

24/8-72

DET NORSKE VIDENSKAPS-AKADEMI I OSLO

GEOFYSISKE PUBLIKASJONER

GEOPHYSICA NORVEGICA

Vol. 29.

June 1972

A Collection of Articles on Cosmic Geophysics Dedicated to the
Memory of Professor L. Harang on the Seventieth Anniversary of His Birth,
19 April 1972

DET NORSKE METEOROLOGISKE INSTITUTT

BIBLIOTEKET
BLINDERN, OSLO 3

OSLO 1972

UNIVERSITETSFORLAGET

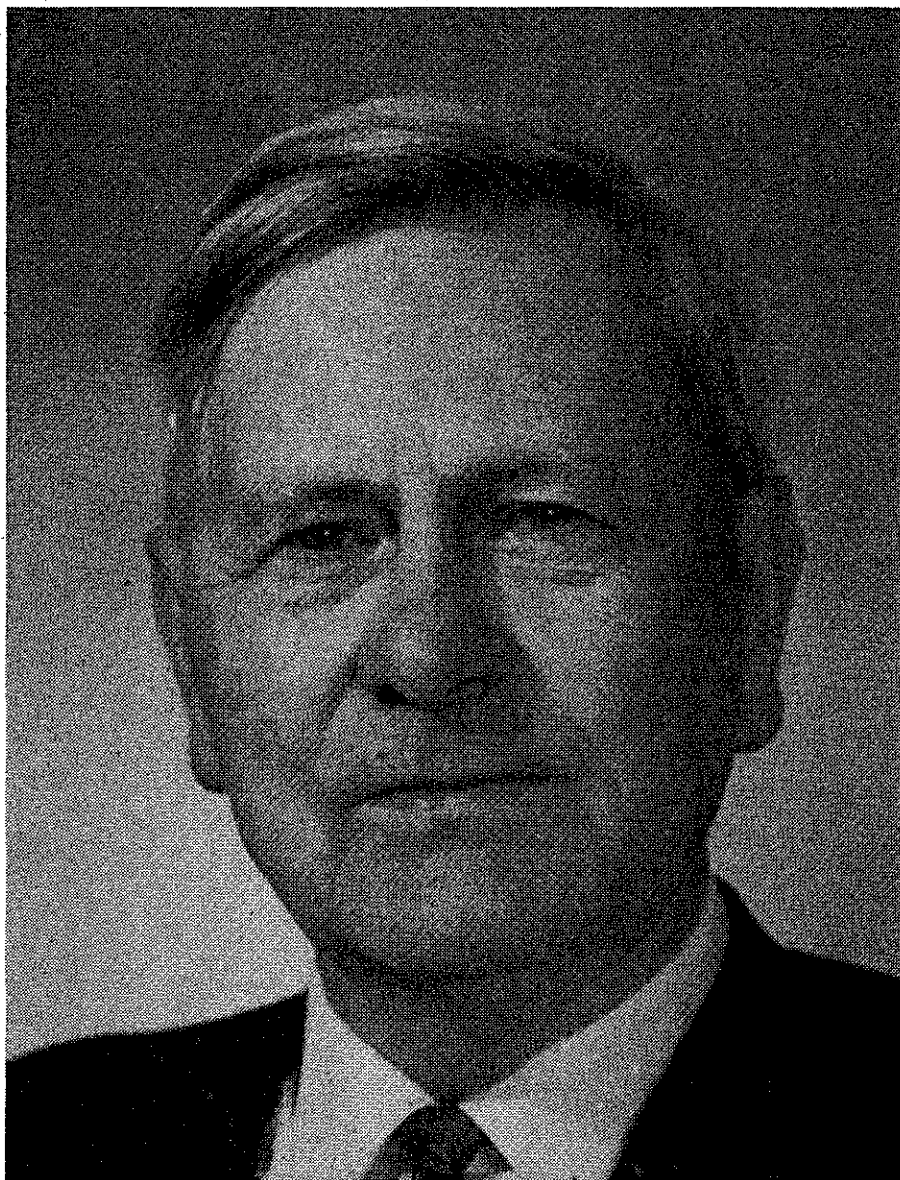


A Collection of Articles on Cosmic Geophysics Dedicated
to the Memory of Professor L. Harang on the Seventieth Anniversary of His Birth,

19 April 1972

Edited by

J. Holtet and A. Egeland



LEIV MARIUS HARANG

Born 19 April 1902, died 21 September 1970.

Professor of Ionospheric Physics
at the University of Oslo and Senior Scientist at the Norwegian
Defence Research Establishment, Kjeller.

DET NORSKE VIDENSKAPS-AKADEMI I OSLO

GEOFYSISKE PUBLIKASJONER

GEOPHYSICA NORVEGICA

Vol. 29.

June 1972

A Collection of Articles on Cosmic Geophysics Dedicated to the
Memory of Professor L. Harang on the Seventieth Anniversary of His Birth,
19 April 1972

OSLO 1972

UNIVERSITETSFORLAGET

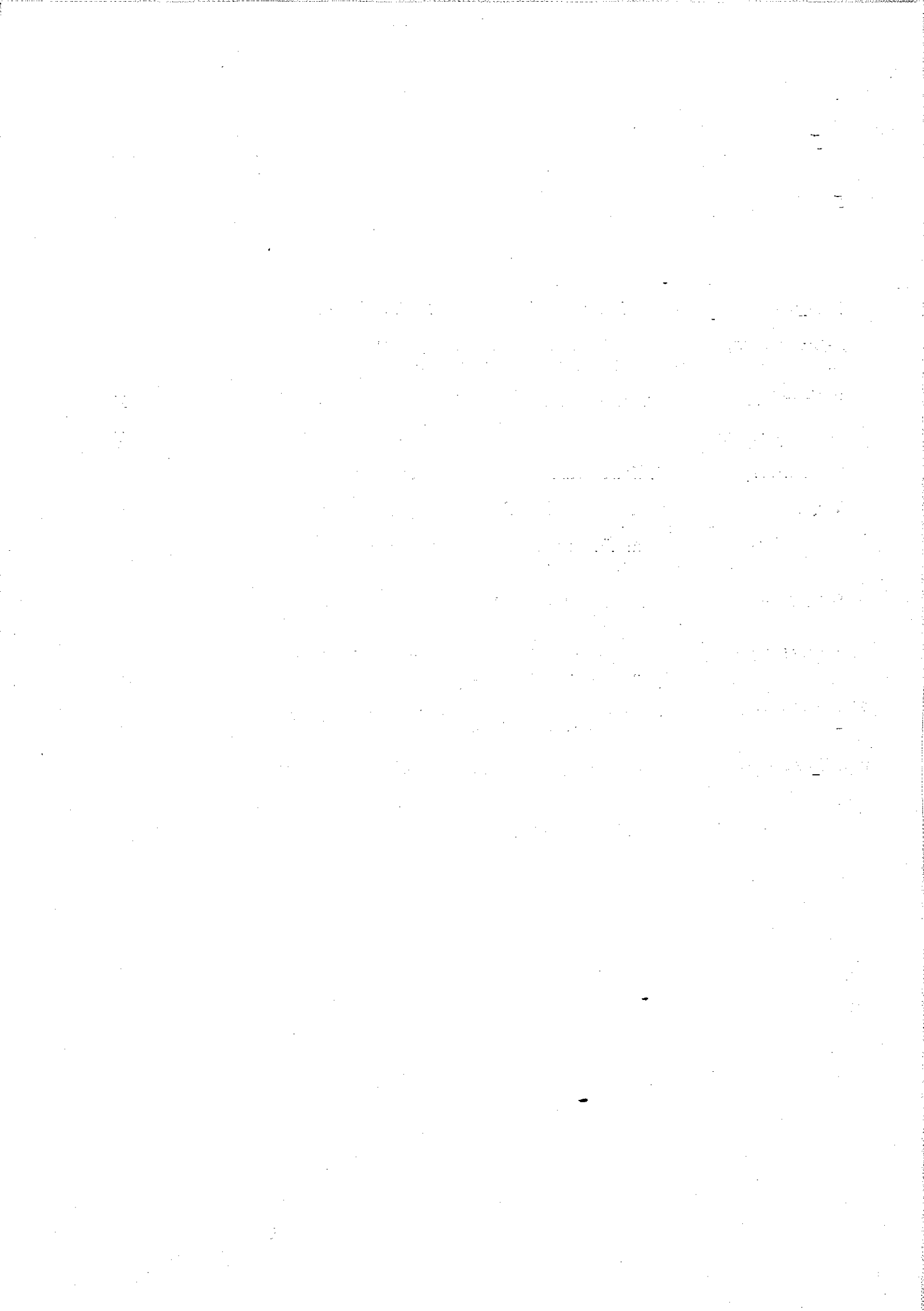
© The Norwegian Research Council for Science and the Humanities 1972

We are deeply indebted to all contributors
for their positive and rapid response, and to the Editor and Publisher of Geofysiske Publikasjoner
for their assistance and support.

A. Egeland and J. Holtet

CONTENTS

F. LIED	Professor, Dr. Philos. Leiv Marius Harang	9
J. A. RATCLIFFE	The formation of the ionosphere; Ideas of the early years (1925-1955)	13
B. HULTQVIST	Auroral particles	27
YA. I. FELDSTEIN	Aurora	57
S.-I. AKASOFU	Midday auroras and polar cap auroras	73
K. LASSEN	On the classification of high-latitude auroras	87
J. P. HEPPNER	The Harang discontinuity in auroral belt ionospheric currents	105
B. H. BRIGGS	Recent work on ionospheric irregularities and drifts	121
A. G. McNAMARA	The occurrence of radio aurora at high latitudes; The IGY period 1957-1959	135
W. STOFFREGEN	Electron-density increase in the E layer below an artificial barium cloud	151
R. E. BARRINGTON	The generation and propagation of VLF emissions	157
	Publications by Leiv Harang	169



FOREWORD

Professor Dr. Leiv Marius Harang died on 21 September 1970.

Leiv Harang combined great personal achievement with the ability to stimulate interest and create a research atmosphere in the best scientific tradition. He held a central position among Norwegian geophysicists, not only because he achieved international renown, but because his work and inspiration was decisive in establishing research activities both in Tromsø at the Auroral Observatory, as well as in Oslo, at the Defence Research Establishment and at the University.

Leiv Harang was a pleasant and modest personality, but he devoted his life to imaginative work at the frontiers of knowledge. The undersigned group of scientists, who started their work under his guidance, and had the pleasure of his continued collaboration and interest for many years, proposed about one year ago that he should be honoured with the dedication of a collection of papers in his areas of interest. We have invited some active scientists to dedicate short contributions to this collection, as a sign of gratitude and respect.

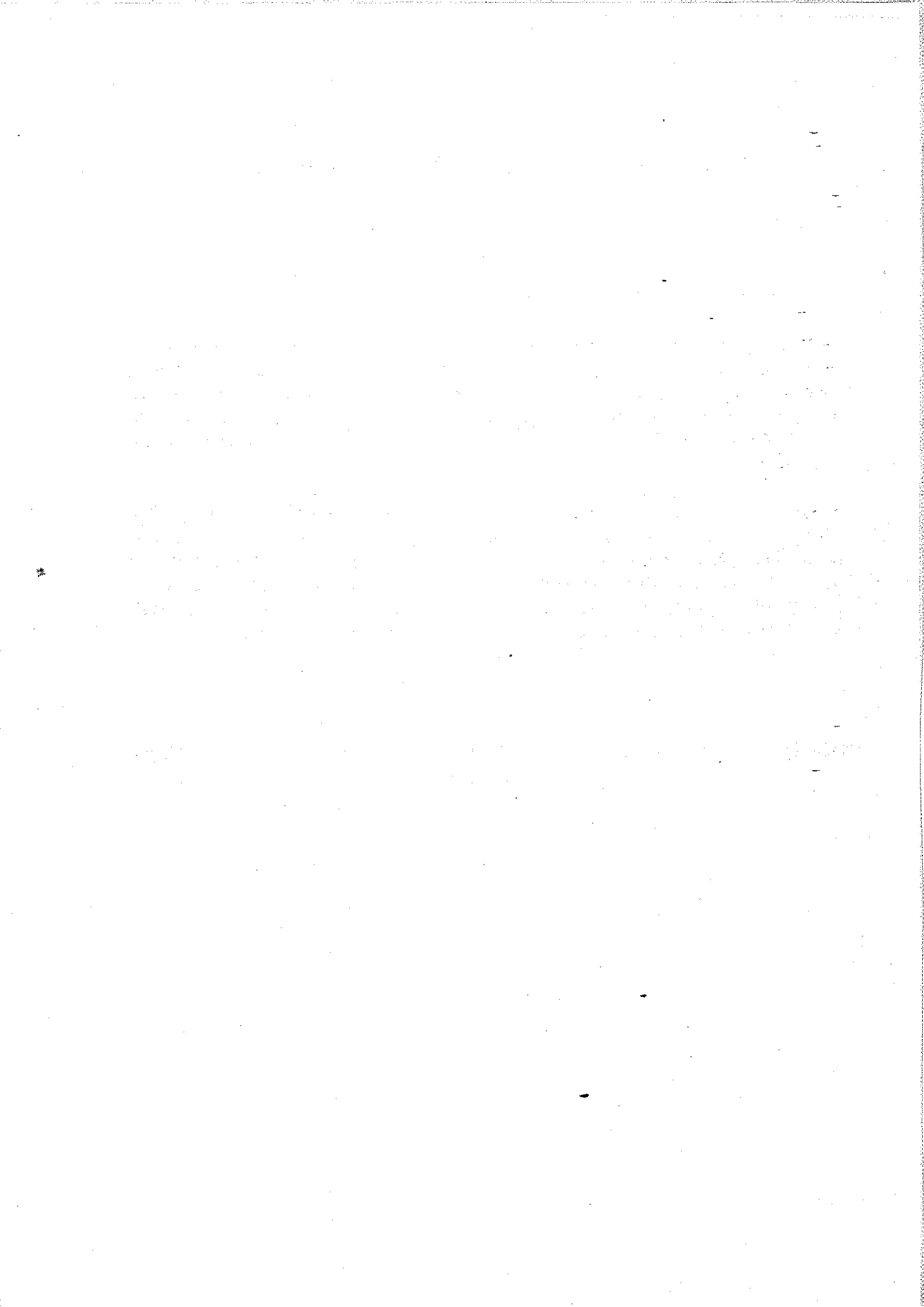
Alv Egeland

Bjørn Landmark

Finn Lied

Anders Omholt

Willy Stoffregen



PROFESSOR, DR. PHILOS. LEIV MARIUS HARANG

BY F. LIED

MINISTER OF INDUSTRY,

ON LEAVE FROM THE NORWEGIAN DEFENCE RESEARCH ESTABLISHMENT,
KJELLER, NORWAY

The man and his work

Professor Leiv Harang passed away on 21 September 1970 at an age of 68. With him the last of the Norwegian pioneers in cosmic and ionospheric physics has left the scene. Kristian Birkeland and Carl Størmer were pioneers of theories on the origin of the aurora and emission of particles from the sun. Lars Vegard pioneered studies of the auroral spectra. Leiv Harang made the study of the ionosphere in polar regions his particular field of interest, attacking problems with a variety of techniques and an unusual and stimulating interest.

Leiv Marius Harang was born on 19 April 1902 in Trondheim. After matriculation at Trondheim Katedralskole, he studied science at the University of Oslo and graduated in 1926. His main subject was physics with X-ray crystallography – a popular subject at that time – constituting his Master of Science thesis. Immediately after graduation he made a brief visit to Göttingen.

The Haldde Observatory, established by Kristian Birkeland on a mountain top near Bosekop, was in operation from 1912 until 1927. It had an extremely inconvenient location from a climatic point of view, and as early as 1925 Professor Vegard approached the Rockefeller Foundation in order to obtain funds for a new auroral observatory in Northern Norway. He was successful and the Foundation made 75,000 dollars available. The Auroral Observatory was built outside the town boundary of Tromsø at that time, not far from the Geophysical Institute in Tromsø which was established in 1918. The Auroral Observatory, together with the Bureau of Geomagnetism in Bergen, became the Norwegian Institute of Cosmic Physics with Professor Vegard as the first chairman of the board.

In 1928 Leiv Harang was offered the position as director of the Auroral Observatory and he held that position for 18 years. The original staff was, in addition to the director, Mr. Einar Tønsberg (physicist) and Mr. Magnus Jacobsen (instrument maker). In addition to auroral observations, the original programme of work included earth-magnetic observations, and studies of earth-electricity and ozone.

Dr. Harang's own work was originally the study of earth-magnetic storms and propagation of radio waves, parts of what later became the field of ionospheric physics. But he was also interested in auroral studies and his first publication in 1931 dealt chiefly with auroral forms and spectra.

