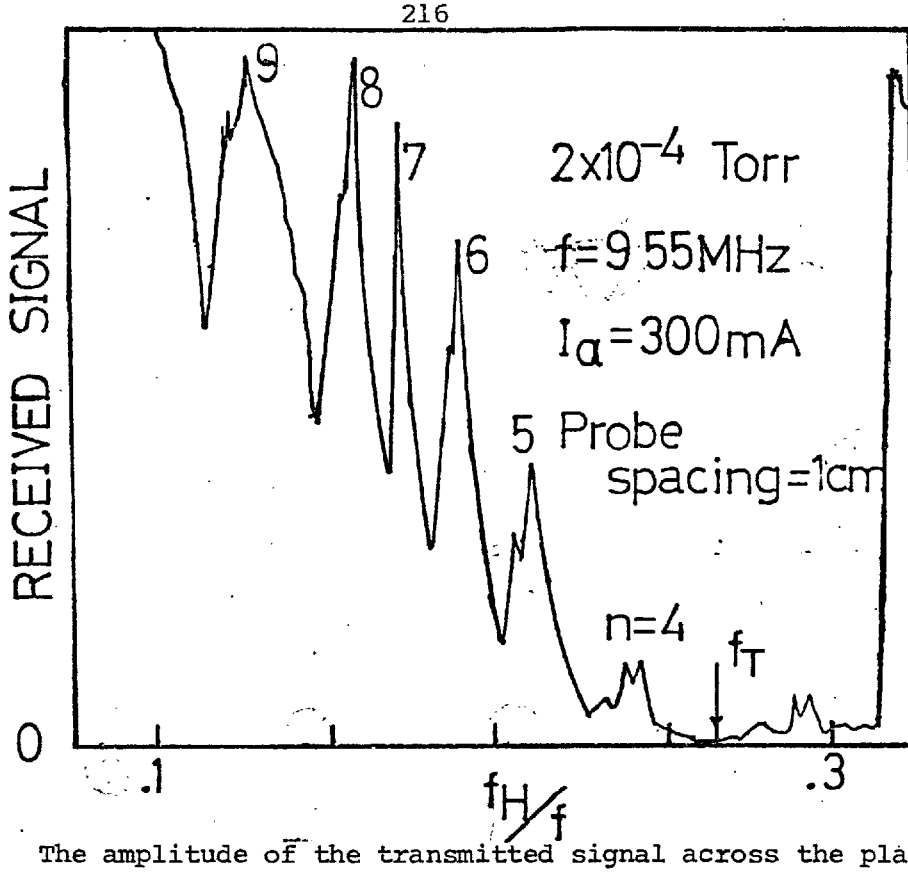


Figure 4.6 The amplitude of the transmitted signal across the plasma or relative plasma admittance as a function of f_H/f. The form of this curve is similar to that of Figure 4.1, the cold plasma admittance. Passbands associated with the gyroharmonics are visible up to n = 9.

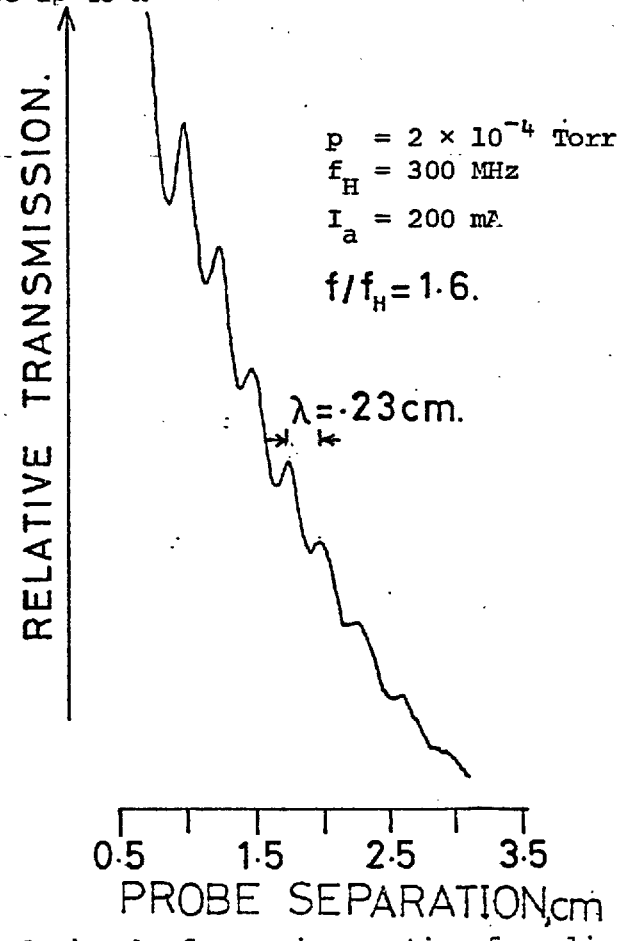


Figure 4.7 The received signal, for a given ratio of applied frequency and gyrofrequency, as a function of antenna spacing. The oscillatory behaviour, here at $\omega < \omega_T$, arises from interference between the capacitively-coupled signal and the single propagating Bernstein mode signal.

parameters. He measured the wave numbers (k_1, k_2) of waves excited corresponding to two operating frequencies (ω_1, ω_2), below the upper hybrid frequency. By inversion of the dispersion relation, it can then be shown that the wave numbers excited are parametric in electron plasma frequency and gyroradius. This can be expressed as

$k_1 = f(\omega_1, \omega_p, r)$ $\qquad$ $k_2 = f(\omega_2, \omega_p, r)$

Thus the two separate wave number measurements provide two linearly independent equations with the two unknowns, ω_p and r. Numerical solutions must be obtained to this problem, but Clinckemaille has succeeded in deriving values of ω_p and r (and hence electron density and temperature) for laboratory plasma situations.

The values of electron density and temperature obtained in this way have been compared with the simultaneous measurements of a Langmuir probe. The results are not in good agreement, especially data for T_e, derived by the different methods. In one experimental case, Andrews (1969), found wave techniques predicted a value of $T_e \sim 3.3$eV. Langmuir probe measurements on identical plasma conditions produced an estimate of $T_e \sim 10$ eV. Thus there appears to be a severe discrepancy between probe and wave measurements of electron temperature. At the present time it is thought the value derived from Bernstein mode experiments is the more reliable, since there is as yet no theory which accurately predicts the behaviour of a probe in a hot magnetoplasma.

Such experiments show, therefore, that laboratory plasma wave methods can, in principle, provide accurate estimates of plasma parameters and are a useful diagnostic tool.

CHAPTER 5

RECENT PLASMA WAVE EXPERIMENTS

5.1 Preamble

The linear theory of plasma wave propagation, in a direction perpendicular to a static magnetic field, and recently performed experiments in this topic, appear in close agreement. Transmission experiments of the type described by Clinckemaille (1970) and Thomas (1975) have reproduced dispersion curves first derived theoretically by Bernstein (1958), confirmed the general form of the plasma admittance predicted by Mantei (1967), and determined that the spatial variation of electric field in a plasma, due to a radio-frequency transmitting antenna, is similar to that given by Buckley (1970). Furthermore, those experiments suggested that such waves could be used as a simple space plasma diagnostic.

However, there are several areas where experimental data are lacking and further work is necessary. One of those areas concerns propagation above the upper hybrid frequency, f_T. As demonstrated in section 4.2, in this frequency range one might expect there to be two propagating Bernstein modes of differing wave numbers. This would have the effect of further modulating the admittance records and the spatial variation of electric field, due to the presence of a third interfering wave. Very few observ ations of simultaneous propagation of two electrostatic signals have been reported. In fact, only Leuterer (1969), who used phase detection techniques, seems to have been successful in exciting both long and short wavelength modes. There are several possible causes for the absence of the second mode. These include high collisional damping and non-linear wave excitation (Thomas et al., 1970) but clearly, further experiment is required to examine the agreement with theoretical mode amplitudes derived by Buckley (ibid).

Sheath effects have been shown to play a prominent part in determining the admittance response of antenna systems in space (Balmain and Oksuitik, 1969). The sheath region, however, is not clearly understood and, as yet, no comprehensive analysis of its structure and behaviour in a hot magnetoplasma has appeared. Experimental data on the sheath are also very sparse. First attempts to probe the region by Crawford et al. (1967) used electron beams, but more recently, attempts have been made to understand the sheath through its interaction with the surrounding plasma.

Thus Crawford and Harp (1964) and Balmain and Oksuitik (ibid) have examined,theoretically and experimentally,the behaviour of the sheath-plasma resonance in the absence of a magnetic field. Very few reports have appeared of similar effects in a magnetoplasma. Andrews (1969) partially attributed anomalies in admittance records to such phenomena and more recently,using a complex probe system,Kist (1976) has determined the dependence of the sheath on such parameters as probe radius and bias.

However, in an anisotropic medium, as results from applying a static magnetic field to a plasma,it is advisable to ensure that probe geometry especially is greatly simplified, so that the part which probe size and form play, in the excitation of waves and admittance of the sheath-plasma system, can be unambiguously identified.

In this chapter some further experiments with Bernstein mode waves are described. This work concentrates on the two areas outlined above. The experimental arrangement described below and, especially, the antenna configuration adopted are purposely kept very simple. In this way purely perpendicular transmission is maintained and the effects of other parameters such as probe dimensions and collisional damping can be adequately assessed.

