DET NORSKE VIDENSKAPS-AKADEMI I OSLO

GEOFYSISKE PUBLIKASJONER

GEOPHYSICA NORVEGICA

Vol. 29. June 1972

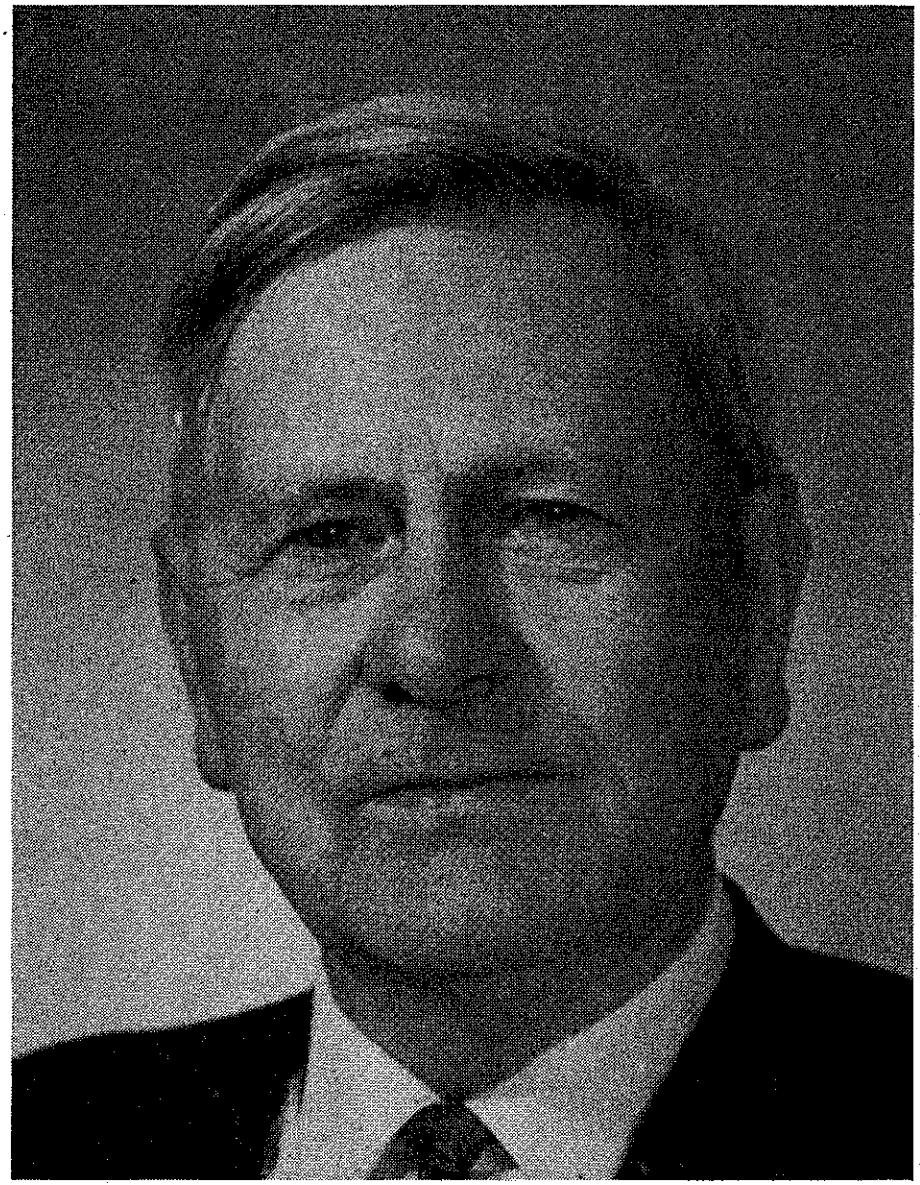
A Collection of Articles on Cosmic Geophysics Dedicated to the Memory of Professor L. Harang on the Seventieth Anniversary of His Birth, 19 April 1972

OSLO 1972
UNIVERSITETSFORLAGET

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

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

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 reseach 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

Geofysiske Publikasjoner, Vol. 29, 1972

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 ozon.

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 Stöhrungen'. 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 clearsighted 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*

$$[i] = \varepsilon_0 m_i \omega^2 / e^2 \text{ or } [e] = \varepsilon_0 m_e \omega^2 / e^2$$

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*, *F*1, and *F*2 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 *F*1 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} cm^2 for O and N. More recent quantal calculations (1939a) led to values of about 10^{-17} cm^2 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} cm^2 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} cm^2.

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} cm^2). 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 Eckersley 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

$X^+ + e \rightleftarrows X + h\nu$ between radiative recombination and photoionization, and had shown that the coefficient α would have a magnitude about 10^{-11} cm^3 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 $\alpha_E = 10^{-8}$ cm^3 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 $\alpha_{F2} = 10^{-10}$ cm^3 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* λ $(= [i^-]/[e])$

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 $\lambda \doteqdot 100$: 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 $0^+ + 0^- \rightarrow 0' + 0''$, 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 2×10^{-7} $cm^3\ s^{-1}$ 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 *F*2-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 *F*2-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 O_2 (1938d)
	» » metals (1942)
	» » NO (1945)
E layer	ionization of O_2 (1938a, d, e)
	» » all gases by X-rays (1948b)
	pre-ionization of O_2 (1945)
*F*1 *layer*	ionization of N_2 (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 ≠

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.	≠Theory of radiative recombination and ionization cross-section
1924	Thomson, J. J.	Phil. Mag., **47**, 337.	Theory of Ionic recombination
1925a	Appleton, E. V. and M. A. F. Barnett	Nature, **115**, 333	Vertical reflection of radio waves
1925b	Breit, G. and M. A. Tuve	Nature, **116**, 357.	
1925c	Bailey, V. A.	Phil. Mag., **50**, 825	Attachment probability
1926a	*Discussion*	Proc. Roy. Soc., A **111**, 1.	≠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.	Two layers E and F
1928a	Hulburt, E. O.	Phys. Rev., **31**, 1018.	Deductions from skip-distance
1928b	Appleton, E. V. and J. A. Ratcliffe	Proc. Roy. Soc., A **117**, 576.	Electrons responsible for radio wave absorption
1928c	Maris, H. B. and E. O. Hulburt	Terr. Mag. Atmos. Elec., **33**, 229.	Ionosphere hot
1930a	Appleton, E. V.	J. Inst. Elec. Eng., **66**, 872.	Result of 1927 eclipse
1930b	Eckersley, T. L.	Nature, **125**, 669.	Loss rate in $F2$
1931a	Chapman, S.	Proc. Phys. Soc., **43**, 26.	Theory of simple layer
*1931b	Chapman, S.	Proc. Roy. Soc., A **132**, 353.	Atomic oxygen above 100 km ≠E produced by corpuscles
1931c	Appleton, E. V.	Nature, **127**, 197.	Penetration frequencies
1932a	Appleton, E. V. and R. Naismith	Proc. Roy. Soc., **137**, 36	≠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.	Time of corpuscular eclipse
1933a	Appleton, E. V.	Proc. Phys. Soc., **45**, 673.	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 $F2$. $\neq$Conductivity of ionosphere
1937b	Pekeris, C. L.	Proc. Roy. Soc., A **158**, 650.	Revised tidal theory
1937c	Massey, H. S. W.	Proc. Roy. Soc., A **163**, 542.	$\neq\lambda = 100$ in E
1938a	Wulf, O. R. and L. S. Deming	Terr. Mag. Atmos. Elec., **43**, 283.	$\neq E$ layer formed at $O_2 - O$ transition level
1938b	Bradbury, N. E.	Terr. Mag. Atmos. Elec., **43**, 55	Height-dependent attachment in F
1938c	Hulburt, E. O.	Phys. Rev., **53**, 344.	$\neq$Reflection from E by ions
1938d	Mitra, S. K., J. N. Bhar and S. P. Ghose	Ind. J. Phys., **12**, 455.	$\neq O_2$ ionized in D and E, N_2 in F
1938e	Bhar, J. N.	Ind. J. Phys., **12**, 363.	$\neq O_2$ ionized in 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.	Quantal calculations of ionization cross section and coefficient of radiative recombination
*1939b	Hulburt, E. O.	In 'Terrestrial Magnetism and Electricity.' (Ed. J. A. Fleming), McGraw Hill, New York.	
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 $\lambda < 1$ in E
*1946b	Green, A. L.	Assoc. Wireless Australia Tech. Rev., **7**, 178.	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 Engs., **98**, 221.	*D*-region echoes
1951b	Helliwell, R. A., A. J. Mallinc -krodt 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 *F*1 at 170 km
1954b	Nicolet, M. and P. Mange	J. Geophys. Res. **59**, 15.	O and O_2 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 *F*2 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 *F*2 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.

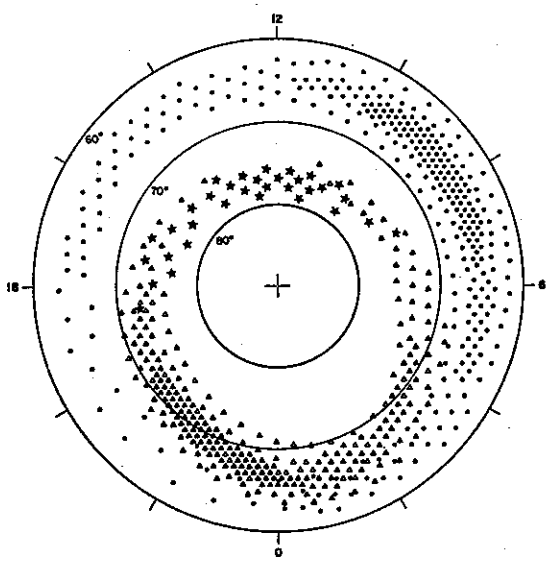


Figure 1. An idealized representation of a three-zoned auroral particle precipitation pattern. The auroral oval distributed precipitation (splash type) is represented by the triangles, the auroral-zone oriented (drizzle type) by the dots, and the polar-cusp precipitation on the dayside by the stars. 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.

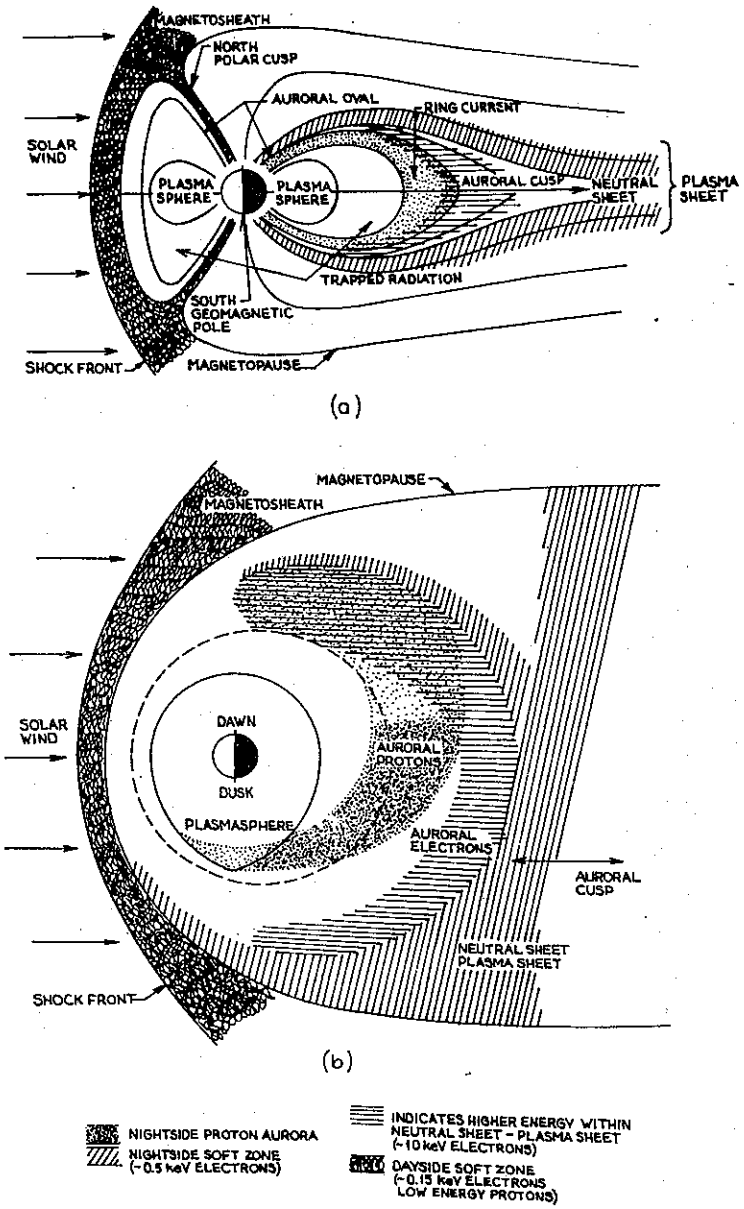


Figure 2. Noon-midnight meridian (a) and equatorial (b) views of the magnetosphere showing the distribution of particles and their precipitation regions (McCormac et al., 1971).

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.

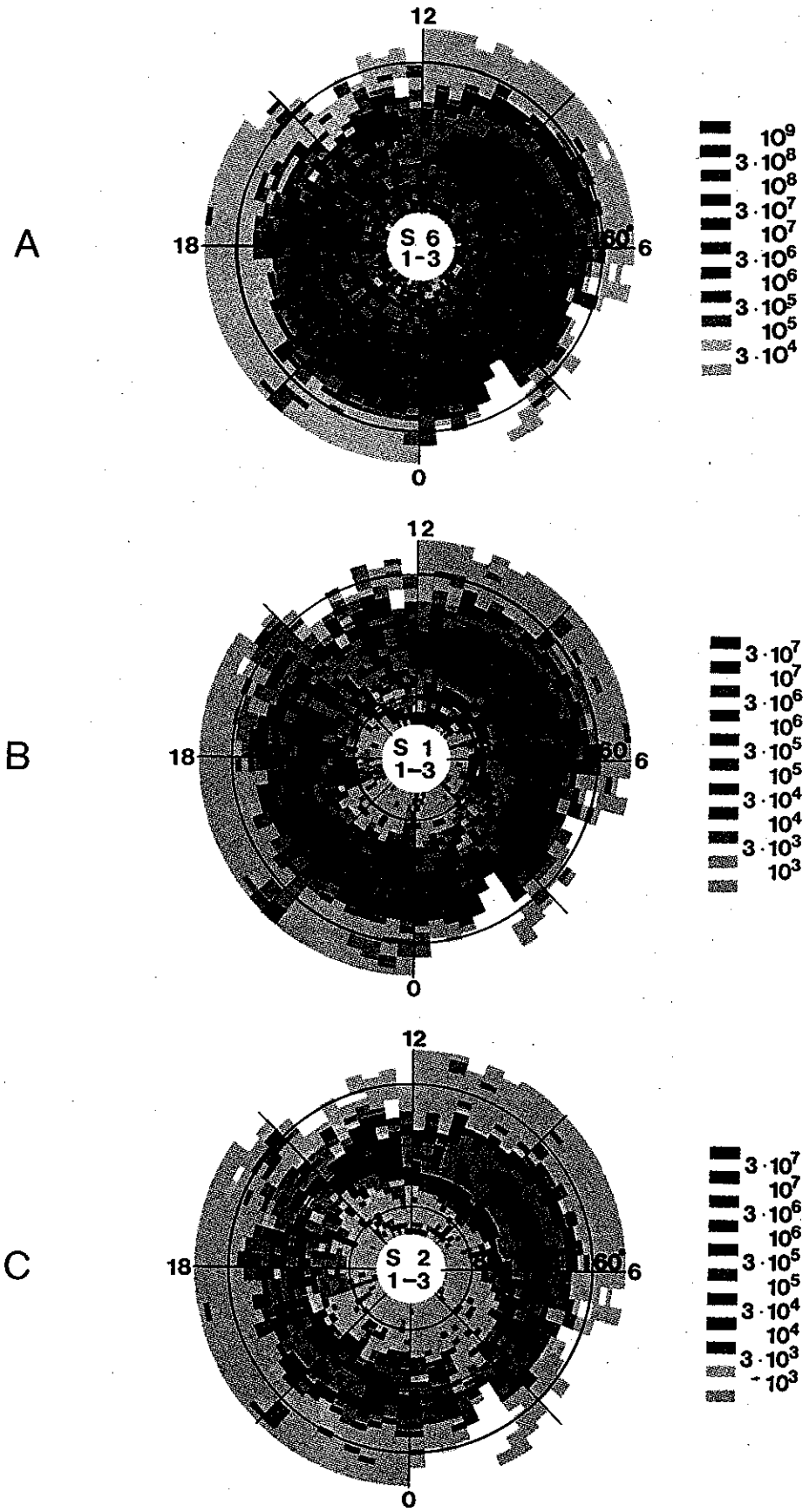


Figure 3. Average fluxes of precipitated electrons of 1, 6 and 13 keV energies for fairly low activity level ($Kp = 1$–3) as measured during a 20 months period from October 1968 to April 1970 with the ESRO 1A satellite. The colour code is given in the figure. (a) is for 1 keV, (b) 6 keV, (c) 13 keV. 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 $\approx$1, $\approx$6, and $\approx$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 3–5° 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

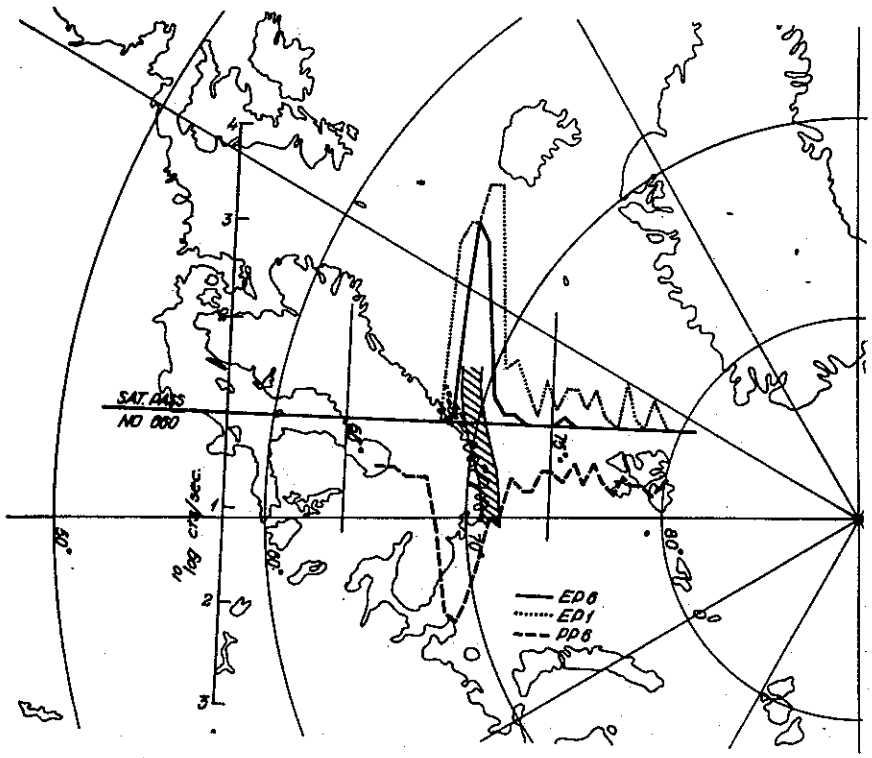


Figure 4. An auroral form measured by all-sky cameras has been projected to the 100 km level for the lower border and to the 130 km level for the upper border and drawn on a map (hatched area). Superimposed are the passage of the satellite and the particle measurements. The latitude values along the 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 $\Lambda = 80°$ 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

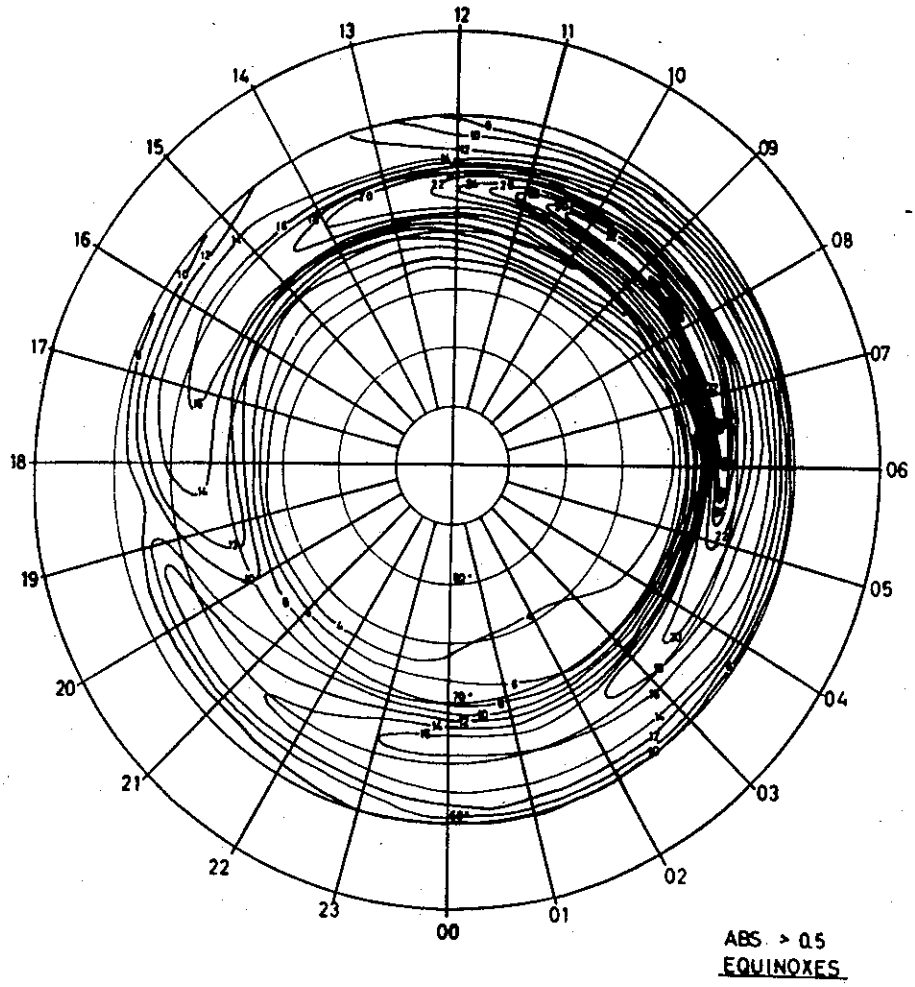


Figure 5. Percentage of total time with absorption exceeding 0.5 dB at different geomagnetic time and latitude during equinox months (Oct. 1958 – Feb./Apr. 1959), (Holt, 1963).

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 (Kp = 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 Kp 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

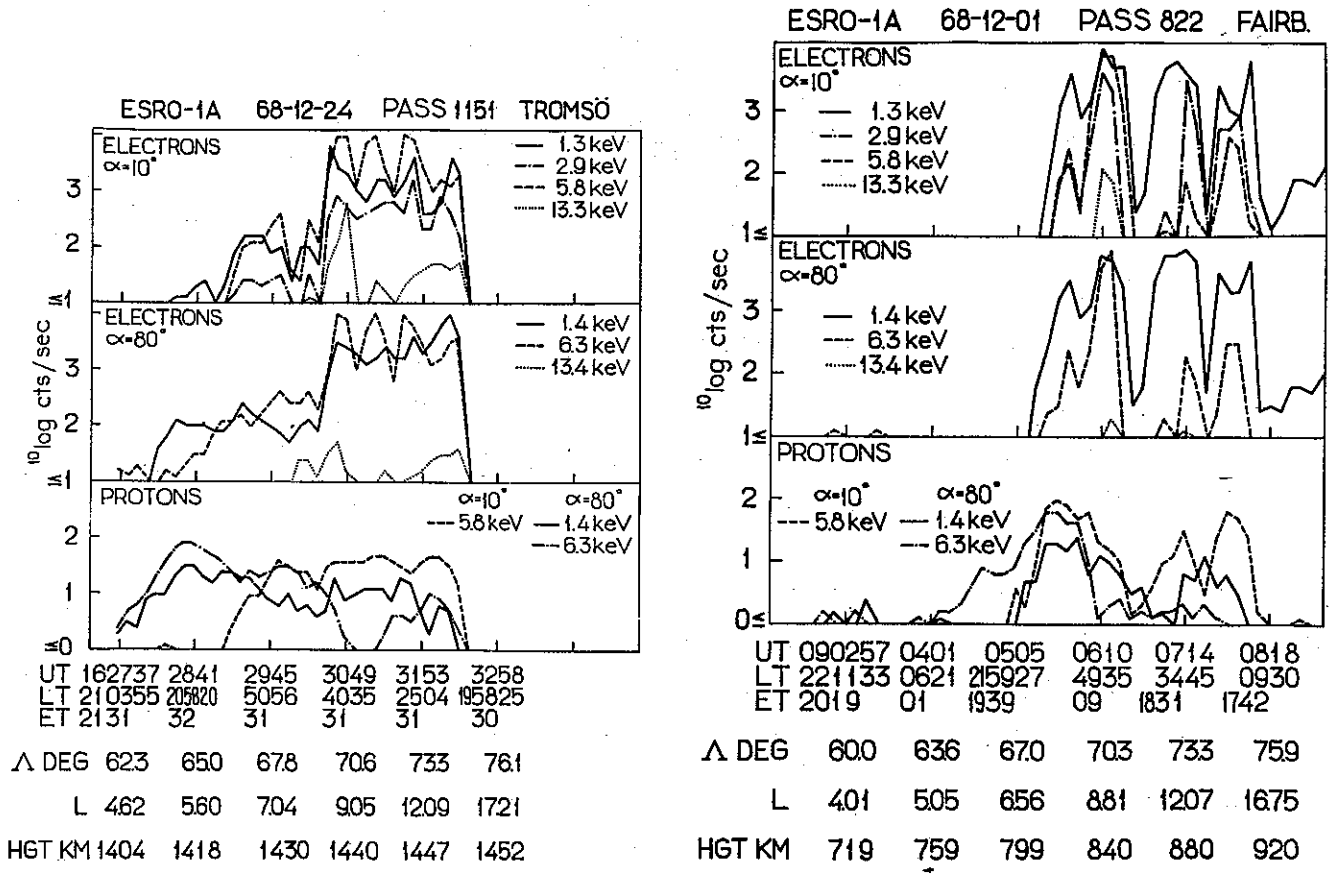


Figure 7. Measurements on board the satellite ESRO 1A of electron and proton fluxes at various energies for two different pitch angles, $\alpha = 10°$ and 80°. The curves connect 8 s averages for the count rates of the detectors. The time figures give hour, minute and second. LT means geomagnetic local time (dipole time) and ET excentric dipole time (in hours and minutes). Λ 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 $\pm 8°$. The telemetry station which received the data is given at the top right (Hultqvist et al., 1971).

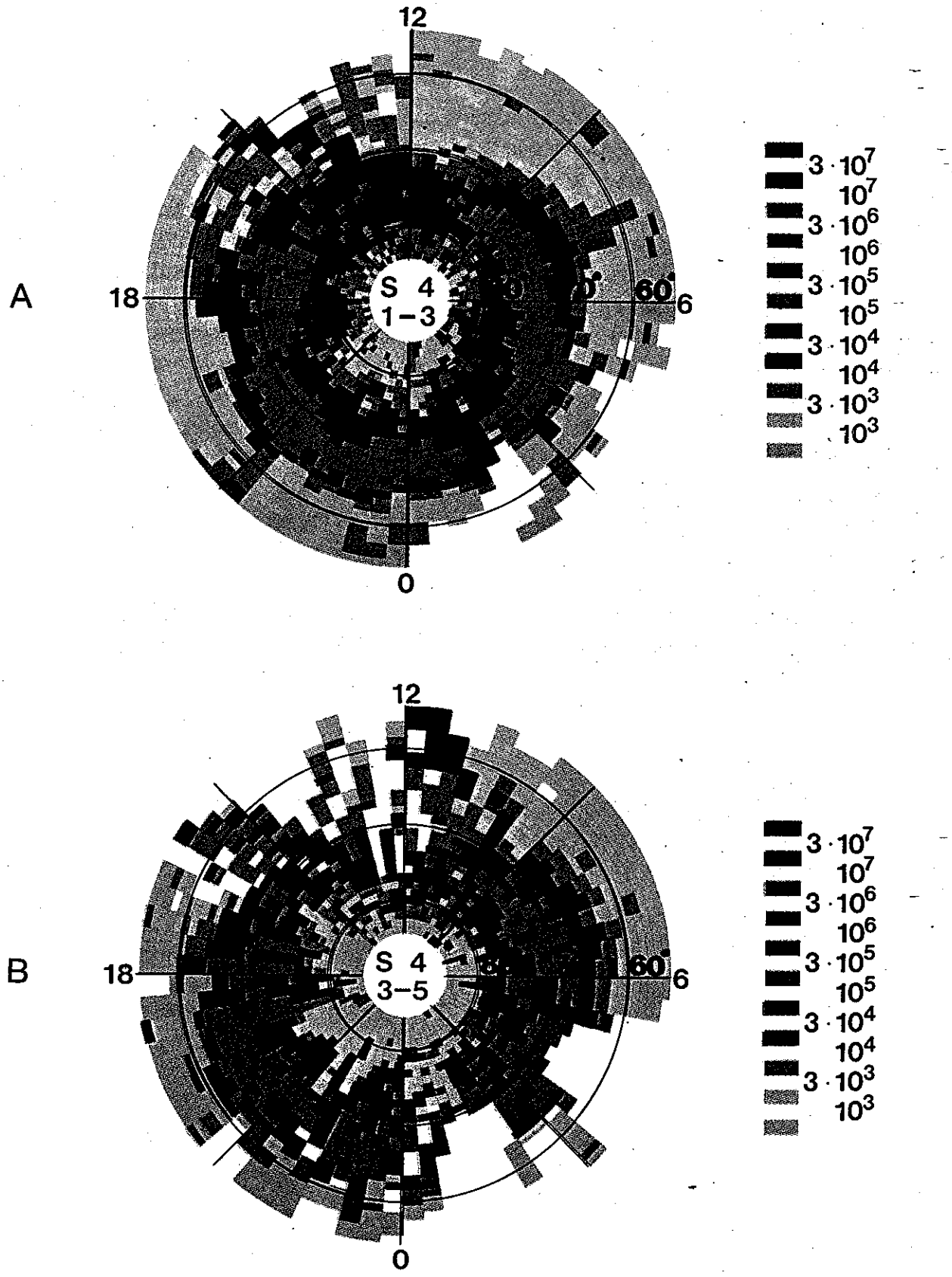


Figure 6. Average fluxes of precipitated (pitch angle 10°) 6 keV protons as measured during a 20 months period from October 1968 to April 1970 with the ESRO 1A satellite. (a) is for $Kp = 1$–3; (b) $Kp = 3$–5. The coordinates are corrected geomagnetic (invariant) latitude and excentric dipole time (Riedler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

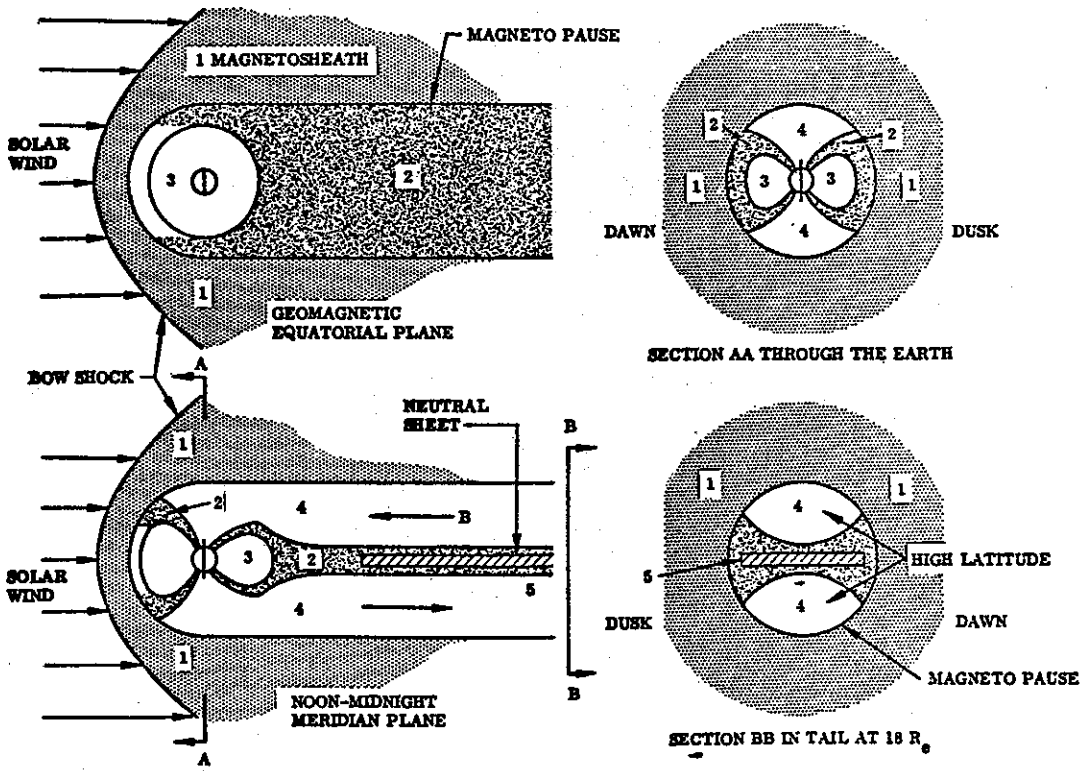


Figure 9. The spatial extent of the plasma sheet and magnetotail. The regions are (1) magnetosheath, (2) plasma sheet, (3) trapping region, (4) high latitude region, and (5) neutral sheet (McCormac et al., 1971).