During the Polar Year 1932-33, Sir Edward Appleton and Mr. R. Naismith visited the Observatory; their visit coincided with that of a German group including Mr. W. Stoffregen, then a young radio engineer, who remained in Tromsø. Both groups were interested in studies of the ionosphere by means of radio waves, and Dr. Harang immediately realised the potential of the new echo-technique. He was able to obtain an ionosonde built at the Radio Research Station in Slough, England, and devoted himself to a long series of experiments, particularly investigating the disturbed ionosphere. In 1937 he submitted his thesis for the degree of Doctor of Philosophy: 'Änderungen der Ionisation der höchsten Atmosphärenschichten während der Nordlichten unter Erdmagnetischer Störungen'. In this very typical study, he tried to illustrate the difference between the normal and disturbed ionospheric conditions in polar regions and even tried to determine recombination coefficients for electrons in the ionosphere. This latter attempt was somewhat premature. But the thesis represented a long leap forward in general knowledge of ionospheric conditions and was the starting point for a great number of investigations by himself and his colleagues.

Between 1931 and the outbreak of the last war, Leiv Harang published nearly thirty papers. During the war, due to other activities, he published only five papers, but in 1945 and 1946 he published nine, some of which had been written during the war. Most of these papers dealt with the aurora and earth-magnetic problems, but his main interest was ionospheric physics and here he was a clear leader in this field in Norway. His efforts were also appreciated by his colleagues abroad, and the Auroral Observatory in Tromsø became internationally known and respected.

Reference also ought to be made to the well-illustrated monograph finished before the war entitled: 'Das Polarlicht, und die Probleme der höchsten Atmosphärenschichten'. This served for many years as an introduction to the physics of the ionosphere.

The war interrupted Dr. Harang's work and he became engaged in clandestine work against the German occupation forces. Together with his colleague, Mr. Reidulf Larsen, he did his country great service. In 1944 he was arrested, and was eventually interned near Berlin. Leiv Harang was a man of considerable courage and his fortitude was an example to all.

In 1946, well after the war and the return of the free Norwegian forces to Norway, the Government decided to establish the Norwegian Defence Research Establishment, an interservice applied research establishment.

Dr. Harang was offered the position as superintendent of one of its five divisions, the Division for Telecommunication. With a handful of young, enthusiastic colleagues and under the difficult conditions then prevalent he established himself as a clear-sighted leader and organized his division in two sections, one for radio wave propagation and one for hardware development. One might think that applied research was far from his interests and field of ability, but the difficult pioneering conditions in Tromsø had turned him into a very practical man and the war had taught him a lesson: Never again!

The work at the Division for Telecommunication resulted in many interesting studies of polarization of reflected waves, absorption in the ionosphere, scattering from irregularities, absorption of extraterrestrial radiation, and photoelectric measurement of auroral phenomena. The combination of Dr. Harang's fundamental interest and the techniques developed during

the war became a happy and fruitful whole and resulted in a great number of publications, chiefly by his colleagues, since Dr. Harang was naturally engaged with establishing and consolidating his division and in 1949 with moving it from Bergen to Kjeller, 25 km outside Oslo.

In 1952 Dr. Harang's work entered a new phase. He was appointed Professor at the University of Oslo on the initiative of Professor Svein Rosseland, but continued his association with the Norwegian Defence Research Establishment. At an age of 50 years, beyond the productive period of most men, Dr. Harang entered perhaps the most creative phase of his life. From 1954 to 1969, when illness brought his activities to a halt, he produced close on 40 scientific publications over a wide field. In addition, he held both undergraduate and graduate courses and guided a number of graduate students, many of them now active in geophysics in Norway.

Dr. Harang's publications from this period dealt not only with his old subjects of radio scattering from auroral forms and drifts of ionospheric irregularities, but also with the then novel application of photoelectronic techniques which he had developed partly in cooperation with his old friend Dr. Willy Stoffregen, then working in Uppsala. His interest was also caught by the VLF emissions from the disturbed ionosphere and correlation between these emissions and the aurora and earth-magnetic disturbances. Studies of the VLF emissions occupied him to the very last and brought him into close contact with workers in many countries. He published from 1964 until 1969 no less than 15 papers on this subject, developing a most ingenious technique to extract the emissions wanted from man-made disturbances.

As a scientist, Dr. Harang bridged the gap between the classical geophysicists and the modern, electronically influenced geophysicists. Perhaps his theoretical understanding was not very deep, but his imagination, his feeling for physics and his ability to interpret confused material were second to none. He was also exceptionally gifted in the art of presentation, and impatient as he could be to get work done, he had exceptional patience in studying and working with the results.

As early as 1940 Dr. Harang became a member of Det Norske Videnskapsakademi and in 1969 he received Framkomiteen's Nansen-award.

Of course a man of Dr. Harang's calibre was called upon to serve on many occasions. He was from 1958 until 1969 a prominent member of the Science Research Council and here he influenced the general developments in physics to a great extent.

But perhaps his greatest influence was through the fact that he placed Tromsø on the scientific map and made the creation of a university in that town a practical possibility. The Auroral Observatory is now an element of the university. Many will remember his efforts in the long and difficult period before the war and will regard the Observatory as one of the pillars on which further happy developments rest.

In addition to being an unusual scientist and public-spirited citizen, Dr. Harang had personal and human qualities that made him a highly valued member of all associations; he had friends all over the world. He drew strength and humour from a happy family life, his wife, Signe Harang, being his devoted and loyal supporter in all circumstances.

Despite the prestige he commanded and the positions of influence he held, Leiv Harang was indeed a humble man. He could argue strongly and effectively, but he was always ready

to listen and willing to revise his opinion. He was a critical man, because for him science was a serious occupation with rules that had to be respected. He guided the young with a firm but always friendly hand, and he delegated responsibility and deliberately displayed trust in his colleagues irrespective of age and position.

But perhaps his friends will above all remember his helping hand when they needed it.

THE FORMATION OF THE IONOSPHERE IDEAS OF THE EARLY YEARS (1925-1955)

BY J. A. RATCLIFFE
CAMBRIDGE, ENGLAND

1. Introduction

Harang was active as a research worker during most of the period from 1925 to 1955 when the ionosphere was being explored by radio waves emitted from the ground, and when ideas about its formation were being developed slowly and with much hesitation. It is the purpose of this article to outline these developments. In the early days attention was concentrated on the simpler aspects of ionospheric behaviour and it was only later that more complicated phenomena could be understood. In particular the behaviour of the polar ionosphere, which Harang did so much to investigate, could not at first be used to establish the fundamental ideas about the formation of the ionosphere. Thus, it is not surprising that the work of Harang himself is not mentioned in this review.

As soon as the experiments (1925a, b) with reflection of radio waves at vertical incidence had demonstrated the existence of the ionosphere and had given a measurement of its height, two separate, but closely related, lines of investigation were started; one to find out how the ionosphere was formed, and the other to discover the nature of the upper atmosphere and of the sun's ionizing radiation. Between 1925 and 1955 ground-based radio measurements provided almost the only evidence, and by 1955 they had led to a fairly complete understanding of the ionosphere and of the atmosphere between the heights of about 90 km (the top of the *D* region) and about 300 km (the peak of the *F* layer).

Experiments with rockets and satellites, started soon after 1950, at first mainly confirmed the previous ideas about the ionospheric *E* and *F* layers, and about the neutral atmosphere up to about 300 km, but by about 1960 they were providing new information above 300 km (the topside ionosphere) and below 90 km (the *D* region). The new ground-based techniques of partial reflection, cross-modulation, and incoherent scatter were also providing information about these previously unexplored regions.

In 1925 little was known about several phenomena that took part in the formation of the ionosphere. They included: the nature and distribution of the atmospheric gases, the nature of the ionising radiations, the nature (ionic or electronic) of the ionospheric charged particles, and the mechanism (recombination or attachment) by which electrons were lost. Knowledge of these separate phenomena advanced simultaneously and because they are interrelated the

story of their investigation is complicated. For simplicity the different topics are here dealt with independently, with cross-reference where convenient. Section 7 shows how the ideas were combined in the early 1950s to suggest in different ways how the ionosphere might be formed. The bibliography (Sect. 8) is arranged chronologically and is accompanied by notes to indicate the development of the subject.

2. The radio evidence

Green (1946b) has given a fascinating account of the early radio investigations. Here only those that provided evidence about the formation of the ionosphere are mentioned. The early experiments (1925a, b) with radio waves reflected nearly vertically from the ionosphere were made on fixed frequencies. From these results it was possible to use the expressions*

$$[i] = \epsilon_0 m_i \omega^2 / e^2 \text{ or } [e] = \epsilon_0 m_e \omega^2 / e^2.$$

to deduce the ion, or the electron, concentrations ($[i]$ or $[e]$) at the height of reflection, but since it was not known whether ions or electrons were operative the concentrations could not be determined without ambiguity.

There was no knowledge of the height of the layer peak or of the concentration at the peak. These quantities were at first deduced from knowledge of the skip distance observed on communication circuits employing different frequencies (1926a, 1927a, 1928a). Later experiments made with vertically reflected waves on a single frequency indicated (1927b) that there were two major ionospheric layers (E and F), and in (1931c) experiments with several frequencies showed how the penetration frequencies could be determined. The F layer was sometimes found to be split, by day, to form the $F1$ and $F2$ layers (1933a, b).

The reflected wave was found (1928b) to be elliptically polarized with left-handed sense in the northern hemisphere and the deduction was made that there were sufficient free electrons to make $[e]/m_e \gg [i]/m_i$ at the height where the waves were absorbed. At first there was no clear evidence that electrons played the major part in *reflecting* the waves, but later (1933d) measurements of the penetration frequencies of magneto-ionically split echoes showed that electrons were responsible for reflection from the F layer. It was, however, not certain that multiple echoes from the E layer were the result of magneto-ionic splitting, and many theorists still supposed that ions were responsible for their return (1932a, 1934a, 1935a, 1938c). Finally, after reception with a polarized receiver had shown (1933e) that, as sunset approached, characteristic waves with opposite senses of polarization penetrate the E layer at well separated times, it was realised that electrons also play the major part in reflecting the waves from the E region. Although it was then known that the ratio $[e]/[i]$ of the concentrations in the E region was much greater than the ratio m_e/m_i of the masses, there was still the possibility, important for the theory of geomagnetism (see Sect. 4), that it might be much less than unity so that ions might be more numerous than electrons.

* Symbols in square brackets $[X]$ denote the concentration (number per unit volume) of the appropriate particle X . Other standard symbols used in the literature of the ionosphere are not defined here.

Ground-based observations made before the war led to a knowledge of two other important quantities – the thicknesses of the layers and the collision frequencies of electrons at different heights. Appleton (1937a) showed that his ionograms were consistent with a model ionosphere in which the *E* and *F* layers had thicknesses in the ratio 1 to 4. Studies of the way in which the amplitude of a reflected pulse depended on its group delay-time led to measurements of collision frequency in the *E*, *F1*, and *F2* layers (1935b, c, 1936b).

Investigation of the *D* region by totally reflected waves has always been difficult because of the long wavelengths that must be used. Some results were, however, obtained during the period under review by observing the echo-times of pulses (1951b, 1959b), and the phases of continuous waves (1951a), reflected from the ionosphere. More recently (1966) the phase-measurements have been interpreted to provide valuable information about electron concentrations at heights down to 60 km.

3. The ionising radiation. Photons or corpuscles?

Appleton's early measurements of vertically reflected radio waves were made on frequencies used in medium-wave broadcasting; they were therefore made between midnight (when broadcasting ceased) and sunrise (when the waves were removed by absorption). They showed that a considerable concentration of electrons (or ions) persisted throughout the night. Most workers thought that free electrons are removed by attachment to neutral molecules so rapidly that they cannot persist through the night at heights near 100 km*. They therefore suggested that the major part of the ionization must be produced by corpuscular radiation that falls on the atmosphere equally by day and by night (1926a (Lindemann), 1926b). Later, when experiments on frequencies received throughout the day showed that the ionization followed the solar zenith angle closely (1932a) it was supposed that the daytime ionizing radiation comes from the sun.

To decide whether the daytime ionizing radiation was produced by photons or by corpuscles, measurements were made on a single frequency during a solar eclipse in 1927. The height of reflection of the waves was found to increase to a maximum at the totality of the optical eclipse, and not later, as might have been expected if a stream of solar corpuscles had been intercepted by the moon (1930a). But later Chapman (1932b) showed that a corpuscular eclipse would be noticed on the earth one or two hours *before* the eclipse of the photon radiation, and because the 1927 eclipse had occurred just after sunrise a possible corpuscular eclipse would not have been observed. Those, like Chapman, who thought that the *E* layer was produced by corpuscles did not, therefore, consider the results conclusive. Finally, however, measurements made during an eclipse during the middle of the day (1933g) provided clear evidence that both the *E* and *F1* layers were ionized by photons (1933f).

* This difficulty was again emphasised by Martyn and Pulley (1936a).

4. Ionospheric conductivity and geomagnetism (1956a)

To explain the regular variations of the magnetic field observed at the surface of the earth, Balfour Stewart (1882) suggested that the conducting upper atmosphere acted as the armature of an 'atmospheric dynamo' when it is moved by the tidal action of the sun and moon, and Chapman (1919) showed that, if the speed of the air movement in the lunar tide is the same at all heights as it is at the ground, the conductivity of the atmosphere must be about 2.5×10^{-5} emu. As different theoretical models of the ionosphere were developed after 1925, calculations were made to see whether they had the necessary conductivity.

Pedersen (1927a) showed that the effect of the permanent geomagnetic field is to reduce some components of the ionospheric conductivity. Because of this reduction it was concluded (1937a) that, whether the charged particles producing radio reflections were ions or electrons, the conductivity of the ionosphere was much too small to provide the currents required by the dynamo theory. About the same time Perkeris (1937b) revised the theory of tides in the atmosphere and suggested that the speeds of the air movements might be 10 or 100 times greater in the ionosphere than at the ground. But even so, the calculated conductivity of the ionosphere was about 10 times too small.

To overcome this difficulty it was supposed that the concentration $[i^-]$ of the negative ions was at least ten times greater than the concentration $[e]$ of the electrons. Because the mass of an ion is so much greater than that of an electron these ions would not play any measurable part in radio wave reflection, but (in the presence of the geomagnetic field) their contributions to the conductivity would be important. For some time, therefore, theoretical work was directed towards deciding whether $[i^-] / [e]$, usually written λ , could be as great as 10 or 100 in the *E* layer (see Sect. 6.3).

But later Martyn (1948a) re-examined the dynamo theory and showed that an ionosphere limited in vertical extent becomes electrically polarized and the resulting electrostatic field increases the conductivity by a factor of about ten. Thereafter there was no need to suppose that negative ions preponderated in the *E* region.

5. Atmospheric constitution and the heights of the layers

In his early masterly review of the ionosphere Pedersen (1927a) drew attention to two important points that had been emphasised by previous workers:

- (a) When radiation ionizes a gas that has an ionization cross-section σ the rate of production of electrons reaches a peak at a height where $N\sigma = 1$, N being the total number of ionizable atoms (or molecules) in a superincumbent column of unit area pointing towards the sun, provided the radiation is not absorbed by any other gas. An alternative way of stating this same result is $H[n]\sigma = 1$ where H is the scale height and $[n]$ is the concentration of the ionizable gas at the height of the peak.
- (b) The rate q_m at which electrons are produced in unit volume at the peak is given by $q_m = CI \cos \chi / H \exp(1)$ where I is the intensity of the incident ionizing radiation, C is the number of electrons produced by unit energy in the radiation, and χ is the solar zenith angle.

These two important relations were re-emphasised by Chapman (1931a) who also calculated the distribution of electrons that would be produced in an equilibrium situation where the rate of production was equal to the rate of loss, assumed to be by recombination with a coefficient that was independent of height. In this calculation he showed how the thickness of the electron layer depends on the scale height of the gas that is ionized.

5.1 *The neutral atmospheric gases*

When attempts were made to use the expression $N\sigma = 1$ (or $H[n]\sigma = 1$) to explain the production of a layer at any particular height, it was necessary to know the distribution of the different atmospheric gases (to give N) and their ionization cross-sections (σ). In the period between 1925 and 1955 a main aim of ionospheric research was to provide this information.

Pedersen (1927a) collated the ideas about the atmospheric gases that had been derived from balloon soundings (up to 30 km), the disappearance of meteors (near 100 km), and the reflection of sound waves (near 100 km). He supposed that the temperature was about 210 K at all heights greater than 12 km and that mixing ceased at a height of 20 km. Above that height the atmosphere consisted of molecular oxygen, molecular nitrogen, and helium in diffusive equilibrium; helium thus predominated above 120 km. Because there was no evidence for hydrogen in the spectra of the airglow or the aurora he supposed it was absent from the upper atmosphere.

For a long time there was doubt about the upper limit of mixing (later called the turbopause) and competent workers suggested a variety of values, for example

1926(a)	Chapmann	20 km
1927(a)	Pedersen	20 km
1938(c)	Hulburt	110 km
1938(a)	Wulf and Deming	20 km
1946(a)	Bates and Massey	250 km
1949(a)	Bates	200-300 km

Finally rocket borne mass-spectrometers showed (1958) that diffusive separation of gases starts near a height of 100 km.