5.2. The propagation experiment

5.2.1. Introduction. To examine propagation of Bernstein modes in a hot laboratory magnetoplasma the discharge tube shown in Figure 5.1 has been developed. (Thomas, 1975). In essence this consists of a glass chamber, 65 cm long and 10 cm in diameter, in which a plasma is generated. Surrounding the chamber a series of coils are arranged to provide a sufficiently uniform magnetic field in the direction of the chamber axis. Wave propagation in the plasma is facilitated by transmitting and receiving antennae immersed in the discharge. These are so arranged that any signal transmission between the probes is radial and hence, perpendicular to the ambient magnetic field due to the coils. External to the chamber but electrically connected to the antenna are suitable transmission and receiving systems.

The probes are made so that they can be moved radially across the plasma and the received signal, as a function of probe separation determined. In this way the system acts as an interferometer with the capacitively-coupled, large wavelength signal taking the part of the reference arm of the instrument.

5.2.2. The plasma. A stable plasma is produced by collisions of electrons with neutral argon gas in the main chamber. The gas enters through a port beneath the probes and the chamber is continuously pumped, to ensure a consistent supply of gas. Pressure is maintained at an ambient level of 10^{-3}-10^{-4} torr.

The source of electrons is a barium oxide-coated nickel disc cathode of diameter ~5cm. When heated with some 200 watts of power, sufficient numbers of electrons are emitted to maintain experimentally suitable

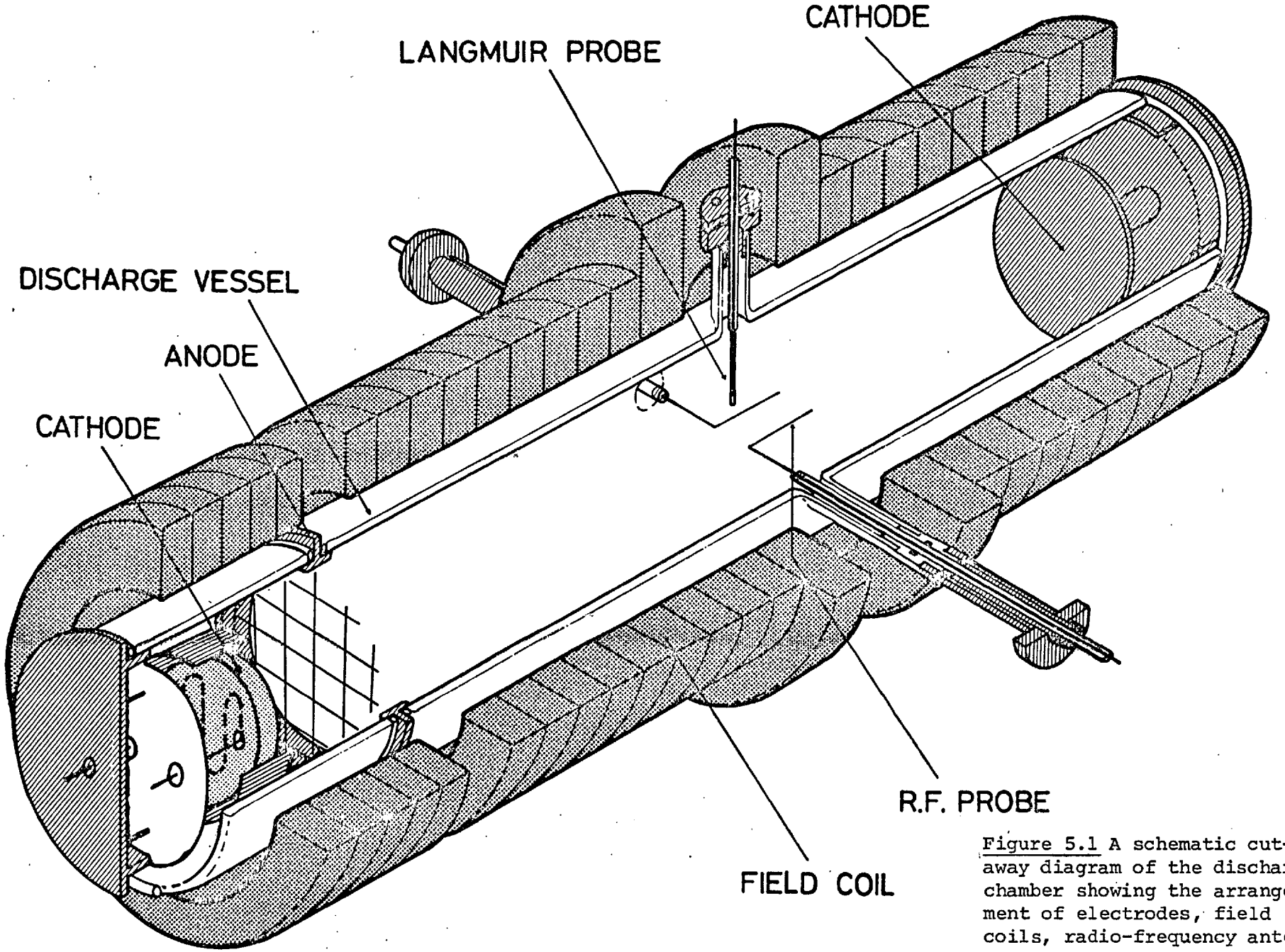


Figure 5.1 A schematic cut-away diagram of the discharge chamber showing the arrangement of electrodes, field coils, radio-frequency antennae and the Langmuir probe.

levels of electron density. Plasma densities are largely controlled by the cathode current and the resulting discharge is operated by a current-regulated supply, which applies a potential of ~50V between cathode and anode with currents in the range 0-500mA. The electrons produced at the cathode are accelerated to 50V energy, and pass through an earthed wire mesh anode, which serves to isolate the main chamber from fluctuations arising near the cathode. In the main chamber the electrons are thermalised by collisions with neutral atoms, the frequency of which,at ambient pressures,is $\sim$ 1MHz.

Measurements of the radial profile with an in-built Langmuir probe have shown that the electron density, over the central region away from the walls, is uniform. Consequently, fluctuations in the region where the probes are situated are small and on a scale comparable with the electron gyroradius.

The levels of electron density found in the tube, easily determined from the transmission null at the upper hybrid frequency, are listed in Table 5.1, along with other important experimental parameters, and the corresponding ionospheric values. The estimated electron temperature given is that obtained from the Langmuir probe measurement but these temperatures are not thought to be reliable and do not agree with Bernstein mode - determined values.

5.2.3. The magnetic field. The magnetic field is supplied by a series of coils powered by a voltage regulated supply which allows the field, and hence the gyrofrequency, to be swept over the range 0-700 MHz,or held constant. Calibration of the field was carried out with a Hall probe gaussmeter, and near the antenna, it was found to be constant to 1% over ±4cm along the tube axis,and largely independent of radial position.

Table 5.1.

Comparison of laboratory and space plasmas

Parameter	Laboratory Experiment	Topside Ionosphere
Electron density (cm^{-3})	$10^9 - 10^{10}$	$\sim 10^5 - 10^5$
Debye length (cm)	.02	$\sim$10
Neutral density (cm^{-3})	$\sim$5 10^{12}	$10^{10} - 10^{11}$
Electron temperature (eV)	3 - 10	$\sim$0.2
Electron gyrofrequency (MHz)	0 - 700	$\sim$0.8
Electron - neutral collision frequency (MHz)	$\sim$1	$\sim 1 \times 10^{-4}$

5.2.4. Signal transmission. Two interchangeable antennae are connected to a receiver and a transmitter. The transmitted signal has a frequency in the range 450-1000 MHz, and an amplitude of ∿0.1V. This signal is amplitude modulated at a frequency of 1000Hz and to a modulation depth of ∿10% of the signal amplitude.

The received signal is passed to a heterodyne detector. Here the signal is first mixed with that from a local oscillator and fed to a narrow band 30MHz amplifier. After detection and appropriate audio filtering, the output can be displayed on a chart recorder or an oscilloscope.

Various antenna types were used in the experiments, including grids and double probes but most measurements were performed with L-shaped tungsten wire probes of diameter .1 - .3 mm and arm lengths ∿ 1-2 cm. Each probe is movable in a radial direction in the plasma through a screw thread of 1mm pitch. One probe is attached to an electric motor by a vibrationless rubber drive to allow automatic variation of probe separation.

5.3. Observations of an Ion Sheath

5.3.1. Background Experimental records of the type represented by Figure 4.6, the variation of relative plasma admittance with the parameter f_H/f, often exhibit anomalous behaviour in the range of of frequencies between f_H and f_T. This effect takes the form of a peak in admittance, not predicted by either the normal cold or hot interpretations of plasma admittance (Mantei, 1967). Furthermore, near the gyrofrequency, f_H, where the theoretical plasma admittance approaches infinity, the measured admittance falls to a minimum.

Several workers have encountered this anomaly but it has received little attention (Mantei, 1967; Clinckemaille, 1970). Figure 5.2

reproduces an example of Mantei's data in which the admittance peak, B, and the minimum close to the gyrofrequency are clearly visible. In this case the maximum of admittance is observed at $f/f_H = f_H/f$. It has been suggested (Clinckemaille, ibid; Thomas, 1975) that such data could be due to the coupling effects of the plasma and the electron depletion or ion sheath regions, which form around the transmitting and receiving antennae immersed in a plasma.

To aid the interpretation of results from current and future propagation experiments in space, a full knowledge of plasma radio-frequency transmission characteristics is required. Anomalous behaviour such as that outlined above represents one aspect of this analysis and the subsequent sections describe an investigation of this admittance peak through the effect of varying antennae dimensions.

5.3.2. Recent theoretical approaches to the sheath

When any metallic probe is placed in a plasma, the faster thermal electrons collide with and, impose on it a negative potential relative to the body of the plasma. This potential repels further incident electrons so that, in equilibrium, the electron flux to the probe is equivalent to the ion flux. Thus there forms, around the antenna, an electron depletion region where quasineutrallity breaks down. The effect, due to the finite temperature of the electrons, is the so-called ion sheath - a little-understood region, where the plasma couples to the electrical signals in the antenna. The size and form of the sheath is the subject of much current discussion and its investigation by both experiment and theory presents a formidable problem.