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 $10^5 - 10^7$ cm^{-2} s^{-1} $sr^{-1}keV^{-1}$.

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° Λ, 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 erg cm^{-2} 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-

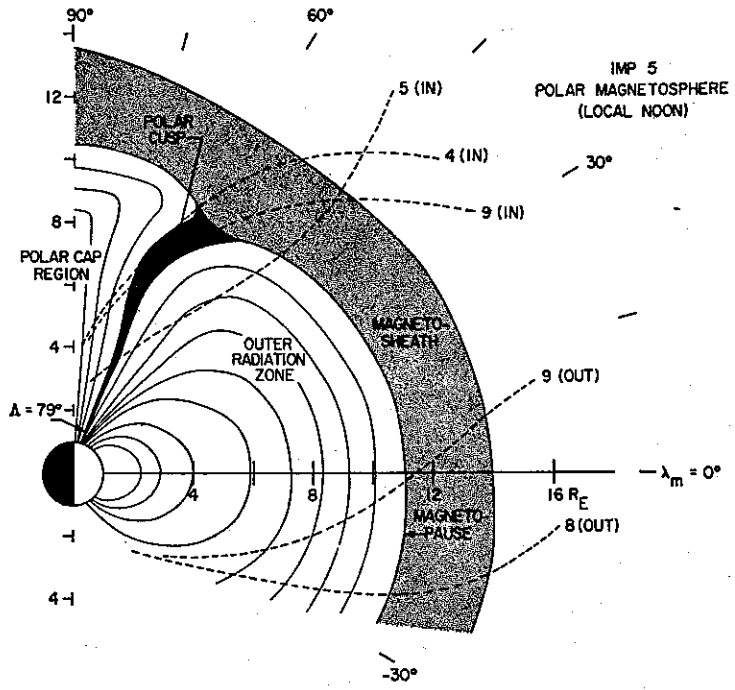


Figure 10. A diagram showing the geometry and location of the polar cusp within the polar magnetosphere in the noon meridional plane during periods of relative magnetic quiescence. The coordinates are geocentric radial distance, R_E, and dipole magnetic latitude, λ_m. The polar cusp intersects the auroral zone at $\Lambda \approx 79°$. 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 participitate 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.

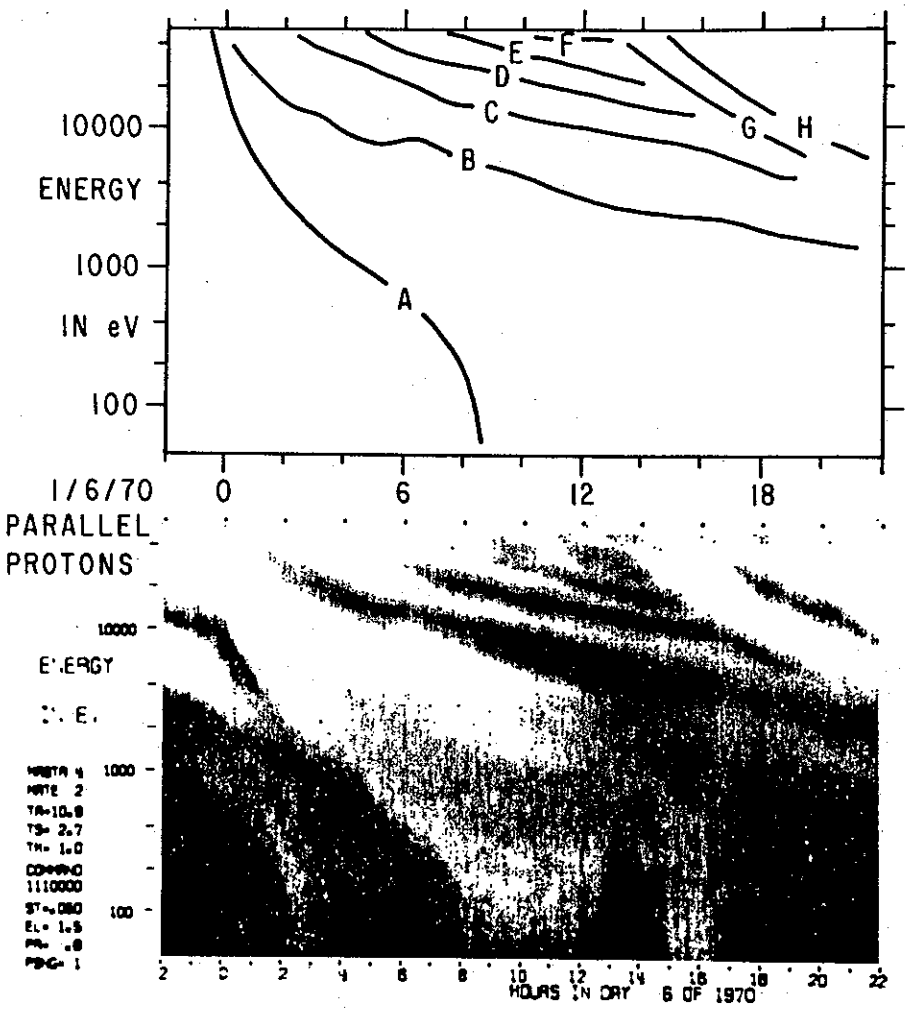


Figure 11. A spectrogram of the proton differential energy flux during 6 Jan. 1970. The pitch angle is 0°. The intensity of the energy flux is given by the brightness of the diagram. The upper half identifies how some of the maxima in the energy spectrum vary with time (DeForest and McIlwain, 1971).

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:

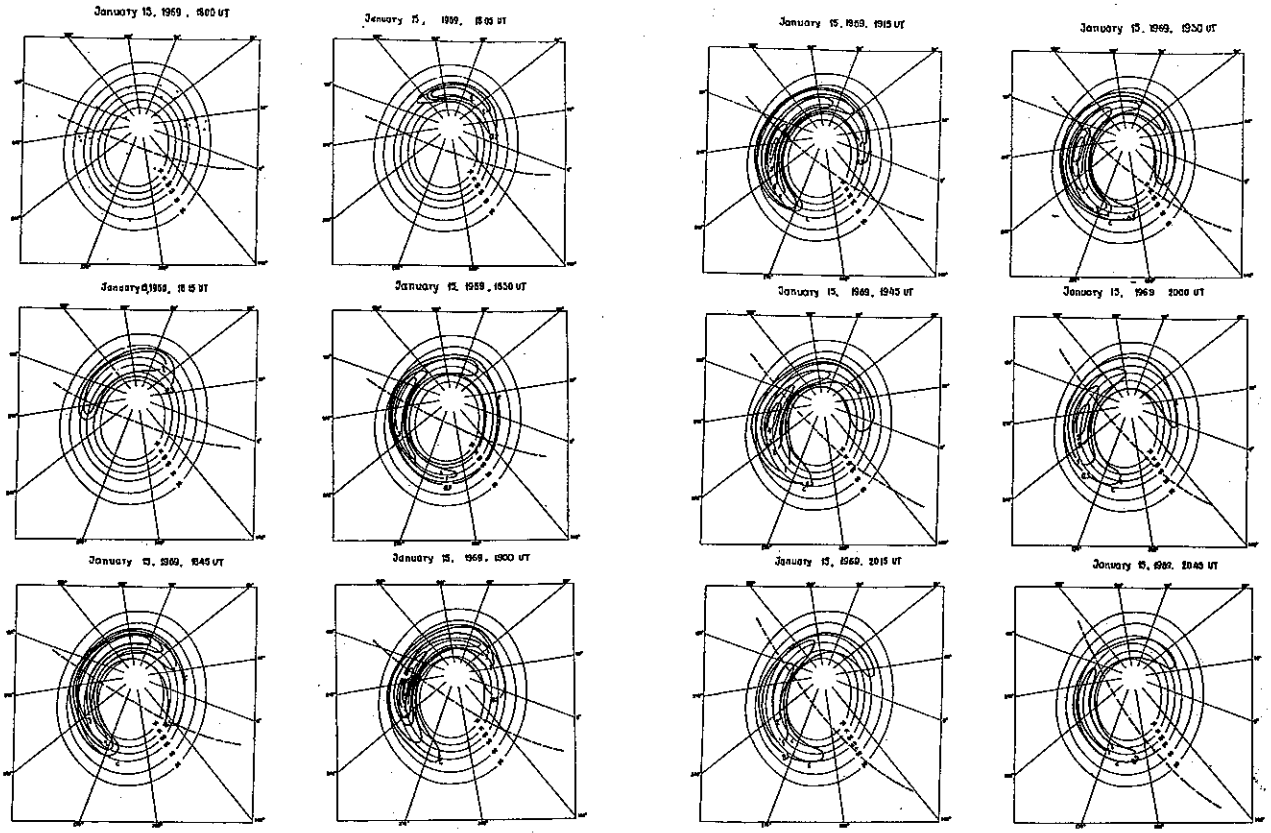


Figure 12. The temporal development of the precipitation pattern for auroral electrons of energies above, say 15 keV as observed by riometers in auroral and polar cap regions. The locations of the riometers which have delivered data to the maps are shown on the first map. The coordinates are geographic 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

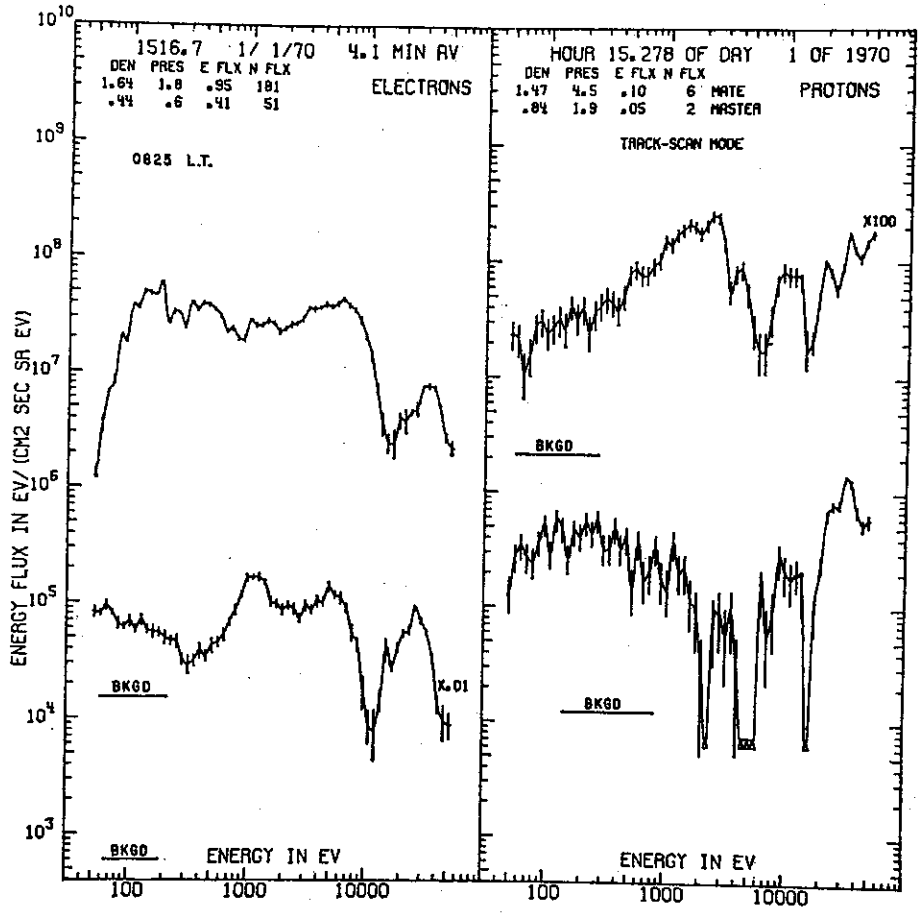


Figure 13. An example of differential energy flux spectra for electrons and protons. The upper curves are for pitch angles close to 90° and the lower ones for pitch angles around 0°. The parallel-to-B electron energy flux has been multiplied by 10^{-2} and the perpendicular-to-B proton energy fluxes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

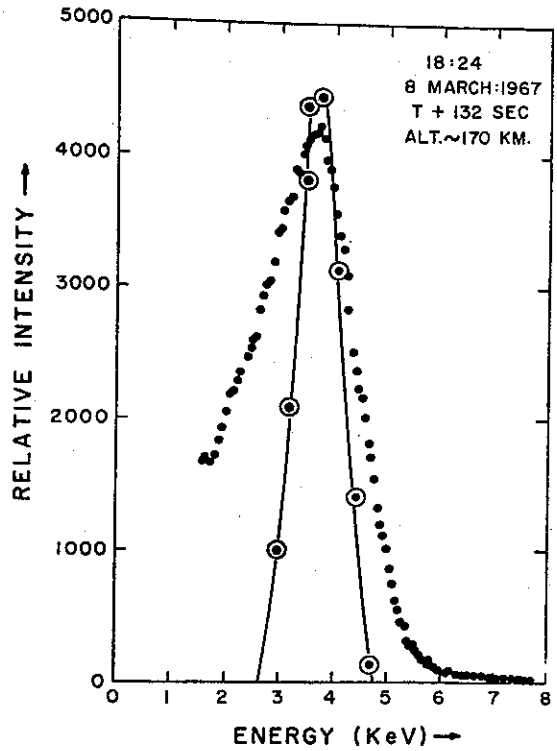


Figure 14. Nearly monoenergetic electron spectrum measured in a pre-breakup aurora of about IBC 1 intensity. The full line gives the response of the detector to a mono-energetic flux. The slope on the high-energy side corresponds to an equivalent *e*-folding energy of the order of a few tenths keV. Hardly 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

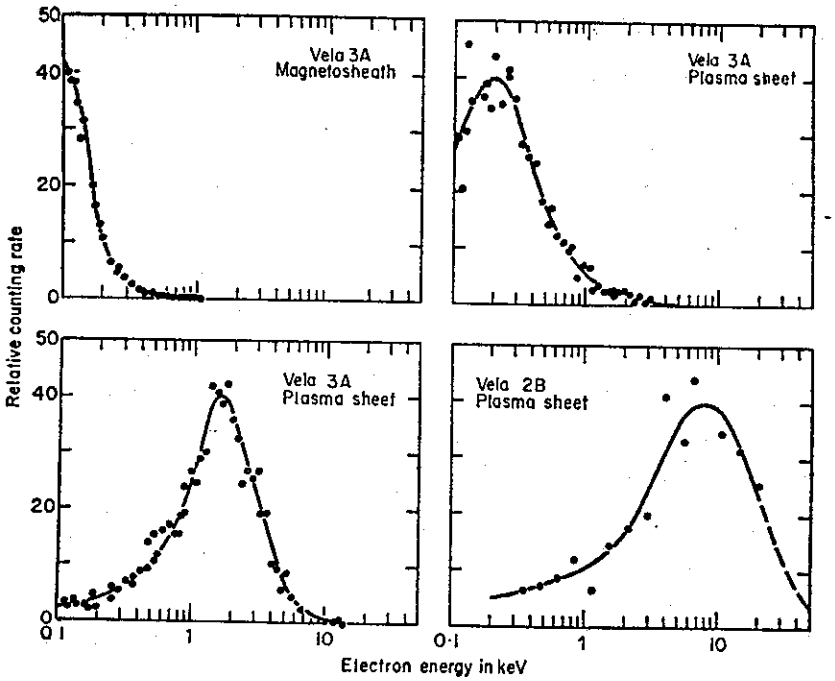


Figure 15. Electron energy spectra observed in the magnetosheath and in the plasma sheet by the Vela satellites (Bame et al., 1967).

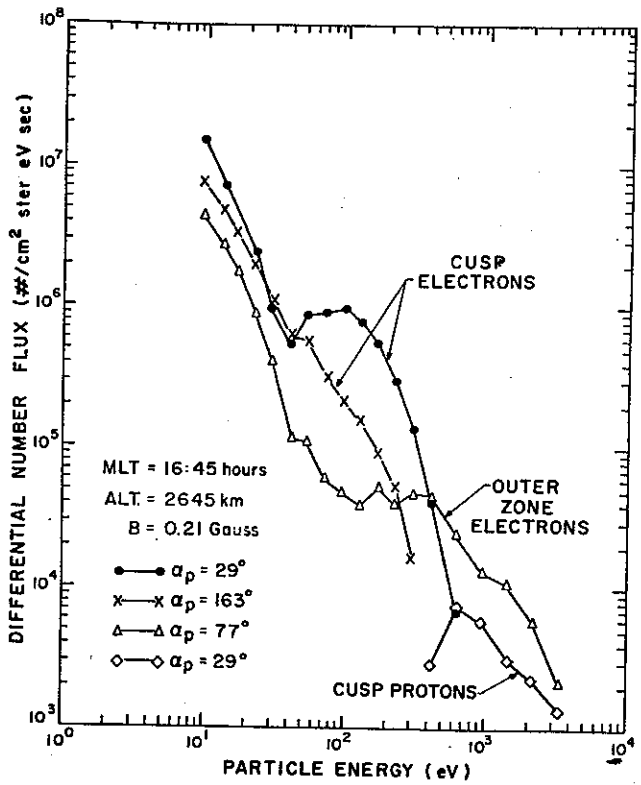


Figure 16. Electron and proton differential spectra from the polar cusp and from the outer radiation belt (Winningham, 1971).

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.

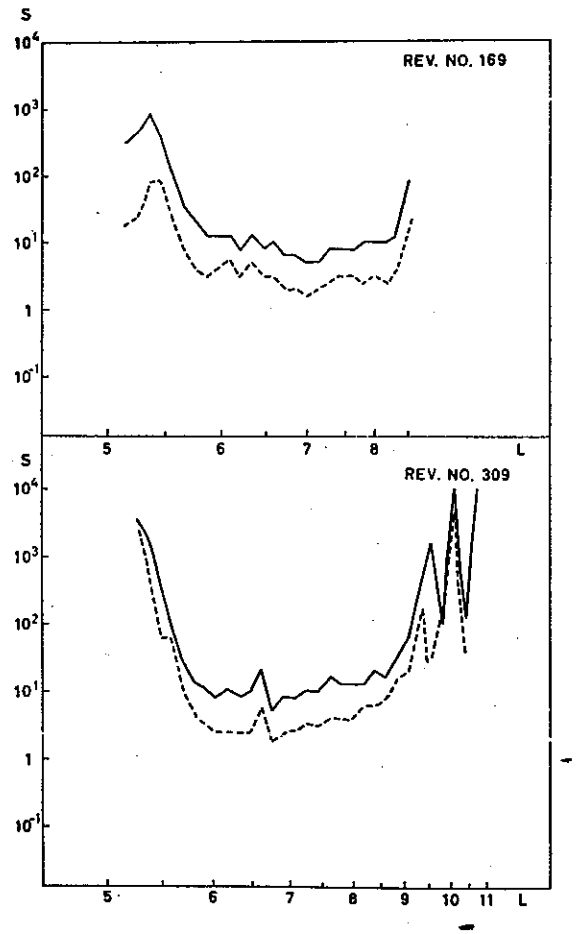


Figure 17. The steepness, *S*, defined as the ratio between electron fluxes at 1 and 6 keV, as function of *L* value for two passes of the satellite ESRO 1A. Notice the 'soft edge' at low *L* values. Solid curves correspond to precipitated electrons, broken to trapped electrons (Liszka et al., 1970).

On the average it seems, however, that the spectrum hardens below a few tens of keV when the intensity increases and it becomes softer above several tens of keV. Sometimes the point of transition from hardening to softening may be around 100 keV or above. In the range 1–100 keV the irregular time variations often seen in auroral measurements are generally stronger, percentage-wise, the higher the energy. This is related to the mentioned hardening of the energy spectrum when the precipitation is intensified.

If there is an electron energy below which the spectrum hardens when the intensity of precipitation increases and above which it softens, one would expect to see a peak in the spectrum in periods of flux intensification. Such an observation has been reported.

4. Pitch-angle distribution

The early high-inclination low orbit satellites as well as the early rocket measurements of auroral particles, mainly above 40 keV, all gave a fairly consistent picture of pitch-angle distributions peaked at about 90° in quiet or stable periods, and approaching isotropy over the upper hemisphere when the intensity of precipitation increased. This is illustrated in Fig. 18. Since the middle of the sixties, when the lowest part of the energy spectrum for the

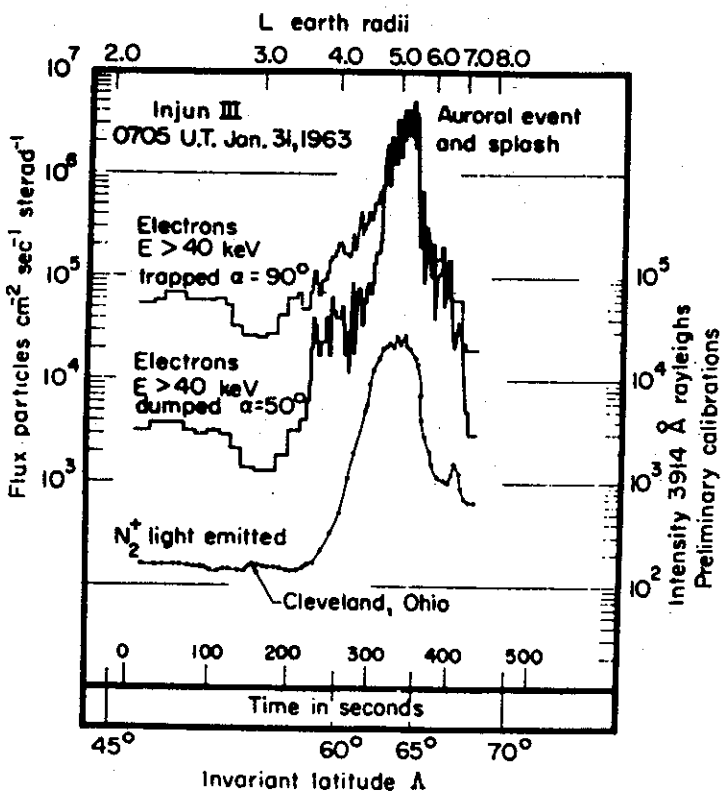


Figure 18. Data from a northbound pass of Injun 3 over North America which show simultaneous detection of an aurora, of the precipitated electrons (>40 keV, pitch angle ~50°), and of trapped electrons. Note the approach to isotropy of the particle flux over the aurora (O'Brian and Taylor, 1964).

auroral particles has also been measured by means of new secondary emission devices, an increasing amount of evidence has been produced for the existence sometimes of pitch-angle distributions peaked at zero pitch angle. This is true for the energy range from a small fraction of 1 keV up to energies of the order of 100 keV. For the higher energies the observations of field-aligned anisotropy are still quite few and have mostly been associated with intensifications in the measured electron flux.

For keV electrons and protons, Fig. 7 illustrates situations seen in the upper atmosphere. The detectors for the same energies at pitch angle 10° and 80° were identical (within some ten percent), so the count rates in Fig. 7 directly provide us with data for computing the ratio of fluxes in the loss cone and approximately perpendicular to it. As can be seen, the cases in Fig. 7 show a depleted loss cone type of anisotropy at the equatorward edge of the particle precipitation zone. This is true for keV electrons as well as for protons. The width of the region with this type of anisotropy may sometimes amount to several latitude degrees for the protons, but is mostly only a fraction of a degree for the electrons. For the protons, the depleted loss cone anisotropy at the equatorward edge is always found on the nightside but practically never on the dayside. The electrons neither show the same consistency nor the large value of the anisotropy as the auroral protons.

Going polewards in the ionosphere through the particle precipitation zone, one reaches approximate isotropy for the keV particles after the above-mentioned anisotropy region has been passed. The pitch-angle distribution in the large majority of cases stays approximately isotropic through the rest of the precipitation zone. That this is not always so is shown by Fig. 7. In fact the cases presented in this figure were chosen to illustrate another type of anisotropy, namely one with the peak flux along the geomagnetic fields lines. As can be seen in all examples in Fig. 7, the protons develop this field-aligned anisotropy in the poleward part of the precipitation zone. In fact it is most frequently seen in a fairly narrow region at the poleward edge. The ratio between fluxes at 10° pitch angle and at 80° may in extreme cases reach values of the order of 100. The observed results of this type of anisotropy may be summarized as follows:

- The anisotropy occurs fairly frequently for 6 keV positive ions. Figure 19 shows the distribution of observations of anisotropies by a factor >10. This kind of anisotropy occurs at all local times.
- It may cover large areas and sometimes be present for hours.
- It occurs in the same regions as strong auroral electron fluxes.

Also for keV electrons, pitch angles with the peak flux at zero pitch angle are sometimes seen. It is, however, a less frequent phenomenon than for the positive ions. The distribution in EDT and invariant latitude of the occurrence is similar to that for protons shown in Fig. 19. Examples of the variation of the ratio of the directional flux at 10° to the flux at 90° with respect to the magnetic field lines (called isotropy factor, Φ) are shown in Fig. 20. The more or less stable value of Φ for 1 keV electrons is not significantly different from unity. Different detectors were used in obtaining the ratio, and deviations from unity by less than a factor two should not be taken too seriously. However, example b in Fig. 20 shows that sometimes

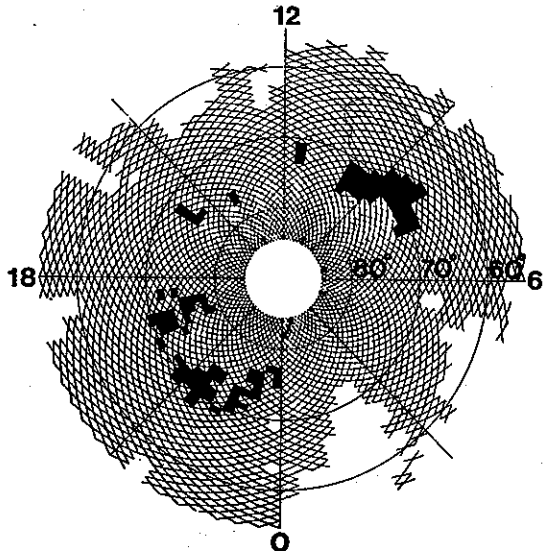


Figure 19. The occurrence probability of field-aligned anisotropy (parallel to perpendicular fluxes $\geq$10) for 6 keV per unit charge positive ions. In the black areas the field aligned anisotropy has been observed in $\geq$5% of all measurements. In the shaded regions measurements have been made but an occurrence frequency of $<$5% has been found. For the white regions no data are available (Borg, 1971).

the flux in the loss cone may be very much larger than that perpendicular to it. Values of this ratio of up to 100 have been seen. Such high values are, however, quite rare. The occurrence probability seems to increase with decreasing energy.

Leaving now the upper atmosphere in the auroral zone for the magnetosphere, it should be underlined that only the particles in the equatorial plane within a few degrees from the magnetic field lines reach the ionosphere. All the others mirror at greater altitudes. The field-aligned anisotropy described above requires, if it is produced in the tail, that there should be a very strong structure in the pitch-angle distribution within the very small loss cone out there. No indications of this have been found experimentally and it seems highly unlikely that the source of the anisotropy is to be found in the tail. We will discuss this further in a later section.

It has been mentioned earlier that the pitch-angle distribution of both electrons and protons in the plasma sheet is approximately isotropic during quiet or stable conditions. When a magnetospheric substorm occurs, field-aligned fluxes are seen there. This does not, however, mean that this anisotropy is so strong that there is strong variation even within a few degrees from the magnetic field lines. On the contrary, observations indicate that the variation of directional flux within a few degrees is quite small. Thus, when these particles reach the ionosphere, they appear with approximately isotropic pitch-angle distribution (unless some special physical processes influence the particles on their way down to the atmosphere).

Also at the geostationary orbit all kinds of anisotropies have been observed – from those with a strong peak perpendicular to the field lines to those with the maximum flux along the magnetic field lines. At the time of writing, there seems to be no clear understanding of what causes these anisotropies, although hypotheses exist.

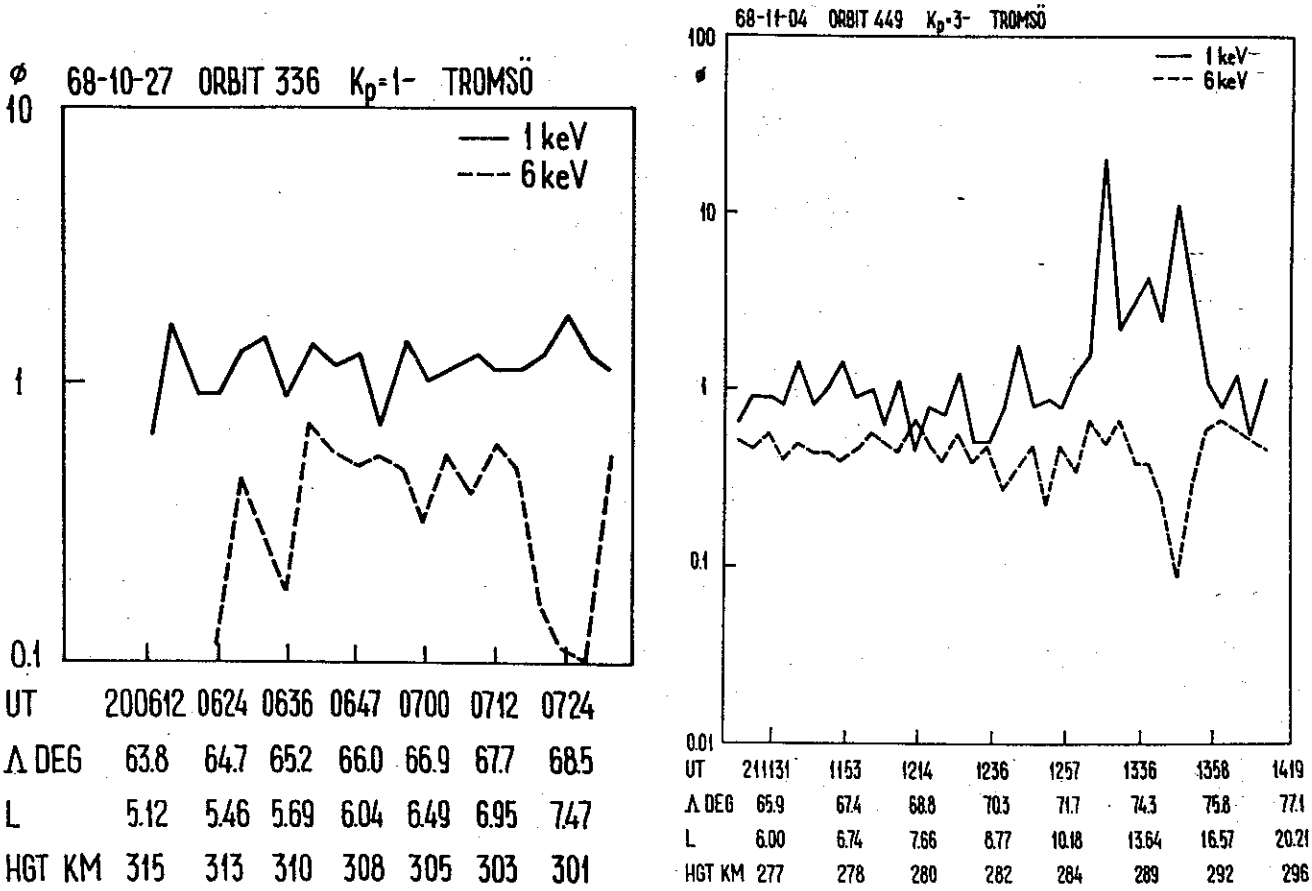


Figure 20. The ratio Φ of the directional electron flux at 10° pitch angle to the flux at 80° observed during four different passes of the satellite ESRO 1A: (a) was obtained during quiet conditions, $Kp = 1$; in (b) a broad field-aligned peak in the 1 keV electron pitch-angle distribution can be seen; $Kp = 3$— (Holmgren et al., 1970).

In the polar cusps the pitch-angle distribution has been found to be isotropic in the upper atmosphere. For greater distances from the Earth some observations of pitch-angle distributions peaked along the magnetic field lines have been reported.

Pitch-angle distributions varying with particle energy have been observed in a few cases. The variation seen has mainly been such that loss cone-depletion has varied with energy, being lower the lower the energy.

5. Acceleration and precipitation mechanisms

We will not present here any complete discussion of all physical mechanisms that may play a role in the acceleration of the solar wind particles to the auroral particle energies and in precipitating them into the atmosphere. Only a few mechanisms which we now think play a major role in these processes will be briefly described. Those are the following: field line merging, Fermi acceleration, betatron acceleration, acceleration in the electric field associated

with the convection of the plasma inside the magnetosphere, electrostatic fields along the geomagnetic field lines, and diffusion of energetic particles in turbulent wave fields.