In the early days it was supposed that oxygen remained in the molecular form up to great heights, but later Chapman (1931b) discussed its dissociation by ultra-violet radiation and concluded that at a height near 100 km there must be a sudden transition from molecules below to atoms above. Nicolet and Mange later (1954b) showed that the relative number of atoms and molecules are determined by diffusion rather than by photo-chemical equilibrium, so that, instead of a sharp transition near 100 km, molecules and atoms are both present above this height distributed each with its own scale height.

Although Pedersen (1927a) had supposed that the temperature of the atmosphere above 100 km was about 210 K, most early investigators accepted the view of Lindemann and Dobson (1922), based on the observed disappearance of meteor trails, that it was about 300 or 350 K. Maris and Hulburt (1928c) were the first to suggest that the ionizing radiation must itself

heat the upper atmosphere to about 1000 K; the deductions drawn from radio measurements, by Martyn and Pulley (1936a) on the one hand, and by Appleton (1937a) on the other, were, however, the first to emphasise the need for high temperatures. Martyn and Pulley pointed out that the electron collision frequency measured by radio methods changed by a factor of only 1.5×10^{-3} as the height increased from 100 to 250 km. Appleton used the measured thicknesses of the *E* and the *F* layers in conjunction with the theory of Chapman to deduce that the scale height changed from about 10 km at a height of 100 km to about 50 km at a height of about 300 km. Each of these facts suggested that the density at 300 km was greater than had previously been supposed. It was therefore suggested that the atmosphere at this height must consist either of helium at a temperature of about 300 K or of atomic oxygen at a temperature of about 1000 K. Helium was ruled out on spectroscopic grounds, and it was concluded that the upper atmosphere was hot.

It is interesting to note that, although this conclusion is still thought to be correct, the arguments that led to it now seem to be invalid (1954a): that based on collision frequencies because electrons in the *F* layer collide more frequently with positive ions than with neutral particles, and that based on the shapes of the layers because the shapes are not those of a 'Chapman equilibrium' layer.

In 1950, just before rocket exploration started, the ground-based observations were thus usually thought to indicate that the important gases below 100 km are molecular oxygen and molecular nitrogen, and that above 100 km oxygen exists in both atomic and molecular forms (and after 1954 they were thought to be distributed each with its own scale height). The turbopause was usually supposed to be near 100 km (although Bates (1949a) still thought that there was mixing right up to 300 km). The temperature was about 300 K at 100 km, increasing to about 1000 K at 300 km; and at these heights it perhaps changed by a factor of 2 or 3 between winter and summer, or night and day.

In the early rocket experiments pressure and density of the atmosphere were measured up to heights of about 220 km, but it was not possible to determine the molecular weight of the air. A 'rocket panel' (1952a) prepared a 'model atmosphere' based on all the known results. They assumed that dissociation of oxygen starts at 80 km and is complete above 120 km. They also assumed (contrary to the ideas of most other workers) that molecular nitrogen began to be dissociated at 120 km and that the dissociation was complete at 220 km. They suggested that helium might predominate at heights a little above 220 km. These results did not seriously modify the ideas about the atmosphere below 240 km that had been deduced from the ground-based radio experiments.

5.2 Ionization cross sections

Once the height-distribution of any particular gas was known it was necessary to know the ionization cross-section (σ) of its molecules or atoms before the peak height of electron production could be calculated. Because the appropriate ultra-violet radiations are strongly absorbed by air, laboratory measurements are difficult, and had not been made in the early days of ionospheric investigations. But theorists concerned with the ionization in stellar

atmospheres had calculated the ionization cross-sections of atoms and an expression derived by Kramers (1923) was generally used, which led to values of about 10^{-16} cm² for O and N. More recent quantal calculations (1939a) led to values of about 10^{-17} cm² for O.

Since about 1953 it has also been possible to measure the cross-section for molecules. That for oxygen has a broad maximum of about 10^{-17} cm² at a wavelength near 1500 Å (1953). In some parts of the spectrum the cross-section varies rapidly with wavelength; at the Lyman alpha line of hydrogen it was found (1955b) to be very small, only about 8.5×10^{-21} cm².

5.3 *The heights of the layers*

At all times investigators have been able to suggest combinations of gases and ultra-violet wavelengths that made $N\sigma = 1$ at heights of 200 or 300 km, so that the formation of the *F* layer at that height by solar radiation seemed reasonable. In the earliest days, when the temperature of the high atmosphere was supposed to be about 300 K the value assumed for *N* was small, but the value of σ was great (10^{-16} cm²). Later, when the temperature and *N* were thought to be greater, σ was supposed smaller (10^{-17}); the two changes cancelled to a large extent and the calculated height of the peak remained near 200–300 km.

As the theory became more accurate, calculated values of σ became smaller and finally (1954a) $N\sigma$ was found to equal unity at a height near 170 km. At that time the theory of the electron loss process (Sect. 6.3) suggested that this was the height of the *F1* peak, although from radio evidence the peak was usually supposed to be at a height near 200 km. A little later, however, improved methods were developed (1959a) for calculating the real height of reflection from the apparent height recorded on ionograms, and when they were applied the radio measurements appeared to agree reasonably well with the theory in showing that the *F1* peak was at about 170 km.

For the *E* layer, however, $N\sigma$ was always found to be much greater than unity, implying that no wavelength could penetrate to 100 km in any of the known gases. Many and varied suggestions were therefore made to account for the occurrence of a peak of electron production at that height. Chapman (1931b) suggested that it was produced by a neutral stream of particles ejected from the sun, but he abandoned this idea after the eclipse experiments of 1932. Another suggestion supposed that the *E* layer was formed when suitable radiation, unabsorbed in the atomic oxygen at greater heights, first encountered the molecular oxygen near 100 km (1938a). This suggestion had to be abandoned when it was realised that the transition from molecular to atomic oxygen was determined by diffusion and was not sharp. Nicolet (1945) suggested that molecular oxygen was pre-ionized by radiation in a part of the spectrum that could penetrate to the *E* region.

Underlying all these early suggestions was the idea that the intensity in the appropriate part of the sun's spectrum (approximating to that of a black body at 6000 K) decreased so rapidly with decreasing wavelength that only the part near the ionizing thresholds need be considered. But Müller (1935d) drew attention to previous suggestions by Eckeršley and Elias that X-rays might be responsible for the *E* layer, and he showed that the appropriate ionization cross-sections were of the right order. Hoyle and Bates (1948b) suggested that heavy ions in the

hot solar corona might emit X-rays in sufficient intensity to produce the *E* layer. Confirmation of these ideas had to await rocket-borne measurements of solar X-rays.

Exploration with long waves emitted from ground stations (1951a, b, 1959b) showed that there was a substantial concentration of electrons in the *D* region, down to heights less than 70 km, and that it changed with the solar angle as though produced by radiation from the sun. It was, however, difficult to suggest a radiation that could penetrate to these depths. It was often supposed that some minor constituent, such as a metal (1942) or nitric oxide might be ionized by a part of the ultra-violet spectrum that could reach these heights: when laboratory measurement showed that the atmospheric absorption of Lyman- α was small it was shown in detail (1945) how that radiation could ionize NO to produce the *D*-region electrons.

6. The rate of loss of electrons

After a combination of gas and ionizing radiation had been found capable of producing a peak of electrons at the observed height, it had to be shown that the observed concentration agreed with that calculated.

The expression $q_m = CI \cos \chi / H \exp (1)$ was used to calculate the peak rate (q_m) at which electrons were produced in unit volume. The spectrum of the ionizing radiation was supposed to be that of a black body at 6000 K and the energy greater than the threshold ionization energy was computed to give I : C was estimated on the supposition that each photon produced one electron. The scale height H at the peak was taken from the assumed distribution of the gas to be ionized, or was calculated by comparing the observed thickness of the electron layer with that deduced from Chapman's theory.

After q_m had been calculated it was necessary to compare it with the value deduced from experimental data. For this purpose the layer was assumed to be in quasi-equilibrium so that the rate of production of electrons balanced the rate of loss, either by recombination at a rate $\alpha[e]^2$, or by attachment to neutral particles at a rate $\beta[e]$. At first it was not known which type of loss process was predominant, or what were the magnitudes of α and β . Investigations of these matters proceeded in three stages in which

- (a) early theoretical values of loss coefficients were used,
- (b) loss coefficients were obtained from the radio observations,
- (c) more advanced theory was used to explain the unexpected results from the radio observations.

6.1 Early theoretical values

Bailey (1925c) had shown that when electrons collide with oxygen molecules they become attached to form negative ions with a probability $p = 10^{-5}$. When the collision frequency is ν the rate of loss of electrons from unit volume is thus $p\nu[e]$: if this is written $\beta[e]$, the attachment coefficient (β) is proportional to ν and hence to the concentration $[O_2]$ of the oxygen molecules.

Two kinds of recombination, radiative and non-radiative, had been considered previously. Kramers (1923), following Saha, had investigated the equilibrium represented by

$X^+ + e \rightleftharpoons X + h\nu$ between radiative recombination and photoionization, and had shown that the coefficient α would have a magnitude about $10^{-11} \text{ cm}^3 \text{ s}^{-1}$.

J. J. Thompson (1924) had investigated a process of mutual neutralization or 'recombination' ($X^+ + Y^- + M \rightarrow X + Y + M$) that occurred in the presence of a third body (M) without the emission of radiation, and had deduced a (pressure-dependent) magnitude for the 'recombination' coefficient; his expression applied equally to atomic and molecular ions.

In the early days it was not clear which of these values of loss coefficient was appropriate. Those who thought that electrons were lost by attachment realised that they would disappear very rapidly and could not last through the night. Those who adopted the theoretical recombination coefficient of Kramers realised that the loss would be slow, and there would be little change through a night. Some (like Pedersen (1927a)) thought that the night decay could be roughly explained by Thomson's recombination theory.

6.2 *The observed rate of electron loss*

Even before it had been decided whether the negative charges in the ionosphere were ions or electrons, their rate of loss was deduced from radio data by observing the changes during a solar eclipse, or the rate of decay at night, or the asymmetry about midday.

It was first necessary to decide whether the loss followed the quadratic law appropriate to recombination, or the linear law appropriate to attachment. For this purpose it was assumed that the peak of electron concentration in a layer was at the level of the peak of production and that there was quasi-equilibrium, so that either $q_m = \alpha[e]^2$ or $q_m = \beta[e]$; and that $q_m \propto \cos \chi$ from Chapman's theory. Observations made for different values of χ , resulting from changes in season or geographical position, favoured the hypothesis of recombination for the *E* and the *F1* layers, but results for the *F2* layer were never sufficiently consistent to lead to a firm conclusion (1932a, 1933f). Later an examination of the total number of electrons in unit column above the *F1* layer led, in a not very convincing way, to the conclusion that electron-loss in the *F2* layer can be described by a linear law (like attachment) (1956b). There has, however, never been a convincing proof; the current belief in the linear law is based on a self-consistent theory of the formation of the *F2* layer.

Once it had been decided which law to follow, the magnitude of the loss coefficient was derived from observations of the time-varying ionosphere during the night, the day, or an eclipse. The value $\alpha_E = 10^{-8} \text{ cm}^3 \text{ s}^{-1}$ was found for the *E* layer (1939c). Although the situation in the *F2* layer was much less certain, observations were often interpreted in terms of recombination. The value $\alpha_{F2} = 10^{-10} \text{ cm}^3 \text{ s}^{-1}$ was found from nocturnal changes in skip-distances (1930b) and from the noon asymmetry of the penetration frequency (1937a).

6.3 *Theories of recombination, attachment, and detachment and the magnitude of $\lambda (= [i^-]/[e])$*

Martyn and Pulley (1936a) pointed out that, with Bailey's value of $p = 10^{-5}$ (Sect. 6.1), the electron concentration in the *F* layer would decay by a factor e^{-1} in five minutes. They therefore suggested that there must be some process capable of detaching the electrons.

Massey, Bates and their collaborators followed up this suggestion in an important series of papers. At first (1937c) Massey thought that in the *E* layer, attachment and collisional detachment were both more rapid than recombination so that a quasi-equilibrium was set up with $\lambda \doteq 100$: this large number of ions was convenient for dynamo theory (Sect. 4). Massey also showed that the effective recombination coefficient (α_{eff}) would be given in terms of the electron recombination coefficient (α_e) and the ion recombination coefficient (α_i) by the expression

$$\alpha_{eff} = \alpha_e + \lambda\alpha_i$$

so that, with $\lambda = 100$ and with the value of α_i that they had calculated for the two-body neutralization process $O^+ + O^- \rightarrow O' + O''$, Massey and his collaborators (1939a) could explain the measured value of α_{eff} for the *E* layer.

Later (1939a), however, Massey and his collaborators pointed out that, because the *E* penetration frequency changes with the sun's zenith angle as described by Chapman's theory, the loss coefficient must be independent of height, whereas $\lambda\alpha_i$ would be height-dependent. They stressed this need for a height-independent α_{eff} .

Later still (1946a), after discussing the balance between collisional and radiative attachment and detachment processes they came to the different conclusion that λ is less than unity at all heights greater than about 90 km. This conclusion, of course, reintroduced the difficulty that the ionospheric conductivity was about 10 or 100 times too small for agreement with the dynamo theory; it also reintroduced the difficulty of explaining large numerical values of α_{eff} . Because no other process seemed to be feasible they speculated that dissociative recombination ($O_2^+ + e \rightarrow O + O$) might prove to be sufficiently rapid to explain the result in the *E* layer (1947a). This suggestion received support from the laboratory measurements of Biondi and Brown (1949b) who measured a two-body recombination coefficient as great as $2 \times 10^{-7} \text{ cm}^3 \text{ s}^{-1}$ at 300 K. The process was thought to be one of dissociative recombination and a theory was developed (1950) which was consistent with the laboratory measurements.

In their discussion of the *F* layer Massey and his collaborators first (1939a) made an improved quantal calculation of the radiative recombination coefficient to O^+ and showed that it was much smaller than had been calculated from the crude formula of Kramers. It thus became clear that radiative recombination could not be the responsible loss process, so they suggested instead a process of charge-transfer or of ion-atom interchange ($O^+ + XY \rightarrow XY^+ + O$ or $O^+ + XY \rightarrow XO^+ + Y$) followed by dissociative recombination ($XY^+ + e \rightarrow X + Y$ or $XO^+ + e \rightarrow X + O$) (1947a). This would lead to a rate of loss that decreased upwards in such a way that there would be an *F2* layer above the *F1* layer, even though there was only one peak of electron production at the height of the *F1* layer; moreover, the rate of loss in the *F2* layer would be proportional to the electron concentration, and would thus simulate an attachment process.

It had been suggested before the work of Bates and Massey that the radiation producing electrons most rapidly at the *F1* peak might also produce the *F2* layer, formed as a kind of 'bulge' on the top of the *F1*. Thus Hulburt (1934a), Appleton (1935e) and Martyn and Pulley (1936a) thought that the 'bulge' might be caused by a greater temperature at greater heights.

Bradbury (1938b) thought that it might occur because the rate of loss of electrons decreased upwards (but he did not quantify his discussion as Massey and his collaborators did).

In the theory of Bates and Massey, or of Bradbury, photochemical equilibrium produces a situation where the electron concentration increases with height: there must be a mechanism that terminates this increase at the height of the *F2*-layer peak. Bradbury supposed that the peak occurred where the upwards decreasing loss-rate became equal to the height-independent loss-rate resulting from radiative recombination. But with the newer quantities this level would be much too high and another reason had to be found for the occurrence of the peak. It was later suggested (1956b, c) that the peak occurs at a height where loss of electrons by 'chemical' processes occurs at the same rate as loss by diffusion. This is the present-day theory (1960) of the *F2*-layer peak. It implies that in the 'topside' ionosphere the distribution of electrons is controlled mainly by diffusion.

7. Deductions from ground-based experiments

In the early 1950s, just before experiments were made in space vehicles, there was wide agreement about the distribution of the atmospheric gases up to a height of about 300 km and about the processes by which ionospheric electrons were removed. Experiments in space vehicles confirmed the essential correctness of these ideas.

Because the intensity of radiation in the solar spectrum was not known, and there was doubt about the ionization cross-sections of some of the atmospheric gases, there were several suggestions about how the different layers were formed. They were collected by Mitra (1952b, p. 290) as follows:

- D region* ionization of O₂ (1938d)
 - » » metals (1942)
 - » » NO (1945)
- E layer* ionization of O₂ (1938a, d, e)
 - » » all gases by X-rays (1948b)
 - pre-ionization of O₂ (1945)
- F1 layer* ionization of N₂ (1938a, d, e)
 - » » O, accompanied by height-decrease of the loss-rate (1947a).

The early rocket measurements of solar UV- and X-radiation (1954a, 1955a), combined with improved measurements of ionization cross-sections, quickly led to the elimination of some of these possibilities and to the present ideas about the formation of the layers. But it is noteworthy that those suggestions in the above list that were most widely accepted as reasonable in 1955 were, in fact, shown to be correct when the rocket experiments were made.

In the preparation of this article I have had considerable help from Professor D. R. Bates, to whom I am most grateful.