The interest in the sheath arises from the need to describe the impedance of antennae immersed in a plasma. Consequently, the treatment involved can vary in complexity according to the model adopted for the plasma, and the frequency range of signals applied to the probe. At high radio

frequencies the presence of ions can be ignored and, on a simple picture, the sheath consists of a vacuum region with an abrupt transition to the plasma body (Figure 5.3a). However, thermal motions will act so as to smooth out this transition and the structure shown in Figure 5.3b, including a linear density increase probably gives a more accurate model (Balmain and Oksiutik, 1970). Early experiments were left to set the size of the vacuum region and these found it to be $\sim 5\ \lambda_D$ (Harp and Kino, 1964; Mayer, 1964) but more recent work (Hall and Landauer, 1971) have estimated it to be much smaller, i.e. $\sim 1\text{-}2\ \lambda_D$. When combined with an expression for the plasma permittivity in the cold approximation, the vacuum gap ion sheath model gives useful results and in particular predicts the sheath-plasma series resonance, observed in an unmagnetised plasma (Crawford, 1970).

A more accurate description of sheath-plasma behaviour is obtained by introducing temperature to the plasma. Initial approaches, using the warm plasma theory, in which particle temperatures are represented by a scalar pressure, retained the idea of the vacuum sheath with an abrupt boundary. The boundary was taken to be a perfect reflector of incident electrons. (Kostelnicek, 1965). Some electrons will, of course, penetrate the sheath and be collected by the metallic surface. This is especially the case when a static positive potential is applied to the probe. If this applied voltage is large enough, the sheath may be completely collapsed and then there is a continuous flux of electrons to the probe. The situation of the probe potential possessing an alternating component has been considered by several workers. One effect of the sheath is to shield the ambient plasma from any electric field due to charge on the antenna. The effectiveness of the shielding is dependent on the frequency of the alternating signal. For frequencies less than the plasma frequency the shielding is almost complete, but for much higher frequencies ($\omega >> \omega_p$) the electric field can penetrate far into the

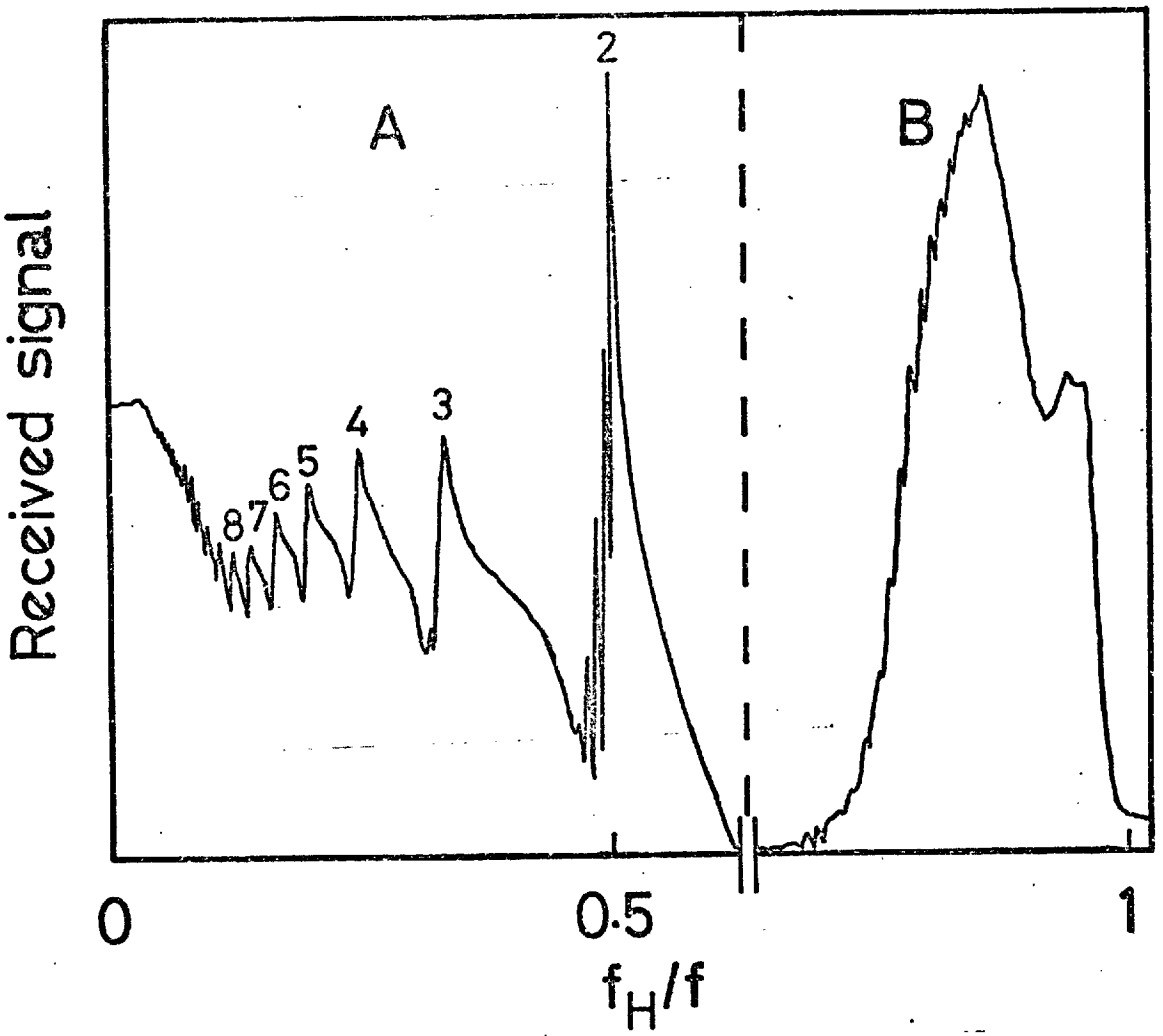


Figure 5.2 (after Mantei, 1967). A specimen record of the variation of received signal (or plasma admittance) with the parameter f_H/f. The section of the curve marked B is plotted on a reduced ordinate scale. This particular plot was obtained at a frequency of 800 MHz, with parallel wire antennae, separated by a distance of 2 cm. The upper hybrid null is apparent at f_H/f = .64 and a prominent admittance peak occurs at f_H/f = .85.

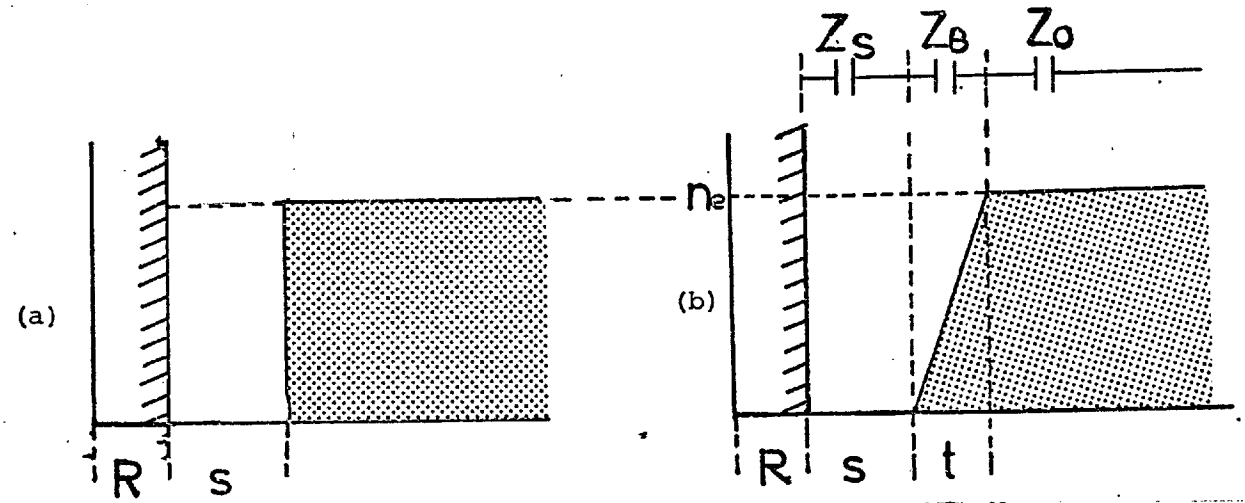


Figure 5.3 Two schematic electron density profiles in the ion sheath region in which the radii and impedances of the vacuum and transmission regions are defined by S, Z_S, t and Z_B. The shaded portion represents an antenna of radius R and the ambient plasma is represented by an impedance Z_o.

plasma. This arises because the oscillating sheath edge cannot accomodate itself to the changes of antenna potential. This problem has been considered by Balmain (1966) in the warm plasma approximation.

However, the most rigorous treatment is that of Buckley (1966). He used plasma kinetic theory and integrated the linearised Boltzmann equation for spherical geometry. His approach assumed an initial potential distribution for the sheath region and determined the frequency response of the sheath-plasma impedance by considering the perturbations to the trajectories of particles close to the probe, due to the partially-screened electric field. As Buckley demonstrated, however, these expressions show little deviation from the earlier simpler theory. Several similar treatments, for different geometries, have also appeared e.g. Laframboise et al. (1975), but in each situation the presence of an ambient magnetic field has been neglected. The complexity of the kinetic approach to the sheath-plasma impedance in a magneto-plasma has, so far, prevented the appearance of any formulation of the problem.