That auroral particles may be accelerated at neutral points of the geomagnetic field was suggested more than 20 years ago, but it was the discovery of the neutral sheet of the magnetosphere in the middle of the sixties that first placed acceleration of auroral particles in neutral regions of the geomagnetic field at the focus of interest. The reason for the importance of neutral regions is that with no magnetic field to deflect the particle motion and cause drift perpendicular to both the electric and magnetic field, an electric field will be free to accelerate the particles.

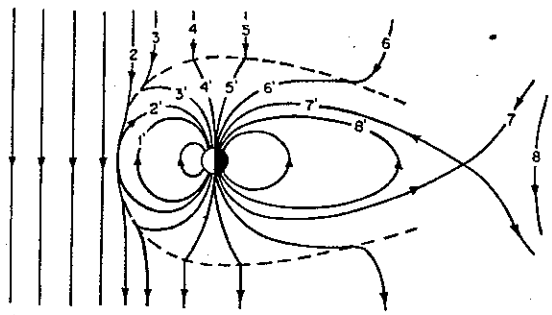


Figure 21. Schematic diagram showing the flow of plasma induced into the magnetosphere because of field-line reconnection at the boundary. The plain numbers show the successive positions of an interplanetary magnetic field line, and the primed numbers show the corresponding successive positions of a geomagnetic field line. Reconnection occurs at the contacts of 2 and 2' and of 7 and 7' (Levy et al., 1963).

Field-line merging (reconnection) is thought to take place in the tail at the neutral sheet. According to Dungey it is a part in the general convection scheme of the magnetosphere, illustrated in Fig. 21. The geomagnetic field lines connect to the interplanetary field lines at a neutral point on the sunward side of the magnetopause. An efficient form of interaction between the solar plasma and the magnetic field is produced in this way. The field lines are moved with the solar plasma outside the magnetospause, whereas inside the magnetosphere they move toward the neutral sheet with a velocity of, at the most, 10% of the Alfvén velocity. When they reach the neutral sheet they may reconnect through the effect of currents in this sheet and convect back toward the dayside.

The reconnection of the field lines through the neutral sheet involves annihilation of magnetic field energy; it is most likely that most of this energy is transferred into kinetic energy of the plasma. A plasma sheet, containing charged particles with energies similar to those characteristic of auroral particles exists, as mentioned earlier, around the neutral sheet. The pressure of this plasma prevents the neutral sheet from collapsing under the magnetic pressure around it. It has been shown that the field-line merging can probably produce enough power to account for an aurora. With reasonable assumptions about particle density and tail field outside the sheet (2 cm^{-3} and 30 γ, respectively) one arrives at an estimated average particle

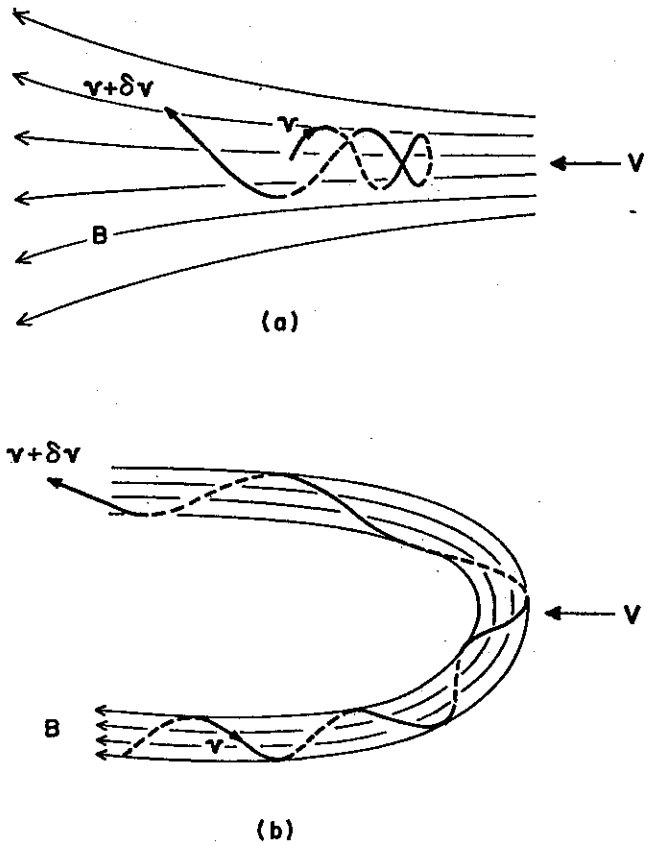


Figure 22. Fermi acceleration of charged particles: Type (a) in which the gyrating particle encounters a converging magnetic field moving with velocity V; Type (b) in which the particle moving in a nearly constant field encounters a region of curved geometry moving with velocity V. Type (b) occurs on closed extended field lines in the magnetotail.

energy of 1 keV and an electron flux of $3 \cdot 10^8$ cm^{-2}s^{-1}sr^{-1}. These values are in good agreement with observations.

The inward convection of geomagnetic field lines in the tail is associated with Fermi and betatron acceleration. This convection is driven by the dawn-dusk electric field caused by the interaction of the solar wind with the geomagnetic field.

The Fermi mechanism can be considered as reflection of a particle from a moving mirror. Two different cases can be distinguished, a and b in Fig. 22. They are mathematically equivalent. For the auroral particle acceleration in the tail, type b is applicable in the first instance.

A rough estimate of the amount of acceleration can be obtained from the constancy of the second adiabatic invariant, $\int v_{||} dl = const$, which we can write $<v_{||}> \cdot l = const$. ($v_{||}$ is the velocity component parallel to the geomagnetic field lines; l is the length of the field line between mirror points). When a volume element of the plasma sheet moves in from say 30 R_E to 7 R_E, the parallel velocity increases by a factor of about 63/17 and the parallel energy by $(63/17)^2 = 14$.

The transverse (betatron) acceleration caused by the same motion is obtained from the

first adiabatic invariant, $\mu = E_\perp/B = const.$ ($E_\perp$ is the energy associated with the motion perpendicular to the field lines). For observed values of B at 30 and 7 R_E in the tail of $\sim 14\,\gamma$ and $\sim 30\,\gamma$, respectively, the ratio of $E_\perp$ at 7 R_E is about twice as large as at 30 R_E. Thus, the Fermi mechanism dominates over the betatron mechanism in the tail at distances greater then ~ 7 R_E.

We have here referred to the accelerations associated with the earthward motion of the tail plasma as Fermi and betatron acceleration. We could also say that it is due to the electric field associated with the convection of the plasma in the magnetosphere. This electric field is caused by the motion of the solar wind relative to the Earth. As seen from the Earth there is an electric field of the order of $\mathbf{v} \times \mathbf{B}$ in the solar wind, where $\mathbf{v}$ is the solar wind velocity and $\mathbf{B}$ is the magnetic field in it. Assuming that the magnetosphere has a diameter of $20R_E$ there is a potential difference of about 40 keV between the dawn and dusk sides of the magnetosphere with the electric field pointing from dawn to dusk. If this electric field penetrates the entire magnetosphere, as it seems to do, it produces, together with the geomagnetic field, a drift motion which in the equatorial plane of the tail is directed towards the Earth. Because of the inhomogeneity of the magnetic field, the more the closer to the Earth, the drift motion is not perpendicular to the electric field, but the particle motion has a component parallel (or antiparallel) to the electric field and thereby the particle gains or loses energy. It is obvious that the potential electric field discussed here cannot increase the particle energy by more than the potential difference over the magnetosphere.

In addition to the potential electric field there may be electric fields for which the curl is not zero but is given by the $\partial B/\partial t$. The importance of such electric fields may possibly be great in certain situations, but very little of a quantitative nature is known about them.

The electric field just discussed is perpendicular to the geomagnetic field lines. In recent years a growing amount of evidence for the existence of electric fields also parallel to the magnetic field lines has accumulated. Field-aligned electric currents and the field-aligned anisotropy of auroral particles described earlier in this chapter are most important in this respect. Potential drops along the magnetic field lines may be created if the resistivity for field-aligned electric currents, which normally is extremely small in the magnetosphere and ionosphere, is strongly increased. This may happen if wave-particle interactions occur, whereby the free motion of charged particles along the field lines is prohibited.

So-called potential double layers, which are well known from laboratory plasma, may possibly exist under certain conditions. They are very thin – of the order of a few Debye lengths – and the potential drop over them may be several kilovolts.

The observed field-aligned anisotropy of auroral protons may be understood as an effect of a potential difference between the hot magnetospheric plasma and the cold ionosphere which, under certain conditions, is determined by the temperature of the hot plasma and, therefore, may amount to several kilovolts. The basic cause for the production of such a potential difference between two plasma of different temperature which are brought in contact with each other is the existence of a higher flux of electrons (ions may be neglected if their temperature is not very much higher than that of the electrons) from the hot plasma to the cold than in the opposite direction. A net negative charge is then built up in the cold

plasma and a stationary state is achieved first when the electrostatic field due to the charge separation has grown to such a value that the electron fluxes in the two directions are equal. This basic mechanism is the same irrespective of whether the electrons propagate between the two plasma in a diffusion process of some kind or move in simple ballistic orbits without any interaction with each other or with field irregularities.

Whereas ordinary collisions play some role in the ionosphere out to distances of a thousand kms from the Earth's surface, the plasma in the plasma sheet is a collisionless plasma, if one considers ordinary particle encounters. It seems clear, however, that the energetic particles interact with waves efficiently enough to create an effective diffusion process.

In order to explain the observation of different locations of the inner boundary of the energetic electron flux in the equatorial plane (inner boundary of plasma sheet) and the inner boundary of the convective motion of the plasma (supposed to be the plasma pause), one has to explain how the convecting field tubes penetrating through the plasma sheet boundary are emptied of its electron content. Since appreciable electron precipitation is known to take place at the inner plasma sheet boundary it seems reasonable to assume that precipitation into the atmosphere removes all the energetic electrons in the field tubes. It is supposed that so-called strong diffusion into the loss cone is the cause of the precipitation. If during the flow of a field tube toward the Earth the life time of the electrons is short compared to the time scale of the flow, many particles will be precipitated and the tube will flow onward depleted of the precipitated particles. The flow time, T_F, and the life time T_M for strong pitch-angle diffusion, have been estimated for a highly simplified situation, which, however, is likely to contain the most important features of the real case. One has found that

$$T_F \approx 10^7/L^2\varphi \qquad \text{[s]},$$

$$T_M^+ \approx \left(\frac{m_i}{m_e}\right)^{1/2} \frac{L^4}{E_i^{1/2}} \qquad \text{[s] for ions}$$

and $$T_M^- \approx \frac{L^5}{E_e^{1/2}} \qquad \text{[s] for electrons}$$

where L is the McIlwain L-parameter, φ is the electric potential drop over the magnetosphere corresponding to the convection, in kV, $E_{i,e}$ is the energy of ions and electrons, respectively, in keV and m_i, m_e are the ion and electron masses.

Since $T_F \propto L^{-2}$ and $T_M^{\pm} \propto L^4$ one expects $T_M^{\pm} > T_F$ for sufficiently large L. Thus, at large distances, precipitation will not affect the density in the field tube significantly. The point at which $T_M^{\pm} \approx T_F$ determines the boundary, since beyond this point particles are depleted very effectively by precipitation before the flow can travel very far.

References

Bame, S. J., J. R. Asbridge, H. E. Felthauser, E. W. Hones and I. B. Strong, 1967: J. Geophys. Res., **72**, 113.

Berkey, F. T., V. M. Driatskiy, K. Henriksen, D. H. Jelly, T. I. Shchuka, A. Theander and J. Yliniemi, 1971: Temporal development of the geographical distribution of auroral absorption for 30 substorms in each of IQSY (1964–65) and IASY (1969), World Data Center A, Data Report.

Borg, H., 1971: Personal communication.

DeForest, S. E. and C. E. McIlwain, 1971: J. Geophys. Res., **76**, 3587.

Eather, R. H. and S. B. Mende, 1971: In 'Magnetosphere-Ionosphere Interactions', (Ed. K. Folkestad), Universitetsforlaget, Oslo.

Evans, D. S., 1968: In 'The Birkeland Symposium on Auroral and Magnetic Storms', (Eds. A. Egeland and J. Holtet), CNRS, Paris, 333.

Frank, L. A., 1971: J. Geophys. Res **76**, 5202.

Gustafsson, G., 1970: Phys. Norv., **4**, 113.

Hartz, T. R., 1971: In 'The Radiating Atmosphere' (Ed. B. M. McCormac), D. Reidel, Dordrecht, 225.

Holmgren, L.-Å., P. Christophersen and W. Riedler, 1970: Phys. Norv., **4**, 85.

Holt, O., 1963: Norw. Defence Res. Est., Rept. 46.

Hultqvist, B., H. Borg, W. Riedler and P. Christophersen, 1971: Planet. Space Sci., **19**, 279.

Levy, R. H., H. E. Petschek and G. L. Siscoe, 1963: AIAA J., **2**, 2065.

Liszka, L., H. Borg and W. Riedler, 1970: Phys. Norv., **4**, 121.

McCormac, B., J. E. Evans, R. D. Sears and R. Berg, 1971: Paper presented at Summer Advanced Institute on Earth's Particles and Fields, Cortina, Aug.–Sept., 1971.

O'Brien, B. J. and H. Taylor, 1964: J. Geophys. Res., **69**, 45.

Riedler, W. and H. Borg, 1971: Paper presented at the COSPAR General Assembly in Seattle, May, 1971.

Winningham, J. D., 1971: Paper presented at Summer Advanced Institute on Earth's Particles and Fields, Cortina, Aug.–Sept., 1971.

AURORA

BY YA. I. FELDSTEIN
INSTITUTE OF TERRESTRIAL
MAGNETISM, IONOSPHERE AND RADIOWAVE PROPAGATION,
MOSCOW, USSR

Abstract

The paper gives a review of recent investigations on auroral space-time regularities and their relation to electron fluxes responsible for the auroral luminescence. In Fig. 3 the position of precipitation regions together with the auroral oval is shown. The following is under discussion: 1) A connection between the auroral oval and the large-scale structure of the magnetosphere (Fig. 4); 2) Variations of inner magnetosphere dimensions (Fig. 5); 3) Magnetic fluxes responsible for magnetic tail formation (Fig. 6) at different phases of the magnetic substorm. Figure 7 represents a general scheme of auroral substorm development in the northern high-latitude region, consisting of three phases, i.e. the creation, expansion, and recovery phase.

On the basis of auroral data, the average energy flux of precipitating electrons in different sectors of the auroral oval (Fig. 8), as well as the relation of aurora with the region of transverse magnetic disturbances (Fig. 9), is obtained.

1. Introduction

Detailed descriptions of morphological characteristics of auroras such as space-time regularities, movements, and geometrical shapes were given in monographs by Störmer (1955) and Harang (1951); these monographs brought together the results of the early classical investigations of Aurora Borealis by Norwegian scientists. A number of regularities in auroral behaviour were revealed by numerous determinations of auroral altitudes by parallactic observations (Egeland and Omholt, 1967). These various works, containing the results of comprehensive observations on auroral morphology prior to the IGY (1957–1958), have served as valuable manuals up to the present time.

The investigation of aurora has been greatly stimulated by space research. Auroral observation provides unique determination of the regions where charged particles penetrate into the upper atmosphere and it allows study of dynamics on a planetary scale. Auroral observation gives information about processes far out from the Earth in the whole domain through which the geomagnetic field lines extend.

2. Auroral morphology

The notion 'auroral morphology' now encompasses a broader field than shape and distribution of the visual aurora; the broadening of the scope grew with the various space researches beginning with the IGY. At present auroral morphology includes: 1) study of the kind, energy spectrum and spatial distribution of particles causing auroral luminescence, 2) the relation of large-scale structure of the magnetosphere with the region of most frequent appearance of zenithal aurora, and 3) auroral dynamics during substorms and the place of the auroral substorm in the general picture of substorm development in the magnetosphere.

The analysis of observational data acquired during the IGY has shown that zenithal auroras occur most frequently along an oval with positions at $\Phi \approx 67°$ near midnight and $\Phi \approx 76°$–$77°$ at noon. In Fig. 1 are given Feldstein's (1963) first results of a statistical analysis of the space-time distribution of zenith auroral forms. The auroral frequency along the oval (shaded) is non-uniform. Furthermore, the oval is divided into two parts, day and night. The bases for this distinction are differences in the auroral forms and their morphological characteristics. At the time these results were first presented, the reasons for distinguishing between day and night parts of the oval were not sufficiently proven. Discontinuities within the auroral oval at morning and evening during magnetically quiet periods were found by Lassen (1963, 1969, 1970) and Feldstein and Starkov (1967). These discontinuitites were filled with rather intense luminescence in the periods of magnetic disturbance when the luminescence was observed continuously along the oval. Variation in the probability of auroras along the oval was shown also by Stringer and Belon (1967) and by Zhigalov (1967).

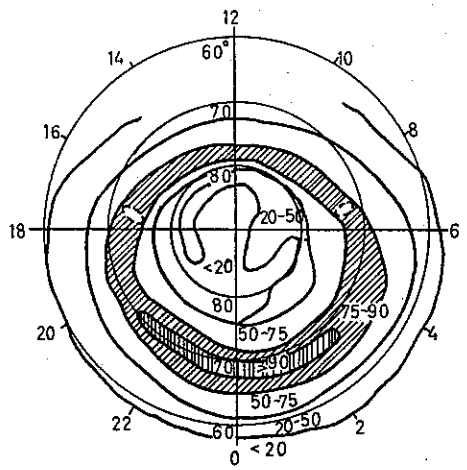


Figure 1. Isoauroras in the Northern Hemisphere from IGY data. The dashed region of the most frequent aurora appearance in the zenith is divided into day and night sections at morning and evening hours.

Khorosheva (1963) has shown that the extended forms of polar auroras are aligned along the oval. This alignment results in a daily variation of the azimuths of extended forms (Davis, 1962; Hultqvist, 1962; Feldstein, 1963). The observation materials on this point were gathered for IGY by Gustafsson (1967) and IQSY by Gustafsson et al. (1969). A detailed time-space distribution of the orientation of extended forms allows for estimations of an auroral arc

position which is in good agreement with the auroral oval (Gustafsson, 1967; Feldstein and Starkov, 1967; Gustafsson et al., 1969). It is shown by Jacka and Bond (1968) and by Bond and Thomas (1971) that the location and orientation of auroras in the Southern Hemisphere at different levels of magnetic disturbances agree nicely with this oval conception, which was based on data from the Northern Hemisphere.

The possibilities of investigations of oval discontinuity and extended form position at high latitudes have recently increased owing to the availability of aircrafts provided with radio and optical equipment. The analysis of observations carried out in 1967–1970 by Buchau et al. (1969, 1970, 1971) and Whalen et al. (1971) showed: 1) The position of discrete forms of auroras between 75°–79° corrected geomagnetic latitude at the noon sector, obtained on individual flights and by statistical processing of data, is in a good agreement with results of ground-based visual observations. 2) The distribution of discrete auroral forms agrees with the existence of an oval belt within the limits of which the auroral arcs are located. 3) The size and position of the oval change with variations in magnetic activity. 4) On moderately disturbed days there is a continuous belt of luminescence along the oval but in magneto-quiet periods discontinuities or sharp luminescence weakenings occur along the oval, and this could be identified with the oval discontinuity in visually observed auroras. However, the broader band of subvisual emissions exists in the same location as the visible forms and continues through the gaps (Buchau et al., 1971).

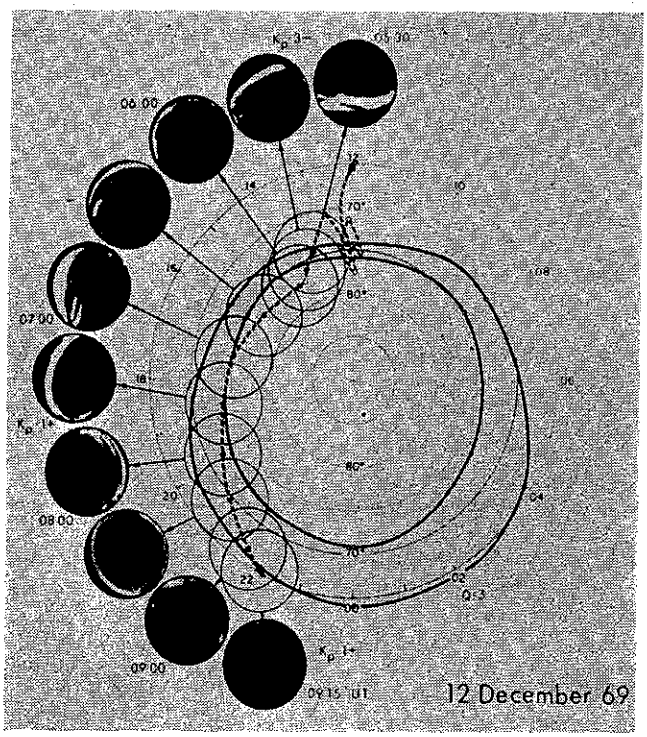


Figure 2. The flight of December 12, 1969 showed a continuous auroral band that extended from the local magnetic noon to the midnight sector. The circles on the flight's trajectory (dotted lines) show the all-sky camera photo in the system of corrected geomagnetic coordinates. The photos are placed so as to keep their orientation with respect to the radius-vector that connects the geomagnetic pole (+) and the zenith of all-sky camera photo (Buchau et al., 1970).

Thus, aircraft observations provide convincing evidence of the existence of an auroral *oval* and not two quasicircular zones (Mishin et al., 1970) at $\Phi' \approx 78°$ and 67°. Mishin et al. (1971) made an attempt to show that the distributions of discrete forms observed from aircrafts correspond better to two quasicircular zones than to an oval. The auroral position of December 6 and 12, 1969, and January 15, 1970 (from Buchau et al., 1970) was projected on the Earth's surface (cf. Fig. 2). However, because of misunderstanding, a reflection of twilight on the southern side of the photograph obtained during day hours of December 6 and 12 was united by Mishin et al. (1971) with the real aurora. On January 15 a reflection of the moon on the northern side of the photograph was taken as polar aurora. But when the extended forms of aurora are projected correctly, they are arranged along the oval.

On the basis of the global distribution of auroras and taking into account that their morphological characteristics differ for the day and night sector (Lassen, 1963; Feldstein and Starkov, 1967; Davis, 1967; Feldstein, 1969; Khorosheva, 1967; Starkov, 1968) planetary schemes of intrusion of electron fluxes with different energy at magneto-quiet and magneto-disturbed periods have been obtained (Feldstein, 1968; Troshichev, 1970). In Fig. 3 is given the planetary scheme by Deehr and Egeland (1971) based on recent satellite observations. In particular, at the night sector the discrete forms of aurora correlate with the penetration of electrons with $E \approx 6$ keV (Gustafsson et al., 1971). Furthermore, aircraft observations of the complex of geophysical events in the upper atmosphere (Buchau et al., 1971) have shown that during day hours discrete auroral forms, *F*layer irregularities and a band of soft precipitation deduced from an enhancement at 6300 Å, are very closely associated and are possibly identical in location. According to the observations the effects on the day side of the oval are

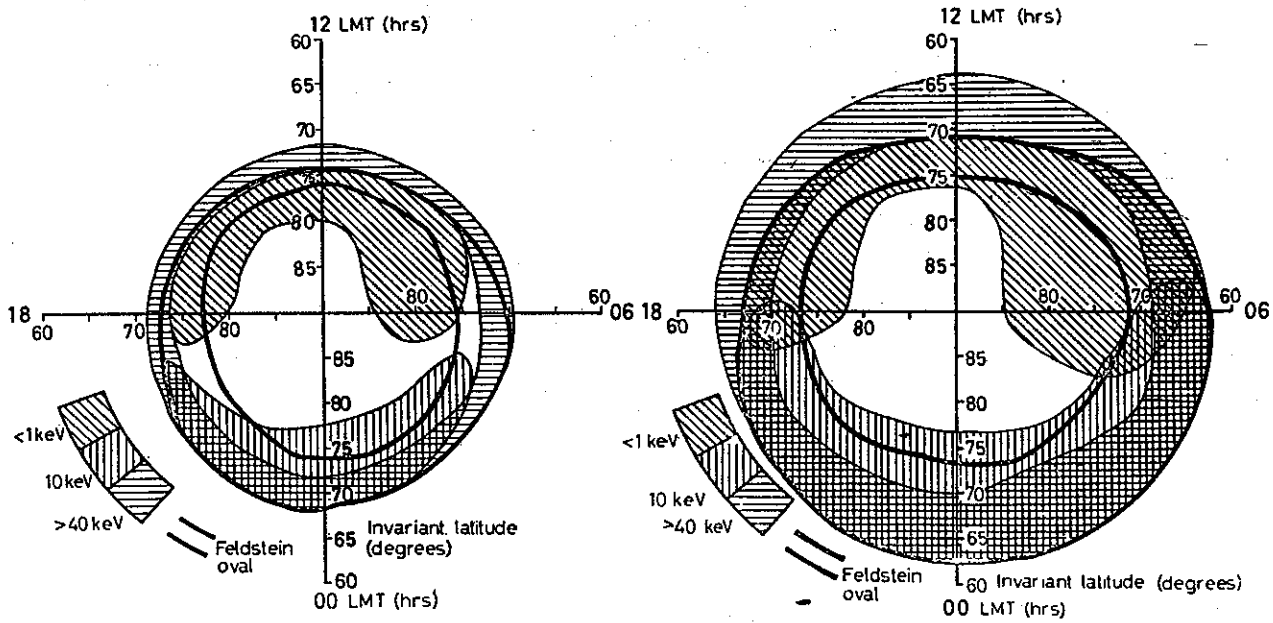


Figure 3. The average position of the regions of precipitation and auroral oval in a) magneto-quiet conditions, b) magneto-disturbed ones (Deehr and Egeland, 1971).

caused by a direct penetration of the magnetosheath plasma to low altitudes (Heikkila and Winningham, 1971; Hoffman and Berko, 1971; Frank, 1971a; Sharp and Johnson, 1971), and on the night side they are caused by penetration of plasma from the plasma sheet of the tail (Vasyliunas, 1970b; Feldstein, 1970). Taking into account the broad region of energetic particle precipitation on the late-morning side and equatorward of the auroral oval, which is responsible for absorption of the cosmic noise and X-rays, it is possible to bring into agreement the present particle measurements in high latitudes (Craven, 1970; Fedorova et al., 1971; Hoffman, 1971) with the data on planetary distribution of penetrating electrons obtained by ground observations of auroral effects.

3. The aurora and its relation to the magnetosphere

It has been shown (Akasofu, 1966; Fairfield, 1968; Shabansky, 1968; Vasyliunas, 1970a; Feldstein, 1969; Feldstein and Starkov, 1970) that there is a close connection between the auroral oval and the structure of the magnetosphere. In particular the association is close between the plasma sheet in the tail of the nightside magnetosphere and the magnetopause projection along field lines on the day side of the Earth. In Fig. 4, a section of the magnetosphere in the noon-midnight meridian plane is schematically given.

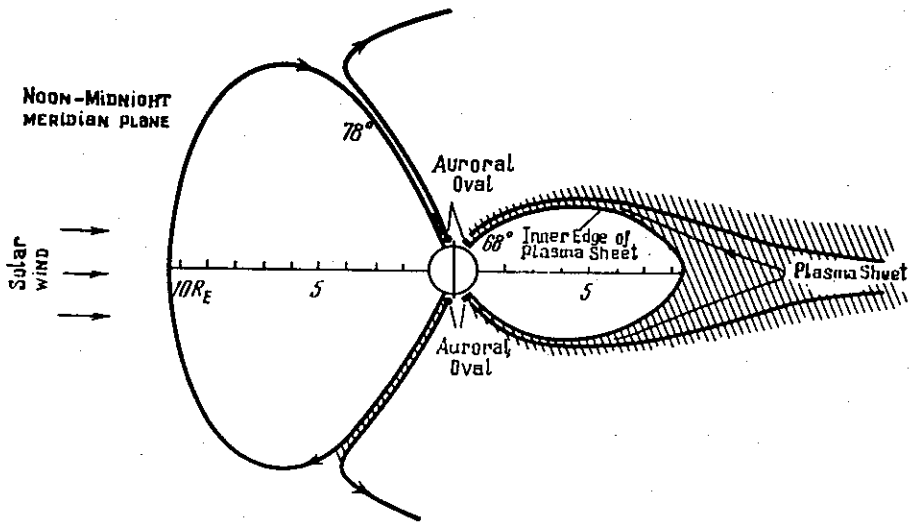


Figure 4. A schematic section of the magnetosphere and the position of the oval in the noon-midnight meridian plane. Shaded regions show the plasma sheet in the tail and plasma fluxes flowing from the tail to upper atmosphere on night side and day side through the polar cusps.

The solar plasma moves from the magnetosheath to the upper layers of the atmosphere at 75°–80° geomagnetic latitude through a gap on the dayside of the Earth (Heikkila and Winningham, 1971; Frank, 1971a). Electron energy of some hundreds of eV is sufficient to produce the luminescence originating at 150–200 km and the energy flux corresponding to that observed in this sector of the auroral oval (Hoffman and Berko, 1971; Hoffman, 1971; Feldstein and Starkov, 1971). The southern boundary of the oval at the Earth's night side is a projection

of the inner side of the plasma sheet along real fields lines (Vasyliunas, 1970a; Feldstein, 1970). Spectra and fluxes of electrons in the central part of the plasma sheet and at auroral luminescence altitudes prove a close connection between these two regions (Vasyliunas, 1970a; Chase, 1969, 1970).

When magnetic disturbance grows, the auroral oval at noon hours is shifted to lower latitudes without a noticeable widening (Starkov and Feldstein, 1967a; Feldstein et al., 1968). A similar shift of some degrees of latitude towards the equator without a large increase in its latitudinal width was noticed for both the day cusp (Frank, 1971a) and the soft high latitude zone of radiation (Hoffman, 1970). The oval position during day hours varies with the change of the inclination angle of geomagnetic dipole axis to ecliptic plane (Feldstein, 1970). The existence of this was deduced from observations of the boundary of trapped energetic electrons by Feldstein and Starkov (1970), and has been confirmed by latest data analysis. (Page and Shaw, 1971). The data on similar variations of the region of soft electron intrusion are very contradictory (Maehlum, 1968; Hoffman, 1971) because of rather limited observations made so far.

The plasma sheet in the equatorial plane of the nightside magnetosphere envelopes the inner magnetospheric region (according to Shabansky (1968) it is defined as a 'core'). On the dayside it links up with the magnetopause (Vasyliunas, 1968b). When the magnetic disturbance grows, the inner boundary of the plasma sheet approaches the Earth and the low latitude boundary of the auroral oval moves equatorwards (Vasyliunas, 1970b; Feldstein and Starkov, 1967). A close connection between the equatorial boundary of the oval on one side and the plasma sheet and magnetopause on the other (Vasyliunas, 1970a; Feldstein, 1970) makes it possible to use data on oval dynamics at different stages of a substorm for estimations of inner magnetosphere dimensions. Whereas the oval position is well known in all sections of the local time during a substorm we cannot say the same about the dynamics of the inner boundary of the plasma sheet (Vasyliunas, 1968a, b, 1969; Frank, 1968, 1971; Schield and Frank, 1970; Freeman and Maguire, 1968).

Figure 5 shows the dimension of the core of the magnetosphere at different phases of the substorm. The shape of the core defined by the projection of the southern boundary of the oval along the real field lines into the equatorial plane. The geocentric distance to the subsolar point of the magnetopause is given by Rudneva and Feldstein (1970). The creation phase of the substorm is characterized by a sharp shift of the plasma sheet boundary in the midnight sector towards the Earth. The smallest size of the core is observed at the highest point of the substorm development. The core's radius reduces by as much as $\approx 6\ R_E$. At the midnight sector, the boundary moves to $\approx 4{,}2\ R_E$ to the Earth. In the recovery phase the plasma sheet boundary again withdraws from the Earth. However, significant fluxes of energetic electrons drifting to the east remain on field lines of the core. These field lines penetrated the plasma sheet in the period of maximum development of a substorm. On their eastward motion electrons are accelerated by the electric field directed across the magnetosphere from dawn to dusk and they precipitate into the ionosphere. Observations of electrons with energy $1 \leq E \leq 25$ keV (Galperin and Fedorova, 1971) are in agreement with this view of dynamics of the plasma sheet during substorms.