8. Bibliography and summary of history

Ideas that were later proved wrong are marked with the sign \neq

Those references that give good reviews of the situation at different times are marked with an asterisk *

1882	Stewart, B.	'Terrestrial Magnetism' in Encyc. Brit., 9th Ed.	} Dynamo theory
1919	Chapman, S.	Phil. Trans., A 218, 1.	
1922	Lindemann, F. A. and G. M. B. Dobson	Proc. Roy. Soc., A 102, 411.	} $T = 300$ K above 100 km
1923	Kramers, H. A.	Phil. Mag., 46, 836.	
1924	Thomson, J. J.	Phil. Mag., 47, 337.	} \neq Theory of radiative recombination and ionization cross-section Theory of Ionic recombination
1925a	Appleton, E. V. and M. A. F. Barnett	Nature, 115, 333	
1925b	Breit, G. and M. A. Tuve	Nature, 116, 357.	} Vertical reflection of radio waves
1925c	Bailey, V. A.	Phil. Mag., 50, 825	
1926a	Discussion	Proc. Roy. Soc., A 111, 1.	} Attachment probability \neq Ionization by corpuscles
1926b	Elias, G. J.	Jahrb. d. drahtl. Tel. 27, 66.	
*1927a	Pedersen, P. O.	'The Propagation of Radio Waves.' Danmarks naturvidenskabelige samfund, Copenhagen.	} Summary of situation in 1927
1927b	Appleton, E. V.	Nature, 120, 331.	
1928a	Hulburt, E. O.	Phys. Rev., 31, 1018.	} Two layers E and F Deductions from skip-distance
1928b	Appleton, E. V. and J. A. Ratcliffe	Proc. Roy. Soc., A 117, 576.	
1928c	Maris, H. B. and E. O. Hulburt	Terr. Mag. Atmos. Elec., 33, 229.	} Electrons responsible for radio wave absorption Ionosphere hot
1930a	Appleton, E. V.	J. Inst. Elec. Eng., 66, 872.	
1930b	Eckersley, T. L.	Nature, 125, 669.	} Result of 1927 eclipse Loss rate in $F2$
1931a	Chapman, S.	Proc. Phys. Soc., 43, 26.	
*1931b	Chapman, S.	Proc. Roy. Soc., A 132, 353.	} Theory of simple layer Atomic oxygen above 100 km $\neq E$ produced by corpuscles Penetration frequencies
1931c	Appleton, E. V.	Nature, 127, 197.	
1932a	Appleton, E. V. and R. Naismith	Proc. Roy. Soc., 137, 36	} \neq Reflection from E by ions Recombination in E and $F1$ Electron concentration depends on solar angle
1932b	Chapman, S.	M. N. R. Astro. Soc., 92, 413.	
1933a	Appleton, E. V.	Proc. Phys. Soc., 45, 673.	} Time of corpuscular eclipse F layer double
1933b	Schafer, J. P. and W. M. Goodall	Nature, 131, 804.	
1933c	Appleton, E. V. and R. Naismith	Proc. Phys. Soc., 45, 389.	

- 1933d Appleton, E. V. and G. Builder Proc. Phys. Soc., 45, 208. Reflection from F by electrons
- 1933e Ratcliffe, J. A. and E. L. C. White Phil. Mag., 16, 125. Reflection from E by electrons
- 1933f Discussion Proc. Roy. Soc., 141, 697. F ionized by photons
- 1933g Henderson Can. J. Res., 8. Mid-day eclipse measurements
- 1934a Hulburt, E. O. Phys. Rev., 46, 823. Reflection from E by ions
- 1935a Hulburt, E. O. Terr. Mag. Atmos. Elec., 40, 193. \neq Two F layers result of temperature difference
- 1935b Appleton, E. V. Nature, 135, 618. } Measurement of collision frequency
- 1935c Eckersley, T. L. Nature, 135, 435. }
- 1935d Müller, E. W. A. Nature, 135, 187. X-rays produce E
- 1935e Appleton, E. V. Nature, 136, 52. \neq Two F layers result of temperature difference
- 1935f Appleton, E. V. and S. Chapman Proc. Inst. Radio Eng., 23, 658. Eclipse results
- 1936a Martyn, D. F. and O. O. Pulley Proc. Roy. Soc., A 154, 455. F layer hot. Importance of detachment
- 1936b Farmer, F. T. and J. A. Ratcliffe Proc. Phys. Soc., 48, 839. Measurement of collision frequency
- 1937a Appleton, E. V. Proc. Roy. Soc., A 162, 451. F layer hot. Thickness of layers. \neq Loss rate in $F2$.
- 1937b Pekeris, C. L. Proc. Roy. Soc., A 158, 650. \neq Conductivity of ionosphere
- 1937c Massey, H. S. W. Proc. Roy. Soc., A 163, 542. Revised tidal theory
- 1938a Wulf, O. R. and L. S. Deming Terr. Mag. Atmos. Elec., 43, 283. $\neq \lambda = 100$ in E
- 1938b Bradbury, N. E. Terr. Mag. Atmos. Elec., 43, 55. $\neq E$ layer formed at $O_2 - O$ transition level
- 1938c Hulburt, E. O. Phys. Rev., 53, 344. Height-dependent attachment in F
- 1938d Mitra, S. K., J. N. Bhar and S. P. Ghose Ind. J. Phys., 12, 455. \neq Reflection from E by ions
- 1938e Bhar, J. N. Ind. J. Phys., 12, 363. $\neq O_2$ ionized in D and E , N_2 in F
- 1939a Bates, D. R., R. A. Buckingham, H. S. W. Massey and J. J. Unwin Proc. Roy. Soc., A 170, 322. $\neq O_2$ ionized in E , N_2 in F
- *1939b Hulburt, E. O. In 'Terrestrial Magnetism and Electricity.' (Ed. J. A. Fleming), McGraw Hill, New York. Quantal calculations of ionization cross section and coefficient of radiative recombination
- 1939c Hulburt, E. O. Phys. Rev., 55, 639.
- *1939d Discussion Q. J. Roy. Met. Soc., 65, 324.
- 1940 Mohler, F. L. J. Res. Nat. Bur. Stand., 25, 507.
- 1942 Vassy, A. and E. Vassy Cahirs de Physique, 9, 28. $\neq D$ region by ionization of metals
- 1945 Nicolet, M. Mem. Roy. Met. Inst. Bel., 19, 1. $\neq E$ by pre-ionization of O_2
- *1946a Bates, D. R. and H. S. Massey Proc. Roy. Soc., A 187, 261. D region by ionization of NO
- *1946b Green, A. L. Assoc. Wireless Australia Tech. Rev., 7, 178. $\lambda < 1$ in E
- History of early radio exploration

- 1947a Bates, D. R. and H. S. Massey Proc. Roy. Soc., A 192, 1. Suggestion of dissociative recombination in *E* and charge exchange in *F*
- *1947b Pande, A. J. Geophys. Res., 52, 375.
- *1947c Mitra, S. K. 'The Upper Atmosphere' (1st edn). The Royal Asiatic Society of Bengal.
- 1948a Martyn, D. F. Nature, 162, 142. Conductivity of thin slab ionosphere
- 1948b Hoyle, F. and D. R. Bates Terr. Mag. Atmos. Elec., 53, 51. X-rays produce *E*
- *1949a Bates, D.R. Mon. Not. R. Astr. Soc., 109, 215. Good survey
- 1949b Biondi, M. A. and S. C. Brown Phys. Rev. 76, 1697. Measurement of dissociative recombination
- 1950 Bates, D. R. Phys. Rev. 78, 492. Theory of dissociative recombination
- 1951a Bracewell, R. S., K. G. Budden, J. A. Ratcliffe, T. W. Straker and K. Weeks Proc. Inst. Radio Eng., 98, 221. } *D*-region echoes
- 1951b Helliwell, R. A., A. J. Mallinckrodt and F. W. Kruse J. Geophys. Res., 56, 53. }
- 1952a Rocket Panel Phys. Rev. 88, 1027. Atmosphere from rockets
- *1952b Mitra, S. K. 'The Upper Atmosphere' (2nd edn). The Asiatic Society, Calcutta.
- 1953 Ditchburn R. W. and D. W. O. Heddle Proc. Roy. Soc. A 220, 61. Measurement of ionization cross-sections
- *1954a Bates, D. R. 'Rocket Exploration of the Upper Atmosphere.' (Eds. R. L. F. Boyd and M. J. Seaton), Pergamon Press, London. 347. Theory shows *F1* at 170 km
- 1954b Nicolet, M. and P. Mange J. Geophys. Res. 59, 15. O and O₂ in diffusive equilibrium
- 1955a Havens, R. J., H. Friedman and E. O. Hulbert Phys. Soc. Conference. 'The Physics of the Ionosphere.' 237. Rocket measurement of UV
- 1955b Po Lee J. Opt. Soc. Amer., 45, 702. Lyman- α little absorbed
- 1955c Bates, D. R. Proc. Phys. Soc., A 68, 344.
- *1956a Chapman, S. Nuovo Cim., 4 Supp. 4, 1385 Survey of ionosphere and geomagnetism
- 1956b Ratcliffe, J. A., E. R. Smerling, C. S. G. K. Setty and J. O. Thomas Phil. Trans., 248, 621. Attachment law in *F*
- 1956c Yonezawa, T. J. Radio Res. Labs. Japan, 3, 1. Diffusion causes *F2* peak
- 1958 Meadows, E. B. and J. W. Townsend Ann. Géophys., 14, 80. Turbopause at 100 km
- 1959a Thomas, J. O. Proc. Inst. Rad. Eng., 47, 162. True height from ionograms
- 1959b Gibbons, J. J. and A. H. Waynick Proc. Inst. Rad. Eng., 47, 160. *D*-region echoes
- 1960 Rishbeth, H. and D. W. Barron J. Atmosph. Terr. Phys., 18, 234. Present-day theory of *F2* peak
- 1966 Deeks, D. G. Proc. Roy. Soc., A 291, 413. *D*-region electron distribution

AURORAL PARTICLES

BY BENGT HULTQVIST

KIRUNA GEOPHYSICAL OBSERVATORY,
KIRUNA, SVEDEN

1. Introduction

By auroral particles we mean in this context superthermal particle populations with energies from less than a hundred eV to a few hundred keV, some of which are precipitated into the atmosphere causing atmospheric excitation and ionization. Such particles exist near the Earth mainly at geomagnetic latitudes above 55° . At greater distances from the Earth they have their largest density in the plasma sheet of the geomagnetic tail and in the so-called polar cusp region on the dayside of the magnetosphere. Also the magnetosheath plasma, between the bow chock and the magnetopause, contains particles in the lower part of the energy interval mentioned. These particles, however, do not interact directly with the atmosphere – except in the polar cusps – and are not dealt with here (with the exception mentioned). Large fluxes of trapped energetic particles in the energy range of auroral particles have also been observed recently in the upper ionosphere above the geomagnetic equator. Their origin is unknown but it is most certainly quite different from that of the auroral particles. They will not be discussed any further.

Practically all auroral particles originate in the Sun. They seem to enter the magnetosphere either along interplanetary magnetic field lines which are connected with the geomagnetic field lines in the tail (true for the majority of auroral particles) or by the polar cusp on the dayside where the geomagnetic field intensity is very low.

More than a decade of direct rocket and satellite observations of the auroral particles in most parts of the magnetosphere have completely changed our concept of the characteristics of the auroral particles and of their acceleration and precipitation mechanisms. The high-energy tail of the energy spectrum (>40 keV for electrons) was first investigated by means of Geiger-Müller tubes and solid state detectors. It is only from the later years of the sixties that the knowledge of the main part of the energy spectrum has been built up, mainly on the basis of measurements with channel multipliers and open photomultipliers.

Synoptic observations of auroral particle precipitation into the atmosphere are provided by the net of all sky cameras and, for the higher energies, by the riometer net. Although such observations only provide rough integral information about different fractions of the energy spectrum of precipitated auroral particles, they are the only means we have available for the

investigation of the dynamics of the auroral particles in the magnetosphere. The dynamical morphology of visual aurora is treated in other chapters in this volume (cf. Akasofu, Feldstein, and Lassen) and will not be dealt with here. We will instead concentrate on the direct particle measurements on board rockets and satellites, but some results on the dynamics of precipitation of the higher energy auroral particles obtained from ground based measurements will also be discussed.

2. The distribution of auroral particles in the magnetosphere, and its dynamics

2.1. General outline

The observations in the Earth's upper atmosphere of a strong asymmetry in the location of the aurora between day and night-sides of the earth, which led to the concept of the auroral oval, have also been verified by the direct satellite measurements of auroral particles below, say 15 keV energy. It has recently been found that on the dayside in this high latitude precipitation region, the characteristics of the energetic particles are the same as those of the warm plasma in the magnetosheath, i.e. the energies are mainly well below 1 keV. One has therefore concluded that these particles penetrate from the magnetosheath into the inner magnetosphere all the way down to the atmosphere along the magnetic field lines in the region of the neutral point of the magnetosphere. This region is, however, not a narrow one, but is a fairly wide and very broad 'cleft' at about 78° invariant latitude stretching from the morning to the evening side. This region is now mostly called the polar cusp.

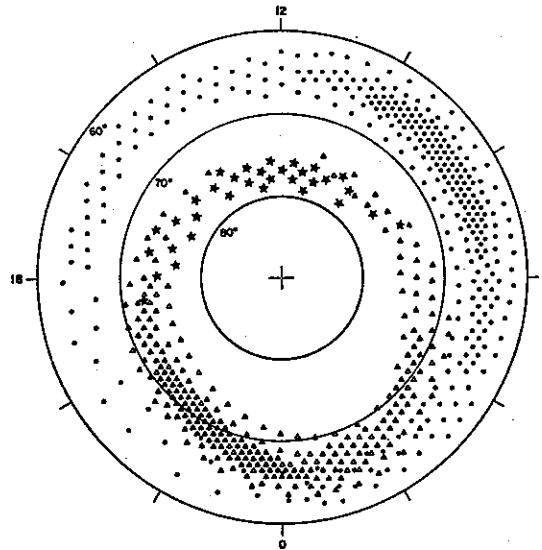


Figure 1. An idealized representation of a three-zoned auroral particle precipitation pattern. The auroral oval distributed precipitation (splash type) is represented by the triangles, the auroral-zone oriented (drizzle type) by the dots, and the polar-cusp precipitation on the dayside by the stars. The average flux is indicated approximately by the density of the symbols. The coordinates are geomagnetic latitude and geomagnetic time (Hartz, 1971).

The high degree of asymmetry between day and night for low energy auroral particles is not seen for the high energy tail of the auroral particle spectrum. Electrons of energies above 40 keV, which are those for which most measurements have been made, show a distribution over the Earth's surface coinciding fairly well with the auroral zone, i.e. they are most frequently seen and have the highest fluxes in a zone approximately centred on the 67° invariant latitude (or corrected geomagnetic latitude) curve around the Earth.

The rate of precipitation of auroral particles into the upper atmosphere may schematically be summarized in a diagram as shown in Fig. 1. The average flux is indicated by the density of the symbols. The dots mainly represent the higher energy auroral particles, the triangles the medium energy ones responsible in the first instance for the visual aurora, and the stars mark the particles entering the magnetosphere through the polar cusp. It has to be remembered that the diagram of Fig. 1 is an idealized one to represent the average situation; actually, new observations suggest rather more overlapping of the three principal zones with a more gradual transition from one to the other than is shown here. We will discuss the various characteristics of these three particle populations later.

Above, the relation of the soft auroral particles, seen at high latitudes on the dayside, to the particle populations in the outer magnetosphere has been outlined and we will now relate the auroral particles on the nightside to the outer magnetosphere particles. During the 1960s we have, in fact, seen a complete change in our picture of the spatial configuration of the hot plasma (i.e. the auroral particles) around the Earth and the morphology for various disturbance phenomena in the high-latitude ionosphere. A schematic diagram showing some of the main relations between certain boundaries and other topographic features (large gradients) observed in the midnight sector in the ionosphere and in the magnetosphere is seen in Fig. 2a.

It has been clearly shown in recent years that the inner boundary of the plasma sheet in the equatorial plane of the magnetosphere is connected with the auroral particle precipitation zone in the nightside ionosphere by means of magnetic field lines. Thus, the hot plasma from the plasma sheet has been observed by satellites practically all the way along the field lines down to the atmosphere. The inner boundary of the plasma sheet connects to the equatorward boundary of the auroral oval.

This inner boundary of the plasma sheet is a boundary for the high energy electrons but not for the energetic protons (and not even for the electron density). The intense proton fluxes characteristic of the plasma sheet reach closer to the Earth than the corresponding electron fluxes. At the inner edge of the plasma sheet, where the electron mean energy sharply decreases with decreasing distance, the proton energy instead increases. The protons may extend almost to the plasma sphere in the midnight sector during magnetic storms. They then form the main part of the asymmetric ring current responsible for the main phase of the storm.

The relations of the energetic particle occurrence in various parts of the magnetosphere according to present views, as summarized in Figs. 2a and b, may possibly be somewhat modified in the future, but not very much.

We will return to the characteristics of the auroral particles in the various regions of the magnetosphere later in this chapter.