5.3.3. Programme of measurements To investigate the cause and behaviour of the admittance anomaly, between f_H and f_T, illustrated by Figure 5.2, many swept-frequency admittance records have been obtained for a variety of experimental conditions. These data were obtained with antennae of simple geometry so that complicated sheath structures were not introduced.

A selection of these admittance curves is shown in Figure 5.4. In these particular experiments the transmitting antennae were of simple L-shaped configuration with their short arms aligned along the magnetic field. The probe material was tungsten. The receiving antenna was of similar shape and material, but constructed of wire of much larger diameter, so that the capacitance of any sheath surrounding the transmitting probe dominated the reactance of the sheath-plasma system. The data set of

Figure 5.4 correspond to constant anode current, gas pressure and radio-frequency applied to the transmitting probe. The parameters varied in this case are the antenna radius and spacing, so that the effect of probe dimensions on the admittance of the system should be visible.

Examination of the records in Figure 5.4 reveals several interesting points. It is clear that in all cases the admittance anomaly between f_H and f_T is visible. However, in this frequency regime, with increasing antenna spacing, and, in contrast to the admittance in the Bernstein mode, pass-bands located near the gyroharmonics (nf_H, $n \geq 2$), the admittance shows a marked increase. This admittance increase is not explained by simple theory, being contrary to the expected fall off of electric field amplitude with distance.

Furthermore, the peak of admittance between f_H and f_T appears to occur at different frequencies for different antenna radii. This shift, to lower values of f_H/f with decreasing probe size would be the effect expected if the anomaly were caused by series coupling of a sheath surrounding the antenna and the ambient plasma.

The maximum of admittance, observed at the electron gyrofrequency in normal records (Andrews (ibid)), is not present in the curves of Figure 5.4. Indeed, at or near this frequency, the admittance falls to a minimum, in a similar way to the result obtained by Mantei (ibid), given in Figure 5.2. Again this decrease in admittance is not predicted by simple theory and the existence of an ion sheath around the antenna must be invoked to explain some of these effects.

Although very few observations of the sheath-plasma series resonance in a magnetoplasma have been reported, Andrews (1969), has suggested that the observed admittance anomaly reported here is, in part, attributable to such coupling of the sheath-plasma system.

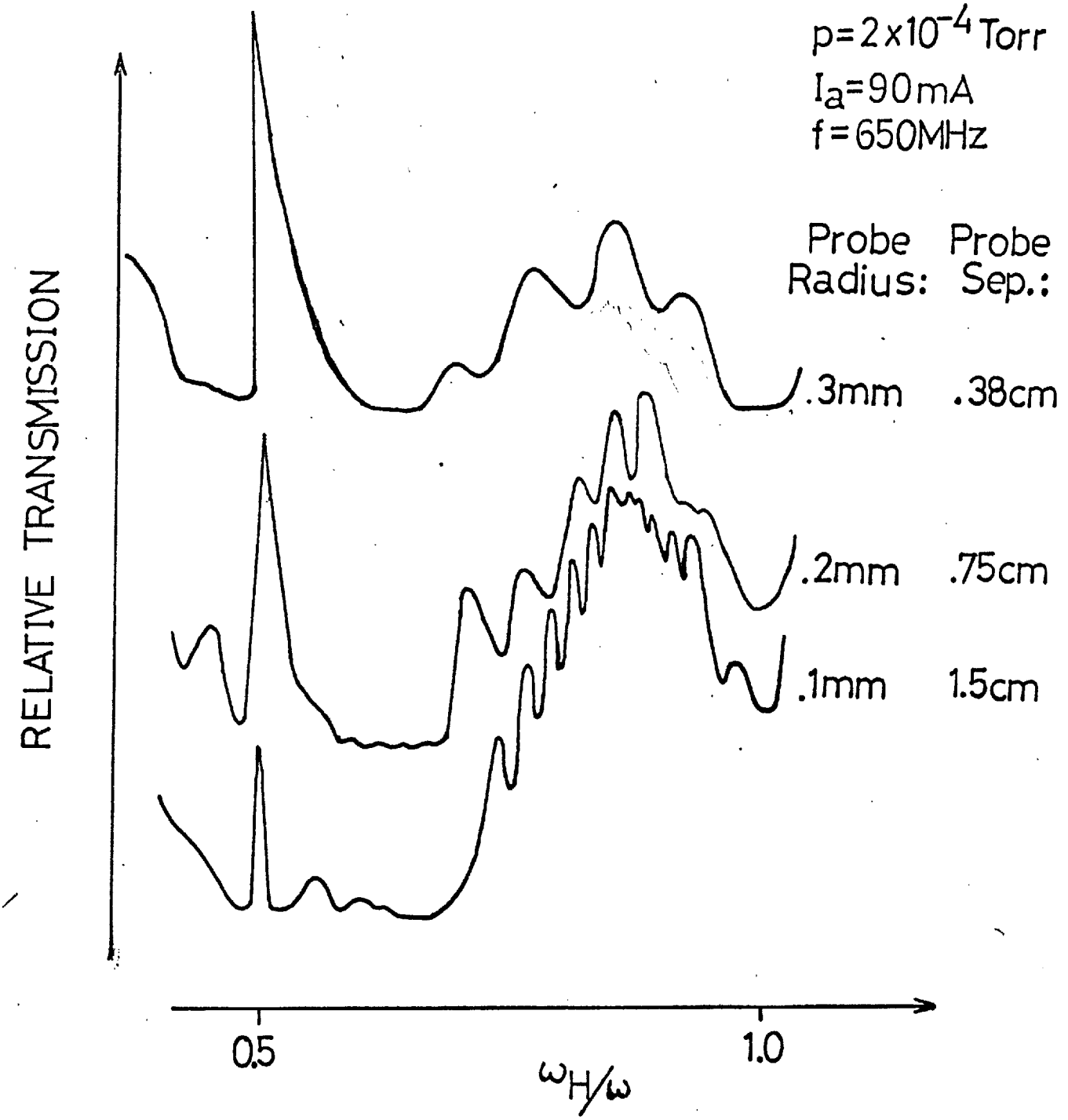


Figure 5.4 A series of admittance records obtained with the constant plasma and external parameters shown. This set of measurements show the effect on the admittance of the sheath-plasma system, of varying the antenna separation and radius only.

5.3.4. Interpretation of experimental records

Generally speaking, the sheath represents a coupling element for any transfer of energy between a probe and a plasma. Therefore it is often represented in terms of a model network, the combined elements of which, sheath capacity and resistance, represent complex physical structures and processes.

The sheath-plasma system may be represented by the network of Figure 5.5, where the plasma impedance between electrodes is Z_m, Z_1 and Z_2 are stray impedances to other grounded electrodes, and Z_{s_i} the capacitive reactances of the sheaths around the electrodes. Ideally $Z_{s_1} = Z_{s_2} = Z_{i_n} = 0$; then Z_m, a true plasma parameter, is given by

$$\left.\frac{V_1}{i_2}\right|_{V_{2=0}} = -Z_m$$

For expected sheath thicknesses, and the given geometry we have

$$Z_1 \sim Z_2 > Z_m > Z_{s_2} \sim Z_{s_1} > Z_{s_3}$$

Then approximately

$$\frac{V_1}{i_2} = -(Z_m + Z_{s_1} + Z_{s_2} + Z_{i_n})$$

Except at frequencies close to f_H, the last three impedances are rather less than Z_m, and they are in any case approximately fixed. Fluctuations in the received current in a low impedance detector, when the system is driven by a constant voltage, therefore represent plasma admittance changes, and any oscillating component is related to cyclotron waves propagating between probes.

In this network the plasma is represented by the simple cold plasma impedance between two cylindrical wires, and the sheath forms are those represented by the electron density profiles of Figure 5.3. In this case the

variation of normalised impedance of the plasma and sheath with f_H/f is shown schematically in Figure 5.6. This indicates that the impedance of the vacuum region sheath remains capacitive at all frequencies, while at frequencies between f_H and f_T, the plasma is inductive (i.e. relative impedance <0). In the linear gradient sheath model (Figure 5.3b), the gradient region contains a small frequency dependent impedance but this is much less than the corresponding variation for the plasma medium.

It is possible, therefore, that a series coupling of the sheath and plasma impedances may occur and the impedances cancel at a resonant frequency, which is located just below the gyrofrequency, in the region where the sheath and plasma impedances are of similar magnitude but of opposite sign. Such a frequency would be expected to shift with the impedance of the sheath region.