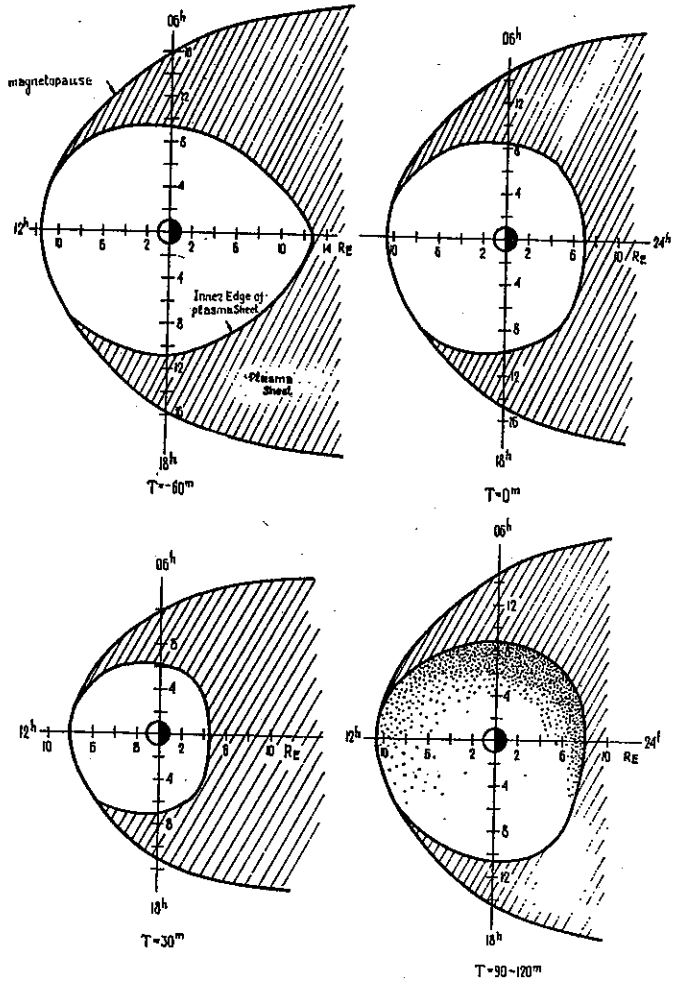


Figure 5. The position of the inner boundary of the plasma sheet in the equatorial plane at different phases of a substorm

1) quiet period and the beginning of creation phase ($T = -60$ min),
2) the beginning of expansive phase ($T = 0$ min),
3) maximum expansive phase ($T = 30$ min),
4) recovery phase ($T =$ up to 120 min).

The plasma sheet is shaded. The region of eastward drifting electrons quasi-trapped in the expansive phase is shown by dots.

The oval dynamics permits one to evaluate variations of magnetic fluxes in remote regions of the magnetosphere. The magnetic flux in the tail is known to grow from $1.3 \cdot 10^{17}$ Maxwells in magneto-quiet periods to $2.3 \cdot 10^{17}$ Maxwells in magneto-disturbed (Piddington, 1967; Starkov et al., 1968; Feldstein, 1969).

Proceeding from the magnetospheric structure and oval position, a supposition may be made that the magnetospheric tail is formed by magnetic flux F_1 which penetrates to the Earth's surface at the high latitude region poleward of the equatorial boundary of the oval. F_1 is partly closed in the neutral sheet of the tail (Mihalov et al., 1968; Behannon, 1970). Assuming that those outer field lines which are closed within the tail and are connected to the Earth all

lie in the plasma sheet in the tail, it is then possible to estimate the magnetic flux F_2 which is reconnected with the interplanetary magnetic field far from the Earth. (The boundary of the plasma sheet is considered the same as the boundary of diffusive luminescence in the aurora.) The magnetic fluxes F_1 and F_2 at different stages of a substorm are given in Fig. 6. Flux F_1 is growing slowly during the creation phase and then at expansion phase sharply increases. The creation phase is characterized by a sharp decrease of the magnetic flux through the neutral sheet, possibly due to the extension of field lines in the antisolar direction.

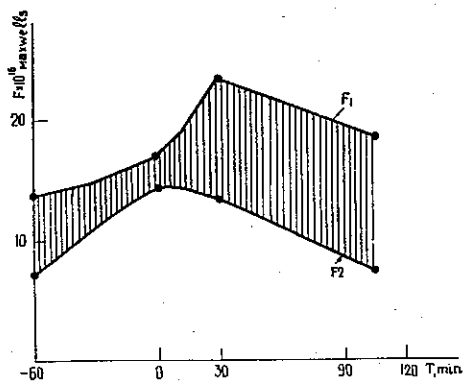


Figure 6. Variations of magnetic fluxes in the tail at different phases of a substorm
$-60 < T < 0$ min – creation phase
$0 \leqq T \leqq 30$ min – expansion phase
$30 < T < 90$ min – recovery phase
F_1 – magnetic flux that formed the magnetosphere tail.
F_2 – magnetic flux outside the plasma sheet of the tail.
Magnetic flux passing through the plasma sheet, closed through the neutral sheet, is shaded.

Thus, at the beginning of an active period of a substorm the field configuration is very unlike the dipole field. At $T = 0$ the magnetic flux F_2 reconnecting with the interplanetary field has its maximum. In the active period of a substorm ($T > 0$) F_2 decreases simultaneously with a sharp increase of the flux closed through the neutral sheet. The field line configuration in the tail is then more similar to a dipole type than it was in the creation phase. These field variations are in agreement with changes in field line configuration in the tail during substorms (Fairfield and Ness, 1970; Fairfield, 1970).

On the basis of the position of the southern boundary of the oval near midnight, the intensity of the magnetic field in the tail was estimated by Feldstein et al. (1968). In contradiction to earlier observations of the magnetic field in the tail (Akasofu, 1968), it was found that this field suddenly increases with substorm development. However, further investigations have shown that magnetic field intensity really increases in the disturbance period (Sugiura et al., 1970; Fairfield and Ness, 1970), and a more detailed analysis has confirmed that the most essential increase of the field is at the beginning of the substorm (Meng and Akasofu, 1971; Camidge and Rostoker, 1970; Russel et al., 1971).

An important goal of auroral investigations is to establish specific features of auroras during substorms, since these reveal the complex character of changes in the upper atmosphere at

the time. At present, auroral dynamics during a substorm is so thoroughly studied that it serves as a basis for analysis of the role the magnetospheric substorm plays in other geophysical processes. Thus the new scheme of auroral substorm development given below is of a broad application.

The scheme of auroral substorm development was advanced by Akasofu (1964, 1965, 1968). It has been one of the most fundamental notions of magnetospheric physics. However, recent investigations have significantly refined the scheme (Ivliev et al., 1970; Pudovkin et al., 1970; Starkov and Feldstein, 1971; Feldstein, 1970; Davis, 1971). Figure 7 is based on recent data obtained by Starkov and Feldstein (1971), and shows the orientation and motion of auroral forms during a substorm. The coordinate system is magnetic time and latitude. The auroral substorm consists of three phases: creation (-60 min $< T < 0$ min), expansion (0 min $\leq T \leq 30$ min), and recovery ($30 < T < 120$ min). $T = 0$ corresponds to the beginning of the second phase: a rapid auroral motion towards the pole on the oval's northern boundary occurs.

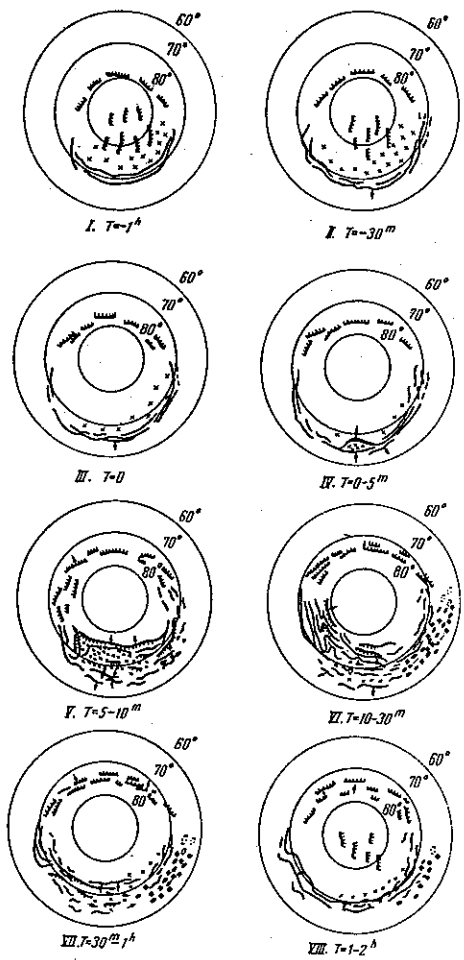


Figure 7. The general scheme of auroral substorm. I, II – creation, III, IV, V, VI – expansion, VII, VIII – recovery phases. Homogeneous forms of auroras are shown by solid lines; rayed forms by solid lines with dashes; pulsating ones by blackened ellipses; luminescence connected with the plasma sheet in the magnetosphere tail by crosses; and those conditioned by electrons drifting to the east in the magnetosphere core by crosses with dots.

(The beginning of the substorm is given in the previous scheme according to Akasofu (1964).) After a quiet period all the three phases may be observed. Occurrence of a substorm will, however, have influence on the size and development of one following immediately after, and also will make difficult recognition of a new creation phase.

In quiet conditions ($T \leq -60$ min) the spatial distribution of aurora corresponds to the scheme advanced by Feldstein (1966). At the night side weak homogeneous or rayed arcs are situated on $\Phi' \approx 70°$. Even in very quiet periods at near-midnight hours there are continuous auroras at these latitudes; this fact agrees with the observations on electron intrusion (Deehr et al., 1971) and radio reflections (Bates et al., 1969). The day side of the oval includes weak rayed forms at $\Phi' \approx 78°$–$80°$. In the polar cap (a section of the near-polar region confined by the aurora oval) transient, weak, Sun-Earth aligned rayed arcs and diffusive luminescence exist. The subvisual luminescence in the polar cap has been revealed by Weill et al. (1965), Eather (1969), Eather and Akasofu (1969), and Eather and Mende (1971) by photometric observations from aircrafts. It is possibly connected with the external region of the plasma sheet in the magnetospheric tail (Vasyliunas, 1970b). Aircraft investigations showed that auroras in the Northern and Southern Hemispheres are exactly conjugated (Belon et al., 1969).

For quiet periods the interplanetary magnetic field is as a rule oriented to the north. When the field orientation changes to south, a creation phase of a substorm starts and lasts about one hour (Pudovkin et al., 1970; McPherron, 1970; Kokubun, 1971). In the creation phase the magnetic field at the near-midnight sector on $\Phi' \approx 67°$ is quiet, but auroras become more active on the day section of the oval (Starkov and Feldstein, 1967b) and are shifted towards the equator (Starkov and Feldstein, 1971; Kaneda, 1971). At the same time a rather slow ($\approx$ 10–15 km/min) shift of arcs and bands in the night sector towards the equator is observed. The shift continues up to $T = 0$ and is, in the midnight sector, approximately 5° of latitude. In the creation phase, no noticeable intensification of arc brightness occurs, nor is any widening of the oval's night sector seen. Discrete forms of auroras gradually disappear in the polar cap.

If the southward orientation of the interplanetary field remains for one hour and the south component of it is more than 6 γ, the creation phase is developed into the expansive one (Nishida, 1971).

The beginning of expansive phase ($T = 0$) is related to increase in near-midnight arc brightness and a sudden movement of the bright auroral band towards the pole (Akasofu, 1964), and away from a rather weak arc in the near-midnight sector. This weak arc remains in the same position or shifts equatorward. A rather intensive diffusive background is left equatorward of the bright band moving towards the pole. The integrated luminescence of this background may, however, be rather pronounced. The discrete forms of auroras in the region of maximum background luminescence are poorly defined. Auroral motion towards the equator is observed during all phases of a substorm. This motion is possibly connected with the electric field existing in the magnetosphere. The poleward motion is of explosive character and is most pronounced near the polar boundary of the oval. In the expansive phase the conjugacy of auroras on the two hemispheres breaks down. At $\Phi \approx 67°$ they are shifted tens of km when comparing with conjugated regions, and at $\Phi \approx 70°$ conjugation sometimes is completely missing (Belon et al., 1969; Davis et al., 1971).

In the recovery phase the auroras withdraw to the starting position. A diffusive luminescence designated as mantle aurora is observed in the morning sector (Sandford, 1964) equatorward of the auroral oval as arc patches and pulsating auroras (Kvifte and Pettersen, 1969; Roldugin and Starkov, 1970). The character of the luminescence, its location only in the morning-noon sectors, their connection with X-rays, and anomalous cosmic noise absorption allows for the supposition that the emission is caused by electrons from the magnetospheric core with energy essentially greater than that in the tail's plasma sheet. The luminescence at the latitudes of the auroral zone in the morning hours is due to the electrons drifting on closed magnetic field shells, that were captured at near-midnight and early morning hours at small L during the expansive phase of the auroral substorm. Auroral pulsations in the morning hours possibly show regularities in strong pitch-angle diffusion of the above-mentioned electrons entering the loss cone when the particles are interacting with waves (Kennel, 1969). Pulsating auroras in the morning sector along the auroral zone are nicely conjugated in both hemispheres (Cresswell and Davis, 1966). From observations of X-rays (Trefall et al., 1970), energetic electrons (Hoffman, 1970; Galperin and Fedorova, 1971) and VLF emissions (Harang and Akasofu, 1969) during the recovery phase, the appearance of hard electrons in the morning hours equatorward of the oval has been found. While precipitation along the oval occurs at any time, quiet or disturbed, the circular zone of intensive electron precipitation at auroral zone latitudes exists only in disturbed periods, enveloping the morning and day sectors. On disturbed days, when substorms follow one after the other, the distribution of precipitating particles may to some extent be presented by a scheme, advanced by Hartz and Brice (1967) and by Hartz (1971). This scheme coincides with the one given in Fig. 3.

Photometric standardization performed on all-sky cameras in the Soviet Union makes it possible to obtain qualitative data on luminescence intensity and subsequently the energy of the effect. In Fig. 8 (according to Feldstein and Starkov, 1971) average intensities of energy flux density of precipitating electrons are given within six time sector along the oval. At noon hours the energy flux increases from 0.45 to 1.4 erg/cm²s. It was noticed earlier (Feldstein,

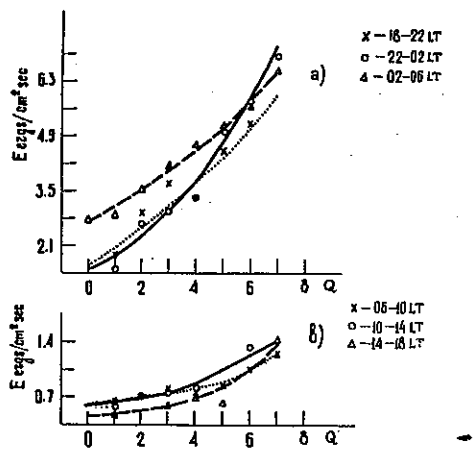


Figure 8. Average intensities of precipitating particle energy at different sections of the oval when magnetic disturbance intensity varies.

1970) that these values for the day sector are sometimes greater than those measured directly by satellites (Johnson and Sharp, 1969; Sharp et al., 1969). The energy of precipitating particles obtained from auroral data on the day side of the Earth is in much better agreement with recent satellite measurements made in the oval's day sector. According to Hoffman and Berko (1971) and Hoffman (1971) the energy flux in the morning hours is greater than in evening and changes from 0.3–0.58 erg/cm^{2}s in morning to 0.1–0.18 erg/cm^{2}s in evening. As to the results of all-sky camera processing, the energy fluxes are found to be 0.6–0.8 erg/cm^{2}s and 0.4–0.5 erg/cm^{2}s in the morning and evening sectors of the oval, respectively. In the noon sector satellite measurements of the energy of precipitating electrons are found to be 0.5 erg/cm^{2}s (Hoffman, 1971; Heikkila and Winningham, 1971) while auroral intensity measurements give 0.45 erg/cm^{2}s.

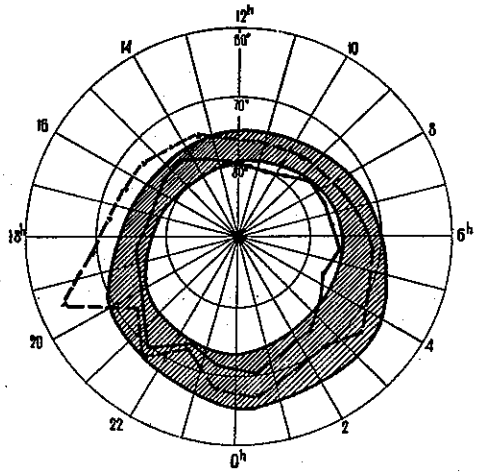


Figure 9. Daily variations of the region of transverse magnetic disturbances in the Northern Hemisphere when $0 \leq Kp \leq 2+$ (dots – the nearpole boundary, triangles – the equatorial boundary). The polar oval with $Kp = 1$ is shaded.

Particle fluxes precipitating along the oval may originate electrical currents. These particles seem to be responsible for field aligned currents that give rise to magnetic disturbances transverse to the direction of the local field at 1100 km altitude (Zmuda et al., 1967; Cummings and Dessler, 1967). In Fig. 9 the position of the region of disturbance is given for $0 \leq Kp \leq 2$ according to Zmuda et al. (1970) together with the oval's position for $Kp = 1$ (Starkov and Feldstein, 1967, 1968). From 2000 until 1400 LT transverse magnetic disturbances are only observed within the limits of the oval, while in the evening sector (1400–2000 LT) they are also observed equatorward of the oval. The intensity of the flux of particles precipitating along the oval is sufficient for the transverse disturbances of the observed intensity (Zmuda et al., 1970; Hoffman, 1971), and almost always the disturbances present are in conformity with constant occurrence of the luminescence along the oval. A wider region of transverse disturbances in evening hours is possibly conditioned by the fact that in this sector the region of vertical currents goes outside the oval. However, unlike vertical currents along the oval caused mainly by electron fluxes and resulting in a westward electrojet, there are intensive fluxes of protons in the evening sector equatorward of the oval which are connected with an assymetric ring

current and eastward electrojet. This idea is in accordance with the data on eastern and western current position when $Kp = 1$, according to Feldstein and Zaitzev (1968).

The region of transverse magnetic disturbances coincides with the auroral oval not only statistically but also in specific cases. For the strong magnetic storm of November 1, 1968, considered by Armstrong and Zmuda (1970), the region of transverse disturbances at 0851 LT shifted to corrected geomagnetic latitudes $61° < \Phi' < 67°$. At that time the oval was at $62° \leqq \Phi' \leqq 67°$ taking into consideration the intensity of the existing D_{st} variation.

References

Akasofu, S. – I., 1964: Planet Space Sci., **12**, 273.

Akasofu, S. – I., 1965: Space Sci. Rev., **4**, 498.

Akasofu, S. – I., 1966: Planet. Space Sci., **14**, 587.

Akasofu, S. – I., 1968: 'Polar and Magnetospheric Substorms', Springer Verlag, New York.

Armstrong, J. C. and A. J. Zmuda, 1970: J. Geophys. Res., **75**, 7122.

Bates, H. F., R. D. Sharp, A. E. Belon and J. S. Boyd, 1969: Planet. Space Sci., **17**, 83.

Behannon, K. W., 1970: J. Geophys. Res., **75**, 743.

Belon, A. E., J. E. Maggs, T. N. Davis, K. B. Mather, N. W. Class and G. F. Hughes, 1969: J. Geophys. Res., **74**, 1.

Bond, F. R. and I. L. Thomas, 1971: Austr. J. Phys., **24**, 97.

Buchau, J., J. A. Whalen and S.-I. Akasofu, 1969: J. Atmosph. Terr. Phys., **31**, 1021.

Buchau, J., J. A. Whalen and S.-I. Akasofu, 1970: J. Geophys. Res., **75**, 7147.

Buchau, J., G. L. Gassmann, C. P. Pike, R. A. Wagner and J. A. Whalen, 1971: Paper presented at the XV IUGG General Assembly, Moscow.

Camidge, F. P. and G. Rostoker, 1970: Can. J. Phys., **48**, 2007.

Chase, L. M., 1969: J. Geophys. Res., **74**, 348.

Chase, L. M., 1970: J. Geophys. Res., **75**, 7128.

Craven, J. D., 1970: J. Geophys. Res., **75**, 2468.

Cresswell, G. R. and T. N. Davis, 1966: J. Geophys. Res., **71**, 3155.

Cummings, W. D. and A. J. Dessler, 1967: J. Geophys. Res., **72**, 1007.

Davis, T. N., 1962: J. Geophys. Res., **67**, 75.

Davis, T. N., 1967: In 'Aurora and Airglow' (Ed. B. M. McCormac) Reinhold Publ. Corp., New York, 41.

Davis, T. N., 1971: Paper presented at the XV IUGG General Assembly, Moscow.

Davis, T. N., T. J. Hallinan and H. C. Stenback-Nielsen, 1971: In 'The Radiating Atmosphere' (Ed. B. M. McCormac, Reidel Publ. Comp., Dordrecht, Holland, 150.

Deehr, C. S. and A. Egeland, 1971: Paper presented at the XV IUGG General Assembly, Moscow.

Deehr, C. S., A. Egeland and F. Söraas, 1971: In 'The Radiating Atmosphere' (Ed. B. M. McCormac), Reidel Publ. Comp., Dordrecht, Holland.

Eather, R. H., 1969: J. Geophys., Res., **74**, 153.

Eather, R. H. and S.-I. Akasofu, 1969: J. Geophys. Res., **74**, 4794.

Eather, R. H. and S. B. Mende, 1971: J. Geophys. Res., **76**, 1746.

Egeland, A. and A. Omholt, 1967: In 'Aurora and Airglow' (Ed. B. M. McCormac), Reinhold Publ. Corp. New York, 143.

Fairfield, D. H., 1968: J. Geophys. Res., **73**, 7329.

Fairfield, D. H., 1970: In 'Proceedings from the IUSTP Symposium, Leningrad', Reidel Publ. Comp., Dordrecht, Holland.

Fairfield, D. H. and N. F. Ness, 1970: J. Geophys. Res., **75**, 7032.

Fedorova, N. I., V. V. Temny and Yu. I. Galperin, 1971: J. Atmosph. Terr. Phys., **33**, 731.

Feldstein, Y. I., 1963: Geomag. Aeronomie, **3**, 227.

Feldstein, Y. I., 1966: Plan. Space Sci., **14**, 121.

Feldstein, Y. I., 1968: Dr. Sci. Thesis, Moscow.

Feldstein, Y. I., 1969: Rev. Geophys., **7**, 179.

Feldstein, Y. I., 1970: In 'Proceedings from the IUSTP Symposium, Leningrad'. Reidel Publ. Comp., Dordrecht, Holland.

Feldstein, Y. I. and G. V. Starkov, 1967: Planet. Space Sci., **15**, 209.

Feldstein, Y. I. and G. V. Starkov, 1970: Planet. Space Sci., **18**, 501.

Feldstein, Y. I. and G. V. Starkov, 1971: J. Atmosph. Terr., Phys., **33**, 197.

Feldstein, Y. I. and A. N. Zaitzev, 1968: Ann. Géophys. **24**, 595.

Frank, L. A., 1968: In 'Physics of the Magnetosphere' (Eds. R. L. Carovillano, J. F. McClay and H. R. Radoski), D. Reidel Publ. Comp., Dordrecht, Holland, 271.
Frank, L. A., 1971a: J. Geophys. Res., **76**, 5202.
Frank, L. A., 1971: J. Geophys. Res., **76**, 2265.
Freeman, J. W. and J. J. Maguire, 1968: Ann. Géophys., **24**, 295.
Galperin, Yu. I. and N. I. Fedorova, 1971: Paper presented at the XV IUGG General Assembly, Moscow.
Gustafsson, G., 1967: Planet. Space Sci., **15**, 277.
Gustafsson, G., Y. I. Feldstein and N. F. Shevnina, 1969: Planet. Space Sci., **17**, 1657.
Gustafsson, G., A. Egeland and C. S. Deehr, 1971: KGO Preprint N71–303.
Harang, L., 1951: 'The Aurorae', Chapman and Hall, London.
Harang, L. and S.-I. Akasofu, 1969: J. Atmosph. Terr. Phys., **31**, 1445.
Hartz, T. R., 1971: In 'The Radiating Atmosphere' (Ed. B. M. McCormac), Reidel Publ. Comp., Dordrecht, Holland.
Hartz, T. R. and N. M. Brice, 1967: Planet. Space Sci., **15**, 301.
Heikkila, W. J. and J. D. Winningham, 1971: J. Geophys. Res., **76**, 883.
Hoffman, R. A., 1970: 'Auroral electron drift and precipitation: Cause of the mantle aurora', preprint GSFC X-646-70-205.
Hoffman, R. A., 1971: 'Properties of low energy particle impacts in the polar domain in the dawn and dayside hours', preprint GSFS, X-646-71-199.
Hoffman, R. A. and F. W. Berko, 1971: J. Geophys. Res., **76**, 2967.
Hultqvist, B., 1962: J. Atmosph. Terr. Phys., **24**, 17.
Ivliev, D. Y., M. I. Pudovkin and S. A. Zaitzeva, 1970: Geomag. Aeronomie, **10**, 300.
Jacka, F. and F. R. Bond, 1968: Ann. Géophys., **24**, 547.
Johnson, R. G. and R. D. Sharp, 1969: In 'Atmospheric emissions' (Eds. B. M. McCormac and A. Omholt), Van Nostrand Reinhold Comp., New York, 219.
Kaneda, E., 1971: Paper presented at the XV IUGG General Assembly, Moscow.
Kennel, C. F., 1969: Rev. Geophys., **7**, 379.
Khorosheva, O. V., 1963: Geomag. Aeronomie, **3**, 363.
Khorosheva, O. V., 1967: 'Space-time distribution of the aurorae', Moscow, N16, 3.
Kokubun, S., 1971: Planet. Space Sci., **19**, 697.
Kvifte, G. J. and H. Pettersen, 1969: Planet. Space Sci., **17**, 1657.
Lassen, K., 1963: Publ. Danske Met. Inst. Rpt. No. 16.
Lassen, K., 1969: In 'Atmospheric Emissions' (Eds. B. M. McCormac and A. Omholt), Van Nostrand Reinhold Comp., New York, 63.
Lassen, K., 1970: Phys. Norv. **4**, 171.
Maehlum, B. N., 1968: J. Geophys. Res., **73**, 2359.
McPherron, R. L., 1970: J. Geophys. Res., **75**, 5592.
Meng, C. I. and S.-I. Akafosu, 1971: Paper presented at the XV IUGG General Assembly, Moscow.
Mihalov, J. D., D. S. Colburn, R. G. Currie and C. P. Sonett, 1968: J. Geophys. Res., **73**, 943.
Mishin, V. M., T. I. Saifudinova and I. A. Zhulin, 1970: J. Geophys. Res., **75**, 797.
Mishin, V. M., V. P. Samsonov, G. V. Popov and T. I. Saifudinova, 1971: 'Issledovania geomagnetisma, aeronomie, physics of the Sun', Irkutsk, N19, 38.
Nishida, A., 1971: Paper presented at the XV IUGG General Assembly, Moscow.
Page, D. E. and M. L. Shaw, 1971: Paper presented at the XV IUGG General Assembly, Moscow.
Piddington, J. H., 1967: J. Atmosph. Terr. Phys., **29**, 87.
Pudvokin, M. I., S. I. Isaev and S. A. Zaitzeva, 1970: Ann. Géophys., **26**, 761.
Pudovkin, M. I., O. M. Raspopov, L. A. Dmitrieva, V. A. Troitskaya and R. V. Shepetnov, 1970: Ann. Géophys., **26**, 389.
Roldugin, V. K. and G. V. Starkov, 1970: Geomag. Aeronomie, **10**, 97.
Rudneva, N. M. and Y. I. Feldstein, 1970: Geomag. Aeronomie, **10**, 804.

Russell, C. T., R. L. McPherron and P. J. Coleman, 1971: J. Geophys. Res., **76**, 1823.
Sandford, B. P., 1964: J. Atmosph. Terr. Phys., **26**, 749.
Schield, M. A. and L. A. Frank, 1970: J. Geophys. Res., **75**, 5401.
Shabansky, V. P., 1968: Space Sci. Rev., **8**, 360.
Sharp, R. D., D. L. Carr and R. G. Johnson, 1969: J. Geophys. Res., **74**, 4618.
Sharp, R. D. and R. G. Johnson, 1971: In 'The Radiating Atmosphere' (Ed. B. M. McCormac), Reidel Publ. Comp., Dordrecht, Holland, 239.
Sharp, R. D., D. L. Carr, R. G. Johnson and Shelly E. G., 1971: J. Geophys. Res., **76**, 7669.
Starkov, G. V., 1968: Geomag. Aeronomie, **7**, 367.
Starkov, G. V. and Y. I. Feldstein, 1967a: Geomag. Aeronomie, **7**, 62.
Starkov, G. V. and Y. I. Feldstein, 1967b: Geomag. Aeronomie, **7**, 367.
Starkov, G. V. and Y. I. Feldstein, 1968: 'Aurora and Airglow', N17, Publ. House Naika, Moscow, 22.
Starkov, G. V. and Y. I. Feldstein, 1971: Geomag. Aeronomie, **11**, 560.
Starkov, G. V., Y. I. Feldstein and A. D. Shevnin, 1968: Kosmitcheskie Issledovanija, **6**, 153.
Störmer, K., 1955: 'The Polar Aurora', Clarendon Press, Oxford.
Stringer, W. J. and A. E. Belon, 1967: J. Geophys. Res., **72**, 4415.
Sugiura, M., T. L. Skillman, B. G. Ledley and J. P. Heppner, 1970: In 'Particles and fields in the magnetosphere' (Ed. B. M. McCormac), Reidel Publ. Comp., Dordrecht, Holland, 165.
Trefall, H., T. Pytte, K. Brönstad, J. Bjordal, I. Singstad and A. Sörhaug, 1970: Phys. Norv., **4**, 161.
Troshichev, A., 1970: Kosmitoheskie Issledovanija, **8**, 915.
Vasyliunas, V. M., 1968a: Geophys. Res., **73**, 2839.
Vasyliunas, V. M., 1968b, J. Geophys. Res., **73**, 7519.
Vasyliunas, V. M., 1970a: In 'The Polar Ionosphere and Magnetospheric Interactions' (Ed. G. Skovli), Gordon and Breach, N. Y., 26.
Vasyliunas, V. M., 1970b: In 'Proceedings from the IUSTP Symposium, Leningrad', Reidel Publ. Comp., Dordrecht, Holland.
Weill, G., M. Fafiotte and S. Huille, 1965: Ann. Géophys., **2**, 469.
Whalen, J. A., J. Buchau and R. A. Wagner, 1971: J. Atmosph. Terr. Phys., **33**, 661.
Zhigalov, L. N., 1967: 'Magnetic disturbances in Antarctic', Gidrometeoizdat, Leningrad.
Zmuda, A. J., F. T. Heuring and J. H. Martin, 1967: J. Geophys. Res., **72**, 1115.
Zmuda, A. J., J. C. Armstrong and F. T. Heuring, 1970: J. Geophys. Res., **75**, 4757.

Geofysiske Publikasjoner, Vol. 29, 1972

MIDDAY AURORAS AND POLAR CAP AURORAS

BY S.-I. AKASOFU
GEOPHYSICAL INSTITUTE
UNIVERSITY OF ALASKA, COLLEGE, ALASKA

Abstract

Characteristics of midday auroras (which occupy the noon sector of the auroral oval) and polar cap auroras (which appear inside the area surrounded by the auroral oval) are described by reviewing earlier studies and also new satellite and airborne all-sky data. The relationships between the midday auroras and polar cap auroras are also discussed. There is little doubt that behaviors of the midday and polar cap auroras will give us many clues on the interactions between the solar wind and the geomagnetic field.

1. Introduction

It is possible to classify auroras into two groups, the *oval auroras* and the *polar cap auroras*. The oval auroras appear within the narrow oval band, the auroral oval, which encircles the dipole pole (Feldstein, 1963, 1966; Khorosheva, 1962; Lassen, 1963; Malville, 1964; Davis, 1962; Sandford, 1964). The auroral oval is eccentric with respect to the dipole pole, and its center is displaced toward the dark hemisphere by about 3°. Auroras which occupy the noon sector of the auroral oval are called *midday auroras*.

Piddington (1965) suggested that the region encircled by the auroral oval should be called the polar cap, rather than the region encircled by the auroral zone. Since the auroral oval delineates approximately the intersection line between the outer boundary of the trapping region and the ionosphere, geomagnetic field lines which originate in the polar cap are supposed to be greatly distorted by the solar wind and form the magnetospheric tail, while those that originate in the middle-low latitude region are imbedded in the trapping region. In the polar cap, another group of auroras occurs. Such auroras are called the *polar cap auroras*. The polar cap auroras appear across the oval and their alignment is approximately parallel to the Sun-Earth line (Mawson, 1925 p. 181; Davis, 1960, 1962; Denholm, 1961).

Sightings of faint auroras in the daytime in very high latitude regions and also near the dipole pole have been reported frequently in records of Arctic and Antarctic expeditions. The report of the Swedish expedition to Spitzbergen (1887) during the First Polar Year (1882/83) includes a number of sketches of the midday auroras. However, the first systematic and extensive study of the daytime auroras was conducted by Lassen (1961).

The purpose of this study is to describe the characteristics of the midday auroras and the polar cap auroras, particularly their dynamical characteristics. These auroras may provide a vital clue in understanding the interactions between the solar wind and the geomagnetic field.