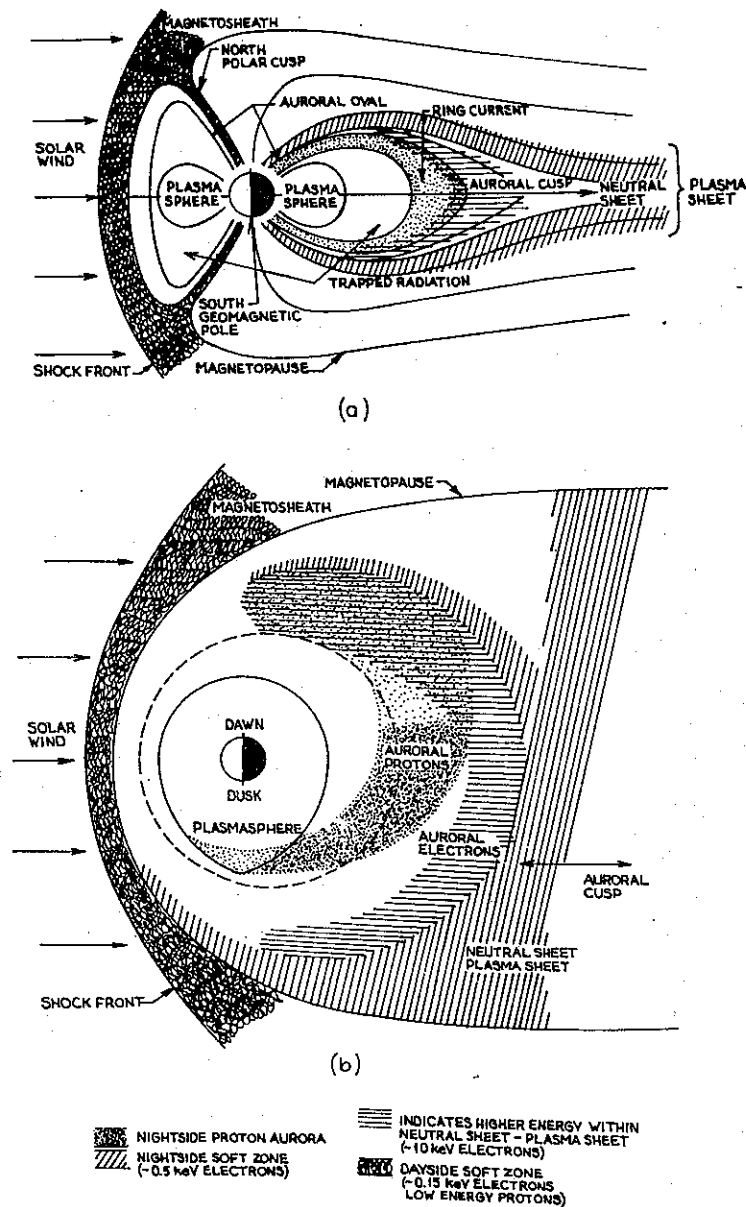


Figure 2. Noon-midnight meridian (a) and equatorial (b) views of the magnetosphere showing the distribution of particles and their precipitation regions (McCormac et al., 1971).

2.2. Distribution over the Earth's surface of particle precipitation – average characteristics

In this section we shall mainly deal with direct particle measurements. However, some groundbased indirect measurements of auroral particle precipitation will also be referred to when there are no data from direct measurements available. This is particularly true of the dynamical aspects of the precipitation which need synoptic data. But before we discuss the dynamics of the auroral particle distribution over the Earth in the next section, we shall outline some average characteristics.

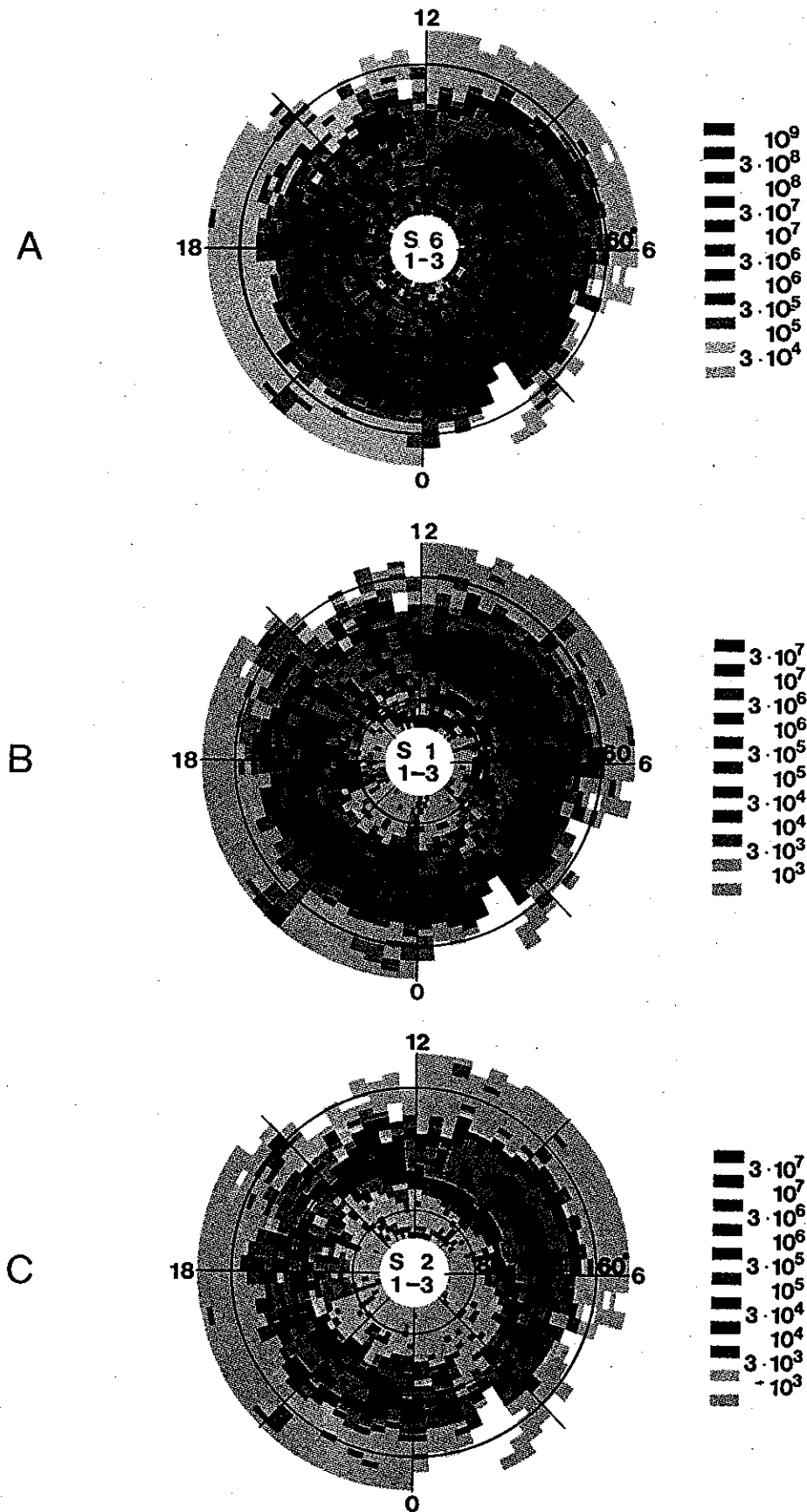
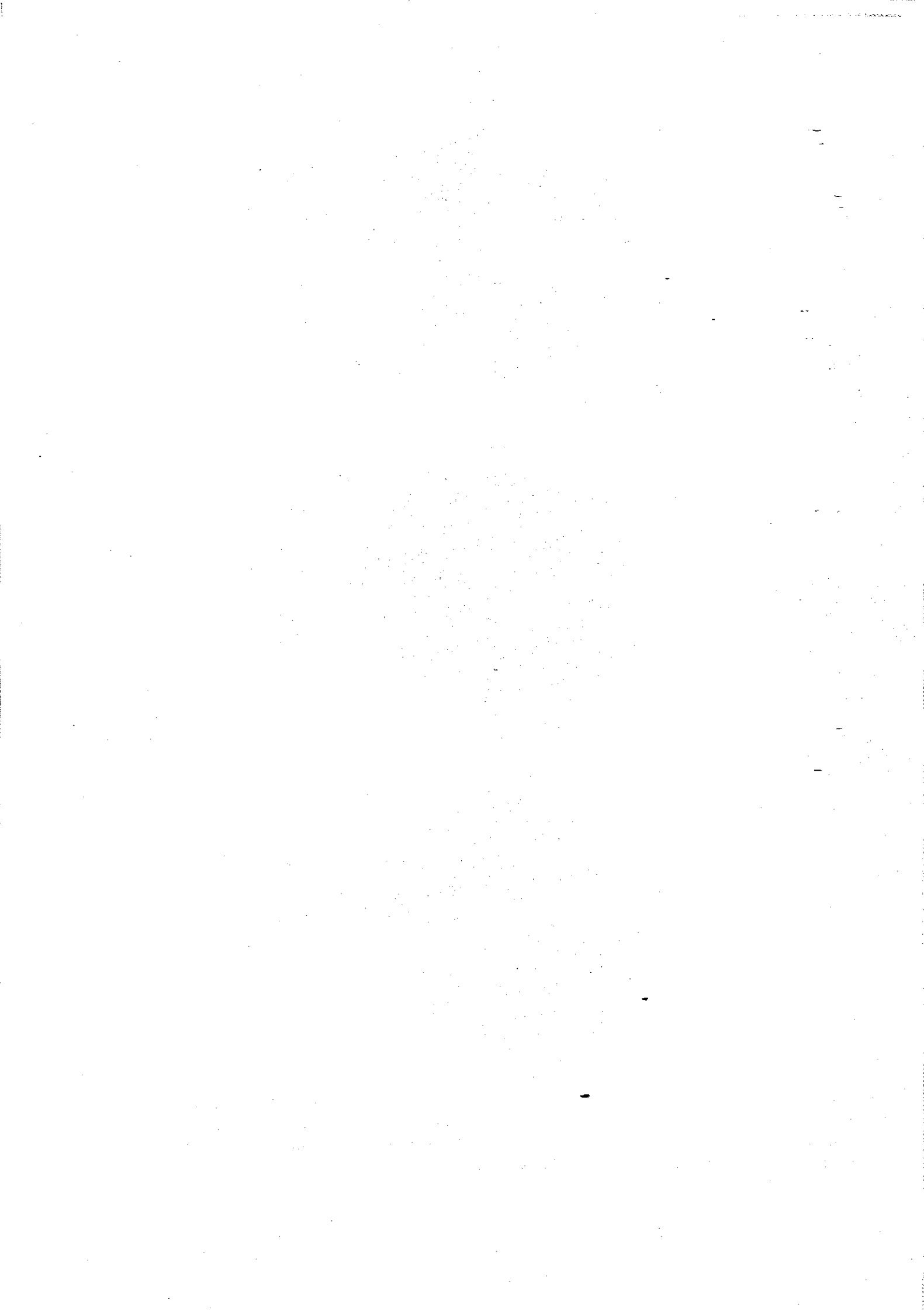


Figure 3. Average fluxes of precipitated electrons of 1, 6 and 13 keV energies for fairly low activity level ($Kp = 1-3$) as measured during a 20 months period from October 1968 to April 1970 with the ESRO 1A satellite. The colour code is given in the figure. (a) is for 1 keV, (b) 6 keV, (c) 13 keV. The coordinates are corrected geomagnetic latitude (practically identical with invariant latitude) and excentric dipole time (Riedler and Borg, 1971).



The distribution over the Earth's surface of the average intensity of precipitation of auroral electrons of ≈ 1 , ≈ 6 , and ≈ 13 keV energies at fairly low activity levels, as observed during a period of 20 months around the latest sunspot maximum on a low orbiting satellite, are shown in Fig. 3.

This figure demonstrates, for instance, that the lower the particle energy, the larger the asymmetry between noon and midnight sectors with regard to polar distance. The difference is, however, not very large between 1 and 13 keV, and the asymmetry is on the whole smaller than expected on the basis of the shape of the auroral oval. Even at the energy 6 keV, which is thought to be near the average of the energy range of those precipitating electrons causing the visual aurora, the impression is that the anisotropy is very small, if existing at all. At 13 keV there seems to be quite good symmetry. One should remember that the auroral oval is generally not thought of as a statistical concept but rather as the locus of the aurora at a given moment. Even so, the satellite results shown in Fig. 3 do not seem to support very well the present views of the dominating role of the auroral oval as the region of most frequent and intense precipitation of keV electrons. This illustrates the fairly considerable difficulties that there are in comparing satellite measurements of auroral particles with observations of auroral luminosity from the ground. The visual luminosity depends not only on energy flux but on the energy flux above a certain threshold and also on energy spectrum and, to a certain degree, on pitch-angle distribution. A few detailed comparisons of ground observations of aurora and measurements of energetic particle precipitation on board satellites have been carried out. They indicate that the electrons of several keV energy generally produce the aurora, as illustrated in Fig. 4.

Figure 3 further shows that the local time distribution of the precipitation intensity for the energy range 1–13 keV electrons has two broad (in local time) maxima centred in late morning and late evening. Between these two maxima there are two pronounced valleys with their bottoms at about 0200 and 1400 EDT (eccentric dipole time). This does not fit too well with the distribution of auroral occurrence which has a maximum around local midnight, or shortly before, in the auroral zone. No good physical explanation of the locations of the minima (valleys) in the local time distribution exists at the present time. We will come back to a discussion of this later in this chapter.

There are observations indicating that there is some amount of discontinuity at about 1900 EDT in the latitudinal distribution of the precipitation intensity, although this is not clearly visible in Fig. 3. The most intense precipitation is located some $3\text{--}5^\circ$ further poleward to the west of this EDT.

Finally, it may be pointed out that the latitudinal gradient of the precipitating flux is much greater on the nightside than on the dayside. The electrons on the low latitude side of the dayside precipitation zone have probably drifted there from the night hemisphere.

The electrons of still higher energies than those represented in Fig. 3 produce ionisation mainly below 100 km altitude, where the collision frequency is high and so, therefore, is the absorption of radio waves. Synoptic data of radio wave absorption provide us with the best information available about the latitude-local time distribution of auroral electron precipitation for energies above, say 15 keV. Figure 5 shows the percentage of the total measuring

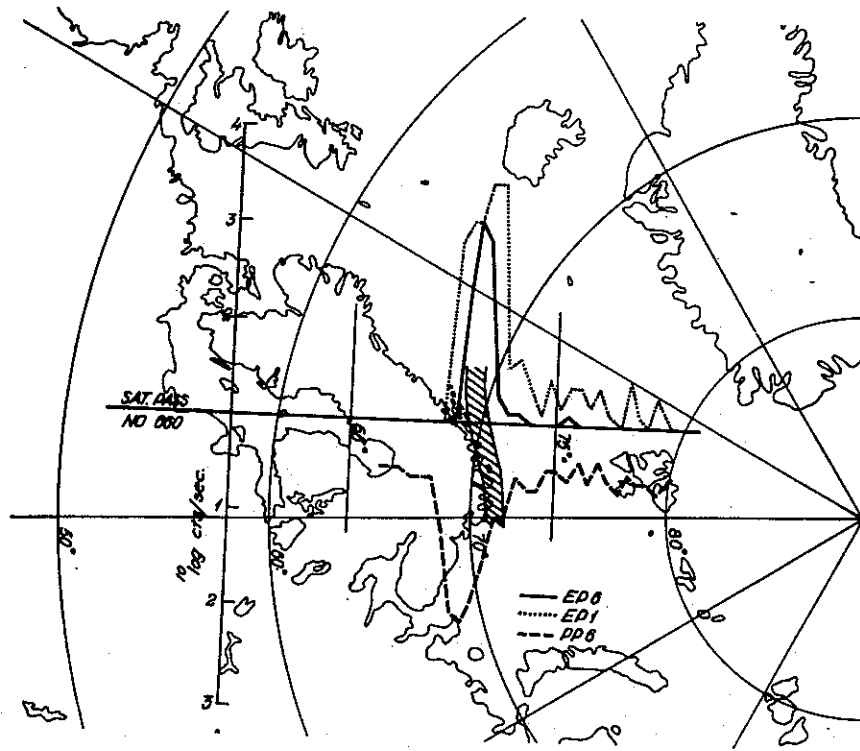


Figure 4. An auroral form measured by all-sky cameras has been projected to the 100 km level for the lower border and to the 130 km level for the upper border and drawn on a map (hatched area). Superimposed are the passage of the satellite and the particle measurements. The latitude values along the orbit refer to the satellite. The meanings of the notations are the following: EP6 – precipitated electrons of 5.8 keV energy, PP6 – precipitated protons of 6.3 keV energy, EP1 – 1.3 keV precipitated electrons (Gustafsson, 1970).

time that auroral radio-wave absorption of 0.5 dB or more occurred at 30 MHz during maximum solar activity. Characteristic is the large maximum in the morning and the minimum at dusk. We will return to the dynamics of this distribution later.

For energies below 1 keV, precipitation is found in general to have a somewhat larger extension, both poleward and equatorward, than at 1 keV. There is, in other words, a softening of the spectrum at the edges of the precipitation zone which will be discussed in a later section.