In the cold plasma approximation the relative permittivity of a collisionless plasma across a static magnetic field is given by

$$\frac{\varepsilon_c}{\varepsilon_o} = 1 - \frac{\omega_p^2}{\omega^2 - \omega_H^2} \qquad 5.1$$

For the case of two cylindrical antennae of radius R, separation 2d, length L, each surrounded by a vacuum region of size S and a regime of linearly increasing electron density of dimension t the impedance of the plasma between the two sheaths, Z_o, is given by

$$Z_o = \frac{\cosh^{-1} d/R}{i\omega\pi\varepsilon_c} \qquad 5.2$$

Furthermore the impedance of the vacuum sheath region, Z_A, is adequately represented by

$$Z_S = \frac{1}{i\omega 2\pi\varepsilon_o L} \ln\left(\frac{R+S}{R}\right) \qquad 5.3$$

The linearly increasing density region can be divided into a series of infinitessimal elements of width Δr, constant electron density and incre-

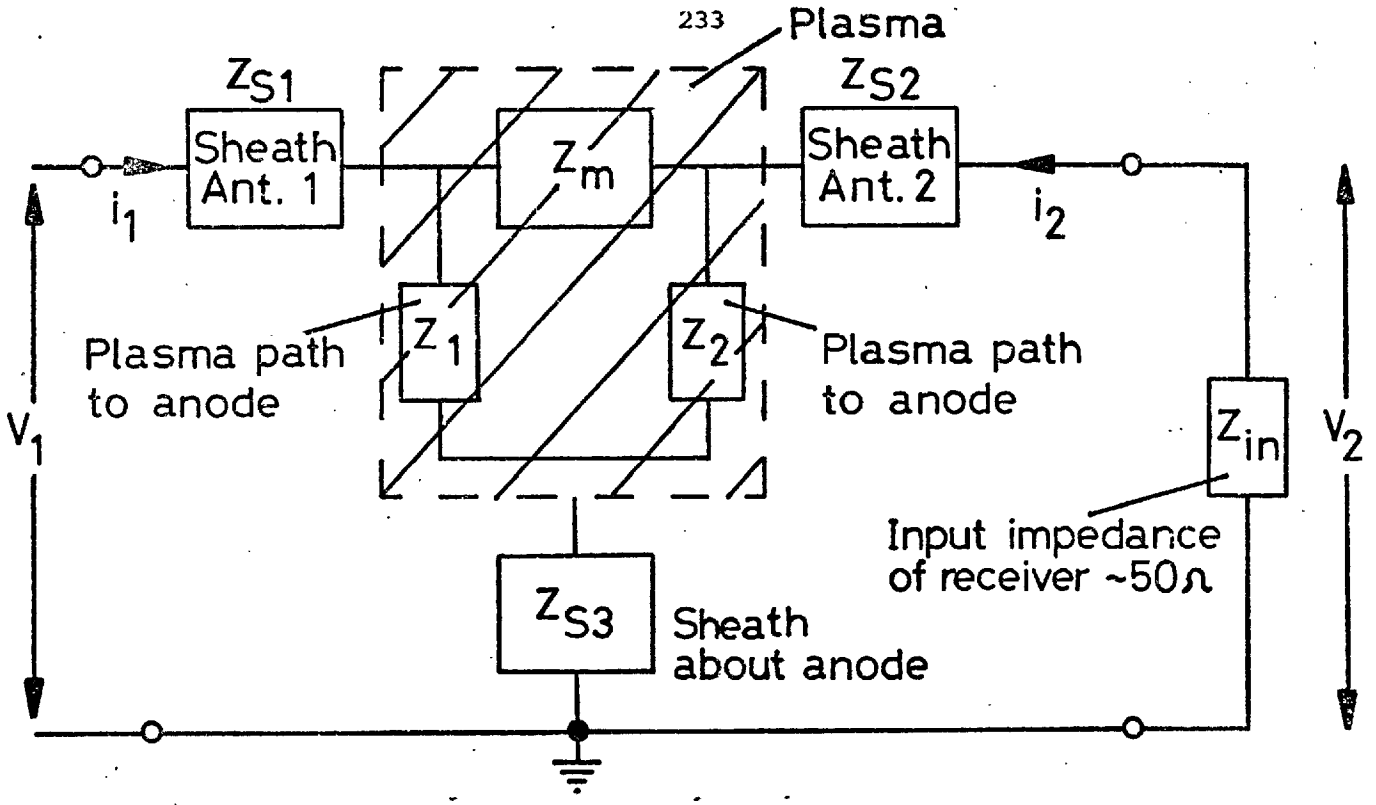


Figure 5.5 A network model of the sheath-plasma system. In essence the antenna, sheath and electrodes in the discharge tube can be regarded as a four terminal circuit. The meanings of the symbols are explained in the text.

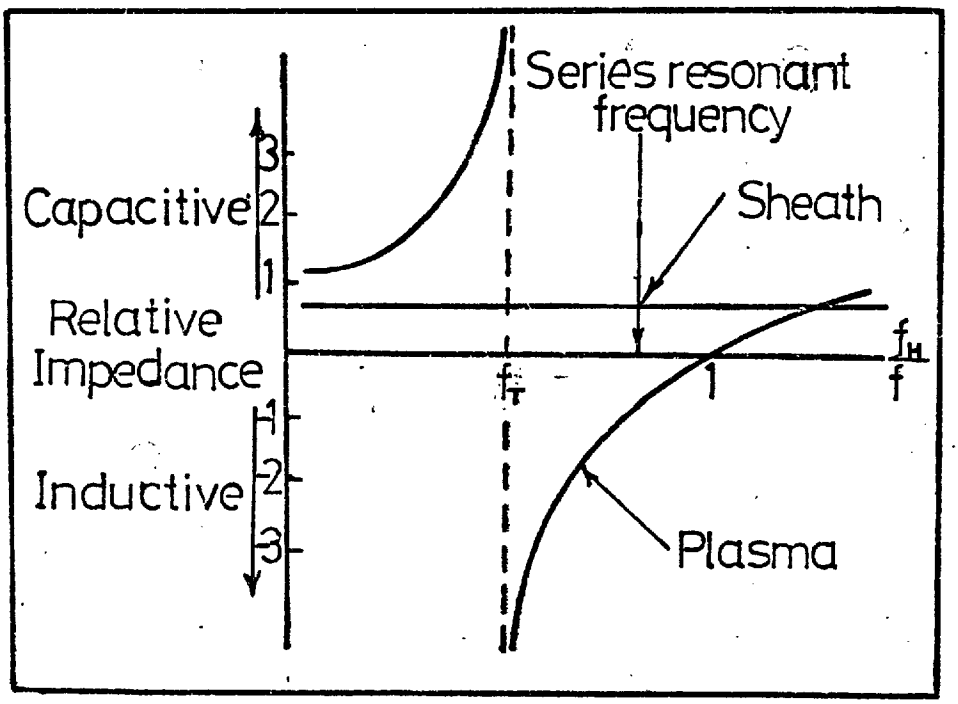


Figure 5.6 Schematic behaviour of the variation of relative plasma and vacuum sheath impedance with the parameter f_H/f.

mental impedance ΔZ_B. The total impedance of the region is then given by $\Sigma\Delta Z_B$. Balmain(1966) has derived the following expression for ΔZ_B:

$$\Delta Z_B = \frac{1}{i\omega 2\pi L \varepsilon_c(r)} \ln\left(\frac{r + \Delta r}{r}\right) \qquad 5.4$$

Here $\varepsilon_c(r)$ is given by equation 4.13 cold plasma considerations in Chapter 4, and is, of course, dependent on distance from the probe through the electron plasma frequency ω_p $(= (Ne^2/m\varepsilon_o)^{\frac{1}{2}})$. Integration of 5.4 yields the impedance of the transition region Z_B. Thus

$$Z_B = \Sigma\Delta Z_B = \frac{1}{i\omega 2\pi L \varepsilon_o} \frac{t}{t + (1-\varepsilon_c/\varepsilon_o)(R+S)} \ln\left(\frac{t + S + R}{\varepsilon_c/\varepsilon_o (R+S)}\right) \qquad 5.5$$

The condition for occurrence of the sheath-plasma resonance is then

$$Z_o + Z_S + Z_B = 0 \qquad 5.6$$

5.3.5. Results and discussion The two simple sheath models illustrated by Figure 5.3 have been adopted and the predicted position of the admittance peak determined for each set of experimental conditions and for vacuum and gradient electron density regions of varying dimensions. These can then be compared with the actual value of ω_H/ω at which the admittance maximum occurs in the records of Figure 5.4.

In agreement with experiment, all sheath models exhibited the same behavioural trend, viz. the admittance peak moved to lower values of ω_H/ω with decreasing antenna radius. In addition a decrease of probe separation forced the admittance anomaly to progressively smaller values of ω_H/ω also. Thus in a qualitative way the sheath-plasma series resonance explains the variation in position of the admittance anomaly. In the frequency range of interest the vacuum region sheath is capacitive while the plasma is inductive. The so-called resonance occurs where a series combination of these impedances is zero. An increase of antenna radius acts so as to decrease sheath impedance and hence the

resonance occurs closer to the gyrofrequency. Decreasing probe separation has the opposite effect, in that the inter-electrode impedance is decreased and the reactance of the sheath region becomes more important in the sheath-plasma system.

Of the sheath models adopted that with a vacuum region of extent $3\lambda_D$, and a linearly increasing density region of size $2\lambda_D$, gave the best qualitative reproduction of the experimental measurements. The agreement between experiment and the calculated values of ω_H/ω at which the admittance peak occurs is given in Table 5.2 and plotted for varying antenna radius and for a fixed antenna separation in Figure 5.7. In general, the agreement seems satisfactory but it is noticeable that the simple theory predicts that the admittance peak should occur closer to the gyrofrequency than is observed in practice. This is most evident at the smallest antenna radius used, and may imply that the sheath size is affected by antenna dimensions, a factor not taken into account in the simple adopted models.

This result seems to be in accordance with that of Kist (1976), who, using non-parallel probes, not identically aligned with the magnetic field, has used the resonances of the sheath-plasma system to probe the sheath region. He found its series resonance to be very dependent on probe radius and observed that it moved to lower frequencies with increasing probe radius.

The explanation invoked here can be tested in a variety of other ways. For example, the effect of a positive D.C. bias on the antenna has been investigated. Application of a large enough potential should collapse the sheath and remove the admittance peak. In this particular case attempts were inconclusive. However, similar admittance peaks in the results of Kist (ibid) were strongly affected by the potential applied to the electrode. These results were as expected from a simple model

of the sheath-plasma system.