2. Midday auroras

In the northern hemisphere, the dipole (dp) sector, which is bounded by dp longitude lines 90° and 270°, is suitably located to study midday auroras (Buchau et al., 1969). Both Pyramida (dp lat 74.5°N) and the Ice Floe T-3 (dp lat 78.3°N) are in this sector. In the southern hemisphere, the South Pole station (dp lat 78.5°S) is an ideal station to observe the midday aurora. These stations are located geographically high enough (so that the 24-hour darkness is secured), but geomagnetically low enough to see the midday part of the oval; Fig. 1 shows the location of the oval when the Pyramida station (marked by a dot) is in the midday sector. Figure 2

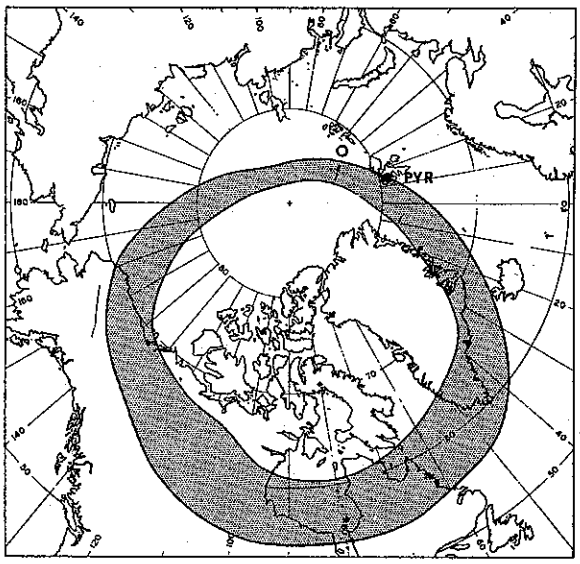


Figure 1. Approximate location of the auroral oval at 08 UT. The direction of the Sun is indicated by a circle; the location of Pyramida station is indicated by a dot.

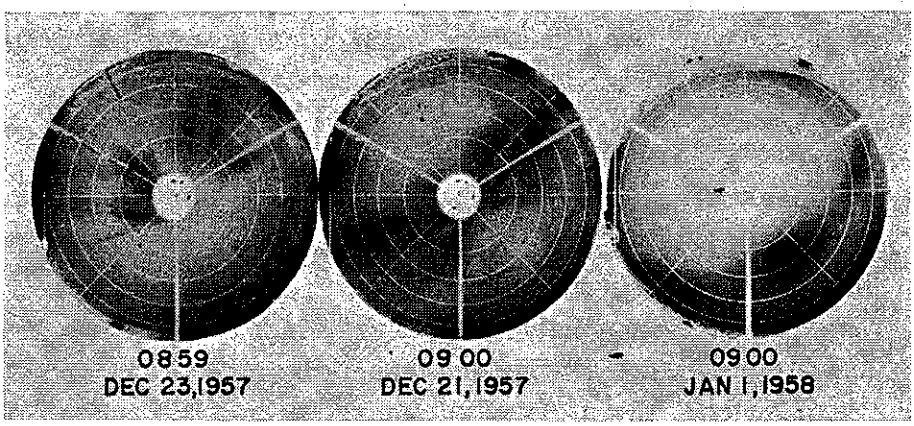


Figure 2. Example (in negative) of the midday auroras observed at Pyramida.

shows three photographs (in negative) of the midday auroras taken from Pyramida. Figure 3 shows a photograph of the midday aurora taken from off the Siberian Coast.

The midday auroras are mostly faint rayed arcs (or bright rayed arcs in the bright background). Feldstein (1966) suggested that they consist of isolated rays (see his Fig. 6), but there were many days during the IGY when the orientation of the rayed arcs could be definitely measured, so that it seems to be more appropriate to treat them as rayed arcs.

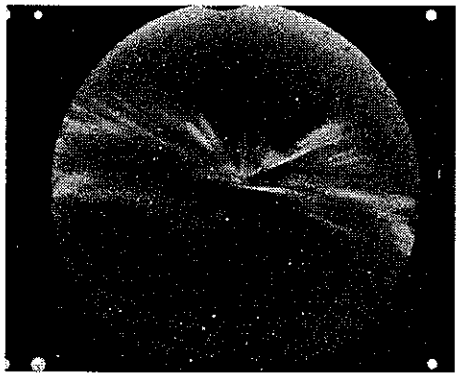


Figure 3. A photograph of the midday aurora taken from an AFCRL aircraft off the Siberian Coast at 0405 UT on December 6, 1969.

In Fig. 2, an arc is seen in the northern sky of Pyramida on December 23, and about the zenith on December 21, and in the southern sky on January 1. The first one (December 23, 1957) corresponds to a quiet condition (Kp = 1+, 1+, 1o, 1+, 1o, 1–, 1–, 1–), the second to a moderately active condition (Kp = 3–, 5–, 4o, 2+, 2o, 2o, 2o, 2o) and the third to very disturbed condition (Kp = 6o, 6+, 5+, 4o, 3+, 4–, 5+, 4o). The D_{st} values (Sugiura, 1963) at the time when the above three photographs were taken were −3, −32 and −103 γ, respectively. It can be seen clearly that the location of the aurora shifts toward lower altitudes as magnetic activity increases. This suggests an expansion of the oval in the day sector (Feldstein, 1966; Zmuda et al., 1966). Akasofu and Chapman (1963) showed that the expansion of the oval in the night-side depends on the intensity of the storm-time radiation belt (the ring current).

Stringer (1966) showed that the midday aurora appeared with a very high occurrence frequency over the T-3 station (practically every day on which observation was possible) even when geomagnetic conditions were so quiet that the midnight auroras were very faint or invisible. Based on ascaplots, Feldstein (1966) obtained the occurrence frequency of approximately 80–85%. When we consider the fact that the auroras appear in the bright twilight sky, this high occurrence frequency is remarkable. Therefore, it is not difficult to infer that the midday auroras are essentially a permanent feature of the Earth.

The midday aurora tends to align along the east-west direction. Figure 4 shows the orientation measured for all the available data obtained during the IGY in Pyramida, together with all that obtained by Feldstein (1961). Stringer (1966) showed the same tendency by using the records taken from the T-3 station. As seen in Fig. 5 the aurora may be single or multiple in form.

During high geomagnetic activity, a violent eastward motion of the auroras occurs in the

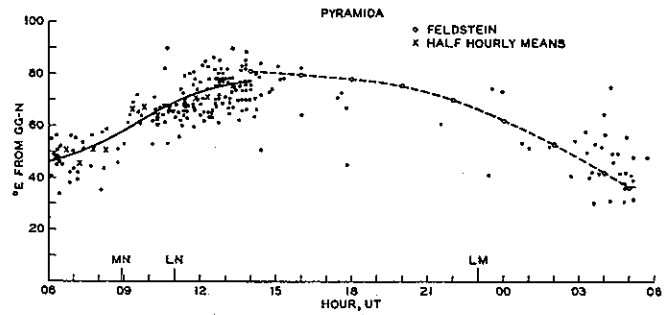


Figure 4. The orientation of auroral arcs observed at Pyramida; MN, LM, LM, refer to the magnetic noon, the local noon and midnight, respectively.

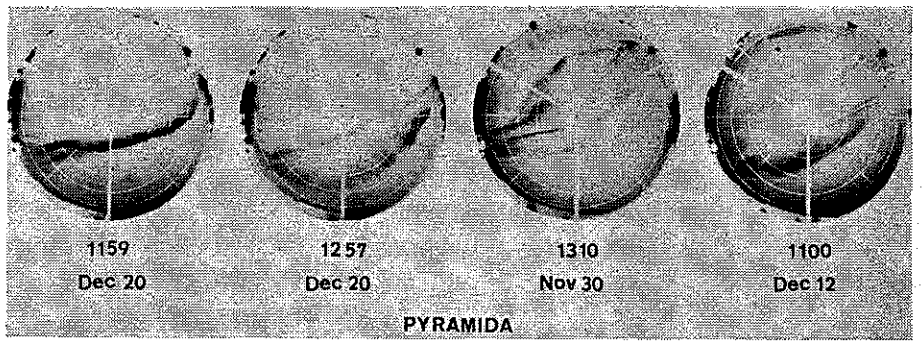


Figure 5. Examples (in negative) of the midday auroras; except the first one (Dec. 20, 1959), they are multiple.

midday sector of the oval. Figure 6 shows an example of the eastward motion of the midday aurora, observed at Pyramida on January 1, 1968. Since the eastward motion is the most common feature of the aurora during the auroral substorm in the morning sector of the auroral oval, it is natural to infer that the eastward motion of the daytime auroras is an extension

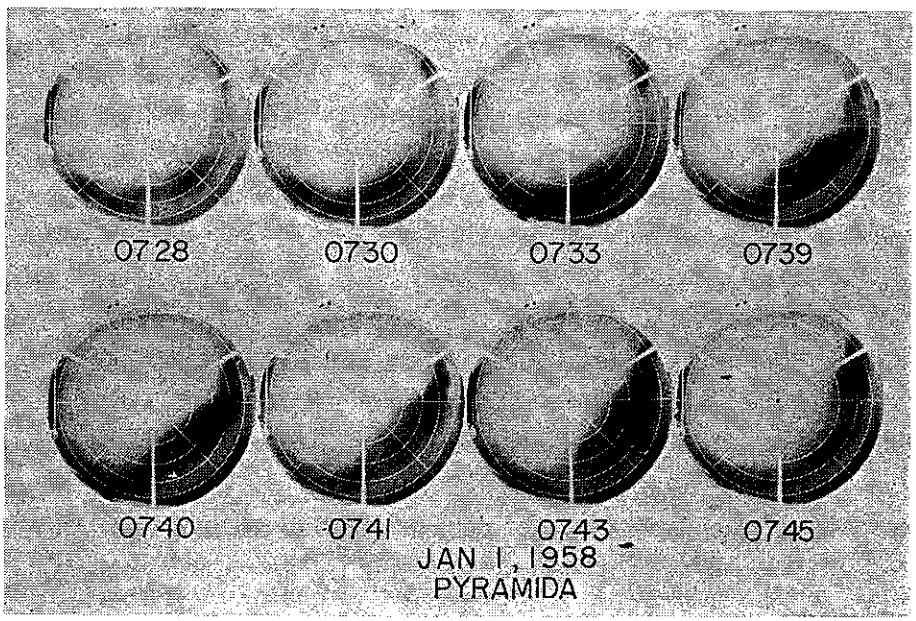


Figure 6. An eastward motion of the midday aurora, observed on January 1, 1958.

from the early morning sector. This suggests that the midday auroras are affected by the auroral substorm activity which originates in the midnight sector of the oval.

It is also worthwhile to mention here auroras which occupy the forenoon and afternoon sector of the auroral oval. Most of the morning auroras are also faint rayed arcs and have been studied most extensively by Lassen (1961), using his visual and all-sky camera observations at Godhavn, Greenland (dp lat 79.9°N). He has found that the percentage hourly occurrence of the morning auroras is very high, 90–100%, and that there are no significant magnetic disturbances associated with the auroras.

Based on the Alaskan north-south chain stations during the IQSY (including the Ice Floe Station T-3 in the Arctic Sea), Stringer et al. (1965) and Stringer (1966) have shown also that auroras appear with a very high occurrence frequency (practically every day on which observation was possible) over the T-3 station (dp lat 78°); in fact, even during an extremely quiet period (both the local *K* and planetary *Kp* indices were zero), morning auroras were observed there, in spite of the fact that auroras in the midnight sector faded below the threshold of a photometer (0.5 kR), as well as that of the all-sky camera.

Lassen (1961) noted that there is a tendency for the overhead aurora to occur less frequently on days of higher planetary disturbance than on other days. This negative correlation of very high latitude auroras with geomagnetic activity was also discussed by Davis (1963). Figure 7 shows some of the all-sky camera photographs (in negative) taken from Resolute Bay (dp lat 84.2°N). Active auroras at such a high latitude region do not occur in the daytime during very disturbed days; in fact, December 28, 1957 was one of the most quiet days during the IGY (*Kp* = 1o, 1+, 1–, 1–, 1o, 1o, 1+, 1+). There is little doubt that this negative correla-

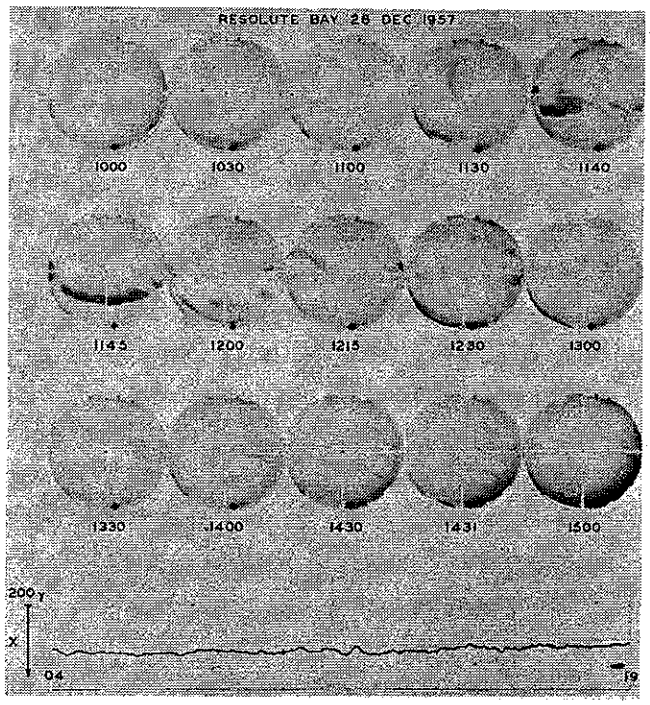


Figure 7. Examples (in negative) of the late morning auroras observed at Resolute Bay on December 28, 1957; the bottom trace indicates the X component magnetic record.

tion is due partly to the contraction of the auroral oval toward the dipole pole during quiet periods and partly to the equatorward expansion during disturbed periods.

Observing the morning auroras at Godhavn, Lassen (1961) noted that most of them tend to appear directly near the zenith, rather than drift toward the zenith. On the other hand, at the T-3 Ice Floe, it is quite common to observe arcs which move perpendicularly to them, both poleward and equatorward, in addition to the familiar eastward drift motion (of speed of order 300 m/s). The poleward motion is far more common than the equatorward motion, so that many arcs move toward the zenith from the southern sky, and also many drift to the northern sky after they appear first near the zenith. Like the midday auroras, their lifetime is short, of order 5–10 min, and their speed is of order 100–500 m/s.

When they are very bright, morning auroras are often associated with very weak changes of the geomagnetic field. For example, in Fig. 7, a slight geomagnetic variation was recorded between 1130 and 1145 UT when the auroras were exceptionally bright. In general, the magnitude of the geomagnetic change is not more than 50γ and lasts only for 15–30 min.

Lassen (1963) pointed out that there is no essential difference between the morning auroras and the afternoon auroras. The afternoon auroras also are mostly rayed arcs. However, evening auroras tend to be in the form of a homogeneous arc, except during the auroral substorm. It is interesting to examine whether or not such a transition occurs gradually and whether or not there exists an indication that the daytime auroras and the nighttime auroras are different systems.

3. **Polar cap auroras**

Very little is known about the polar cap auroras except that they tend to align parallel to the Sun-Earth line and that they appear to fade out or do not appear during intense geomagnetic activity (Davis, 1963; Akasofu, 1964). Their Sun-Earth line alignment was first demonstrated by Mawson (1925) and confirmed later by Davis (1960, 1962), Denholm (1961) and others on the basis of extensive all-sky film records. Figure 8 shows examples of the polar cap auroras (in negative) which appeared over Resolute Bay at different UT times (0600 UT = 0000 LT); the direction of the Sun is indicated by a dot in each photograph.

Motions of the polar cap auroras, which can be studied with reasonable accuracy by the present all-sky camera, are those which are perpendicular to the arcs. Figure 9 shows examples of the motions of polar cap auroras. In general, the speed is of order 100 m/s (Danielsen, 1969), and there does not seem to be any systematic change of the direction as a function of local time; it may be noted, however, that if polar cap auroras are fixed with respect to the Sun, and the Earth rotates underneath, there will be an apparent motion (because of the eccentricity of the oval with respect to the geographic pole); at Resolute Bay, it is a westward motion with a speed of order 30 m/s in the midnight sector.

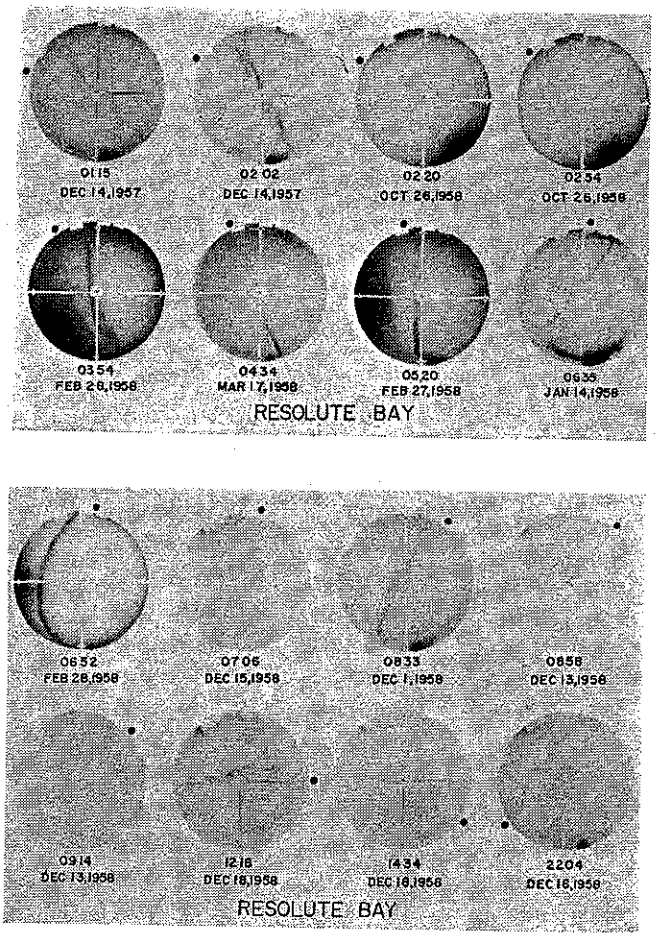


Figure 8. Examples (in negative) of the polar cap auroras observed at Resolute Bay at different UT. The direction of the Sun is indicated by a dot.

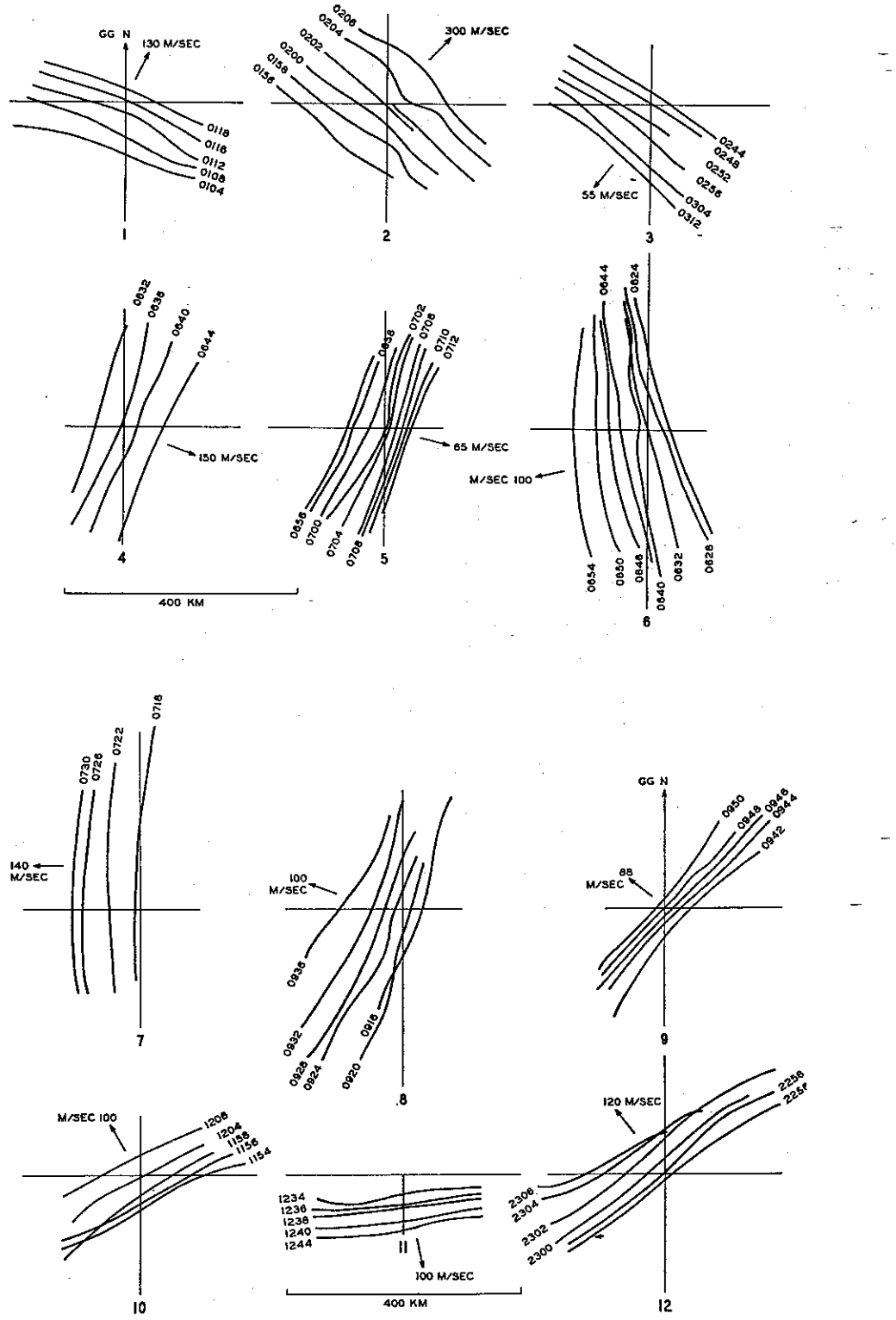


Figure 9. Examples of the motions of the polar cap auroras observed at Resolute Bay at different UT hours. (1) December 14, 1957; (2) December 14, 1957; (3) October 26, 1958; (4) January 14, 1958; (5) February 28, 1958; (6) February 28, 1958; (7) December 14, 1958; (8) December 13, 1958; (9) December 13, 1958; (10) December 15, 1958; (11) December 15, 1958; (12) December 15, 1958.

4. Relationships between the midday auroras and polar cap auroras

In all-sky films, the polar cap auroras appear to have the same features as the daytime auroras. Therefore, if they appear near the oval in the mid-morning or mid-afternoon sectors, it is not possible to distinguish them from the oval auroras.

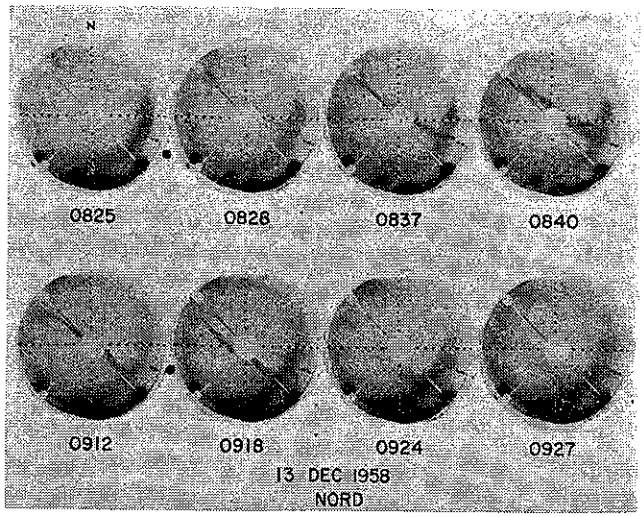


Figure 10. All-sky camera photographs (in negative) showing the co-existence of oval auroras and polar cap auroras taken from Nord, Greenland, December 13, 1958. The top of the photograph is toward the geographic North Pole. The directions of the Sun at 0825 UT and 0912 UT are shown by dots.

There are some indications that the polar cap auroras are 'connected' to oval auroras in the day sector. Figure 10 shows all-sky photographs taken from Nord (dp lat 80.8°), Greenland on December 13, 1958. Figure 11 shows the ground projection of auroras observed, at 0746 UT on the same day, in the field of view at Nord and Alert. The radius of the circle around Nord and Alert is 800 km. The approximate direction of the Sun and the location of the auroral oval are indicated.

This particular example occurred during the long initial phase of the geomagnetic storm of December 13, 1958 (cf. at 0001 UT); the *Kp* index was 3+ during 06–09 UT on December 13. Inside the polar cap, the Sun-Earth aligned arc was also seen at Alert (dp lat 85.9°) between 0830 UT and 0925 UT.

During the same period, Alaskan stations, i.e., Barrow and Bettles, were in the midnight sector of the auroral oval. Only very faint auroras were seen between 0727 UT to 1300 UT; and no polar substorm activity was observed until 1303 UT.

Figure 12 illustrates another example of a photograph which shows both the midday aurora and polar cap aurora. This photograph was taken from Nord on January 26, 1959.

From these observations, it appears that the polar cap auroras joined perpendicularly to the oval aurora. However, there are many frames to suggest that the polar cap aurora is simply a segment of the oval auroras, which loops away from the oval at a certain point toward the cap along the Sun-Earth line. In fact, although it is not obvious in the figures, an inspection of the films with a movie projector shows that as the polar cap aurora moves, the

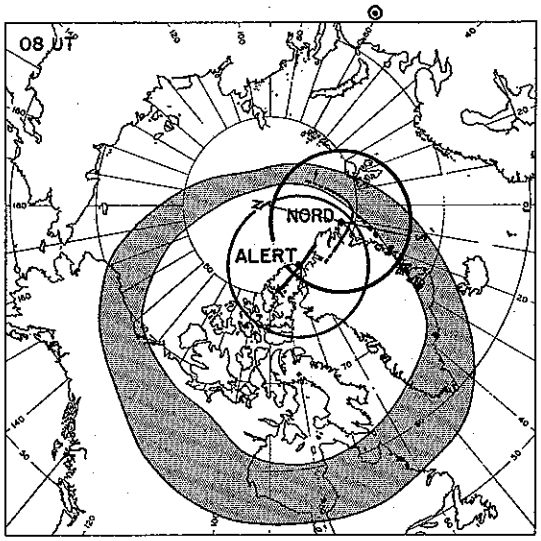


Figure 11. The locations of Nord and Alert together with a rough location of the auroral oval at 08 UT. The direction of the Sun in indicated by a dot within the circle. The ground projection of auroras at 0746 UT on December 13, 1958 and the field of view of all-sky camera at Nord together with the approximate observed Sun-aligned arc orientation and field of view of the camera at Alert are also shown.

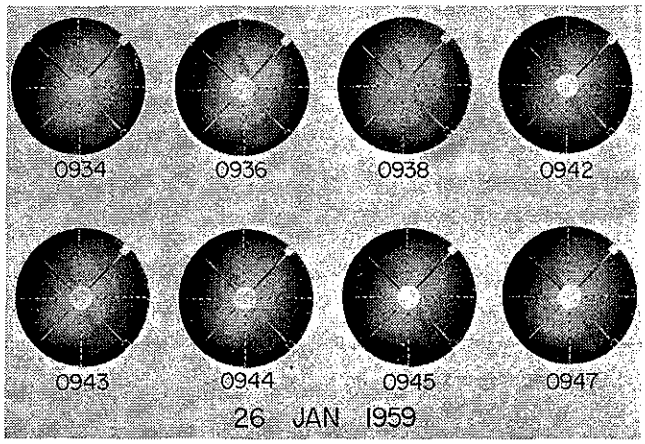


Figure 12. All-sky camera photographs (in negative) showing the co-existence of oval auroras and polar cap auroras taken at Nord, Greenland, January 26, 1959. The top of the photograph is toward the geographical North Pole.

'inflection point' also moves with it. It is not certain whether such an observation confirms the suggestion made by Gustafsson (1967).

It should be mentioned that such a geometrical relation between the 'oval auroras' and polar cap auroras does not seem to exist in the midnight sector. In spite of the fact that we scanned a considerable number of all-sky films, we have been unable to find a single frame to show the 'connection' of both types of auroras.

5. Discussion

During the last few years, there has been a considerable progress in the studies of the midday auroras and polar cap auroras. This is partly because intense fluxes of soft electrons ($\approx$ 300 eV) above the location of the midday aurora have been detected by OGO-4 (Hoffman and Berko, 1971), ISIS-1 (Heikkila and Winningham, 1971), Injun-5 (Frank and Ackerson, 1971) and IMP-5 (Frank, 1971) and because those electrons have an energy spectrum which is very similar to that of electrons in the magnetosheath. Such observations suggest strongly that magnetosheath electrons penetrate into the magnetosphere, perhaps through the neutral line on the magnetopause.

Before such an important satellite finding, airborne observations of the midday auroras had been carried out by Buchau et al. (1969, 1970), Whalen et al. (1971), Eather and Mende (1971), and Akasofu (1970) over the Greenland sea and off the Siberian Coast. These studies revealed that the midday part of the oval is a 'gigantic subvisual red belt' and its spectrum is characterized by intense emission of the [OI] 6300 Å. The magnetosheath electrons have a suitable energy to produce such an enhanced [OI] 6300 Å emission compared with other familiar auroral emissions, such as the [OI] 5577 Å and $[N_2^+]$ 3914 Å emissions.

The belt has the north-south width of a few hundred kilometers. Within the belt, there is a single or multiple bright curtain and such curtains have a form similar to those of nighttime auroras. The difference is that the 6300 Å emission is more intense than the 5577 Å emission (Heikkila et al., 1971). Polar cap auroras have similar spectral characteristics. Further, when such red curtains become very bright, the green (5577 Å) emission appears near the lower edge. Thus, there must be more energetic electrons in the curtains than in the belt. An acceleration process similar to that of nighttime electrons may thus exist in the day sector, as well as in the night sector.

It has been noted by the satellite observations mentioned earlier that the magnetosheath electrons are seen just on the poleward side of the trapping boundary. Therefore, the midday auroras give visible information as to the location of the intersection line between the outer boundary of the trapping region and the polar ionosphere.

Thus, behaviors of the midday auroras should provide extremely useful information on the amount of the magnetic flux which is transferred from the day sector of the magnetopause to the magnetotail.

Indeed, it has been found by Akasofu (1972) that during an early phase of a substorm the midday auroras move equatorward, indicating that the trapping boundary moves also equatorward and thus that some 'closed' field lines in the day sector become 'open' and are transferred to the magnetotail. It has been suggested that such an 'erosion' of the dayside magnetosphere is caused by the merging of interplanetary and geomagnetic field lines when the former has a southward component. It is of great importance to compare north-south motions of the midday auroras with the simultaneous interplanetary magnetic data, since the range of the north-south motion should provide quantitative information of the amount of the transferred magnetic flux.

The geomagnetic field lines which originate in the polar cap 'constitute' the high latitude magnetotail, that is, the magnetotail region outside the plasma sheet. The plasma density in

this region is far less ($< 0.01/cm^3$) than that in the surrounding regions, such as the plasma sheet ($\approx 0.1/cm^3$) and the magnetosheath ($\approx 10/cm^3$). It is of great interest to examine why magnetosheath electrons appear along a group of field lines in such a way that those field lines form a sheet oriented in the north-south direction; note that the spectral characteristics of the polar cap aurora are very similar to those in the midday aurora, so that it is quite likely that polar cap auroras are produced by magnetosheath electrons. At a great distance, say $r_{eq} \sim 100$ earth radii, some of the field lines in the high latitude magnetotail must lie in the magnetopause or in the magnetosheath. Further, the polar cap auroras move in the direction perpendicular to the Sun-Earth line. What does this motion tell us? There is little doubt that the behaviors of the polar cap auroras will tell us many clues on the interactions between the solar wind and the geomagnetic field.

Acknowledgement

The author would like to thank Dr. G. Gassmann and Mr. J. Buchau, AFCRL, Dr. C.-I. Meng, and Mr. D. S. Kimball for their discussion during the preparation of this paper. This work was supported by the Atmospheric Sciences Section of the National Science Foundation through grant GA-28101.