Figure 3 represents fairly low geomagnetic activity levels. When the activity increases the precipitation zone grows equatorwards but the main features seen in Fig. 3 are retained.

Over the central polar cap the fluxes of precipitated auroral particles are generally very much lower than in the auroral region. Also the average energy is much lower and particularly below 80° invariant latitude (Λ) there is much more spatial fine structure in the flux. Above $\Lambda = 80^\circ$ the intensity of auroral electrons is very low and has again less spatial variation. The precipitation intensity there appears to be anticorrelated with the magnetic disturbance level.

Sporadically intense precipitation of higher energy (>40 keV) auroral particles occurs over the polar caps. The geographical extension of such precipitation events is quite limited and they therefore show up as spikes in measurements on board low-orbit satellites. These spikes

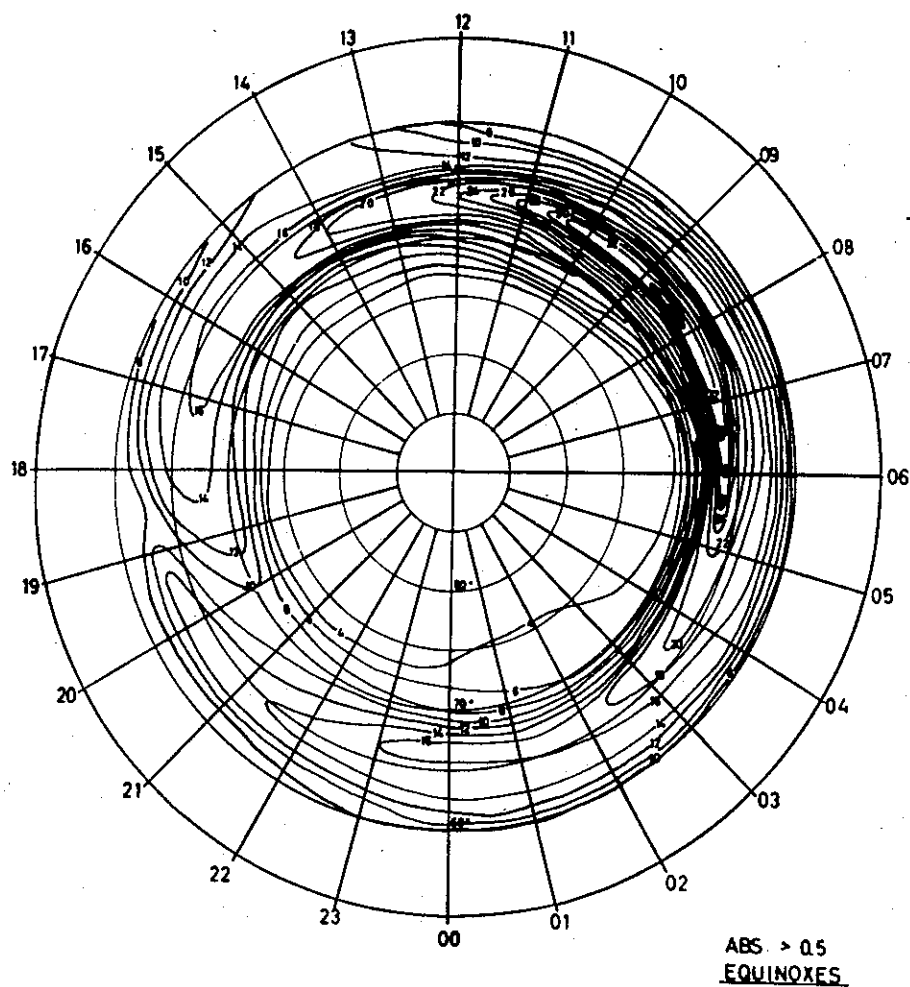


Figure 5. Percentage of total time with absorption exceeding 0.5 dB at different geomagnetic time and latitude during equinox months (Oct. 1958 - Feb./Apr. 1959), (Holt, 1963).

are probably related to the particle 'islands' sometimes observed in the distant magnetosphere in and outside the plasma sheet; they will be discussed later.

We have hitherto only described the electron precipitation. For auroral protons the amount of observational data is more limited. Figure 6 shows the distribution of precipitating 6 keV auroral protons for activity levels varying from fairly low to fairly high ($K_p = 1-5$). We see that the distribution resembles very much the electron distributions in Fig. 3.

The relative intensity in the noon sector is, however, lower than for electrons. There is a discontinuity at about 1900 EDT. There is a marked minimum at 0200 EDT, and the latitudinal gradients of the flux are greater on the night side than in the day, as it is for electrons. The anticorrelation between precipitated fluxes over the polar cap and K_p is even more pronounced than for the electrons. The distribution of the 6 keV proton flux well outside the loss cone, i.e. of protons which are not on their way down into the atmosphere but will mirror and return towards the other hemisphere, looks largely the same as that for the precipitating protons.

A comparison between Figs. 3 and 6 gives the impression that the auroral electrons and protons of several keV energies are precipitated together. Investigations of individual cases show that this generally happens, but frequently the proton precipitation on the evening side reaches latitudes a little lower than the auroral electrons. For the midnight-dawn sector the equatorward proton boundary is often well poleward of the electron boundary. Two examples of latitudinal distributions of electron and proton precipitation as observed in the dusk-midnight sector by a low orbiting satellite are presented in Fig. 7.

It is not protons of energies below 10 keV that are mainly responsible for the $H\alpha$ and $H\beta$ emissions in aurora, but somewhat higher energy protons penetrating to ordinary auroral heights of 100–150 km in the atmosphere. Direct experimental data of the kind shown in Fig. 6 have not been published for these proton energies. Only indirect information about the

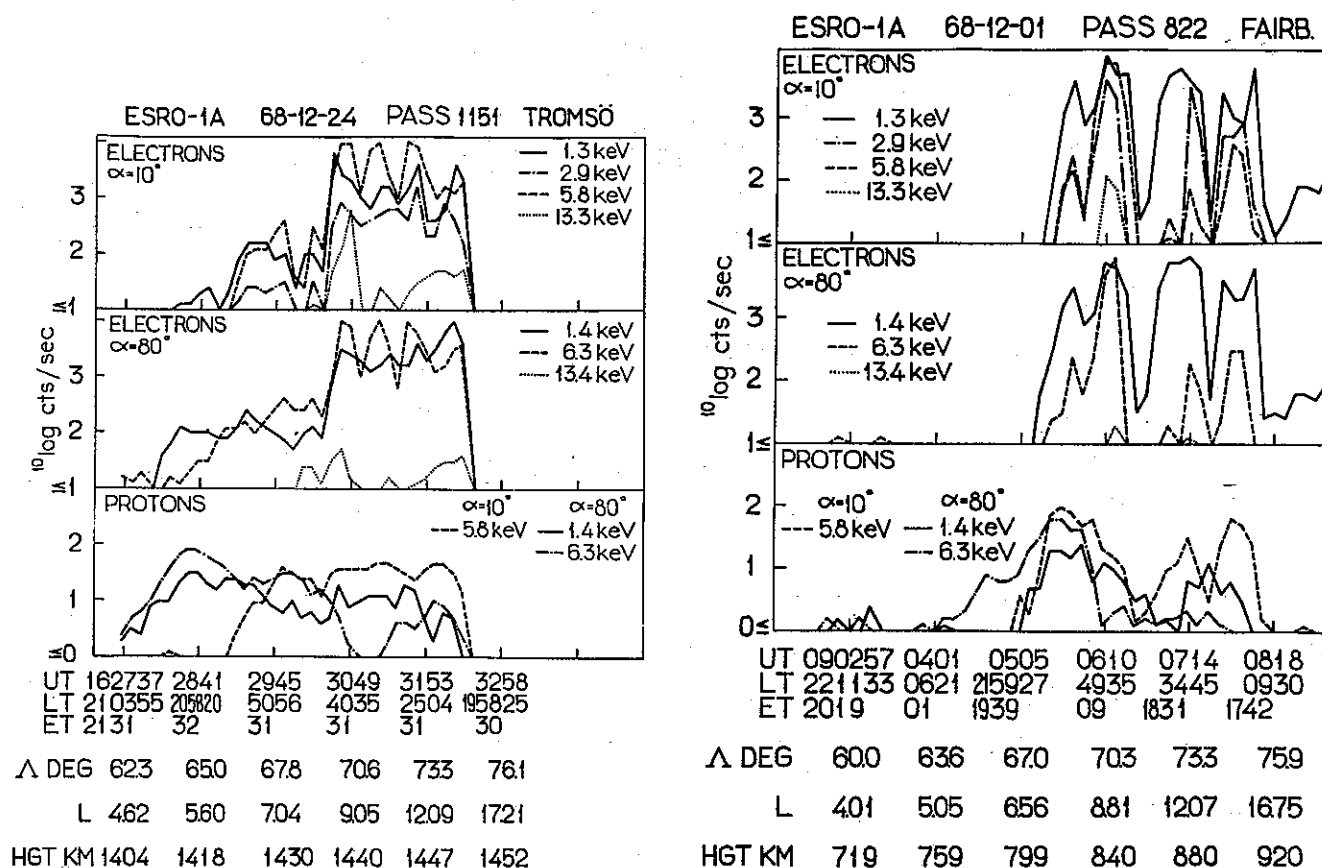


Figure 7. Measurements on board the satellite ESRO 1A of electron and proton fluxes at various energies for two different pitch angles, $\alpha = 10^\circ$ and 80° . The curves connect 8 s averages for the count rates of the detectors. The time figures give hour, minute and second. LT means geomagnetic local time (dipole time) and ET excentric dipole time (in hours and minutes). Δ is the invariant latitude (in practice identical with corrected geomagnetic latitude) at the Earth's surface derived from the L -value for the satellite location. The energy windows of the detectors had a width of about 10% of the energy value. The field of view was about $\pm 8^\circ$. The telemetry station which received the data is given at the top right (Hultqvist et al., 1971).

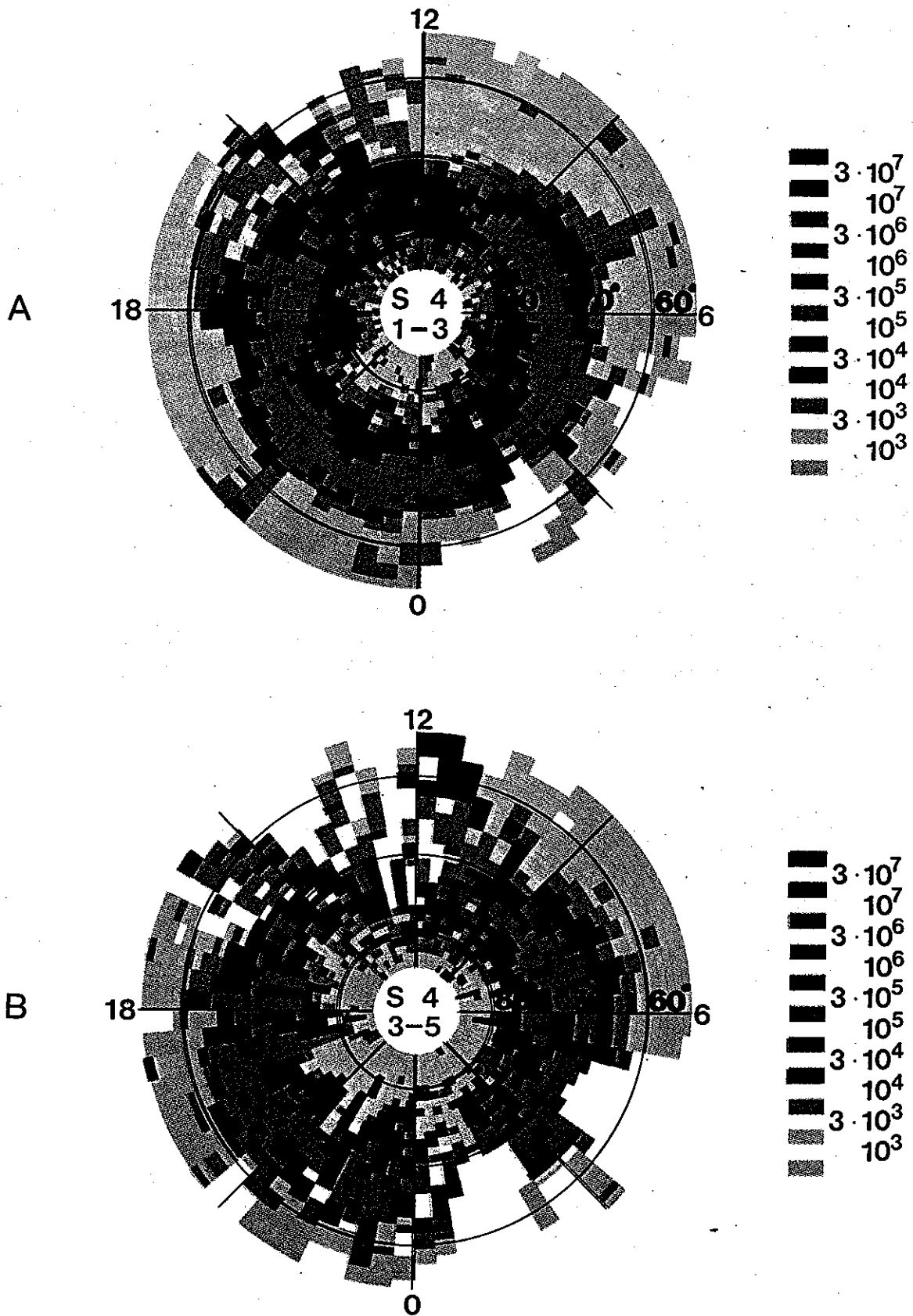


Figure 6. Average fluxes of precipitated (pitch angle 10°) 6 keV protons as measured during a 20 months period from October 1968 to April 1970 with the ESRO 1A satellite. (a) is for $K_p = 1-3$; (b) $K_p = 3-5$. The coordinates are corrected geomagnetic (invariant) latitude and excentric dipole time (Riedler and Borg, 1971).

One reason for this difference is a time-smoothing that has been introduced in the satellite data. Eight second averages have been used in Fig. 7. But even if a time resolution of 50 ms is used, corresponding to a spatial resolution of about 350 m for a low-orbiting satellite, the gradients seen do not correspond to the visual impressions. This is mainly because the eye can see the aurora only when the luminosity is above a certain limit (depending on the contrast). The distinct auroral forms are often the only luminosity in the sky recognized by the eye (and the all sky camera), but in fact they are just the peaks of a widespread luminosity, like the visible part of an iceberg. The satellite instruments see the entire 'iceberg' and therefore the visible peaks do not look so very spectacular.

2.3. Auroral particles in the magnetosphere – average characteristics

As indicated in Fig. 2, particle populations contributing to the excitation and ionization of the upper atmosphere fill up large parts of the outer magnetosphere. We may classify the regions as the plasma sheet, the ring-current belt (also called the storm-time belt) and the polar-cusp regions. The energetic particle density in the various parts of the magnetosphere varies strongly with the disturbance level. In this section the more stationary characteristics will be outlined and the dynamical aspects will be dealt with in the next one.

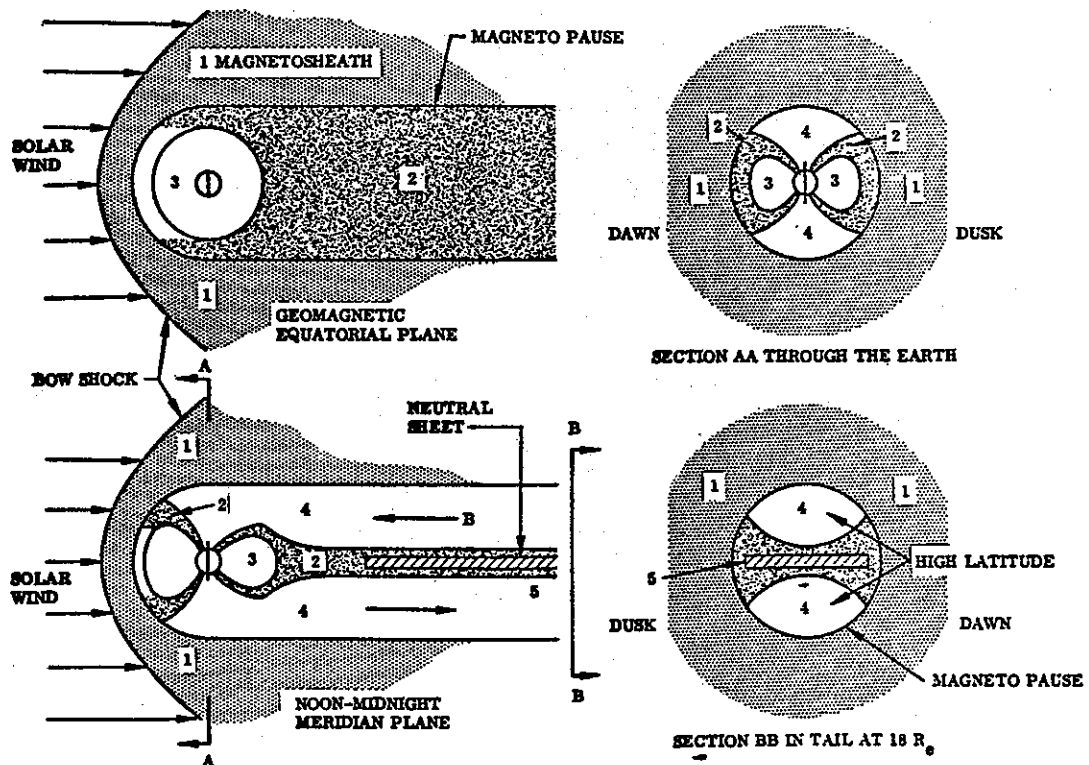


Figure 9. The spatial extent of the plasma sheet and magnetotail. The regions are (1) magnetosheath, (2) plasma sheet, (3) trapping region, (4) high latitude region, and (5) neutral sheet (McCormac et al., 1971).