The qualitative success of the sheath-plasma resoncnace in explaining the admittance anomaly suggest a possible use of the admittance curves as a diagnostic. Such curves, through location of the upper hybrid transmission null, provide an accurate and swift measure of the electron density. The position of this null remains unaffected by the presence of a sheath. This arises because of the impedance infinity at this frequency. However, it is possible that further parameters, such as sheath size and Debye length, may be obtainable from admittance records containing the admittance peak. To do this it must first be assumed that the peak is due to the sheath-plasma resonance. From the position of the resonance peak it is then possible to determine the sheath capacitance and, in turn, using a simple sheath model compute its size. Furthermore, such measurements, together with an accurate knowledge of the form of the sheath region, in terms of the Debye length, from theoretical sheath models, can yield measurements of the electron temperature, via the Debye length. This has been attempted recently by Holt and Troim (1970). It would appear, however, that, before accurate values of temperature can be obtained, further work on the form of the sheath is necessary. For although the simple models used here give a satisfactory quantitative agreement with experiment, there is a systematic deviation from current theory of the frequency of occurrence of the admittance peak at large antennae radii. Present knowledge of the sheath contains no account of such behaviour.

5.4 Experiments With Moving Probes

5.4.1. Frequencies below the upper hybrid. At frequencies below the upper hybrid frequency the amplitude of the plasma wave mode seems large and propagation is easily obtained. Figure 5.8 shows a typical example. of the variation of received signal with probe spacing. The curve closely

Table 5.2

Probe Radius $x10^{+4}$m	Probe Spacing $x10^{+2}$m	Position of admittance peak	
		Theory	Observation
1.0	1.50	0.89	0.87
2.0	1.50	0.91	0.90
3.0	1.50	0.91	0.94
1.0	0.75	0.88	0.86
2.0	0.75	0.90	0.88
3.0	0.75	0.91	0.88
1.0	0.38	0.76	0.82
2.0	0.38	0.78	0.84
3.0	0.38	0.81	0.85

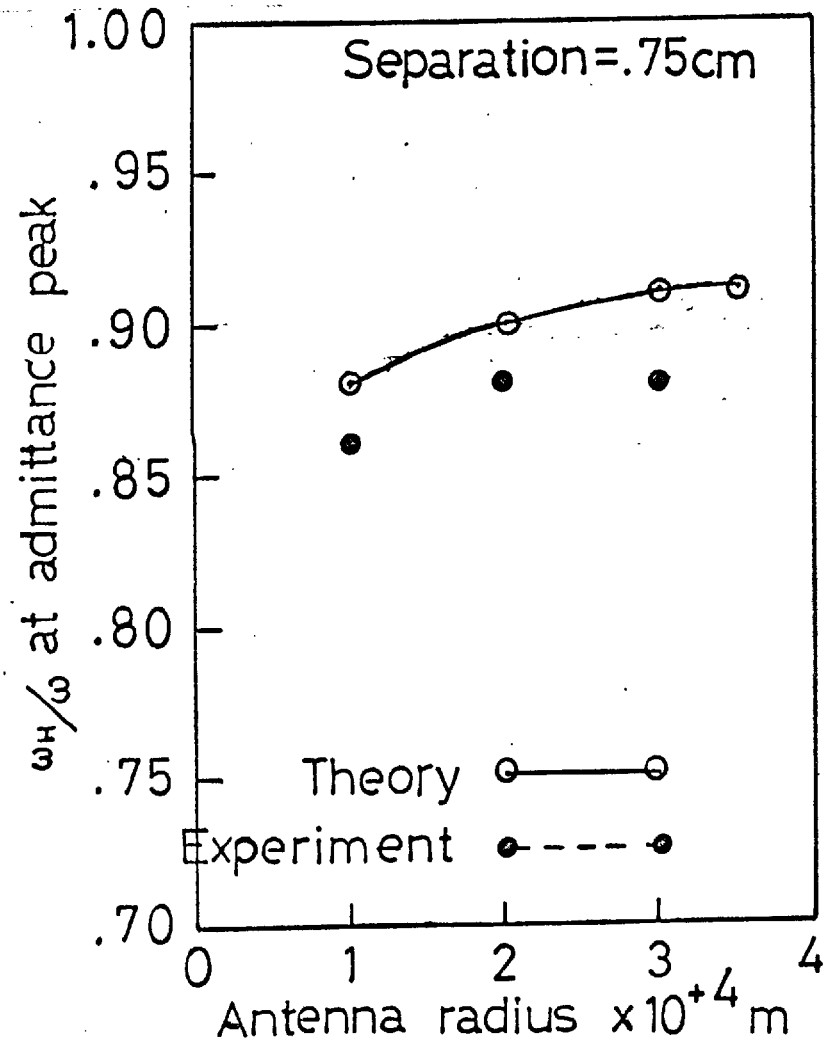


Figure 5.7 The theoretical and experimental movement of the admittance anomaly with antenna radius. The open circles were calculated using the gradient sheath model described in the text, and the solid circles represent experimental observations obtained at a probe separation of 0.75 cm.

follows the theoretical form of 4.28, with the addition of a spatially decaying component. As outlined in the theory, a value of the Bernstein mode wave-length can be read off directly from such a record. These measurements, when combined with a swept-frequency admittance plot to identify the upper hybrid frequency from the transmission null, and a set of computed dispersion curves, can be used to estimate the plasma electron temperature.

Thus, in the example given, the plasma frequencey was initially found to be 656 MHz, signifying an electron density of 5.9×10^9 cm^{-3}. Using this value and the known gyrofrequency, f_H, to fix the density factor f_N/f_H, dispersion curves, showing the relationship between the normalised variables f/f_H and kr (Figure 4.3), were computed assuming a Maxwellian form for the particle velocity distribution of the plasma. Such curves, and the value of f/f_H, corresponding to the input frequency and ambient magnetic field, provide the parameter kr of the excited wave. k is known from the direct measurement of wavelength from the transmission record and hence an electron temperature, contained in the expression of gyroradius r, can be derived. In this example (Figure 5.7) the indicated temperature is 2.7 eV.

This value is considerably lower than the value of $\sim$ 10 eV inferred from use of a Langmuir probe. However, it is thought that parameters derived from such methods are strongly dependent on the high-energy component of the plasma particle velocity distribution and consequently greater confidence can be placed in diagnostic measurements utilising electrostatic waves.

5.4.2. Frequencies above the upper hybrid. In the situation where the input frequency exceeds the plasma upper hybrid frequency, theory predicts that two Bernstein mode waves are excited. With two modes present, the variation, with probe spacing, of the signal at the receiving probe is given by

$$2E_o^2 \left(k_o + R_1 \sin (k_1 x - \phi_1) + R_2 \sin (k_2 x - \phi_2) + R\cos (k_1 - k_2)\, x\right)$$

where R, R_1 and R_2 are amplitude factors depending on the amplitudes of the individual modes, and ϕ_1 and ϕ_2 are similarly dependent phases. There are thus three oscillatory components in the spatial variation of received signal and the wave numbers k_1 and k_2 are not directly measurable.

Therefore, on simple theory, in the propagation bands above the upper hybrid frequency, a further modulation of the received signal should be apparent. Buckley (1970), using a vanishingly small collision frequency, has computed the relative mode amplitudes. He found that, above the second gyroharmonic the long wavelength or low-k mode amplitude was greater than the high-k mode amplitude by a factor varying between 2 and 10. Experiments, using amplitude modulation techniques, have, however, not detected simultaneous propagaion of two waves. This null result appears to be due to the low group velocity of the high-k mode which results in high collisional damping of this wave. The ratio of mode amplitudes is consequently much greater than that given by Buckley.

However, by working at low ambient pressure and aligning the probes and receiver for maximum sensitivity, it has proved possible to observe simultaneous excitation of two Bernstein modes in the propagation band above the second gyroharmonic. An example of such a record is shown in Figure 5.9. The existence of the short wave-length mode is clearly suggested by contrast of these experimental data with Figure 5.8, in which only one mode was excited.

The application of spectral analysis techniques to such records to facilitate comparison of theoretical and experimental relative mode amplitudes is under investigation. In this way it may be possible to derive information on the collisional damping process and attenuation of the wave in a plasma.

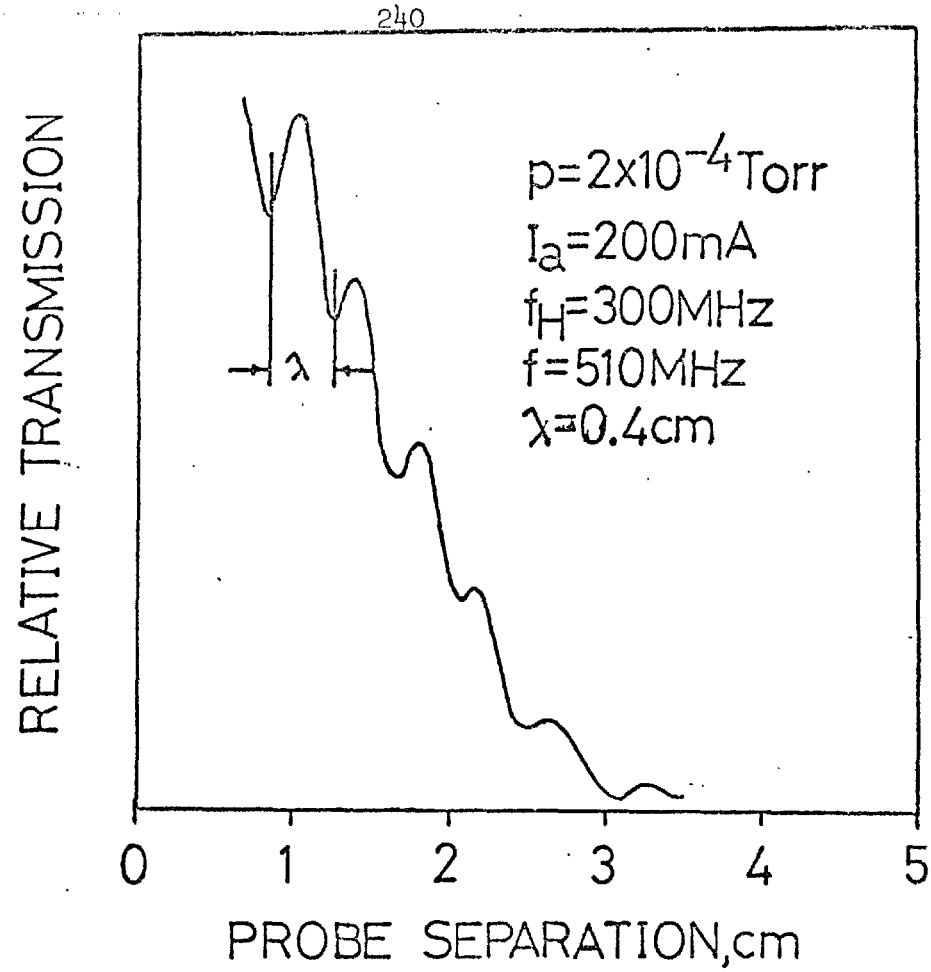


Figure 5.8 The variation of received signal with probe spacing, at frequencies below the upper hybrid. The pattern arises from the superposition of the capacitively-coupled signal, the single plasma wave mode and a decaying background component.