References

Akasofu, S.-I., 1964: Planet. Space Sci., **12**, 273.
Akasofu, S.-I., 1970: Trans. Amer. Geophys. Union, **51**, 370.
Akasofu, S.-I., 1972: J. Geophys. Res. (in press).
Akasofu, S.-I. and S. Chapman, 1963: J. Atmosph. Terr. Phys., **25**, 9.
Buchau, J., J. A. Whalen and S.-I. Akasofu, 1969: J. Atmosph. Terr. Phys. **31**, 1021.
Buchau, J., J. A. Whalen and S.-I. Akasofu, 1970: J. Geophys. Res., **75**, 7147.
Carlheim-Gyllenskiold: 'Exploration Internationale des Regions Polaries' 1882/86, Exp. Suedoise, Aurorae Boreale, Stockholm.
Danielsen, C., 1969: Danish Met. Inst., Geophys. Papers R9.
Davis, T. N., 1960: J. Geophys. Res., **65**, 3497.
Davis, T. N., 1962: J. Geophys. Res., **67**, 75.
Davis, T. N., 1963: J. Geophys. Res., **68**, 4447.
Denholm, J. V., 1961: J. Geophys. Res., **66**, 2105.
Eather, R. H. and S. B. Mende, 1971: J. Geophys. Res., **76**, 1746.
Feldstein, Y. I., 1961: 'Investigations of the Aurorae', Aurorae and Airglow, No. 4.
Feldstein, Y. I., 1963: Geomag. Aeronomy, **3**, 183.
Feldstein, Y. I., 1966: Planet. Space Sci., **14**, 121.
Frank, L. A., 1971: J. Geophys. Res., **76**, 5202.
Gustafsson, G., 1967: Planet. Space Sci., **15**, 277.
Heikkila, W. J. and J. D. Winningham, 1971: J. Geophys. Res., **76**, 883.
Heikkila, W. J., J. D. Winningham, R. H. Eather and S.-I. Akasofu, 1971: J. Geophys. Res., (submitted).
Hoffman, R. A. and F. W. Berko, 1971: J. Geophys. Res., **76**, 2967.
Kawasaki, K. and S.-I. Akasofu, 1971: J. Geophys. Res., (submitted).
Khorosheva, O. V., 1962: Geomag. Aeronomy, **2**, 696.
Lassen, K., 1961: Pub. Danske Met. Inst., No. 15.
Lassen, K., 1963: Pub. Danske Met. Inst., No. 16.
Malville, J. M., 1964: J. Geophys. Res., **69**, 1285.
Mawson, D., 1925: 'Australian Antarctic Expedition, 1911–14', Sci. Rep., Ser B, II, 1.
Piddington, J. H., 1965: Planet. Space Sci., **13**, 565.
Piddington, J. H., 1965: Planet. Space Sci., **13**, 363.
Sanford, B. P., 1964: J. Atmosph. Terr. Phys., **26**, 749.
Stringer, W. J., 1966: M. S. Thesis, Geophysical Institute, University of Alaska.
Stringer, W. J., A. B. Belon and S.-I. Akasofu, 1965: J. Atmosph. Terr. Phys., **27**, 1039.
Sugiura, M., 1963: Ann. I.G.Y., **35**, 9.
Whalen, J. A., J. Buchau and R. A. Wagner, 1971: J. Atmosph. Terr. Phys., **33**, 661.
Zmuda, A. J., J. H. Martin and F. T. Heuring, 1966: J. Geophys. Res., **71**, 5033.

Geofysiske Publikasjoner, Vol. 29, 1972

ON THE CLASSIFICATION OF HIGH-LATITUDE AURORAS

By Knud Lassen
DANISH METEOROLOGICAL INSTITUTE,
GEOPHYSICAL SECTION II,
CHARLOTTENLUND, DENMARK

Abstract

Polar graphs showing the distribution of mean auroral frequencies as observed from the network of all-sky cameras in Greenland are presented for values of *Kp* from 0 to 5. The grouping of the frequencies in the plots and the dependence of this on *Kp* as well as on the sunspot activity is discussed. Details are presented from single nights, and finally, with support from a survey of the position in the polar graph of observed arcs and bands at $Kp = 1$, the following model of the auroral grouping is inferred; the main precipitation curve is a spiral-like figure, situated in the morning at 75°–80° between 0300 corrected geomagnetic time and the noon sector. From this it decreases monotonically in latitude through the afternoon and evening hours down to 65°–70° at 0200.

This precipitation is mixed up with a night-system between 70° and 75° in the midnight sector, as well as with a system of polar cap arcs converging from the whole night sector at 70°–80° towards the cusp region in the noon sector. The model presented here implies a discontinuity in the main precipitation curve in the early morning; it is valid for quiet magnetic conditions.

1. Introduction

During the International Geophysical Year 1957–59, four all-sky camera stations were operated in Greenland (Lassen, 1963). In 1963 the stations were established again, and in 1964 two new stations were added (Lassen and Rud Laursen, 1968). Thus, every winter since 1964 a network of six stations, well distributed between corrected geomagnetic latitudes 68° and 86°, has been operated with exposures once every minute.

Results of statistical studies of the photographic observations supplemented by visual observations from the pre-IGY period will be discussed in the following sections.

2. Grouping of the auroras

For the six stations in Greenland polar graphs were constructed for each of the years 1964-65

to 1968-69 and for each value of *Kp* from 0 to 5. The polar graphs are represented as corrected geomagnetic latitude (CGL) vs. time dials; the parameter shown in the diagram is the percentage occurrence frequency of zenithal auroras (zenith distance $< 60°$), defined as the ratio within each hour of the number of quarter-hourly intervals during which auroras were observed to the number of quarter-hourly intervals during which observation was possible. The polar graphs for a given value of *Kp* did not differ significantly from each other from year to year. To increase the accuracy of the plots, all years were therefore combined to show the variation

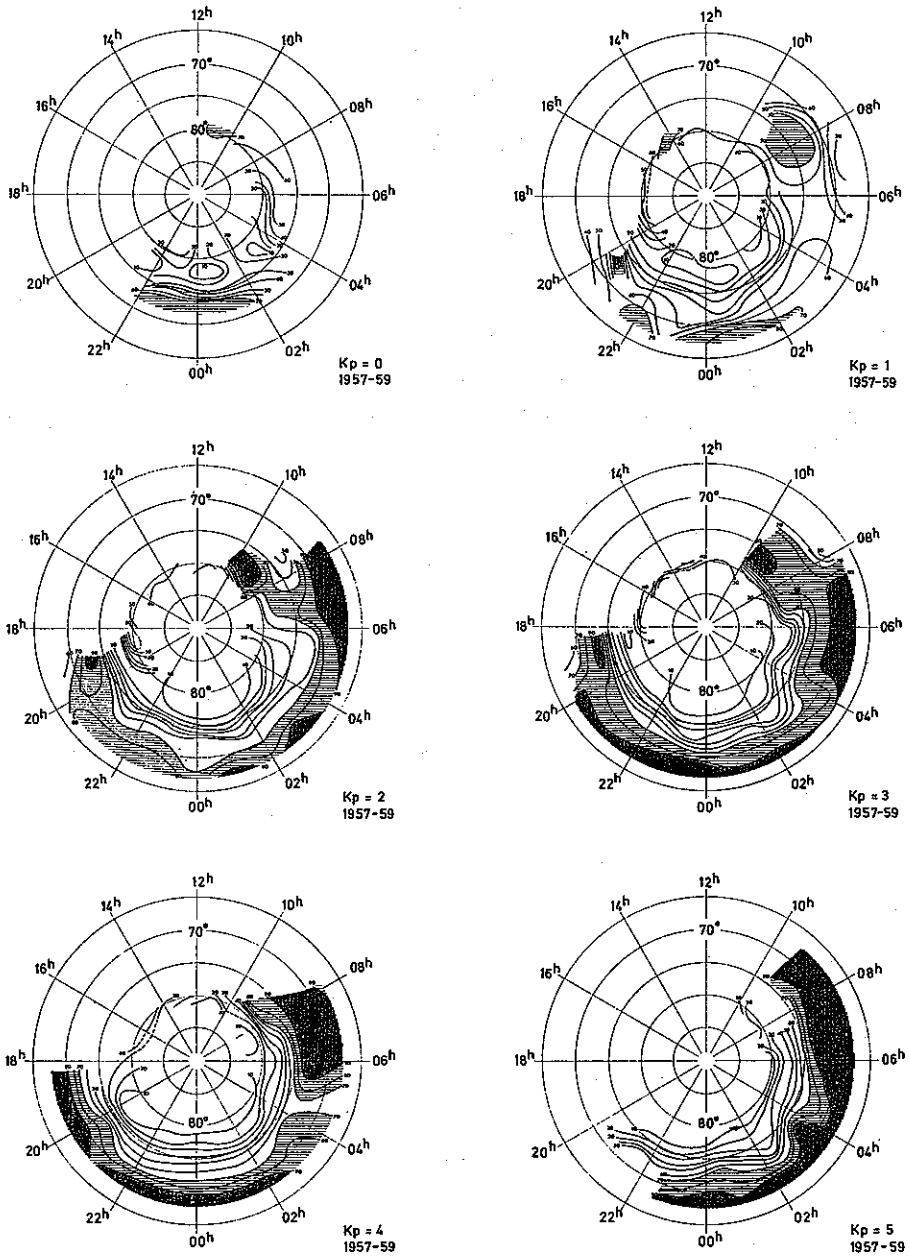


Figure 1. Polar graphs showing distribution in corrected geomagnetic coordinates of occurrence frequency of auroras for $Kp = 0$ through $Kp = 5$. Period: 1964–69. Isoauroral lines are based on zenithal frequency and drawn for multiples of 10 per cent. Frequencies greater than 90% are indicated by dense hatching; frequencies between 70% and 90% are lightly hatched.

with Kp for the whole period 1964–69. The resulting diagrams are shown in Fig. 1. Isolines with an interval of 10% have been drawn in the graphs to give a survey of the distribution of the frequencies.

The graphs are only valid for latitudes higher than 67°. They are empty between 1000 and 1800 CGT at latitudes below 80°. In December and January observations are possible down to about 75° (from Nord), but since they are not zenithal, they have been omitted here.

The polar graphs in Fig. 1 confirm earlier findings (Lassen, 1970). For $Kp \geq 3$ the isolines form an oval as demonstrated by Tromholt (1880), Feldstein (1963), and Lassen (1963). But for lower values the existence of several individual precipitation areas is clearly indicated. Thus, for $Kp = 0$ the absolute maximum of occurrence frequency is situated in the late morning hours at 78°. The maximum value exceeds 90%, and the auroras producing the maximum occupy an area in the polar graph which follows the 78° latitude-curve till early morning. A symmetrically situated distribution in the afternoon may be connected with the morning area in the midday hours, thus forming one horseshoe-shaped day-time distribution, but although midday auroras are known to exist at the proper latitude, there is still a possibility that the morning and the afternoon groups are two separate populations.

The well known night auroras are less frequent at low Kp-values. Thus, whereas during magnetically quiet intervals auroras are present on practically all mornings, night auroras are at $Kp = 0$ observed with a maximum frequency not greater than $\approx 50\%$ shortly after midnight. In the graphs, the night area is overlapping with the day distribution(s) in the evening and early morning, but except for high degrees of geomagnetic disturbance there exist separate night, morning, and afternoon peaks. Combining this with the fact that day-time auroras differ morphologically from night-time auroras, in the present discussion it is assumed that the auroral oval, although admittedly formally and in daily speech a practical concept which is not easily replaced by any other expression, is not a physical unity but a combination of several distributions of auroras, each of which may be formed by particles of characteristic energies and arriving from different parts of the magnetosphere. On the polar side of the oval distributions, polar cap auroras are situated in two areas at the morning and evening side of the polar cap, respectively. The mutual connection between these two distributions and between them and the oval auroras is not clear and has so far not been discussed in the literature.

3. Night auroras

In the frequency plots the group of night auroras is centred at about 0100 CGT. For $Kp = 0$ the peak is situated at 72.5° CGL with a maximum value $\approx 50\%$, increasing rapidly to at least 90% at $Kp = 2$ at the same time as the peak latitude decreases to a latitude estimated to be just below the 67° border of the polar graph. Also, with increasing geomagnetic activity the night distribution rapidly expands towards morning as well as evening hours where it overlaps with the daytime auroras.

The night auroras are the usual auroral zone auroras including the brilliant substorm phenomena. They are discussed here only in connection with their possible relation to the day-time auroras.

4. Morning auroras

The existence of a population of morning auroras at 75°–80° CGL was reported by Lassen (1959a, b), who also published a detailed description of their morphology (Lassen, 1961a, 1963).

The auroras were not correlated with the simultaneously recorded geomagnetic noise. Only in a few cases of extremely high brightness could a small perturbation be demonstrated in the magnetograms. The missing magnetic effect has been regarded as a hint of the fact that morning auroras are generated by electrons of lower energy (less penetrating into the atmosphere) than those generating night auroras. This view was supported by the ionospheric effects accompanying the auroras. Invariably the *F* layer was enormously disturbed; during the more active part of the auroral display auroral E_s was observed, but generally not in virtual heights below 120–150 km. No appreciable absorption could be demonstrated in connection with this type of auroras.

The greater height of the day-time auroras was demonstrated directly by triangulation using two all-sky camera stations at Spitzbergen (Starkov, 1968). The distribution with height showed a main maximum at 150 km and a secondary maximum at 175 km. Using a similar technique for a pair of stations in Greenland, Lassen (1969) confirmed that morning auroras are situated at greater heights than night auroras, but the maximum was situated at 130 km. The discrepancy between the two results appears to be explained by the fact that most of Starkov's measurements were taken later in the forenoon than Lassen's (Starkov, personal communication). Taken together, the two results may thus indicate a gradual increase of the auroral height from early morning towards noon. From his height measurements, Starkov concluded that the morning auroras could be generated by electrons of energy 0.5–1.0 keV.

The early observations from Godhavn (77.5° CGL) showed a slight decrease of the occurrence frequency with planetary magnetic disturbance. From Fig. 1 this is seen to be due to an equatorwards displacement of the distribution rather than to a decrease in the maximal frequency of occurrence. Whereas the latter appears to be greater than 90% for any *Kp*, the latitude of the maximum at 0900–1000 CGT gradually decreases from a centre value of $\approx$ 77.5° at $Kp = 0$ to $\approx$ 75° at $Kp = 4$ and even lower (72°?) at $Kp = 5$ at the same time as the morning distribution is developing farther and farther towards the midnight hours. The latitude variation is in reasonable agreement with observations by Feldstein and Starkov (1967). It may in part explain the negative correlation with the sunspot number reported on the basis of visual and photographic observations as well as ionospheric effects. An increased number of magnetically disturbed days at sunspot maximum will displace the peak equatorwards from Godhavn and thereby lower the zenithal frequency at that station. This reasoning is based on the assumption that the distributions with *Kp* are independent of the sunspot number. The similarity between the single years of the period 1964–69 seems to confirm this. A comparison with the extreme sunspot maximum period 1957–59 (Fig. 2) shows that the assumption is essentially correct. In spite of the fact that the 69°-camera (Julianehåb) was operating only in the first winter and a camera at 75°, used in the construction of Fig. 1, was missing during IGY, it may be concluded from Fig. 2 that the morning peak is situated at approximately the same latitude as in Fig. 1, but concentrated in the latest morning hours. The average

frequency in the morning hours is thereby reduced; the deep minimum in the sunspot variation at Godhavn, observed in 1957–59, appears to be due to a combination of this effect with the equatorwards displacement of the peak with increasing *Kp*.

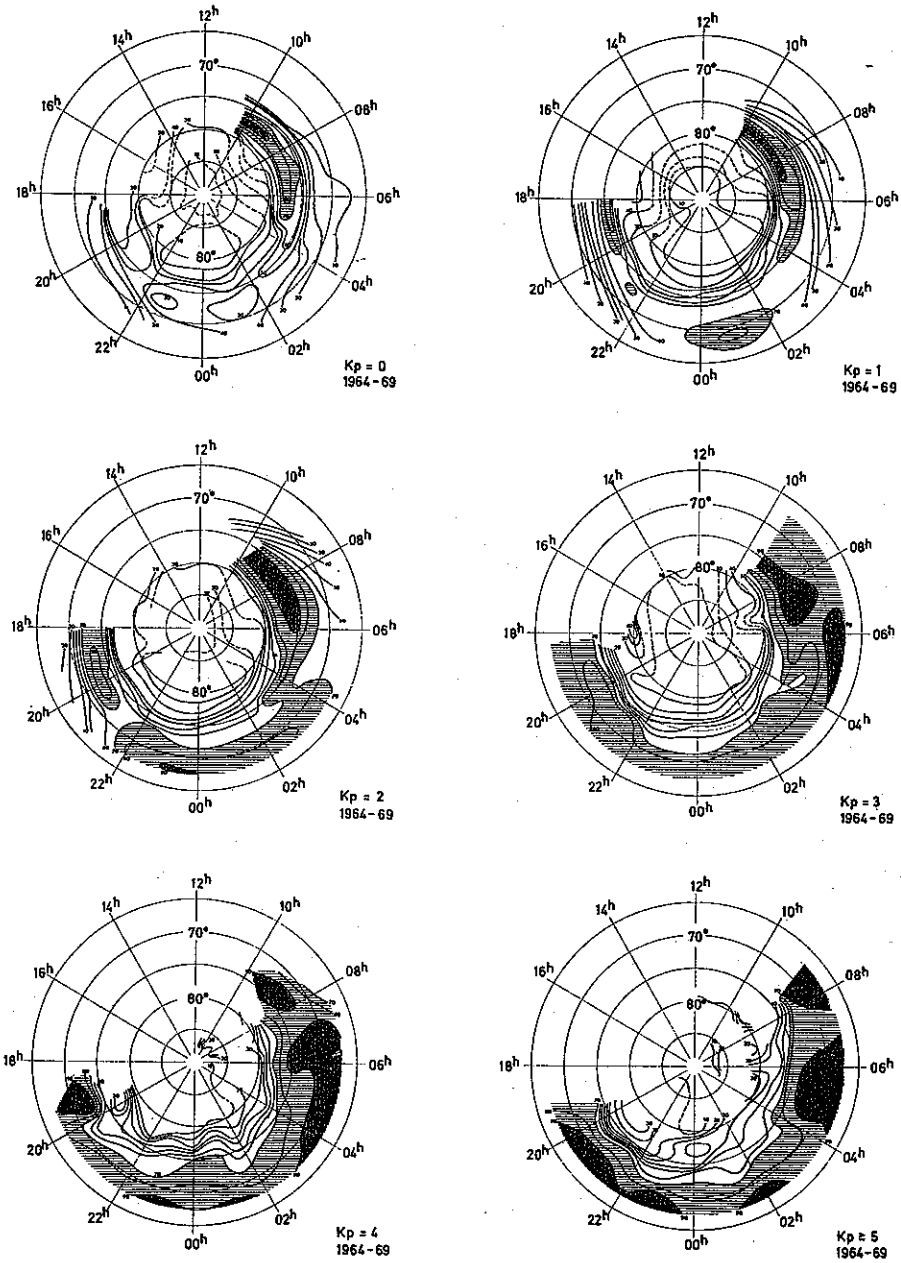


Figure 2. Polar graphs showing distribution in corrected geomagnetic coordinates of occurrence frequency of auroras for $Kp = 0$ through $Kp = 5$. Period: 1957–59. Isoauroral lines are based on zenithal frequency and drawn for multiples of 10 per cent. Frequencies greater than 90% are indicated by dense hatching; frequencies between 70% and 90% are lightly hatched.

The low value during IGY of the early morning frequency at 75°–78° is likely to have played an important role for the hypothesis that the auroras form one complete oval-shaped belt, even at low *Kp*-values.

Feldstein and Starkov (1968) have demonstrated that the polar distance of the 'oval' is dependent on the ring current, expressed by the equatorial D_{st}. This parameter reaches substantially higher values during 1957–59 than during 1964–69; however, the average values of D_{st} for the periods for which the polar graphs have been drawn, do not differ sufficiently to cause greater differences between Figs. 1 and 2. An estimated 1957–59 decrease in latitude of the night peak of about 1° is in agreement with the result of Feldstein and Starkov as well as with the difference between IGY and Second International Polar Year 1932–33, determined by Lassen (1963).

Lassen (1959, 1961a) observed that morning auroras lit up locally at 75°–80° without any connection with the night auroras in lower latitude. He therefore called the group of auroras an inner (day-time) auroral zone, in which auroras are present every day. On some mornings, however, the onset of the morning auroras was hidden by quiet rayed bands which gradually approached the station (Godhavn, 77.5°) from the south. These bands were regarded as a special type of auroras not belonging to the 'proper' morning auroras. They differed from these mainly in their being accompanied by moderate magnetic perturbations and ionospheric absorption. The occurrence of this type of auroras appeared to be relatively rare during the period of observation; however, it should not be neglected, since it will contribute to the difficulties in deciding, statistically as well as on single nights, whether the night and morning groups are independent distributions or not.

The position and evolution of the night and morning peaks and the surrounding isolines in Figs. 1 and 2 give the impression that the distributions are independent, developing along parallel, but slightly different latitudes in the early morning (cf. e.g. Fig. 1, $Kp = 2$). On the other hand, the statistical plots do not show a clear separation between the distributions, only a trough corresponding to a decrease in frequency of about 10% near 0300 CGT. Thus it is possible, when judged from the statistics, that the distributions are in general overlapping, or that their limits show a variation from day to day which makes the statistical pattern less distinct. In this connection it should be remembered, too, that the smoothing of quarter-hourly zenithal observations implies that every hourly value in the plot summarizes approximately two hours coverage at 75° latitude, so that even sharp limits between the distributions will have a tendency to be blurred by the procedure itself. Finally, it has been observed that late substorm auroras or the special bands mentioned above may fill out the expected gap between the distributions.

Thus, the interpretation of Figs. 1 and 2 does not promise that a discontinuity in the auroral distribution at about 0300 CGT can be found at any instance. Therefore, observations of auroral forms across the morning trough on a limited number of mornings (Buchau et al., 1970) give no proof of the existence of a continuous oval in the sense that the precipitation follows the border between the open and closed field lines all around the pole, unless it is proved, too, that the transition from night auroras to morning auroras takes place gradually with a smooth increase in latitude.

On the contrary, clear observations of discontinuities in latitude between night and morning auroras, as they are observed occasionally from the ground stations, are difficult to explain from the continuity hypothesis, whereas they are in complete agreement with the interpreta-

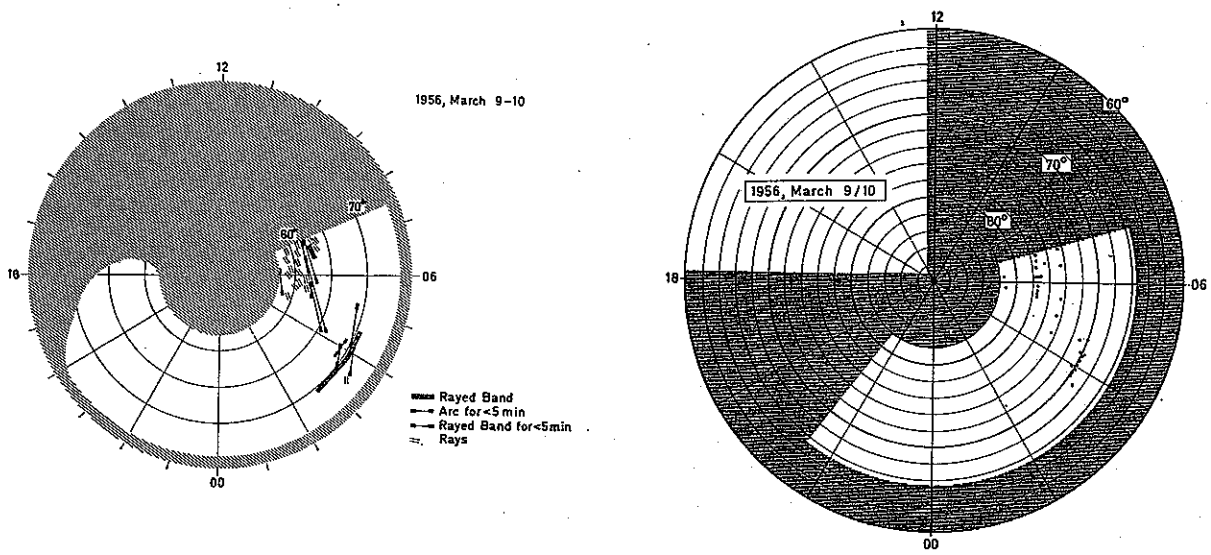


Figure 3a. Polar graph showing position in corrected geomagnetic coordinates of auroras observed from Godhavn on the morning of March 10, 1956.
Figure 3b. Same situation as in Fig. 3a, but only the position in the meridian of the auroral forms has been plotted.

tion given here of the frequency plots in Figs. 1 and 2. Examples of such observations are presented in Figs. 3–6.

Figure 3a shows in a representation introduced by Buchau et al. (1970) the situation on March 9 to 10, 1956. Observations were made visually by the author at Godhavn (77.5°). The visibility was good, and auroras could be observed from 15° above the northern horizon down to the southern horizon (cf. Lassen, 1961a). In plotting the auroras a height of 105 km for the lower border was assumed.

Observations which began at about 2130 CGT and were performed about once per hour until about 0230 showed complete absence of auroras. The same was the case at 0300 and 0315. From 0330 observations were made continuously till dawn (0705 CGT).

Auroras were first seen 0344 CGT as raybands low at the southern horizon in international brightness class (IBC) 1–2. They were slowly pulsating in intensity; often they disappeared completely or were reduced to a single ray. The last ray disappeared at 0445 CGT.

At 0520 CGT faint single rays began to appear low in the northeast. Slowly and in low intensity the aurora was elongated across the zenith until finally it reached the landscape about 5° above the southwestern horizon, all the time showing the gradual change of intensity between the individual rayed forms which is characteristic of the morning auroras. During a few shorter intervals IBC 3 was observed, but as a whole the intensity was low, and auroras were rare on the southwestern sky and at the zenith.

The *Kp* index at the time of the auroras was 3_0,3–. A plot of the type shown in Fig. 3a presents the position in the polar graph of all auroras observed from the station. Since a wide area is covered at any instant auroras may appear close to each other in the diagram even if

they are observed at rather different universal times, as about 0400 CGT in Fig. 3a. Although this representation is correct it is possible to get a less confusing graph if for every term of observation only the position of the auroras in the corrected geomagnetic meridian is plotted. By such a procedure the shape of the auroral oval or any other precipitation curve should easily be traced. In Fig. 3b this has been done for the data plotted in Fig. 3a. On this morning, the figure shows that auroras occurred at two fixed latitudes, i.e. first at (and possibly below) 70°, and shortly after at 75°–78°. The discontinuity between night and morning auroras is obvious.

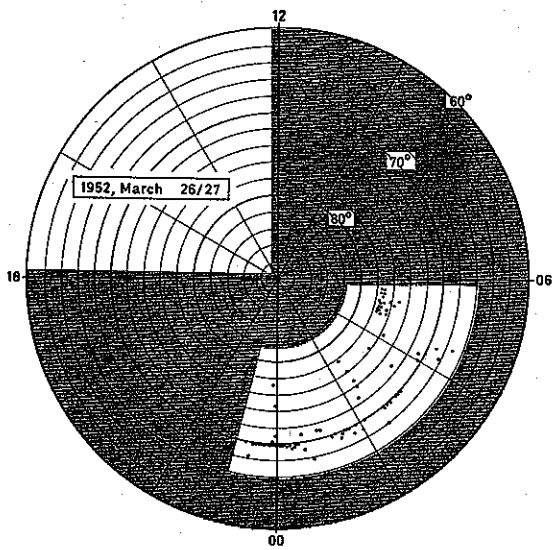
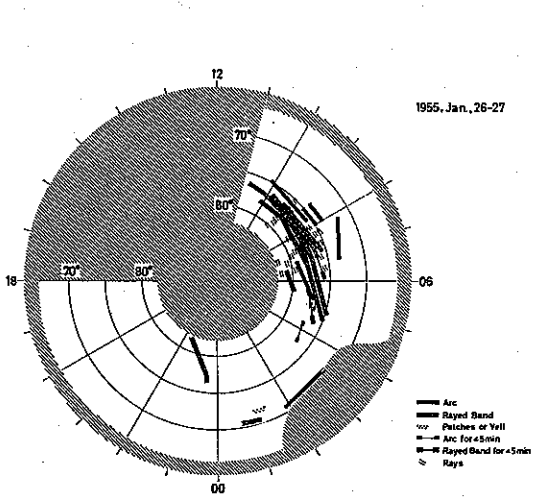


Figure 4. Meridian position of auroral forms on the morning of March 27, 1952. Observations from Godhavn.



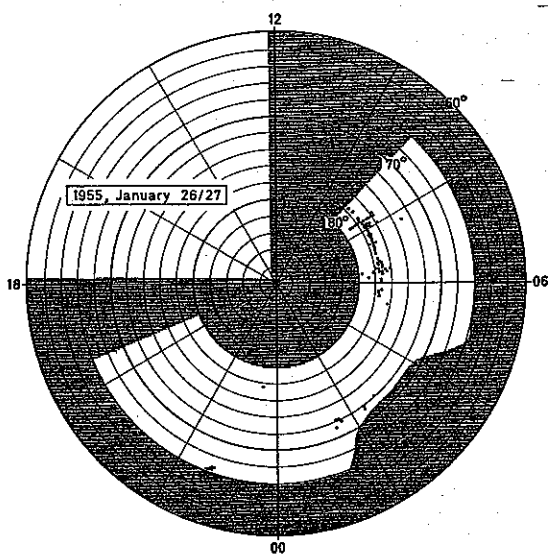


Figure 5a. Position of auroras observed from Godhavn on the morning of January 27, 1955.
Figure 5b. Meridian position of auroral forms on the morning of January 27, 1955.

In Fig. 4 visual observations from Godhavn 1952, March 26 at $Kp = 4$ show the presence of auroras around 70° from before midnight to 0430. Morning auroras appear isolated at

76°–78° from 0430–0500 CGT. The radially directed row of dots at about 0230 is due to a polar cap aurora. Figure 5 based upon visual observations on 1955, Jan. 26–27 at $Kp = 0_0,1+$, 2+ again shows night-time auroras along a fixed latitude parallel (72°), after a break of more than two hours being followed by morning auroras spontaneously appearing at 76°–78°.

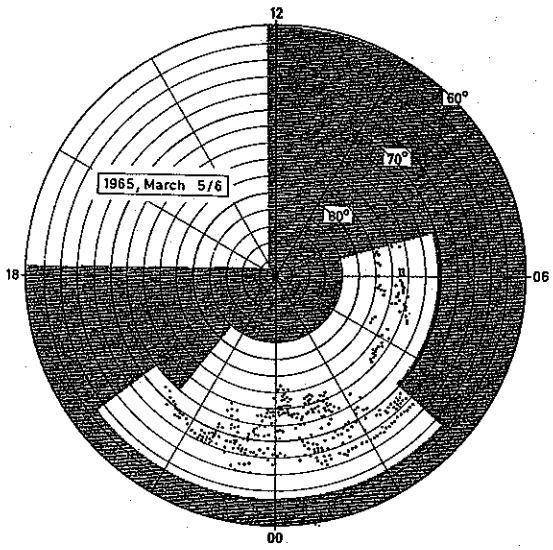


Figure 6. Meridian position of auroral forms on the night of March 5–6, 1965, based on all-sky camera photographs from Godhavn, Sukkertoppen, and Narssarssuaq.

Finally, Fig. 6 summarizes observations on March 5–6, 1965, from three all-sky cameras at the same meridian, namely Godhavn (77.5°), Sukkertoppen (75.3°), and Narssarssuaq (68.8°). The cameras cover the latitude interval from 63° to 82° along the meridian, and observations have been plotted with intervals of five minutes. The *Kp* values between 2100 and 0900 UT were 2–, 1_0, 1_0, 0+.

The night distribution is in this case situated at 66°–70°, with several poleward expansions appearing in connection with magnetospheric substorms from the first break-up around 2230 CGT to the last retreat between 0200 and 0230 CGT. Morning auroras at 74°–78° are first observed shortly after 0300 CGT at the same time as the night distribution at 67°–70° is being obscured by clouds.

The above mentioned examples are clearly in favour of the idea of two distinct distributions which may be completely separated.

It has been mentioned, however, that a complete separation is not to be expected at any time. This is illustrated by Fig. 7 which shows the situation on March 15–16, 1955, according to visual observations from Godhavn (Lassen, 1961a).

Night auroras were observed at 68°–72° until observation was discontinued shortly after midnight. When continuous observation began at about 0300 CGT rayed bands were seen approaching the station from the south, thus affording a connection between the night and day areas. At the same time auroras are still visible at 67° at 0400 CGT. The rayed bands differed from the 'proper' morning auroras in several respects. Their direction formed an angle of about 15° with the usual direction of the first morning auroras, and they were accom-

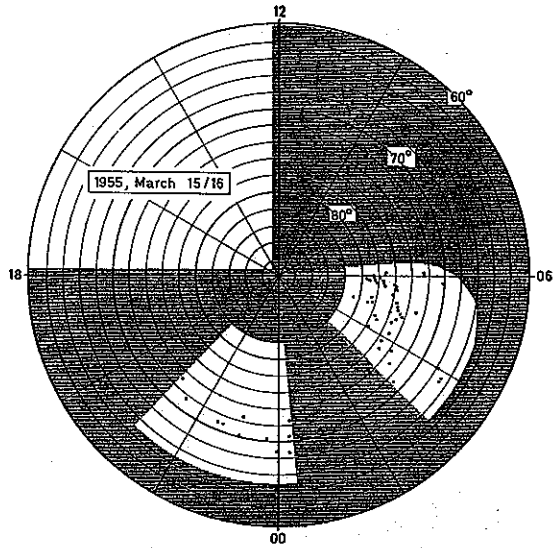


Figure 7. Meridian position of auroral forms on the night of March 15–16, 1955. Observations from Godhavn. No observations in the hatched area.

panied by magnetic perturbations as well as increased absorption in the ionosphere. They were therefore regarded as belonging to a special class, different from the morning auroras (Lassen, 1961a). This class is possibly identical with the polar cap auroras.

It is concluded that a discontinuity may be observed in the morning about 0300 CGT. For several reasons the discontinuity will not always be found, but this fact should not in itself be regarded as a sufficient proof of the existence of a continuous oval from evening after midnight till noon.