The *plasma sheet* in the magnetospheric tail is centred on the neutral sheet. While the latter is quite thin, the plasma sheet has normally a thickness of a few earth radii at the centre and 2–3 times this value at the magnetopause. The topography of the plasma sheet and its connections to the ionosphere are shown in a schematic way in Fig. 9. In the plasma sheet the density of the hot plasma (auroral particles) is of the order of 1 cm^{-3} . The typical energy spectrum of the plasma sheet electrons has a broad maximum at an energy ranging from a fraction of 1 keV to more than 10 keV, with the mean energy mostly of the order of 1 keV near the centre of the plasma sheet and $1/3$ to $1/2$ of that value near its high-latitude boundaries. The spectra, which fairly often do not seem to be much different from a Maxwellian distribution, may be characterized by an equivalent temperature of the order of 10^7 K for an average particle energy of 1 keV. The proton spectra have a shape similar to that of the electrons, but the mean energy generally exceeds the electron mean energy, sometimes by a factor of up to 10. The angular distributions of both electrons and protons are mostly isotropic except in certain disturbance situations which will be discussed later.

As mentioned above, precipitation spikes of fairly high energy electrons are sporadically observed over the polar caps. In a number of cases it has been possible to relate these spikes to intense bursts, called *particle 'islands'*, in the tail. Many of these bursts represent sudden hardenings of the electron spectrum in the plasma sheet. In other cases the islands have been located outside the plasma sheet.

As mentioned before, the protons may extend almost to the plasma sphere in the midnight sector during magnetic storms, and even, as indicated in Fig. 2, into the plasma sphere at dusk. They then form the *ring current* responsible for the main phase of the storm. The ring current belt is highly asymmetric with the largest density on the evening side. The proton fluxes in the energy band 50 eV–50 keV are generally found in the range $10^5 - 10^7 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ keV}^{-1}$.

Satellite data taken in 1969 and 1970 have established the existence of two cusp-like regions in the dayside magnetosphere and the penetration of magnetosheath plasma to low altitudes through them as illustrated in more detail in Fig. 10. The particle flux in the *polar cusps* shows less variations with magnetic activity than other auroral particle fluxes in the magnetosphere. There is, however, some structure in the electron flux in the cusps; bursts are superimposed on a background continuum (for electrons but not for protons). The width of the cusp is generally 2–3 latitude degrees in the ionosphere but may occasionally be larger. The cusp is mostly found at $75-79^\circ \text{ A}$, but has been observed as far equatorward as 67° during strong magnetic storms. The longitudinal extension of the cusps is quite large; generally from about 0800 to 1600 EDT. There are thus two large clefts in the dayside magnetosphere, extending over the whole sunlit side, through which large amounts of magnetosheath plasma penetrate. This energy influx is mostly of the order $0.1 \text{ erg cm}^{-2} \text{ s}^{-1}$ which is sufficient to explain the dayside aurora. The particle spectrum is much softer in the cusps than in the tail. The average energy is usually of the order of 100 eV.

The discovery of the *radiation belts* in the late fifties created a boom of speculations about the role of these belts in the auroral phenomena. For several years the idea was prevalent that the precipitated particles were dumped from the large storage of energetic-trapped par-

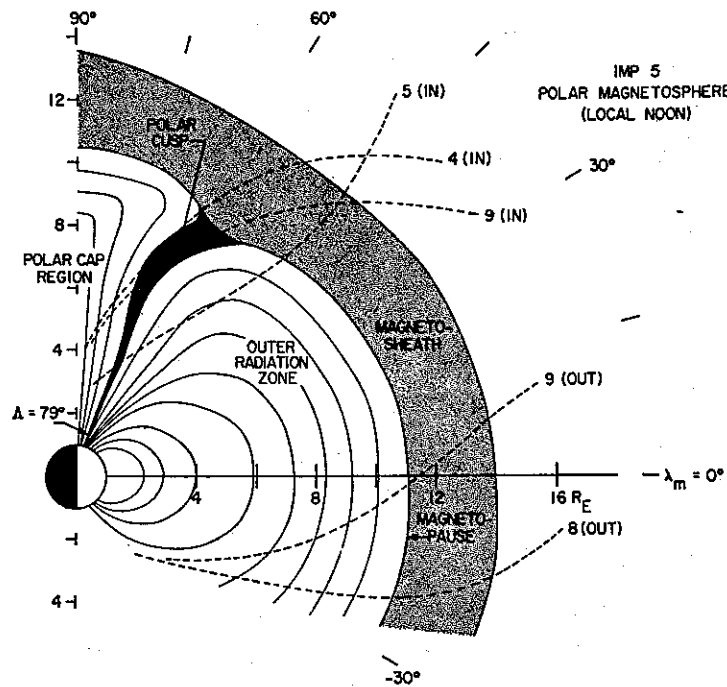


Figure 10. A diagram showing the geometry and location of the polar cusp within the polar magnetosphere in the noon meridional plane during periods of relative magnetic quiescence. The coordinates are geocentric radial distance, R_E , and dipole magnetic latitude, λ_m . The polar cusp intersects the auroral zone at $\Delta \approx 79^\circ$. Several sample trajectories of IMP 5 through the dayside magnetosphere are also shown (Frank, 1971).

ticles in the magnetosphere through the influence of disturbing effects, until Lunik 2 and Explorer 12 measurements demonstrated that the earlier values of the particle flux in the radiation belt were too high by several powers of 10. It was also found that the flux of trapped radiation increases, not diminishes, when auroral particle precipitation takes place, indicating that the radiation belts are fed with particles by the same mechanism that causes precipitation into the atmosphere.

Closely related to the question of the relations between radiation belt particles and auroral particles is the question of whether auroral particles occur on 'open' or 'closed' field lines. Even if by 'closed field lines' we mean field lines that are permanently closed and do not participate in the convection from the dayside to the tail and back again, it is clear from the fact that worldwide auroras occur down to $L \approx 1.3$ that at least sometimes aurora occurs partly within the region of permanently closed field lines. If, on the other hand, we define closed field lines only by their connecting the two hemispheres, even though they may be drawn out quite far into the tail, as illustrated in the auroral cusp region of Fig. 2, a large fraction of the auroral particles (giving rise to the most intense aurora) are found on closed field lines. In fact, the maximum flux of precipitated electrons greater than 40 keV has been found at the same L value as that of trapped electrons in the same energy range. On the other hand, at least the soft particles at the poleward edge of the precipitation zones and over the polar caps are on open fields lines – open in the sense that bouncing between hemispheres and drift of particles cannot occur.

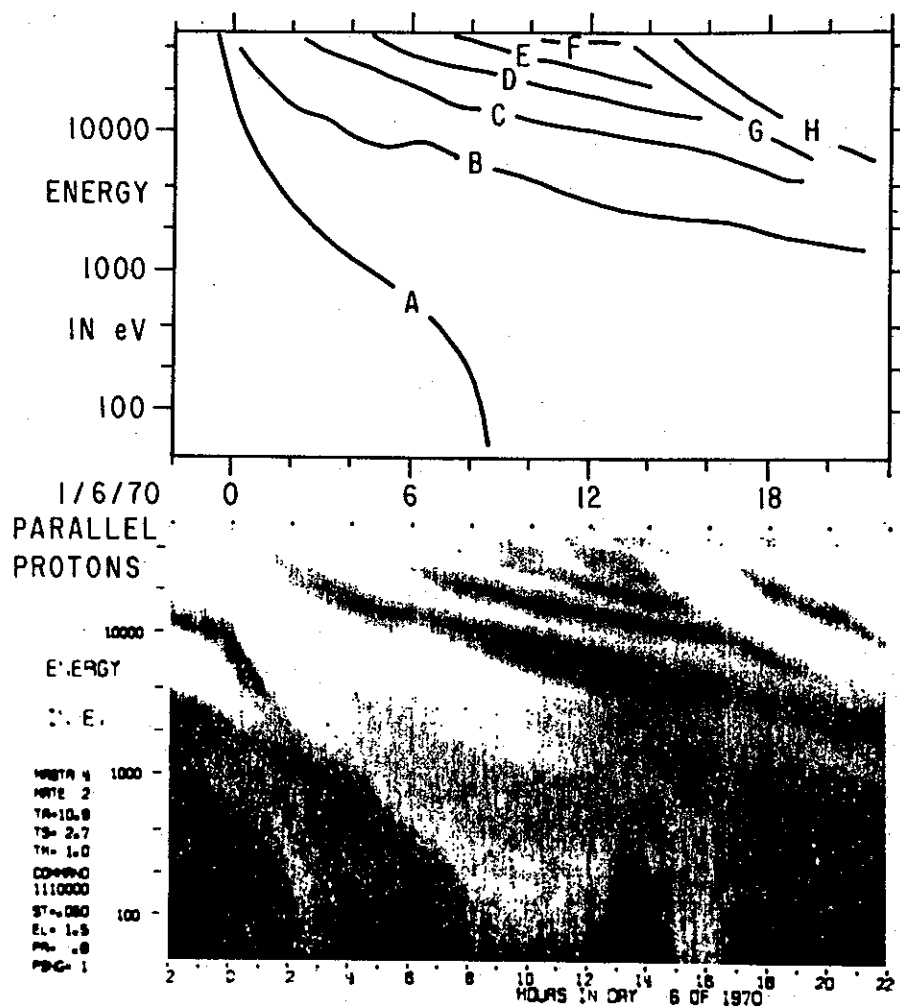


Figure 11. A spectrogram of the proton differential energy flux during 6 Jan. 1970. The pitch angle is 0° . The intensity of the energy flux is given by the brightness of the diagram. The upper half identifies how some of the maxima in the energy spectrum vary with time (DeForest and McIlwain, 1971).

2.4. Dynamics of auroral particle morphology in the magnetosphere

The auroral particles on the dayside, which intrude the magnetopause through the polar cusps, show relatively little variations with magnetic activity level. The process is more complicated on the nightside, where most auroral particles apparently originate in the plasma sheet. This, in turn, is fed by the solar wind in a way which is not quite clear. The solar wind plasma may be heated in the tail by merging of the geomagnetic field lines which have been convected from the dayside of the magnetosphere. Such a process may increase the average energy of the solar particles to values of the order of 1 keV, as has been observed in the plasma sheet. But even after this heating, an acceleration of part of the particle population is needed. This acceleration process is in fact not a steady one but involves a characteristic sequence of events called a substorm.

A substorm may be divided up into a number of stages of which the first one is usually called the growth or development phase. This seems to start when a south pointing component in the interplanetary magnetic field develops. At the beginning of the substorm growth phase the plasma sheet is generally several earth radii thick even at magnetic midnight. During the growth phase, which lasts typically 1–2 hours, the plasma sheet gradually decreases in thickness and the plasma moves towards the Earth. The inner boundary of the plasma sheet in the equatorial plane moves several earth radii inwards and at the same time a quiet auroral arc is seen to move slowly equatorward.

At some stage a violent eruption occurs. What triggers it is not known. The eruption is known as the expansive phase of the auroral substorm. Highly directed particle flows are observed in the plasma sheet at this time and after a while the plasma sheet thickens again. The very break up of the aurora at the initiation of the expansive phase shows up in the tail only as a little spike in the flux of fairly energetic auroral particles.

The hot plasma which is brought closer to the Earth during the growth and expansive phases seems to be introduced in a fairly narrow local time sector, located in most cases close to magnetic midnight. From this injection sector the auroral particles then drift around the Earth in the combined disturbance electric field and the magnetic field dominated by the stationary geomagnetic field. Measured (in local time) at some distance from the injection region, the spectrum of the energetic particles drifting by is strongly peaked at any moment, the energy value of the peak decreasing continuously as time goes by and slower and slower particles reach the detector. Presented in a three-dimensional diagram with time and particle energy as axes and the measured energy flux given by a grey scale (the higher flux the brighter) the results obtained at the geostationary orbit look as in Fig. 11. Each of the sloping light bands represents a substorm injection of particles into the inner magnetosphere.

In the ionosphere the dynamics of the keV auroral particles are best shown by the dynamics of the visual aurora during a substorm. It has been described in the papers by Akasofu and Feldstein (this volume) and will not be dealt with here.

The auroral particles of energies above, say 15 keV show a somewhat different behaviour during substorms than do the keV particles. At these energies the longitudinal drift caused by gradients in the geomagnetic field dominates other kinds of drift, and therefore these particles follow invariant latitude circles when they drift – the more accurately the higher the energy – instead of the auroral oval type of geometry. The temporal development of a substorm event as seen in the precipitation flux of this high energy tail of the auroral particles is illustrated in Fig. 12. The contour lines in the figure represent isoabsorption lines for radio waves at 30 MHz. The maps are based on measurements at 44 riometer stations all over the auroral regions and the polar cap. There is one map for every 15 minutes of the substorm in Fig. 12. As can be seen, the injection of the energetic particles takes place in the midnight sector and the electrons then start drifting eastward. While drifting the precipitation rate increases; the maximum auroral absorption moves eastward. The event shown in Fig. 12 is a typical one. There is, however, a large dispersion in the behaviour of different substorms. Some general conclusions on the basis of studies of many substorms may be stated as follows:

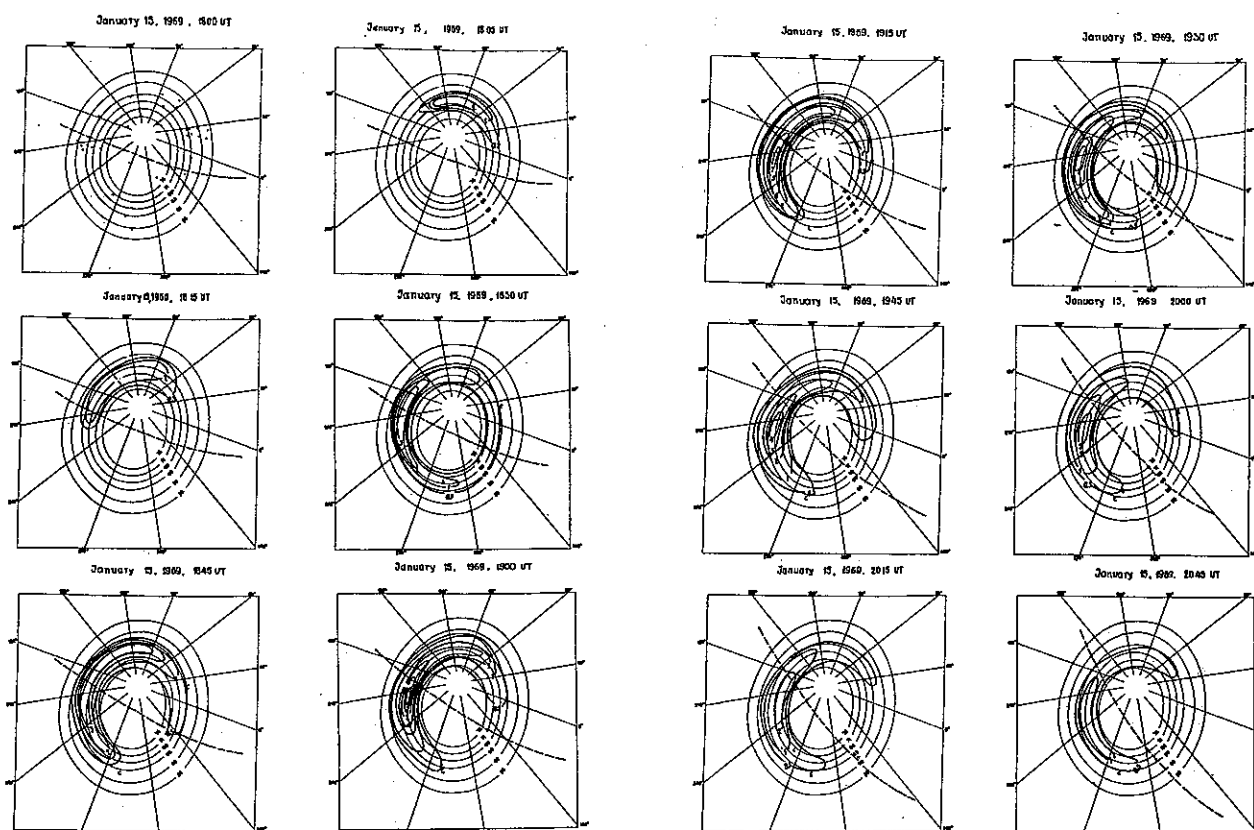


Figure 12. The temporal development of the precipitation pattern for auroral electrons of energies above, say 15 keV as observed by riometers in auroral and polar cap regions. The locations of the riometers which have delivered data to the maps are shown on the first map. The coordinates are geographic longitude and corrected geomagnetic (in practice identical with invariant) latitude. The isoabsorption contours are given for 0.3, 1, 2 etc. dB. The dashed line cutting across the maps shows the location of the boundary between sunlit and dark hemispheres at a height of 80 km (Berkey et al., 1971).