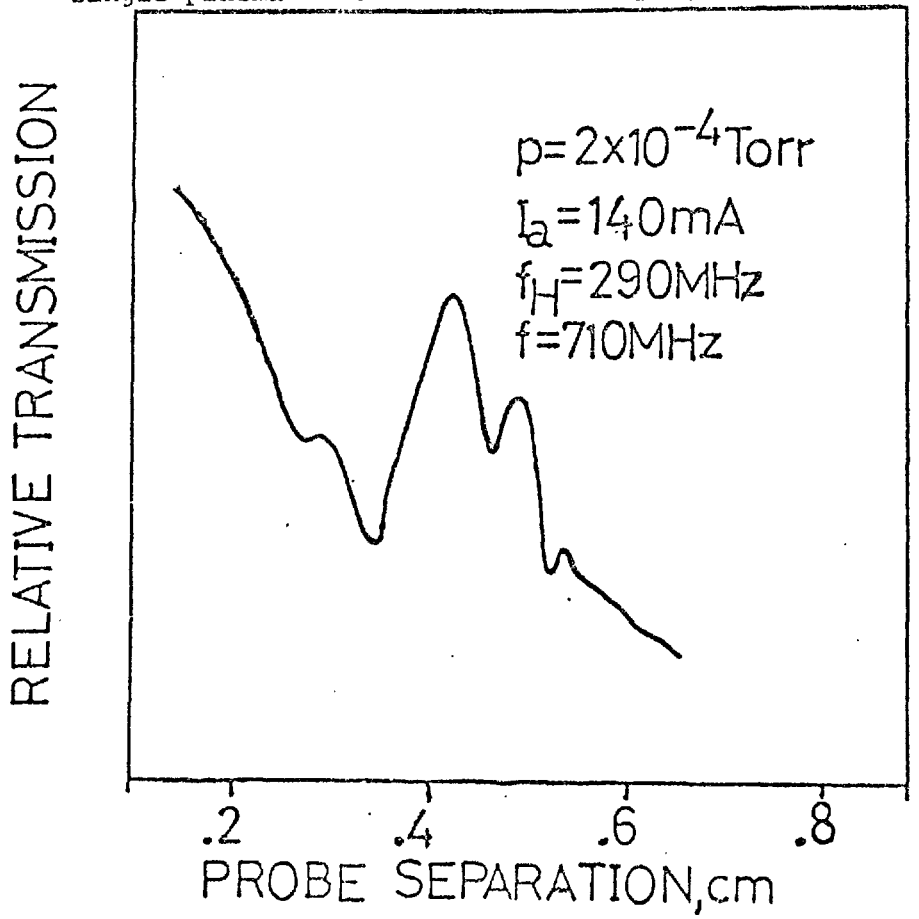


Figure 5.9 Propagation above the upper hybrid frequency. The spatial variation of received signal in this record is distinct from those obtained at frequencies below the upper hybrid, and suggests the presence of a second Bernstein mode.

5.5 Plasma Wave Experiments in Space

Of the wide variety of plasma wave experiments performed in the laboratory several are suitable, with modification, for examination of wave propagation in the Earth's plasma envelope. Furthermore, transmission experiments which have been used for laboratory plasma diagnostics may yield accurate values of space plasma parameters.

Until recently the majority of diagnostic space plasma experiments have been of the self-admittance and resonance relaxation types. The former swept-frequency measurement suffers the disadvantage of lack of rigorous theory for a probe in an anisotropic medium; (Andrews and Hall, 1970). The latest application of the resonance relaxation technique is on the geostationary GEOS satellite. By suitable choice of frequency and receiver bandwidth, the frequencies and fine structure of the characteristic plasma resonances will be examined. (Petit and Beghin, 1970). Under some circumstances, this fine structure, due to interference of two waves of slightly different frequency, can be used to estimate ambient electron temperature (Warnock et al, 1970).

Assemblies of more than one probe have also been proposed, for example the quadrupole probe. In this experiment the mutual impedance between two pairs of probes is determined as a function of frequency. In this way many of the difficulties associated with electrical coupling to the spacecraft body are overcome. A similar type of measurement has been suggested by Andrews and Hall (ibid). They envisage a transmission experiment in which the swept-frequency admittance between two parallel probes is determined. A transmitted signal of 1-5 MHZ excites both capacitive and electrostatic wave modes. As the frequency varies the admittance changes in the characteristic manner described in Section 4.4.3. Thus the electrostatic wavelenths are obtained only indirectly from the subsidiary frequency variations of the admittance.

These techniques suffer several disadvantages. Most notable of these is the role of the plasma sheath. Laboratory experiments (Section 5.3) illustrate the importance of the sheath in determining the overall admittance behaviour of the probe-plasma system. In the ionosphere the sheath is small ($\sim$10cm), but values of 20 m are not uncommon at geostationary orbit. Thus transmission experiments may not be possible at high altitudes due to restrictions on the length of the transmission path, which must be many times the sheath dimension. In addition probes must be deployed away from the spacecraft sheath.

In the lower ionosphere, however, such restrictions are not critical and transmission experiments may be employed. These may be especially interesting at frequencies above the upper hybrid where, perpendicular to the magnetic field, two cyclotron harmonic waves propagate. Laboratory experiments have failed to detect both modes with any consistency, recent data suggesting that this is due to strong collisional damping of the short-wavelength mode. In space where the electron-neutral collision frequency is much lower it is probable that simultaneous detection of two Bernstein mode waves can be easily observed. Furthermore, it would be an advantage if the method yielded direct measurements of wave number, k, corresponding to each exciting frequency.

Of tested laboratory methods, that of moving probes is the most direct measurement of wave number. However, at present such probe arrangements seem impractical for space applications. Although this situation may change with the advent of the space shuttle, an electronic method is required for measurement of wave number. Some workers (Tsutsui et al, 1974) have suggested that this may be accomplished by transmission of a frequency modulated wave. Adapting this suggestion to the experimental arrangement of Figue 5.1, a frequency modulated wave of centre frequency ω_o and modulation frequency Ω is applied to transmitting atenna.

Thus,

$$v_T = A \sin \left(\omega_o t + \frac{\Delta\omega}{\Omega} \sin \Omega t\right)$$

If the probes are separated by a distance z and the frequencies under consideration are below the upper hybrid, then the received signal will consist of a capacitively-coupled signal and a single electrostatic mode. Hence the signal in the receiving antenna is given by V_R, where

$$v_R = B \sin \left(\omega_o t + \frac{\Delta\omega}{\Omega} \sin \Omega t + \phi_1\right)$$
$$+ C \sin \left(\omega_o t + \frac{\Delta\omega}{\Omega} \sin \Omega t + (k_o + \Delta k \cos \Omega t)\, z + \phi_2\right)$$

Here k_o is the wave number excited by the frequency ω_o and Δk is the amplitude of wave number variation corresponding to a change in frequency $\Delta\omega$.

Figure 5.10 shows a schematic experiment for deriving the wave number amplitude from such a received signal. The signal is first multiplied by the transmitted frequency, offset by a frequency derived from a local oscillator. The instantaneous output frequency, ω_i, after filtering, then contains information concerning the plasma wave number and is given by

$$\omega_i = \omega_s + (\Delta k . z . \Omega) \sin \Omega t$$

Thus wave number values can be derived from the output frequency of the low-pass filter or, alternatively, as shown in Figure 5.10, commonly-used frequency demodulation methods can be applied and the wave number variation recorded as an amplitude modulation.

The practical difficulties associated with a space experiment do not seem too complex. The centre or carrier frequency would need to be in the range of 1-5 MHZ with a modulation frequency of 1 kHz. By sweeping or stepping the output frequency over this range, measurements are possible in the passbands near the higher order gyroharmonics. The amplitude of the frequency modulation is restricted by signal generation

considerations but this does not seem critical because of the excellent signal to noise ratio characteristics of FM transmission. In the plasma of the lower magnetosphere the wavelength of modes excited will take values .2 - 1 m and, since probe dimensions must not be less than the scale length of the transmissions, this will fix the dimensions of the antennae. Thus a probe size of 2m seems appropriate, with an orientation as close to the magnetic field as possible. However, before such experiments are performed in space a full laboratory investigation must first be carried out. This can be accomplished with the discharge tube shown in Figure 5.1, with suitable transmitting and receiving systems attached to the probes.