5. Afternoon and midday auroras

In Fig. 1, the afternoon peak for $Kp = 0$ is centred at about 76° with a value of 50%. The maximum value increases rapidly with Kp at the same time as the peak latitude decreases to 72° at $Kp = 3$ and less than 70° at $Kp = 4$. The peak is gradually extending farther towards the midnight hours.

During IGY (Fig. 2) the latitude variation is very nearly as in Fig. 1. This is in disagreement with Feldstein and Starkov (1967) who found a much smaller dependence on Kp with the northern edge of the area situated between 1800 and 2200 near 71° for all degrees of disturbance. The reason for the discrepancy is not clear, unless the measurements of Feldstein and Starkov refer to the night area alone, cf. the $Kp = 3$ plot of Fig. 2.

Midday auroras are observable from Nord in Northeast Greenland and from Spitzbergen. Their variation with Kp was studied by Feldstein and Starkov and to some extent by Lassen (1961b). Whereas the Russian authors found a variation in accordance with that of the other parts of the oval, Lassen found an increased frequency at Nord on internationally disturbed days of IGY which he interpreted as the result of a broadening of the oval, assumed to be situated some degrees to the south of the station. The interpretation does not fit well with the Russian results; the cause of this is not known.

Buchau et al. (1971a, b) have studied the auroral phenomena in the noon sector intensively on aeroplane flights. They found a peak of discrete auroras at 75°–78° CGL. On some flights visual auroras were absent, but in these cases subvisual auroras were shown photometrically to fill the gap between the last observed discrete arcs on the morning side and the first occurring arcs in the afternoon sector. They therefore concluded that the morning and afternoon groups are connected through the noon sector. It seems possible, however, that there are two frequency maxima, one on either side of the noon meridian.

Direct evidence of the continuity through the noon sector was presented by Khorosheva (1963). Using all-sky camera photographs from the Soviet network of stations for the season 1957–58, she was able to follow a single rayed arc from about 1000 to 1800 CGT. The latitude of the arc was 74° CGL at noon and 73° at 1800, *Kp* was 3. This observation seems to indicate that the morning and afternoon auroras form one single daytime group.

The afternoon peak during IGY is more elongated towards the midnight hours than in Fig. 1. For the morning distribution the opposite was found to be the case. Thus, if the two areas are actually two branches of one distribution, this must have been shifted counter-clockwise during the high sunspot maximum years 1957–59.

Buchau et al. (1971b) have found evidence of a hardening of the precipitating particles in the afternoon branch from early afternoon towards the midnight sector. The counter-clockwise shifting of the daytime distribution may therefore be an indication of the fact that the energy spectrum of the auroral particles was harder during IGY than during 1964–69.

At low disturbance levels a trough is found between the afternoon and night distributions in Figs. 1 and 2. A search for a discontinuity in the form of a gap between the distributions at 2100–2200 is not likely to be successful except in extreme cases. Whereas the night and morning arcs at 0200–0300 are approximately parallel a study of the direction of arcs in the evening shows that the arcs in the afternoon group form an angle with the direction of the night arcs. The two systems are often overlapping thereby probably giving rise to an increased frequency in the hours 2100–2300 where overlapping takes place. Under these circumstances the hypothesis of two independent auroral groups does not necessarily imply the presence of a hole in the evening part of the oval.

6. Polar cap auroras

The frequency maps of Fig. 1 show that auroras occur quite frequently over the inner polar cap in an area which may be characterized as consisting of two tongues stretching from the noon hours towards the early morning and late evening, respectively. The only areas where auroras are really rare phenomena are the high latitude midnight sector and a narrow area around the pole of the corrected geomagnetic system.

The polar cap auroras on the evening side seem to have a maximum near 80° at about 1800. The peak intensity is quite constant in 1964–69; in IGY it decreases with *Kp* from 70% (?) at $Kp = 1$ to 40% at $Kp = 3$. The latitude changes from approximately 82.5° to 80°. Whereas the change of the peak intensity in Fig. 1 is very small, the flank of the distribution at higher latitudes retreats rapidly towards the peak latitude with increasing *Kp*. Thus,

in the afternoon and evening hours the zenithal frequency at latitudes greater than 80° is seen to be negatively correlated with *Kp*.

On the morning side the distribution shows the occurrence of polar cap auroras directed from 80°, 0200 towards 80°, 1100–1200. The gradual increase in frequency towards the morning peak might suggest that the auroras are merely morning auroras broadening towards the pole in the hours of maximum frequency. The curvature of the isolines may, however,

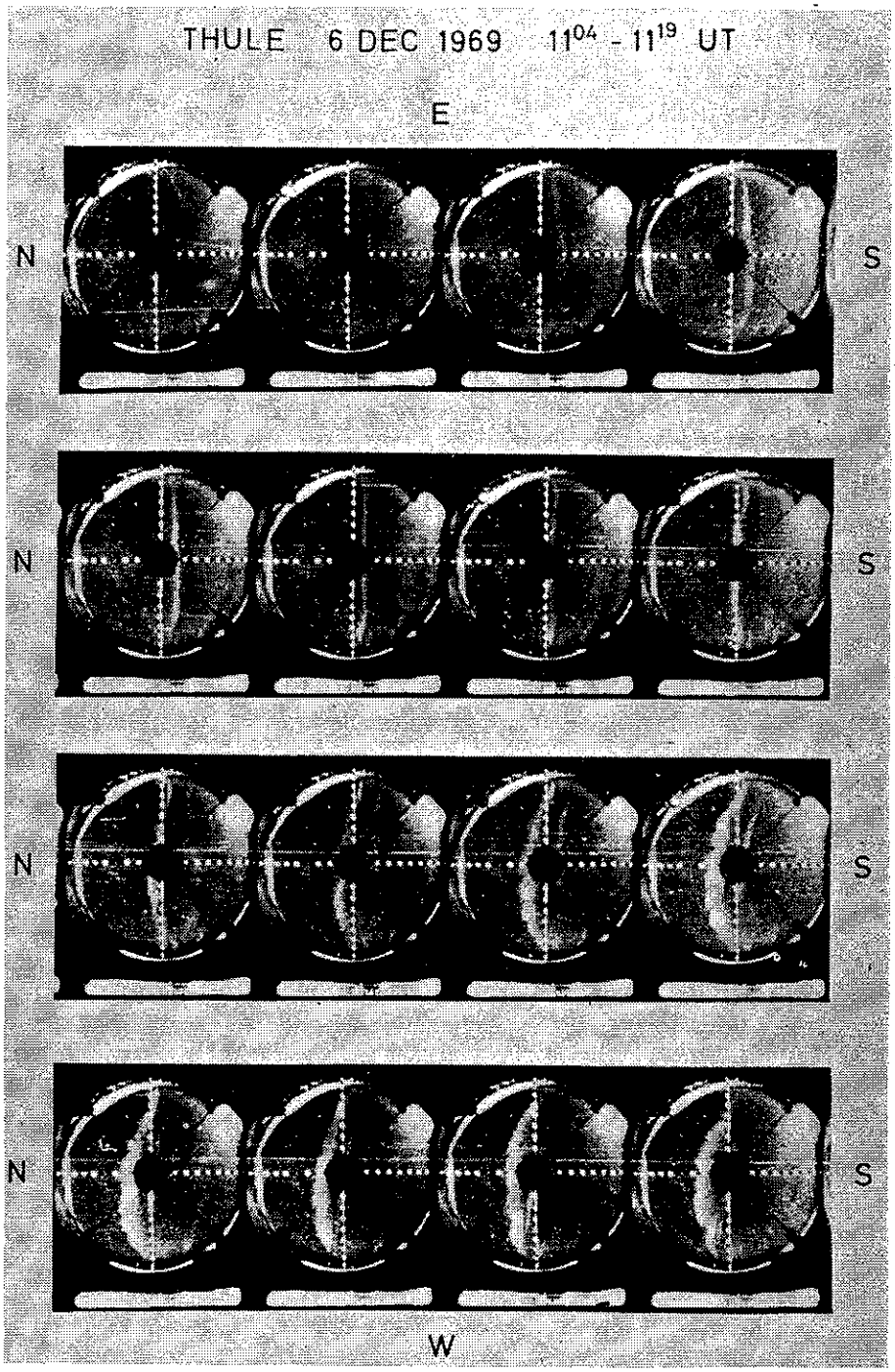


Figure 8. Polar cap aurora (rayed arc), Thule 1969, December 6, 1104–1119 UT.

indicate that a special group of auroras is present, which is near 80° overlapping with the morning distribution. The variation with *Kp* is rather complex, but could be in agreement with the hypothesis of two overlapping distributions. The frequency is at 83°–87° independent of *Kp*; at 80°–82° it is anticorrelated with *Kp*, presumably as a consequence of the equatorwards shifting of the morning peak.

Morphological studies of polar cap auroras including discussions of the direction of arcs have been published by Mawson (1925), Davis (1962), Weill (1958), Eather and Akasofu (1969), and Danielsen (1969). The auroras are described as faint rays, patches and fragments of arcs with a lifetime of a few minutes only, but well developed arcs of longer duration are also observed. The arcs and fragments of arcs tend to align along the Sun-Earth line. Danielsen reports that the arcs often move transversely to their direction. The direction of the movement is from the right to the left when looking along the auroral form and facing the Sun. The velocity of the movement is typically 300–400 m/s. The arcs may drift across the pole of the corrected geomagnetic system; thus this point is not a singular point (see also Fig. 9). A similar drift has been observed in sun-aligned arcs over Godhavn around midnight. The movement is believed to be an important expression of some dynamic effect in the magnetosphere. This is also indicated by the fact that an auroral movement nearly always started after the formation of rays or an increase of the intensity. Sometimes the auroral bands vanished and then returned after 5–10 minutes in such a place that it seemed as if the movement had continued without interruption, while the intensity of the aurora was temporarily below the threshold of visibility (Danielsen, 1969). A typical polar cap arc is shown in Fig. 8.

Systematic classification of the auroras occurring over the polar cap has not yet been attempted. Significant differences between types do exist, but our knowledge about the phenomena is not sufficient to show whether it will be reasonable from a physical point of view to try to separate the auroras into subgroups.

It has already been mentioned that some auroras are short-lived and faint, whereas others are more stable and sometimes very bright. The auroras observed at Thule are in some cases formed in the field of view of the camera, in others they drift in from lower latitudes. Of the locally formed auroras recorded during 1964–65 more than 50% had a lifetime shorter than 10 minutes. The total duration of the 31 observed arcs was less than ten hours. Auroras coming from lower latitude may be of a different origin. For example, they may be the northermost outposts of the morning distribution, although no transverse movement has ever been observed from Godhavn at the peak latitude.

According to our observations the sun-alignment of the forms is only valid as a first approximation. Actually, the arcs on each side of the pole form such angles with the Sun-Earth line as if they were directed towards the cusp region environment of the noon sector. The mean direction over 24 hours will then be that of the Sun-Earth line. Arcs near the midnight meridian are approximately sun-aligned. For the present it is not known whether it makes sense to distinguish between evening and morning side arcs. Besides the difference in direction, the difference in dependence on *Kp* should be recalled.

The latitudes in which the arcs begin and end are expected to be important for the understanding of the processes involved in the formation of polar cap auroras. One might expect

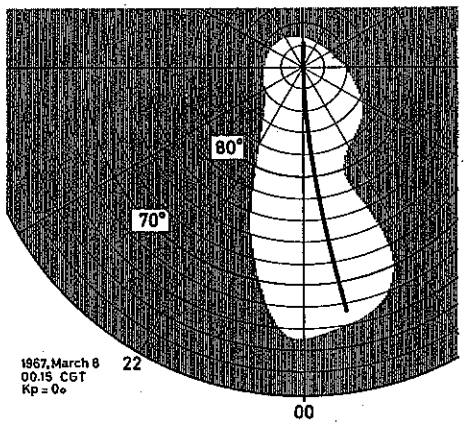


Figure 9. Polar cap arc, March 8, 1967, 0145 UT. The arc was photographed near zenith from Thule, Godhavn, Sukkertoppen. Narssarssuaq was overcast.

the arcs to be restricted to the polar cap itself, i.e. to the area on the polar side of the closed field lines. An observation, reproduced in Fig. 9, may, however, demonstrate that this is not always the case. The figure shows an arc photographed near midnight from three stations. The arc could be followed from about 2° colatitude in the noon meridian to 67°–68° in the midnight sector. Polar cap arcs may, accordingly, be of considerable length, in this case more than 2000 km. The *Kp* value was 0₀, and the latitude in which the arc became invisible is therefore assumed to be lower than the limit between open and closed field lines. Near this limit, on the night side, sun-aligned arcs seem to be rare. Harang (1931) has published an unusual observation from the Auroral Observatory at Tromsø of such an arc, which was directed approximately at right angles to the usual direction of arcs in the area. The time of observation was 0200 CGT, and the height of the lower border was 107 km. The day was moderately disturbed (international magnetic activity figure *C* about 1.1). Brilliant auroras were observed before midnight, but no or only weak auroras were seen during the last two hours preceding the observation of the sun-aligned arc. It is not possible from the description to judge whether the arc traversed the limit between open and closed field lines or not.

Polar cap arcs may even be present together with arcs of the oval system as shown in Figs.

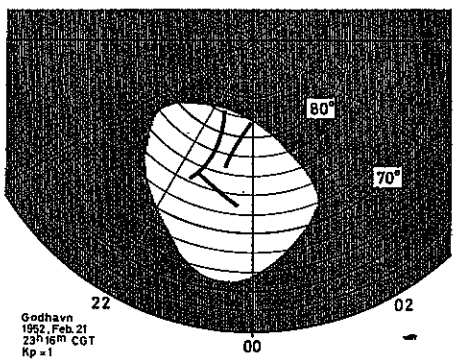


Figure 10. Polar cap aurora crossing an evening band. Observation from Godhavn, February 22, 1952 0046 UT.

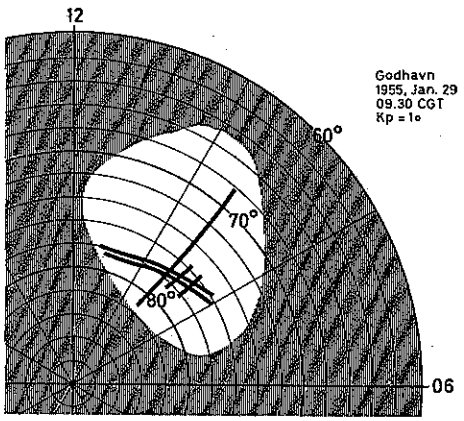


Figure 11. Intense polar cap rayed band crossing morning rayed arcs. Observation from Godhavn, January 29, 1955, 1100 UT.

10 and 11. The latter illustrates the crossing of usual morning auroras with one long and two short polar cap draperies. The observation is believed to be an outstanding one; the long drapery could be traced after twilight during more than one hour after the fading of the morning auroras; its brightness must therefore have been very strong. The position over Godhavn was nearly unchanged through $1\frac{1}{2}$ hours, thus indicating a drift in the ionosphere in the direction reported by Danielsen and with a velocity of about 160 m/s. The direction of the drapery is parallel to the polar cap arcs on the evening side. The clear crossing of morning arcs and polar cap draperies must be explained by any theory on the formation of polar cap auroras.

Starkov (1968) reports that all auroras with anomalously great heights ($H > 300$ km) were found to be similar to auroras of the polar cap type. On the other hand, Heikkila et al. (1971) observed in the noon sector bright arcs and bands which might well be polar cap auroras with a green lower border, suggesting that those auroras were excited by more energetic electrons than more faint arcs observed at an earlier flight. This observation, the bright drapery of Fig. 11, and the rayed bands mentioned earlier in this paper, which may fill the gap between night and morning auroras, and which are accompanied by minor geomagnetic and ionospheric effects (absorption), suggest the existence of more than one type of auroras over the inner polar cap.

Direct as well as indirect height measurements have shown that at least parts of these high-latitude auroras are generated by electrons of energy less than 1 keV. It is interesting to compare the distributions in Figs. 1 and 2 with satellite observations of the quiet day electron density at 1000 km height, published by Sato and Colin (1969) and Nishida (1967). Both papers report density concentrations resembling the tongues in the frequency plots; in Nishida's paper the two tongues radiate from a maximum area at about 78° immediately after noon. Both this distribution and the direction of the arcs in the two tongues suggest a relation between the dayside cusp and (at least some of) the polar cap auroras.

7. Direction of arcs

In the preceding sections the direction of arcs has occasionally been mentioned. From a more detailed study of this subject, which is being performed, a map of the prevailing directions is

presented in Fig. 12. For the two winter months December and January of 1965–66 all photographed arcs and bands were plotted in corrected geomagnetic coordinates, assuming a height of the lower border of the auroras of 105 km. Observations from all six camera stations were plotted in one polar graph which was then so crowded with auroras that it was found less suitable for reproduction. Therefore, a more rarified copy was drawn, in which the essential features of the original figure have been preserved. This copy is presented as Fig. 12. The original figure confirmed what was expected beforehand that the arcs are directed along the isofrequency lines except where two systems are overlapping. This makes it easy to compare Fig. 12 with Figs. 1 and 2.

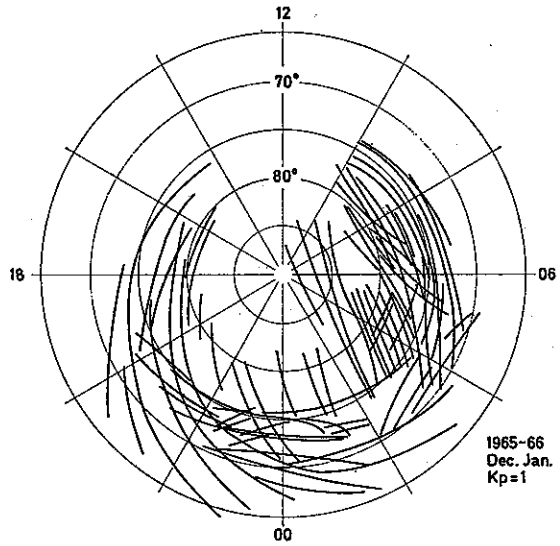


Figure 12. Position of auroral arcs observed from the network of stations in Greenland during December–January 1965–66 at $Kp = 1$. The diagram is slightly schematic.

The afternoon auroras are recognized as long spiral shaped arcs. In the final drawing, these arcs are very extended due to a good fit of independent observations from different terms of observations; it has not been possible, however, to follow extended arcs in the same way as done by Khorosheva or in Fig. 9; the network of stations is not suitable for this purpose. The afternoon system may be followed until about 0200 near 65° latitude.

Morning arcs follow the curvature of the latitude circles. They are situated close to the polar cap arcs which are stretching from the 0000–0600 quadrant at 75° towards about 75° latitude in the noon meridian. Polar cap arcs are visible before midnight, too. In the winter of 1965–66 they were rare on the evening side of the polar cap at $Kp = 1$. In the present figure they are not clearly distinguishable from the afternoon system.

A system of arcs is situated in the midnight sector between 76° and 70°. The arcs are nearly parallel to the latitude circles. This system of arcs is less obvious than the extended afternoon spiral; at $Kp = 0$ they are even less obvious. They are interpreted as the arcs of the night distribution of Fig. 1.

8. Conclusion

The auroral groupings in the frequency polar graphs combined with the representation of the position of arcs and bands observed from the stations in Greenland lead to the suggestion that the auroral precipitation curves consist of a spiral-like distribution which in magnetically quiet periods is situated between 75° and 78° from about 0300 til nōon, after which time it decreases monotonically in latitude down to 65°–70° at 0200. This precipitation area interferes with a system situated between 70° and 75° from about 2000 to about 0500 as well as with at least one system of arcs crossing the polar cap from 70°–80° on the night side of the graph towards the cusp region in the noon sector.

References

Buchau, J., J. A. Whalen and S.-I. Akasofu, 1970: J. Geophys. Res., **75**, 7147.
Buchau, J., J. A. Whalen and R. Wagner, 1971a: J. Atmosph. Terr. Phys., **33**, 661.
Buchau, J., G. J. Gassmann, C. P. Pike, R. A. Wagner and J. A. Whalen, 1971b: Symposium on Aurora and Airglow, IAGA, Moscow, August 1971. To be published in Ann. Géophys.
Danielsen, C., 1969: Danish Met. Inst., Geophys. Papers R-9.
Davis, T. N., 1962: J. Geophys. Res., **67**, 75.
Eather, R. H. and S.-I. Akasofu, 1969: J. Geophys. Res., **74**, 4794.
Feldstein, Y. I., 1963: Geomag. Aeronomy, **3**, 183.
Feldstein, Y. I. and G. V. Starkov, 1967: Planet. Space Sci., **15**, 209.
Feldstein, Y. I. and G. V. Starkov, 1968: Planet. Space Sci., **16**, 129.
Harang, L., 1931: Zs. Geophys., **7**, 271.
Heikkila, W. J., J. D. Winningham, R. H. Eather and S.-I. Akasofu, 1971: J. Geophys. Res. (submitted).
Khorosheva, O., 1963: Geomag. Aeronomy, **3**, 294.
Lassen, K., 1959a: Danske Met. Inst. Comm., Magn. 24.
Lassen, K., 1959b: Nature **184**, 4696, 1375.
Lassen, K., 1961a: Danske Met. Inst., Meddelelser No. 15.
Lassen, K., 1961b: Nature, **192**, 4800, 345.
Lassen, K., 1963: Danske Met. Inst., Meddelelser No. 16.
Lassen, K., 1969: In 'Atmospheric Emissions' (Eds. B. M. McCormac and A. Omholt), Van Nostrand Reinhold Comp., New York, 63.
Lassen, K., 1970: Phys. Norv. **4**, 171.
Lassen, K. and O. Rud Laursen, 1968: Danish Met. Inst., Geophys. Papers, R-4.
Mawson, D., 1925: 'Australian Antarctic Expedition 1911–14', Sci. Rep., Ser. B, II, 1, Sydney.
Nishida, A., 1967: J. Geophys. Res. **72**, 6051.
Sato, T. and L. Colin: J. Geophys. Res. **74**, 2193.
Starkov, G. V., 1968: Geomag. Aeronomy, **8**, 28.
Tromholt, S., 1880: Inst. Met. Danois, Copenhague.
Weill, G., 1958: Comptes rendus, Academie des Sciences, Paris, **246**, 2925.

Geofysiske Publikasjoner, Vol. 29, 1972

THE HARANG DISCONTINUITY IN AURORAL BELT IONOSPHERIC CURRENTS

BY J. P. HEPPNER
GODDARD SPACE FLIGHT CENTER,
GREENBELT, MARYLAND,
USA

1. Introduction

In 1946 L. Harang published a study of high latitude magnetic disturbances which ranks with the works of Birkeland (1908, 1913), Chapman (cf. Chapman and Bartels, 1940), Vestine (cf. Vestine et al., 1947), and their co-workers in setting the stage for describing the latitude vs. local time pattern of the disturbances and the inferred ionospheric currents. Although for many years eclipsed in journals by references to Chapman and Vestine and their co-workers, and later Nagata (e.g., 1963) and his associates and students, Harang's study was singularly significant in recognizing and emphasizing the midnight (or pre-midnight, see later discussion) discontinuity in the auroral zone. His statements in the 1946 paper indicate that he expected the discontinuity in the magnetic disturbance and inferred currents to appear also as a discontinuity in the auroral activity. For example, he states 'it would be of special interest to follow the variations in the auroral luminosities during the hours when the storminess in H changes from positive to negative values...' and 'it would be of great interest to decide whether there is a parallel discontinuity in the mean movements of the auroras.' These expectations were borne out. Heppner (1954) demonstrated that the reversal of E-W currents as indicated by the magnetic disturbance was in space-time coincident with a transition from comparatively quiet, usually homogeneous, arcs to more active rayed forms with the major change in activity occurring at the time the lowest latitude auroral arc near midnight broke into bright, rapid moving, rayed aurora (i.e., auroral break-up). Davis (1962) demonstrated that the discontinuity was in space-time coincident with a reversal in auroral motions from E to W at local times preceding the discontinuity, to W to E at local times following the discontinuity.

Harang (1946) also expressed views on ionospheric closure of the electrojet currents. Relative to return currents through the lower latitude ionosphere 'the hypothesis that perturbing currents flowing in the ionosphere are closed is untenable' and relative to closure over the polar cap 'the material is, however, not sufficient to support the view that the perturbing currents as a whole are closed over the polar cap.' These are modern views that mesh with journal articles being published today. Particularly, because it has become recognized that the most dynamic processes within the magnetosphere (e.g., partial collapse of the magnetospheric

tail, major particle energizations, redistributions of trapped particle populations, etc.) occur coincident with sudden changes in the midnight discontinuity, often referred to as substorm onsets, Harang's (1946) diagrams for the disturbance patterns merit more attention than they have received. These patterns can be identified with convective motions; in fact, the basic correctness of the form of Harang's patterns is indirectly confirmed by numerous recent electric field measurements. Nevertheless, only a few of the many investigators concerned with high latitude phenomena and their magnetospheric implications have recognized or ascribed any importance to the configuration of the discontinuity where auroral break-up, and associated changes, first appear.

The principal intent of this paper is to direct attention to Harang's contribution by bringing together observations which illustrate the reality and form of the discontinuity near midnight, which we have appropriately named the 'Harang discontinuity.' In presenting this subject the author is going to draw heavily on his own, and co-workers', observations. A literature review is not attempted. Hopefully, others will be motivated to compare their observations with Harang's patterns.

2. Harang's discontinuity

High latitude magnetic disturbances and auroral displays have never been found to be completely identical on any two days. Nevertheless, at any single observatory repetitious similarities as a function of local time are immediately obvious. Harang analyzed this diurnal variation in magnetic records as a function of latitude by using a latitudinal distribution of eleven observatories in a small longitudinal sector covering Scandinavia. To see if the diurnal characteristics were dependent on the degree of storminess he separated days into four classes, I–IV, based on the daily sum of the hourly values of the perturbation in the horizontal component, H, at Tromsö. Diurnal curves for each observatory within each class were obtained for each component (H, D, Z) by averaging hourly means. To illustrate the diurnal variation graphically, equal values of the perturbations (ΔH, ΔD, ΔZ) were contoured on a conical projection of geomagnetic latitude vs. local time. Figure 1 illustrates the result for one of the four levels of disturbance analyzed. In more recent years it has become well known that longitudinal differences in the diurnal pattern are largely removed by using geomagnetic local time (MLT) in place of solar local time. Thus, to make Harang's diagrams applicable to other longitudes, the time scale has to be changed. In Fig. 1 the change is indicated by tick marks giving the MLT at Tromsö (magnetic latitude = 67.2°). As MLT is not independent of latitude the change in the local time scale is not exact for other magnetic latitudes but the differences are small enough so as not to affect major characteristics.

The principal feature that made Harang's disturbance pattern distinct from others is that it showed that the reversal in the sign of ΔH in the nightside auroral belt occurred at different local times as a function of latitude as shown in Fig. 1. The coincident pattern in ΔZ was consistent with interpreting the ΔH discontinuity in terms of ionospheric currents flowing eastward and westward, respectively, in the southern and northern parts of the auroral belt prior to the reversal becoming complete across the entire belt between 2200 and 2300 MLT.

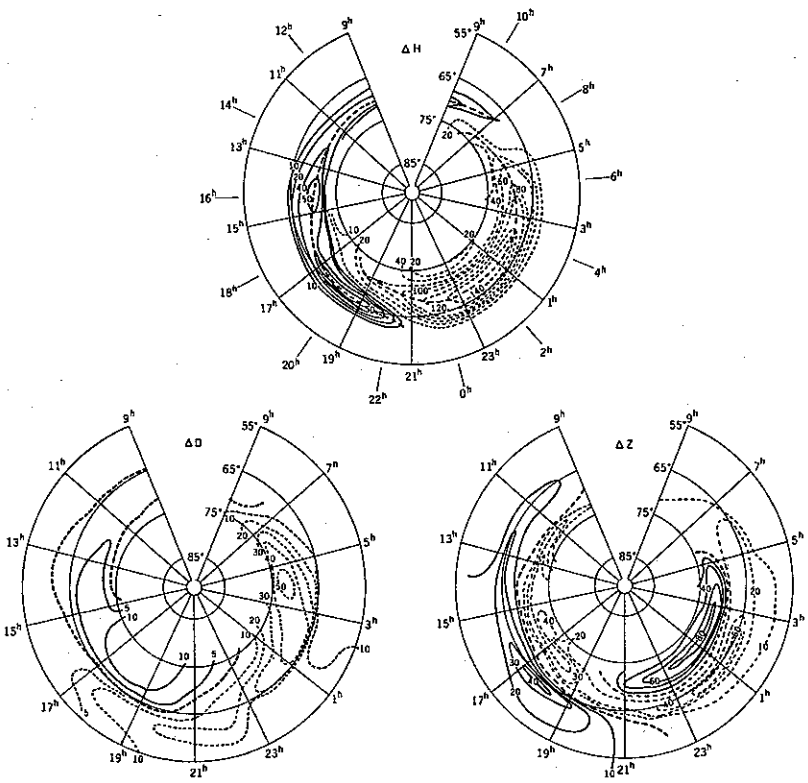


Figure 1. Contours of equal ΔH, ΔD, ΔZ for Harang's (1946) Range III disturbance levels: Solid lines ($+\Delta H$, $+\Delta Z$, West dec.), dotted lines ($-\Delta H$, $-\Delta Z$, East dec.), heavy dashed lines for ΔH and ΔZ indicate the center of the current system. Original coordinates are magnetic latitude and solar local time. The magnetic local time scale is added for ΔH.

A similar but less striking discontinuity appeared near 1000 MLT on the dayside. Harang (1946) did not emphasize the 1000 discontinuity, possibly because of the weaker magnitudes and its location above 70° magnetic latitude where his observatory coverage was more limited. Thus, in naming the E-W current reversal the 'Harang discontinuity' the term is applied here only to the night-time discontinuity. Also, recent investigations of both electric and magnetic field variations have shown the reversal near 1000 MLT to be highly complex, usually involving multiple reversals, such that a simple latitudinal overlap of E-W currents derived from averages is not likely to be as representative of individual cases as it is during the night hours.

It is appropriate to ask why most other analyses of high latitude magnetic disturbance patterns have not shown the latitudinal overlap of + and $-\Delta H$ regions (or equivalent E-W currents) in the pre-midnight sector. The answer lies in the averaging and smoothing techniques used relative to how combinations of factors such as the observatory distribution, the mixing of days with different degrees of disturbance, sampling intervals, the universal time distribution of the data, etc. relate to the large spatial asymmetries in magnitude and abrupt changes in sign of the parameters being analyzed. This is easy to recognize; it is thus remarkable that the Silsbee and Vestine (1942) analysis has been so widely referenced. Their analysis contained most of the obvious pitfalls and thus is illustrative; for example, between 44° and

88° magnetic latitude only 5 observatories were used, individual samples were 3 hour averages, the analysis was based on 9 hours of Universal Time corresponding to times when no auroral belt data below 70.8° was available in the local time sector 1800–2400 and, as Silsbee and Vestine (1942) state, the selection of bays for study was biased toward large values of $-\Delta H$. As to be expected, the resulting Silsbee and Vestine (1942) current pattern does not provide a representation of auroral belt disturbance that fits other data in the afternoon and pre-midnight sectors of the auroral belt. Many analyses appear to miss the form of the Harang discontinuity simply because the data is greatly smoothed in space and time; however, some suggestion of Harang's discontinuity has appeared even in highly smoothed representations (e.g., Nagata, 1963). There are also numerous analyses of selected times which give the impression that the pattern of high latitude disturbances is extremely variable and thus might suggest that average patterns such as Harang's are without meaning. In most cases detailed examination of these analyses reveals that the instantaneous distributions are greatly affected by the observatory distribution and the interpolation between widely spaced observatories when representation is in terms of equivalent currents closed within the ionosphere. Examined critically, the patterns drawn by Grafe (1969) illustrate this point; the techniques in many other analyses have been similar.

The next question is 'how well do Harang's patterns represent the spatial distribution of individual disturbances?' From magnetograms it is obvious that neither positive nor negative bays reach maximum intensities at identical magnetic local times on different days even when the days are selectively grouped as in Harang's analysis. Thus the smoothing effect of averaging different days produces a local time distribution with less contrast in magnitude as a function of local time than is present in individual cases. A more subtle effect occurs relative to the location of the nighttime $+\Delta H \rightarrow -\Delta H$ discontinuity. Because negative bay, post-break-up values for $|-\Delta H|$ are in most cases 2 to 8 times greater than ΔH in the pre-break-up positive bay sector, the discontinuity that appears in averages is disproportionally biased toward earlier local times by a small number of individual cases where the discontinuity occurs at an abnormally early MLT. Thus, in the majority of individual disturbances the $+\Delta H \rightarrow -\Delta H$ discontinuity will be located at a somewhat later MLT than the discontinuity that appears in Harang's patterns. Similarly in the pre-midnight sector where $+\Delta H$ and $-\Delta H$ regions overlap in latitude (e.g., between 2000 and 2200 MLT in Fig. 1) the discontinuity is shifted slightly towards lower latitudes in the averaging process. These biasing effects, illustrated in more detail by Heppner (1954), are unavoidable when averages are used. A more recent study (Heppner, 1967), involving computer-generated movies of the simultaneous disturbance vectors at 25 observatories at 2.5 minute intervals throughout several weeks, showed that the biasing effects in Harang's analysis do not appreciably alter the principal features of the distribution (e.g., to represent a greater fraction of individual disturbances the shift in the MLT of the $+\Delta H \rightarrow -\Delta H$ discontinuity would be approximately one hour). In fact with allowance for these biases and ignoring local enhancements in magnitude (which are obviously obliterated in Harang's averages), the movies show that Harang's average patterns for classes II, III, and IV provide a remarkably good representation of the distribution of the instantaneous disturbance most of the time.