- (a) There is in general injection of electrons on the nightside. They then drift eastward.
- (b) But there is often also an extension westward which cannot be due to drifting electrons (and certainly not to drifting protons either).
- (c) The location of the injection area varies also very much from case to case and can occur anywhere between evening and forenoon (but appears most frequently around midnight).
- (d) The velocity of the eastward expansion can sometimes be extremely high (up to 140° in 5 min.).
- (e) The distribution in general follows the auroral zone fairly well, but deviations from it occur, some of which indicate the auroral oval. On the whole the oval has very little to do with these particles, however.
- (f) The absorption maximum may stay in the injection area, but most frequently it moves eastward. However, it has also been seen moving westward. It may move more than 270° around in eastward direction.

In summary, although drift motions are very important for the temporal development of these precipitation events they cannot explain all characteristics described. Direct injection from the outer magnetosphere over expanding sectors seems to be needed during the expansive phase of substorms.

Before leaving the subsection on the dynamics of auroral particles, a few words about time variations in the particle precipitation. It is not possible to distinguish time and space variations in measurements by satellites, moving fast with regard to the magnetospheric reference system. Even most sounding rocket measurements present serious problems in this respect. Visual observations of active aurorae suffice to demonstrate fast time variations of periods down to a second or even less. Auroral X-rays, measured with balloons, which are produced by electrons of energies above some 25 keV, show time variations with periods as low as 0.1 s. There are indications that the temporal variability increases with increasing energy of the particles.

3. Energy spectra of auroral particles

The energy spectrum of the auroral electrons may be divided into three parts of different characteristics with regard to detailed spectral shape, spatial distribution, and relation to upper atmosphere phenomena.

- In the lowest energy range above thermal (<0.5 keV), the spectrum is mostly, but not always, smooth. The pitch-angle distribution is often peaked along the magnetic field lines. The precipitation is spatially widespread; it does not give rise to distinct auroral forms but carries field-aligned electric currents.
- The particles causing the auroral forms are found in the energy range 0.3–20 keV. Thus, they carry most of the energy of the particle precipitation and show strong spatial structure. The energy spectrum in this band mostly contains one or more peaks in the pre-breakup aurora but are smoother after breakup. The pitch-angle distribution is in general nearly isotropic.
- The high-energy tail of the auroral particle spectrum (>15 keV) is unimportant as energy and current carrier. These particle fluxes are very dynamic both spatially and temporally. The energy spectrum is mostly smooth but not always. The pitch-angle distribution shows a depleted loss cone when the precipitation intensity is not too high, but generally approaches isotropy in intense precipitation events. In exceptional cases, field-aligned anisotropy seems to occur.

The fairly limited data on the auroral proton energy spectrum that are available generally correspond to an e -folding energy of 10–50 keV below 100 keV and still higher values in the high-energy tail.

It is not possible to give any statistically meaningful average spectra, particularly not for the lower part of auroral particle energy range, as only a few of such data have so far been published. Instead, some examples from various parts of the magnetosphere and ionosphere will be presented below. For the high-energy tail of spectrum (>40 keV), statistical spectral data

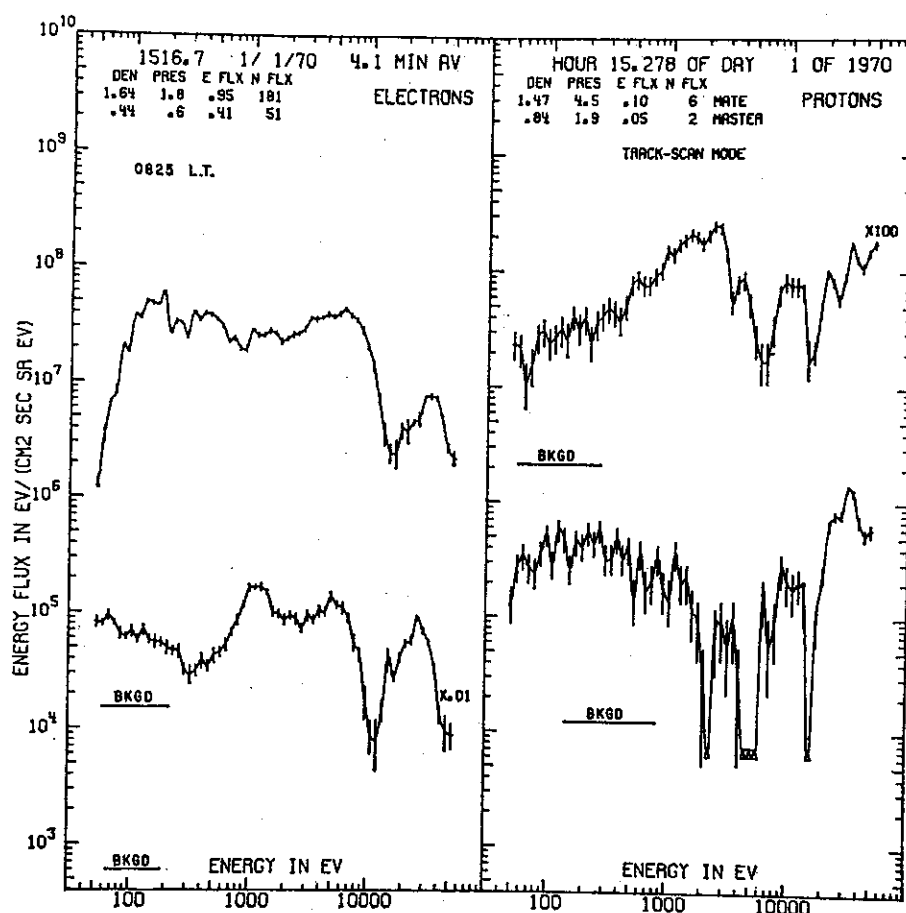


Figure 13. An example of differential energy flux spectra for electrons and protons. The upper curves are for pitch angles close to 90° and the lower ones for pitch angles around 0° . The parallel-to- B electron energy flux has been multiplied by 10^{-2} and the perpendicular-to- B proton energy fluxes by 10^2 . All values are 4.1 minutes averages (DeForest and McIlwain, 1971).

have been published. The spectra have, however, been obtained from integral flux measurements and are therefore quite rough. They may be related to e.g. statistical data on average auroral absorption of radiowaves.

More insight into the physics involved is provided by the dynamic spectra shown in Fig. 11, which have been obtained on a geostationary satellite. The magnetic field lines of the geostationary orbit reach the Earth's surface in the auroral zone. The dispersion of the hot plasma clouds injected during substorms has the effect that at a given local time one expects the energy spectrum to vary widely with time from the injection. The energy spectrum generally shows one or more peaks and the energy values of the peaks may vary from below 1 keV to about 50 keV. An example of the differential energy flux (particle flux multiplied by energy) of electrons and protons is given in Fig. 13. As can be seen, there are several strong peaks particularly at high energies. A case of almost monoenergetic electron spectrum observed in a pre-breakup aurora is shown in Fig. 14.

It seems that the high degree of fine structure of the energy spectrum at $L = 6.6$ (geostationary orbit) does not exist in the plasma sheet. Some examples of electron spectra measured

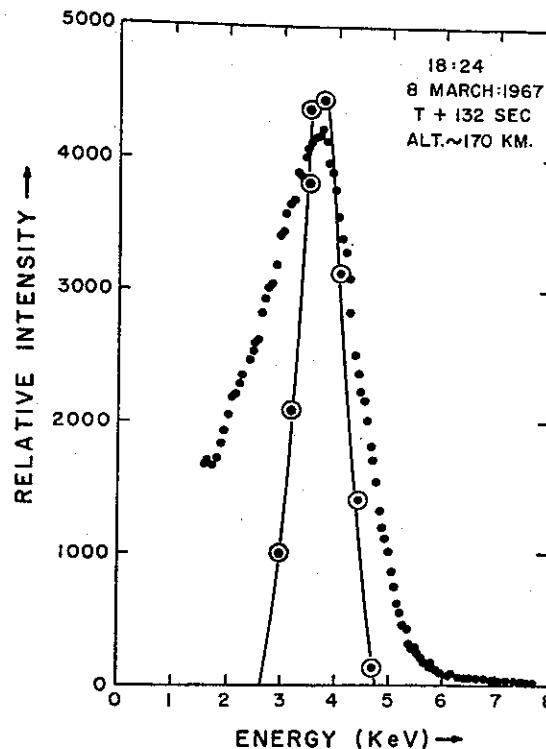


Figure 14. Nearly monoenergetic electron spectrum measured in a pre-breakup aurora of about IBC 1 intensity. The full line gives the response of the detector to a mono-energetic flux. The slope on the high-energy side corresponds to an equivalent e -folding energy of the order of a few tenths keV. Hardly any variation of the energy value of the peak (3.8 keV) occurred over a period of 150 sec (Evans, 1968).

there at a distance of 17 earth radii (R_E) are shown in Fig. 15. Not only are the spectra generally smoother in the plasma sheet than in the upper atmosphere but also the fluxes are lower. It therefore seems clear that an acceleration takes place during the inward transportation of the hot plasma from the outer magnetosphere. An acceleration process seems to be needed between the equatorial plane and the ionosphere too, because the particle fluxes seen in the atmosphere tend to be somewhat higher than at the middle of geomagnetic field lines, as mentioned.

Some representative examples of energy spectra from the polar cusp region can be seen in Fig. 16. The energy spectrum shows a latitudinal variation for keV electrons. In the main part of the precipitation zone the 'hardness' does not vary very much, but at the edges the spectra soften considerably. On the poleward side and up into the polar cap the spectra are much softer than in the main auroral precipitation region. This soft zone may have quite a large latitudinal extension. On the equatorward side it is just the edge of the precipitation zone that shows a softer spectrum. This is illustrated in Fig. 17. On a much smaller scale there also exist large variations in the hardness of the spectrum. The most important regularity in this connection is that the core of an auroral form has a harder spectrum than its edges.

For the lowest part of the energy range of auroral electrons the energy spectrum has been found to be generally softer on the nightside of the Earth than on the dayside (except in the

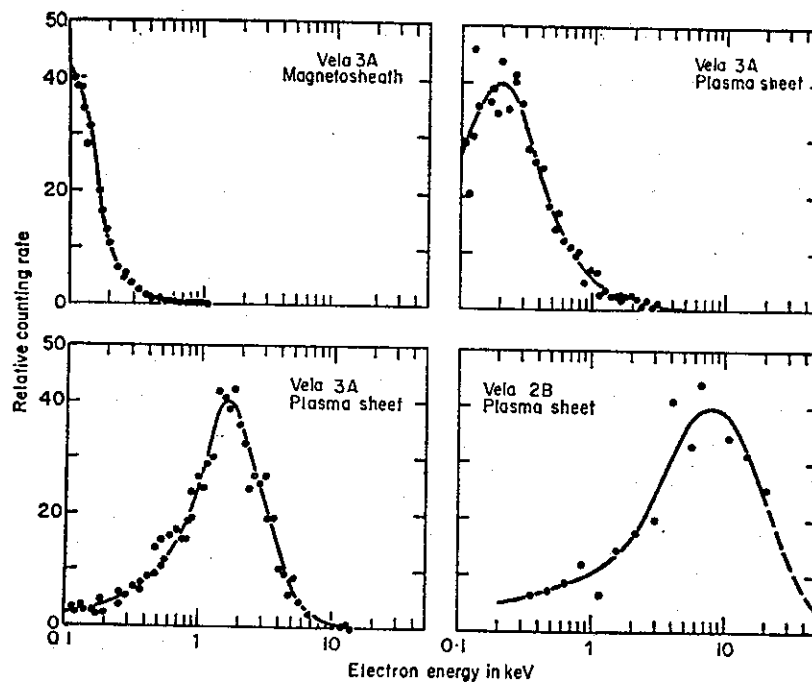


Figure 15. Electron energy spectra observed in the magnetosheath and in the plasma sheet by the Vela satellites (Bame et al., 1967).

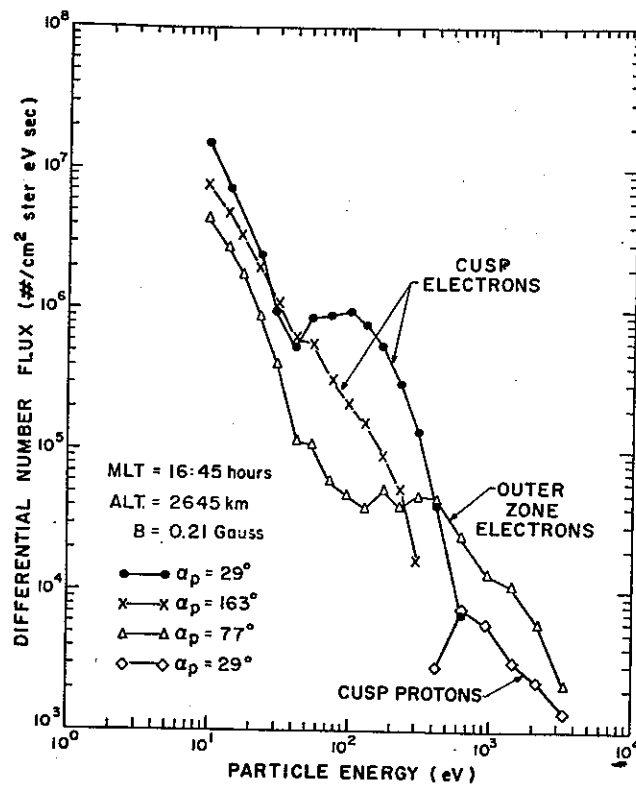


Figure 16. Electron and proton differential spectra from the polar cusp and from the outer radiation belt (Winningham, 1971).

polar cusps). For electron energies above some 25 keV the hardness of the precipitation spectrum has a minimum a few hours after magnetic midnight and a maximum shortly before midnight.

It has been known for some years that in the energy range 40–250 keV there is a tendency for softening of the electron spectrum, i. e. lowering of the average energy with increasing Kp . On the other hand the measuring results for energies below 40 keV show that part of the spectrum for precipitated electrons tends to harden when Kp increases. It thus seems that the greatest relative increases in the spectrum occur somewhere fairly close to 40 keV.

It may be worthwhile to emphasize that the intensity and spectrum of auroral electrons very often show an astonishing degree of stability. Most electron measurements in aurora by rockets in the last few years have shown such a high degree of stability. But measurements have also been reported in which the rocket appeared to pass through beams of electrons, and intensity changes of three orders of magnitude in less than a second have been observed.

In many cases it has not been possible to find any clear correlation between changes in intensity and changes in the spectrum and angular distribution. Strong variations in flux need not be associated with any variation in spectrum shape at all.

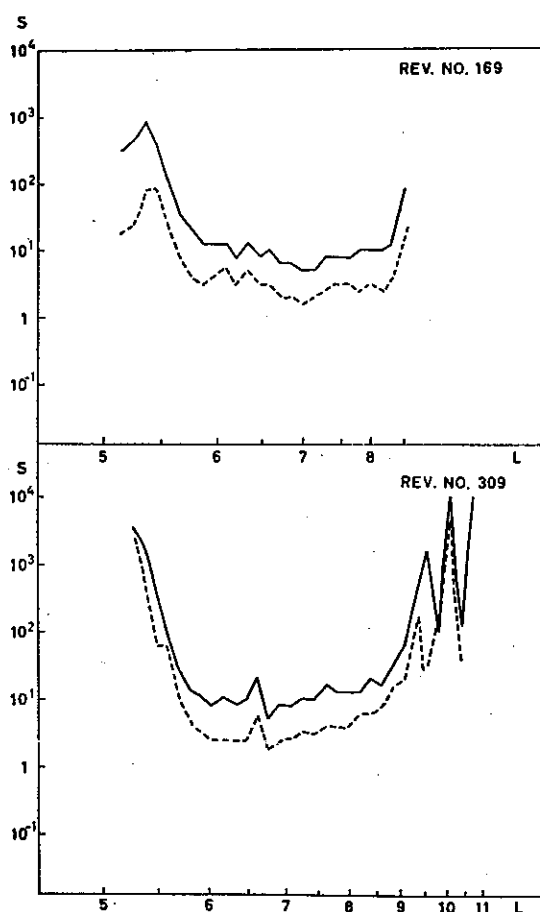


Figure 17. The steepness, S , defined as the ratio between electron fluxes at 1 and 6 keV, as function of L value for two passes of the satellite ESRO 1A. Notice the 'soft edge' at low L values. Solid curves correspond to precipitated electrons, broken to trapped electrons (Liszka et al., 1970).