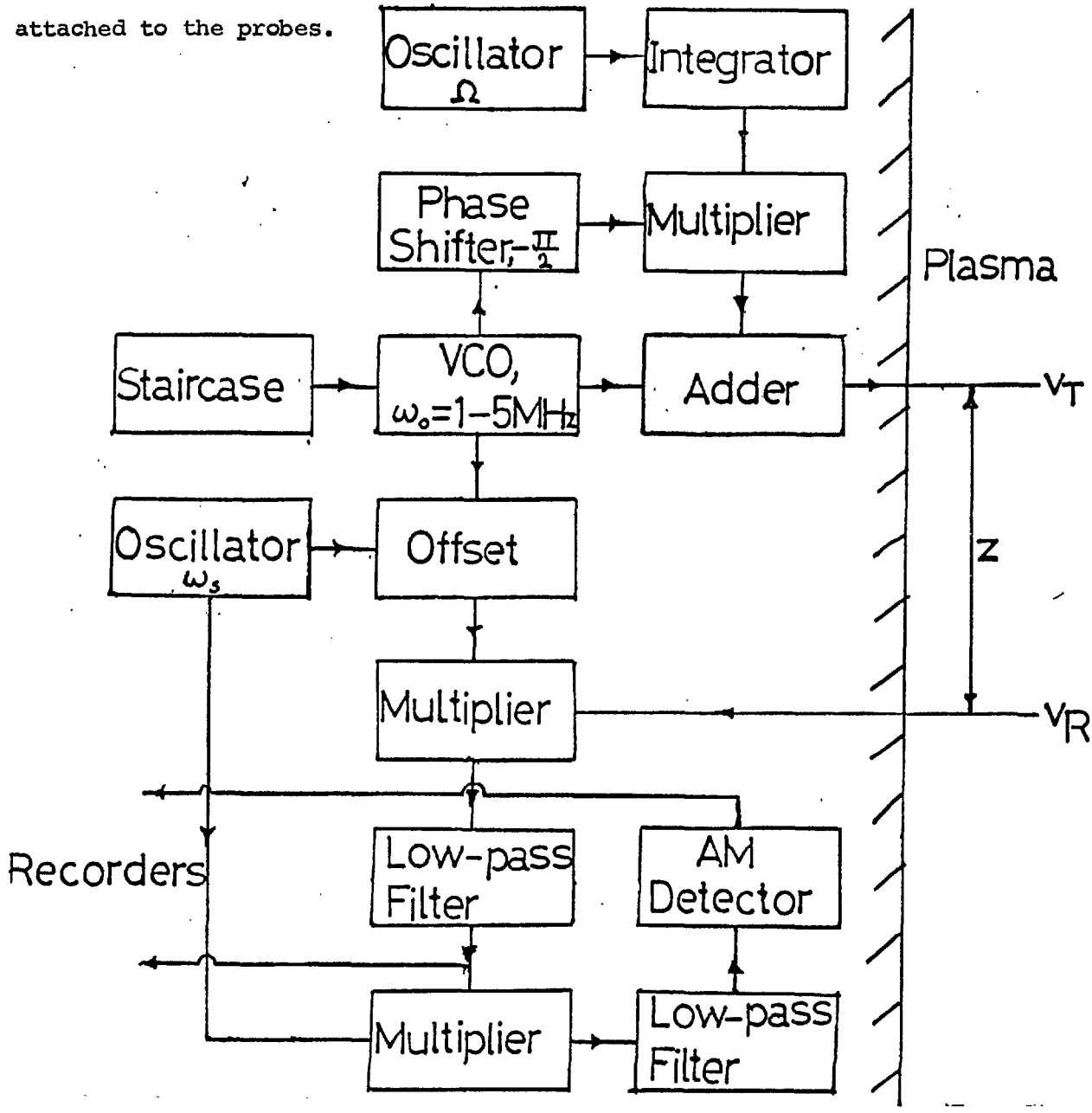


Figure 5.10 Details of an experimental arrangement designed to measure Bernstein mode dispersion characteristics by frequency modulation techniques. The method can be used in the laboratory or adapted for space experiments.

5.6 Conclusions

The use of plasma waves as a laboratory plasma diagnostic seems well-established and the good qualitative, and satisfactory quantitative explanation of the admittance anomaly, by assuming the presence of an ion sheath, suggests a further application of these waves. Very few measurements of the sheath in a magneto-plasma exist, but it appears possible that swept-frequency admittance curves, obtained with the current experimental arrangement, can provide information on its size. To confirm that the admittance anomaly is due to series coupling of the sheath and plasma, further experiments can be performed. In this respect, repetition of inconclusive tests with biassed probes and records obtained with electron temperature as a variable, would be useful. Such investigations are valuable in regard of current magnetospheric probing by satellite instruments, in which the sheath may play an important role.

Measurements with moving probes indicate that simultaneous propagation of the two Bernstein modes, in the existing discharge tube, is possible. Collisional damping, however, makes such observations difficult, since it affects waves with low group velocity. Therefore, it is possible that a satisfactory understanding of propagation above the upper hybrid frequency must be carried out in other discharges or space, where the electron-neutral collision frequency is low. Of recently-proposed experiments, a frequency modulation technique, for measurement of the dispersion relation, appears the most direct way of deriving information on plasma parameters. Adaption of this method to a laboratory plasma, such as that used here, would be a first step in assessing its usefulness in space.

REFERENCES (Part II)

Andrews, M.K., Ph.D. Thesis, University of London, 1969.

Andrews, M.K. and T.A. Hall, Imperial College Internal Report, SP T102-70, 1970.

Balmain, K.G., Rad. Sci. 1, 1, 1966.

Balmain, K.G. and G.A. Oksiutik, In Plasma Waves in Space and Laboratory, Vol. 1, eds. J.O. Thomas & B.J. Landmark, Edinburgh University Press, 1970.

Bernstein, I.B., Phys. Rev. 109 (1), 10, 1958.

Bitoun, J., Rad. Sci. 9, 17, 1974.

Buchsbaum, S.J. and A.A. Hasegawa, Phys. Rev. Lett. 12, 685, 1964.

Buckley, R., Proc. R. Soc. A 290, 186, 1966.

Buckley, R., in Plasma Waves in Space and Laboratory, Vol. 2, eds. J.O. Thomas & B.J. Landmark, Edinburgh University Press, 1970.

Buckley, R., Imperial College Internal Report, SP T01-74, 1974.

Christopoulos, C. and P.J. Christiansen, Plasma Phys. 16, 499, 1974.

Clemmow, P. and J.P. Dougherty, Electrodynamics of Particles and Plasmas, Adison-Wesley, London, 1969.

Clinkemaille, A., in Plasma Waves in Space and Laboratory, Vol. 2, eds. J.O. Thomas & B.J. Landmark, Edinburgh University Press, 1970.

Crawford, F.W., in Plasma Waves in Space and Laboratory, Vol. 2, eds. J.O. Thomas & B.J. Landmark, Edinburgh University Press, 1970.

Crawford, F.W. and R.S. Harp, Int. Memo. No. 1174, Microwave Lab. Stanford University, 1964.

Crawford, F.W., Kino, G.S. and H.H. Weiss, Phys. Rev. Lett. 13, 229, 1964.

Dougherty, J.P., In Plasma Physics, ed. B.E. Keen, The Institute of Physics, London, 1974.

Hall, T.A. and G. Landauer, Rad. Sci. 6, 967, 1971.

Harp, R.S., Appl. Phys. Lett. 6, 51, 1965.

Harp, R.S. and F.W. Crawford, J. Appl. Phys. 35, 3436, 1964.

Harp, R.S. and G.S. Kino, Proc. 6th Int. Conf. on Ionisation Phenomena in Gases 3, 45, 1964.

Holt, O. and J. Troim, In Plasma Waves in Space and Laboratory, Vol. 2, eds. J.O. Thomas & B.J. Landmark, Edinburgh University Press, 1970.

Kist, R., Preprint, 1976.

Kostelnicek, R.J. Aeronomy Rep. No. 8, University of Illinois, 1965.

Laframboise, J.G., Rubinstein, J. and F.H. Palmer, Rad. Sci. 10, 773, 1975.

Landau, L.D., J. Phys. Moscow 10, 25, 1946.

Leuterer, F., Plas. Phys. 11, 615, 1969.

Leuterer, F., Plas. Phys. 14, 499, 1972.

Mantei, T.D., SUIPR Rep. 194, 1967.

Mayer, H.M., Proc. 6th Int. Conf. on Ionisation Phenomena in Gases, Paris, 1964.

McAfee, J.R., J.G.R. 73, 5577, 1968.

McAfee, J.R., J.G.R. 74, 802, 1969.

Petit, M. and C. Beghin, EST/RSR Tech. Note No. 53, (S301, S304), 1970.

Phelps, A.D.R., Imperial College Internal Report, SP T101-71, 1971.

Stix, T.H., Theory of Plasma Waves, McGraw-Hill, New York, 1962.

Storey, L.R.O., Aubry, M.P. and P. Meyer, In Plasma Waves in Space and Laboratory, Vol. 2, eds. J.O. Thomas & B.J. Landmark, Edinburgh University Press, 1970.

Tataronis, J.A. and F.W. Crawford, J. Plas. Phys. 4, 231, 1970.

Thomas, J.O., Andrews, M.K. and T.A. Hall, Imperial College Internal Report, SP T105-70, 1970.

Thomas, J.O., Phil. Trans. R. Soc. Lond. A 280, 193, 1975.

Tsutsiu, M., Matsumoto, H. and S. Migatake, Rep. Ion. & Space Res. in Jap. 28, 55, 1974.

Warnock, J.M., McAfee, J.R. and F.L. Thompson, J.G.R. 75, 7272, 1970.

Wharton, C.B., Proc. 4th Int. Conf. on Ionisation Phenomena in Gases, Upsala, 1960.