In addition to the four classes, or levels, of disturbance I–IV, Harang (1946) analyzed the nine most intense storms at Tromsö recorded between 1932 and 1937. He found no fundamental differences between these storms and his Class IV days. In effect, because the basic form of his patterns changed very little between Classes II, III, and IV the pattern was basically the same for all Classes II and above. His Class II corresponds roughly to $Kp = 3-$. His Class I pattern (corresponding roughly to $Kp < 2-$) differs in that overlapping $+\Delta H$ and $-\Delta H$ regions do not appear and the MLT of the nighttime discontinuity is displaced toward earlier hours. Except for this apparent difference for Class I, his finding that the form of the pattern is independent of disturbance level was confirmed by the movie analyses of Heppner (1967). As the movie analyses did not indicate any basic differences for very weak disturbances, and electric field directions in the auroral belt between 1800 and 2200 MLT appear to be independent of disturbance level, it is likely that Harang's averages for ΔH for very weak disturbances, Class I, may have been influenced by a small baseline error or one of numerous other factors which can influence averages when dealing with small quantities. This is suggested also by the fact that his ΔD and ΔZ diagrams for Class I do not differ appreciably from those for other classes. Thus, no significance should be attached to his ΔH difference for Class I. Like all other analyses Harang found that the region of disturbance extended to lower latitudes with increasing disturbance.

3. The auroral discontinuity

Much of the apparent complexity of magnetic disturbances at auroral latitudes becomes less mysterious when the simultaneous auroral behavior is observed in detail. Heppner (1954) was able to demonstrate repetitive relationships between the magnetic disturbance and sequences of occurrence of different auroral forms and the distribution of these sequences in local time and latitude. The typical diurnal behavior during the night hours between 60° and 70° magnetic latitude was illustrated in terms of two patterns, Fig. 2. Pattern I was regarded as the basic pattern; pattern II was included to cover cases, usually of weak disturbance, in which the disturbance and aurora temporarily died out instead of being continuous through the night. Although not drawn in an idealized pattern, highly complex disturbances were shown to result when new bay activations occurred before a previous disturbance died out. The auroral symbolism describes the visible auroral forms which characterize the auroral activity in terms of magnetic local time and latitude and the related stage of development of simple positive and negative bays. The dominant auroral transition is clearly the break between homogeneous forms, usually appearing comparatively stable, and rayed forms, usually brighter, more rapidly moving, and more transient in duration. In Fig. 2, Pattern I, this break is indicated by the dotted line *AB*. Davis (1962) used a similar representation to show the local time vs. magnetic latitude distribution of directions of auroral motions. Figure 3 is an example of a highly disturbed day in which the aurora extended to latitudes $< 60°$. The character of the pattern is clearly dominated by the reversal of east-west motions along the line *AA'* which Davis identified as corresponding to Heppner's line *AB*, indicated in Fig. 2. Davis (1962) found that this reversal was characteristic of every auroral display observed independent of the level of

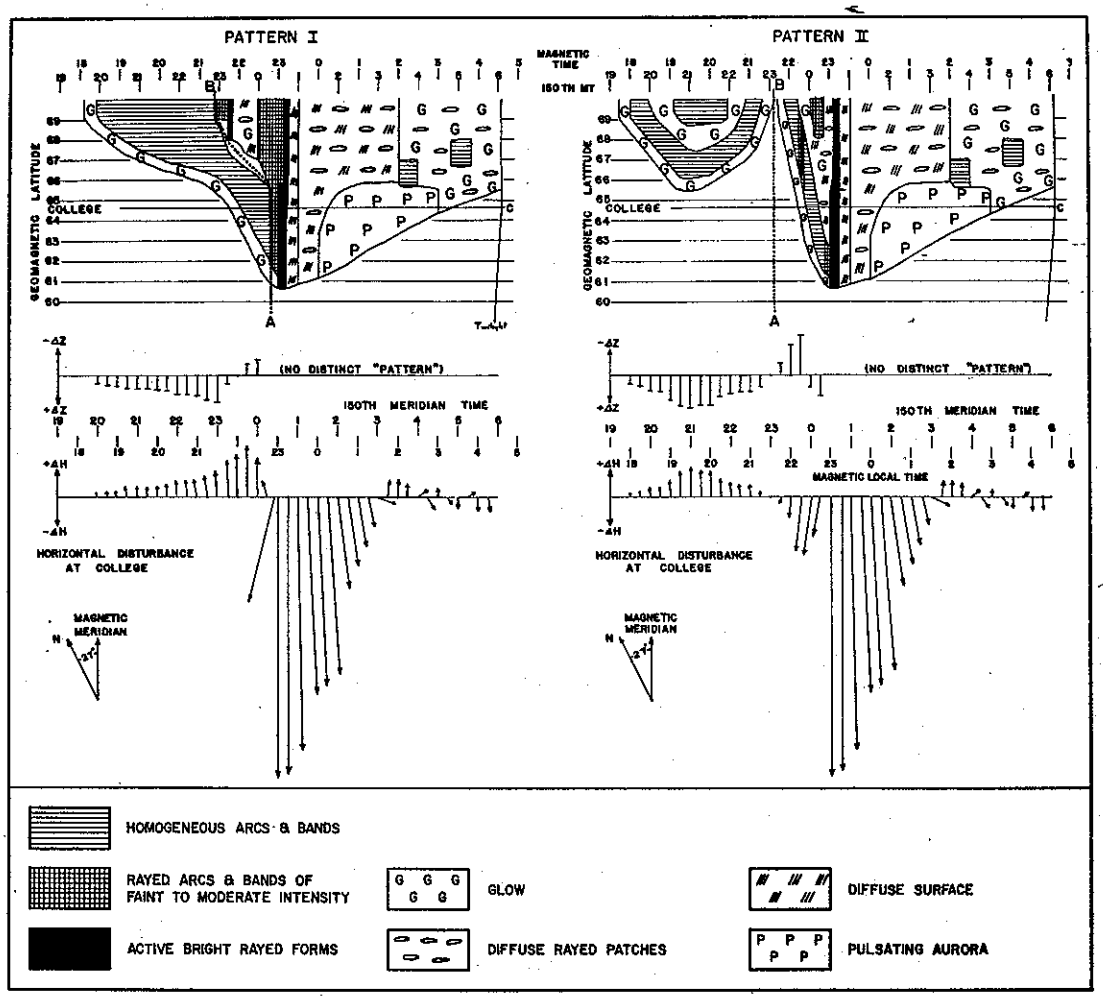


Figure 2. Idealized patterns of simultaneous auroral and magnetic activity from Alaska observations of Heppner (1954). The line *AB* (Pattern I) represents the Harang discontinuity.

activity. Over 44 nights at College, Alaska (mag. lat. = 64.7°) the total range of local times for the reversal was nearly 6 hours with the average 'approximate geomagnetic time' of the reversal falling near 2300. Davis (1962) also found large variations in the alignment of auroral forms associated with the reversal.

Both Heppner (1954) and Davis (1962) found that the simultaneous magnetic disturbance was consistent with ionospheric currents directed eastward before and westward after the local time of the *AB* (or *AA'*) auroral discontinuity. A station such as College, Alaska, located near but slightly to the south of the center of maximum activity (most often 66°–68° in the midnight hours) is frequently situated so that the magnetic disturbance observed between the local times of *B* (or *A'*) and *A* (Figs. 2 and 3) is a superposition of effects from westward currents to the north and eastward currents overhead and to the south. As the westward current in the high latitude part of the auroral belt (following *AB* or *AA'*) is usually more intense than the eastward current in the low latitude part of the auroral belt (preceding *AB* or *AA'*), the superposition usually gives a small, but variable, $-\Delta H$ at these times. The College magneto-

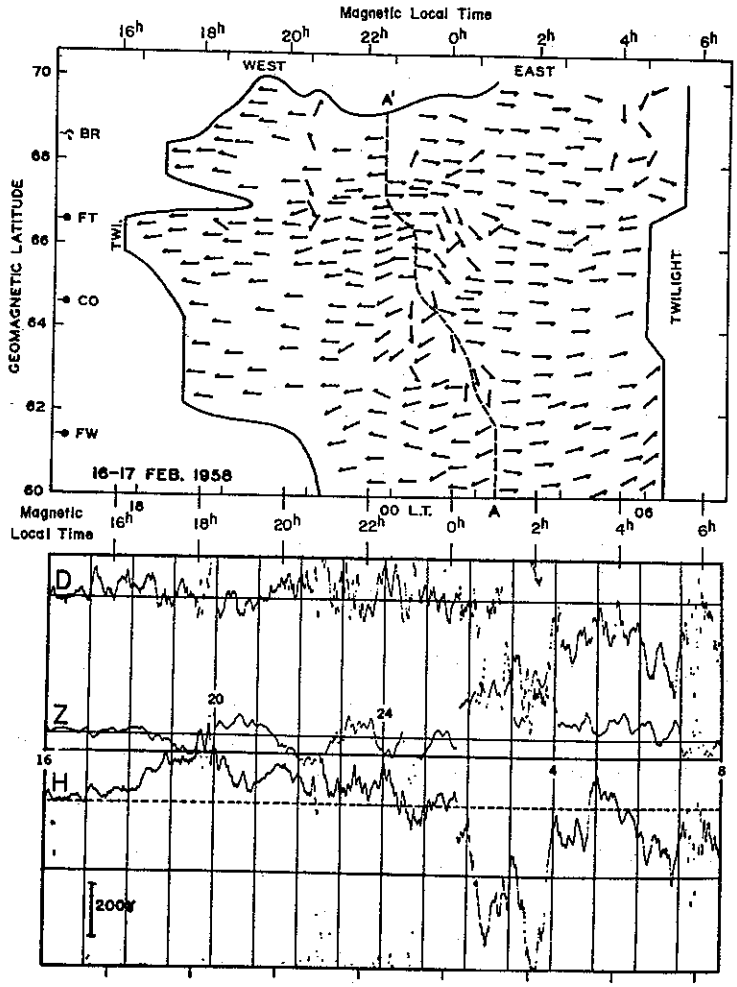


Figure 3. Directions of auroral motions with simultaneous College, Alaska magnetogram for one night from Alaska observations of Davis (1962). The line *AA'* corresponds to the Harang discontinuity.

grams, *H* trace, between magnetic times 2300 and 0000 (or local 150th meridian times near 0100) in Fig. 3 provides a typical example. This tendency for the superposition to produce a slightly negative ΔH illustrates the latitudinal bias in Harang's averages, noted in the previous section.

The analyses which showed the *AB*, or *AA'*, discontinuities coincident with the ionospheric current reversals were performed for a large number of individual nights and disturbances. Thus, they illustrated the reality of Harang's discontinuity independent of statistical averages.

4. Convection at the Harang discontinuity

The motions of visual aurora, described by Davis and others, and the motions of auroral ionization irregularities by radio techniques described by Kaiser (1958) and Harang and Tröim (1960), were used by Axford and Hines (1961) as observational evidence for their convective model of high latitude disturbances. This is equivalent to stating that the ionospheric currents

are Hall currents and numerous electric field measurements, beginning with the Ba^+ motion studies of Föppl et al. (1968) and Wescott et al. (1969), have since confirmed that the auroral electrojets are Hall currents. Thus, it would appear that convection patterns can be deduced from either the magnetic disturbance (see e.g., Heppner, 1969) or the auroral motion (see e.g., Davis, 1971). The limitations of these procedures are, however, evident. In the case of the magnetic disturbance it is well known (Haerendel and Lüst, 1970; Westcott et al., 1970) that the intensity of the ionospheric current is more closely related to the ionospheric conductivity than to the magnitude of the electric field. When this consideration is coupled with the fact that the surface magnetic observatory is seeing the integrated effect of the regional ionospheric current it is clear that abrupt changes within the convection pattern cannot be spatially resolved. Contributions from non-ionospheric currents also introduce uncertainties, and in regions displaced from the principal auroral currents, such as the polar cap, it has been found (Heppner et al., 1971) that the magnetic disturbance cannot be related to the overhead ionospheric convection. The limitations in deducing convection from auroral motions are quite different and fall into two general categories: (a) ionospheric irregularities associated with aurora can arise from a variety of mechanism and without knowledge of cause one cannot be sure that the motions are truly convective, and (b) the particles producing visual aurora are subject to non-convective drifts in the magnetosphere (e.g., from magnetic field curvature and gradients) and thus the precipitation patterns are not necessarily along convective shells and changes in magnetospheric magnetic fields, precipitation energies, etc. will cause apparent motions of aurora that are not closely related to the convection.

Barium release experiments have clearly shown both the applicability and the pitfalls of using magnetic disturbance vectors and auroral motions to deduce convective patterns. For example, there have been a number of cases (Wescott et al., 1969 and unpublished observations) where the motions of barium ion clouds were almost exactly parallel to auroral arcs and perpendicular to the surface magnetic vector. However, there are other cases (Wescott et al., 1970 and unpublished observations) where the motion of auroral forms has been at a large angle to the Ba^+ cloud motion and their paths have crossed. These obvious exceptions have been observed in association with break-up aurora and transitions in the magnetic disturbance that can be identified with the Harang discontinuity. Thus, they illustrate that the discontinuity is a region of instability whose instantaneous form and location is frequently shifting as discussed in Sect. 5.

Because of the shifting of the discontinuity, the uncertainties in using auroral motions and magnetic vectors, noted above, and the limitation of having a finite number of Ba^+ clouds to observe at one time, it is not possible to obtain a complete instantaneous picture of convection at the discontinuity. However, the Ba^+ motions observed slightly before, during, and shortly after, times when the presence of the discontinuity is indicated by the local magnetic disturbance and aurora, suggest several features that may be general (i.e., each has been observed more than once and each case involves multiple Ba^+ clouds). (1) At local times slightly before (e.g., 0 to 30 min.) the local time of the beginning of a negative bay, convective motions are westward and closely parallel to lines of constant invariant latitude. (2) In space-time coincidence with the beginning of a $-\Delta H$ disturbance and the development of bright rays,

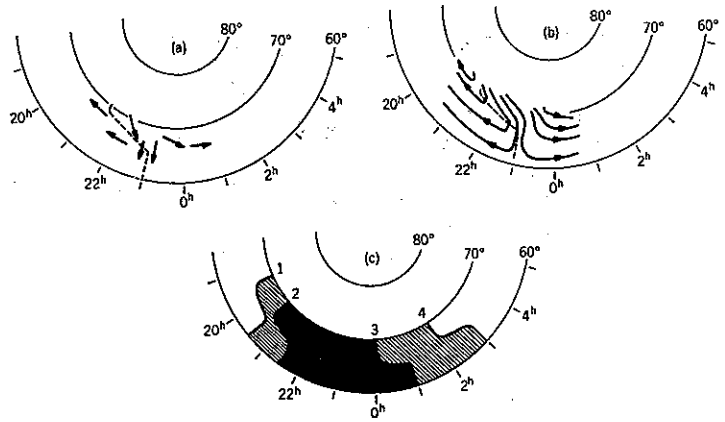


Figure 4. (a) Convective motions of Ba^+ clouds relative to the Harang discontinuity, idealized from northern Norway observations of Wescott, Stolarik, and Heppner (see text), (b) convective continuity indicated for the region of the Harang discontinuity, (c) ranges of variability in the location of the Harang discontinuity (see text): Coordinates are magnetic local time and invariant latitude.

rapidly moving within unstable, moving auroral forms, convective motions toward the equator are observed. (3) At local times slightly after the local time of the beginning of a negative bay, while $|-\Delta H|$ is increasing, eastward convective motions are dominant, but there is also a significant equatorward component transverse to lines of constant invariant latitudes (examples in Wescott et al., 1969). The observations are idealized in Fig. 4a by referencing them in local time to a discontinuity arbitrarily drawn to resemble *AB* in Fig. 2 (Note: plotting actual times and mixing observations from different nights would introduce an unreal randomness relative to the discontinuity location as a consequence of the discontinuity displacements between nights of observation). In Fig. 4b the idealization is carried one step further by drawing continuity for the convection in this region. Using Ba^+ observations at earlier and later MLT's and in the polar cap, together with the extensive electric field measurements of OGO-6 (Maynard, 1972; Heppner, 1972) the convection pattern could be extended to show its typical form at all local times. This is beyond the scope of the present paper but a point to note is that one of the principal uncertainties in the pattern occurs at local times and latitudes (usually $\geq 69°$) where the auroral belt discontinuity merges with the anti-solar polar cap convection. The OGO-6 data in this region frequently suggest the existence of eddy-like flow structures with dimensions of several tens of kilometers to several hundred kilometers. The double reversal of the highest latitude Ba^+ track in Fig. 5a could be related to such an eddy structure rather than distinct time shifts in the discontinuity as discussed for the examples in Fig. 5b and 5c in the next section. Harang (1946) was perhaps prophetic when he stated 'we get the impression that the currents producing the geomagnetic disturbances on the southern edge of the zone move regularly, whereas on the northern, or inner, edge of the auroral zone the currents move more irregularly often forming systems of whirls.'

Despite the short-comings of deducing the convection near the Harang discontinuity from magnetic vectors, auroral alignments, and auroral motions, noted above, the results resemble Fig. 4b (see e.g., Davis's (1962) contours). In essence this means that they yield statistically

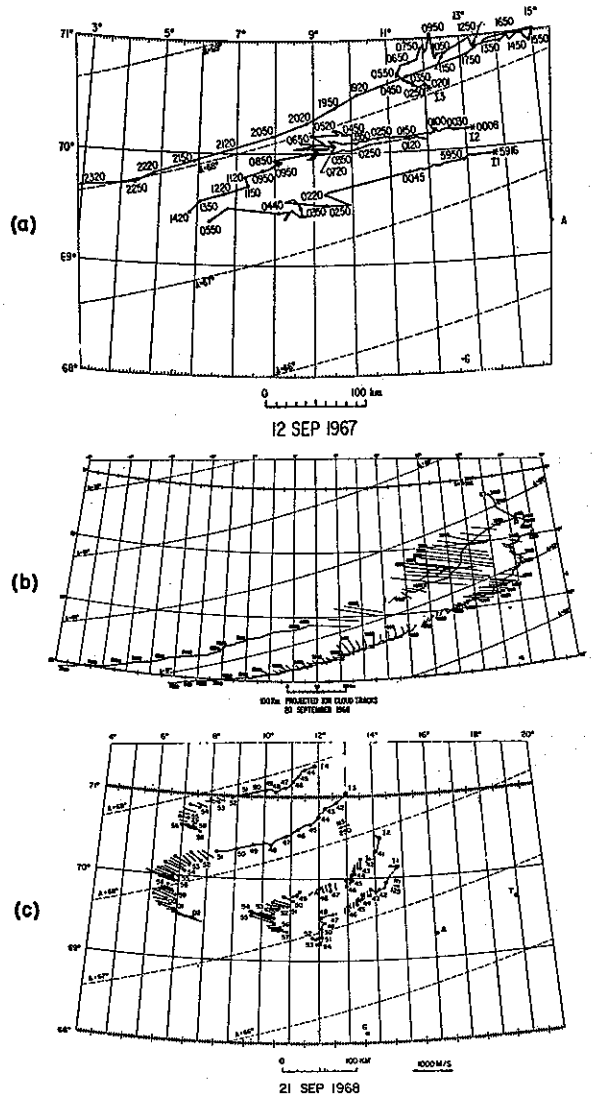


Figure 5. Horizontal tracks of Ba^+ clouds projected along magnetic field lines to the 100 km altitude level. Invariant latitude lines (Λ) are superimposed on the geographical grid. Numbers along tracks are minutes of the hour in (c) (2 digits) and minutes and seconds in (a) and (b) (4 digits). From rocket release experiments at Andenes, Norway by Wescott, Stolarik, and Heppner.

reasonable results when allowance is made for spatial resolution in the case of magnetic field analyses and the possibility of misleading auroral motions and alignments in individual cases.

5. Time variability of the Harang discontinuity

The previous sections have emphasized a picture in which the Harang discontinuity appears to have fixed location and form in magnetic time and latitude. Dynamic characteristics of the discontinuity were noted primarily to explain how individual events bias averages or typical distributions. The 'fixed' picture is necessary to show the form and the most common location of the discontinuity. Next, the variability, particularly as related to initial stages of negative

bay development, has to be examined (Note: the term 'substorm' or Birkeland's term 'polar elementary storm' can be used interchangeably with 'bay'. The term bay can be used somewhat more precisely to describe individual activity increases within a stormy interval and when needed reference can be made separately to positive and negative bays. It is, of course, even more precise to refer to ΔH directly and avoid the semantic confusions that have accompanied these names.)

Most descriptions of the near midnight auroral and magnetic activity center on the dynamics accompanying the sudden onset of a large negative bay and a visually impressive auroral break-up (e.g., the extensive descriptions of Akasofu et al., 1964–1966). The attention given these events is understandable because correlations with other observations are the most distinct. However, clearly identified sudden onsets seldom occur more than eight times per 24 hours of UT and the associated duration of outstanding change (i.e., time for $|-\Delta H|$ to reach large values) is usually less than one hour. This means that these descriptions apply less than 8 out of 24 UT hours, and more typically less than 4 out of 24 hours. Thus in describing the variability of the Harang discontinuity it would appear that distinctions might be needed between: (a) universal times displaced from the UT intervals of maximum change accompanying a sudden activation, (b) universal times when magnetograms collectively show growing bay disturbances of considerable magnitude, either or both + and —, but when there is a lack of evidence that the growth involves an identifiable activation time, and (c) universal time intervals embracing the UT's of sudden activations but restricted to the periods of maximum change. In principal the (b) intervals are included under (a); however, it is quite likely that some of the (b) intervals are related to (c) but an activation time is not recognized as a consequence of the discontinuity (i.e., onset location) not being near an observatory at the UT of onset. The ambiguity of (b) has the same implication as noting that there are cases where onsets are not detected and treating (b) as part of (a).

Despite expectation that the sudden bay activations, (c) above, would produce larger shifts in the discontinuity location than shifts at other times, this was not borne out in the movie analyses of ΔH by Heppner (1967). Particularly, with recognition that the bias effects in a ΔH analysis are largest at the time of negative bay increases, only a questionable tendency for larger shifts in MLT was evident. In essence, this confirms Harang's discovery of relatively little difference in pattern as a function of disturbance level, but applies it to a dynamic condition (i.e., the variability during times of large change is not greatly different than the variability observed at other times). The most characteristic behavior in the movie analysis was that the transition from + to $-\Delta H$ disturbance progressed rather smoothly, moving successively from one auroral belt observatory to the next as the Earth rotated. In several longitudinal sectors the observatory separations were 1 to 2 hours in MLT which meant that the MLT shifting of the discontinuity was confined to less than 2 hours of MLT much of the time. However, intermittently superimposed on this general behavior, discrete jumps, in which the shifting between observatories took place within 10 to 20 minutes, were evident. In total these observations indicated that there could be considerable small scale motion of the discontinuity, apparently independent of disturbance level, but that major sudden shifts over ranges as large as 3 hours in MLT must be rare. A different form of exception to a completely

fixed pattern was noted following some of the sudden shifts to earlier MLT's; in effect, the discontinuity gradually returned to later MLT's at approximately the Earth's rate of rotation. This implies that to some extent the pattern tries to rotate with the Earth but that the forces involved are apparently weak compared to the forces that give the basic pattern.

The variability of the discontinuity in terms of auroral forms and motions (Heppner, 1954; Davis, 1962) appears compatible with the ΔH movie analyses. Segments of the lines AB or AA' in Figs. 2 and 3 in some cases correspond to a slow east to west progression of the auroral transformation across the observer's sky (e.g., most apparent when the east end of a homogeneous arc bends northward to form a partial loop open toward the west). In select cases the transformation within an east-west arc appears to take place over its visible length within periods as short as 10 minutes. The latter cases suggest sudden shifts of the discontinuity $\geq$ 2 hours in MLT at the arc latitude; however, they must be interpreted with caution in view of recent findings of non-convective auroral motions during the break-up phase (noted below). In general, however, auroral observations used with the magnetic data improve the resolution for studying the shape of the discontinuity (i.e., its trace in magnetic local time vs. latitude projection) over that possible using only magnetic data. In extreme cases MLT differences ranging up to 5 hours have been observed between points B and A in representations like Fig. 2 (Pattern I) with A occurring well after magnetic midnight. A speculative point for investigation is whether or not these cases are indicative of the tendency for the pattern to rotate with the Earth, as mentioned previously. The other extreme, that of zero time difference between B and A, cannot be proven in terms of magnetic field and auroral observations as it involves prediction in regions where aurora is absent (e.g., Fig. 2, Pattern II). It is significant, however, that there are not any documented cases of B (or A') (Figs. 2 and 3) occurring at a later MLT than A. Neglecting the uncertain zero time difference cases, values of 30 minutes to 3 hours between B (or A') and A are characteristic. Figure 5c is an attempt to be more statistical between magnetic latitudes 60 and 70°. Based primarily on the observations of Heppner and Davis, a conservative estimate is that at least 66% of the time both points B (or A') and A fall between the lines 2 and 3 and at least 95% of the time between lines 1 and 4.

Two examples of the motions of Ba^+ clouds (Fig. 5, b and c) illustrate apparent stability and shifting, respectively, of the Harang discontinuity. Both illustrate non-convective auroral motions. In Figure 5b the two lowest latitude clouds, labelled I1 and I2 at the release points, moved southward as well as westward in invariant latitude during the first 10 minutes of observation. Simultaneously: (a) auroral arcs in the same vicinity showed rapidly moving rays and progressed poleward across the magnetic shells of the Ba^+ clouds and (b) the magnitude of the existing $+\Delta H$ disturbance at an observatory at the right edge of the figure near $\Lambda =$ 66.6° decreased to values near zero. After this initial period the clouds moved westward parallel to invariant latitude lines. The higher latitude cloud, I3, moved principally southward throughout observations. These motions suggest a highly stable discontinuity location passing through I3 but with I1 and I2 on the westward fringe of the discontinuity such that the initial westward component progressively moved them away from the discontinuity.

Figure 5c, in contrast to Figure 5b, illustrates a shifting discontinuity. In this case the dominantly westward early motion (beginning at release points I1, I2, I3, and I4) turned south-

ward between 49^m (I1) and 54^m (I4). At 51^m: (a) ΔH at the earth's surface began a rapid change toward large $-\Delta H$ values, and (b) auroral arcs to the south brightened and moved poleward. The poleward moving aurora later crossed the magnetic shells of the Ba^+ clouds as in the previous case. The latest observations shown (e.g., I3 at 02^m) indicate that the motion was turning from southward to eastward. As the initial westward motion was faster than the Earth's rotation the discontinuity had to shift rapidly toward an earlier MLT to reach the Ba^+ cloud positions.

6. Continuity of current at the Harang discontinuity

As quoted in the Introduction, Harang (1946) considered it untenable that the auroral belt currents closed their circuit through lower latitudes and stated that closure over the polar cap could not be supported in terms of the limited data available. These opinions came from examination of the pattern of current vectors using Birkeland's technique of drawing these vectors perpendicular to the horizontal disturbance vectors with magnitudes proportional to the disturbance. He did not speculate further on the closure and as late as 1966 (personal communication) still regarded this as an unresolved problem. In 1946 he did, however, pursue the question one step beyond the current vectors by drawing the indicated directions of current flow to and from the auroral belt. As shown in Fig. 6, for Class IV disturbances this flow was drawn relative to center lines of the evening eastward and morning westward currents. As indicated in his other diagrams, center lines for all classes of disturbances were constructed from both ΔH and ΔZ and tested for compatibility. The dominant characteristic of Fig. 6 is the outflow of current on the nightside with the distribution of outflow centered on the region of latitudinal overlap of eastward and westward currents (i.e., the discontinuity).

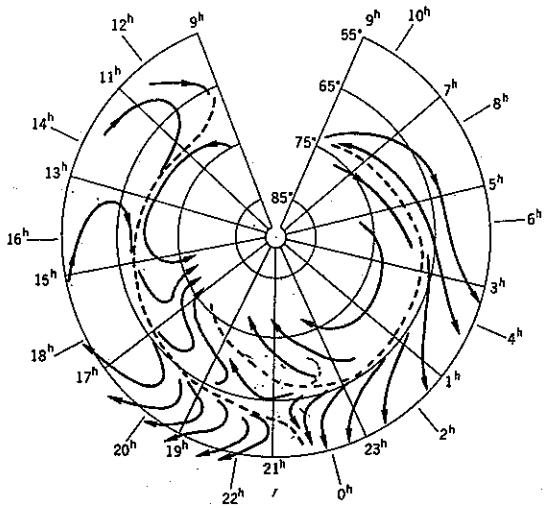


Figure 6. Harang's (1946) representation of directions of in-flow and out-flow of currents from the center of the auroral belt currents, indicated by the dashed lines. His coordinates were magnetic latitude and solar local time. An approximate magnetic local time scale has been added (outer time scale).

Recently, in seeking explanation for disagreement between the convective motion of Ba^+ clouds in the polar cap and the existence of current closure via the polar cap ionosphere Heppner et al. (1971) concluded: (a) that neither the middle latitude night-side ionosphere nor the polar cap ionosphere was conducting sufficiently to provide continuity for the auroral electrojects, (b) that variations in the ratio of (grad N)/N to (grad E)/E in sub-zones of auroral precipitation were such that the ionospheric Hall currents would close via field-aligned current flows to and from the magnetosphere (N = lower ionosphere plasma density, E = electric field intensity), (c) that the magnetic disturbance vectors in both the polar cap and regions adjacent to the low latitude boundary of auroral activity couls be explained in terms of two sheet currents flowing 'out of' and 'into' the auroral belt ionosphere, respectively, in the magnetic local time sectors 2000–2400 and 0800–1200, and (d) that a distribution of precipitation which would give maximum ionization in the region of electrojet current reversal would also make this region the mean location of a net field-aligned current. In essence, when idealized further the magnetic time vs. latitude trace of the Harang discontinuity became the horizontal intercept of a field-aligned current sheet. The sheet associated with the Harang discontinuity represented a net inflow of electrons. A net outflow of electrons (or inflow of ions) in the 0800–1200 sector was required for continuity and explanation of the ΔH vector in adjacent regions but the sheet configuration was less certain. Between these sectors field aligned currents were flowing 'in' and 'out' but the net current was small compared to that near the regions of current reversal when viewed in terms of widespread effects.

Unfortunately, the authors (Heppner et al., 1971) had forgotten Harang's 'lack of continuity' diagram (Fig. 6) at the time of the above study and it is not referenced. It would have geometrically supported the conclusions reached. Converted to field-aligned flow the arrows on the night side would leave the ionosphere in the regions where they depart significantly from E–W flow. This gives outflowing currents distributed toward earlier and later MLT, respectively, in the higher and lower latitude portions of the auroral belt (i.e., a mean distribution following the discontinuity).

7. Related topics

Detailed studies contrasting the energy spectra and composition of particles precipitating at and on opposite sides of the Harang discontinuity have apparently not been carried out for ionospheric altitudes. Through association with stages of auroral and magnetic activity some information is potentially available from rocket and satellite measurements. Indirectly from radio-wave absorption, sporadic-E, and auroral photometric studies it is known that the precipitation during and following break-up must be more intense and penetrating than prior to break-up; however, more subtle differences should be sought. At synchronous satellite altitudes, the reality of a discontinuity has been apparent in a number of studies (e.g., Leznaik and Winckler, 1970; DeForest and McIlwain, 1971) but its L vs. MLT form is not revealed at a constant distance.

Because the magnetosphere and ionosphere form a feedback system it has been impossible to distinguish whether events such as sudden bay onset have a magnetospheric or ionospheric

origin point. Thus, magnetospheric events can be treated as the logical consequence of an ionospheric short-circuiting (e.g., Heppner et al., 1967) in the reversal region but this does not specify the cause of the short-circuiting which is likely to be closely tied to conductivity changes resulting from a change in the precipitating particle flux from the magnetosphere. One important point, consistent with the feedback characteristic, is that sudden bay onsets of appreciable magnitude occur only when aurora is already present. Another important point is that electric field measurements (Heppner, 1972) are not consistent with invoking major changes in the global electric field as a cause of bay activations. Similarly, as discussed here, the basic properties of the Harang discontinuity do not change with the level of disturbance. The Harang discontinuity is, however, the ionospheric locus of initial sudden change when a bay activation occurs. This suggests that the search for instability mechanisms should be directed to the conditions found at the location of the discontinuity. Equatorward convection, transverse to L shells, and the presence of auroral forms predominantly aligned parallel to L shells are two such conditions. Thus, at the Harang discontinuity the conditions are optimal for obtaining a mixing of plasma populations having distinctly different energy spectra. Detailed time-space mapping at and near the discontinuity for all particle energies, including the cold plasma, is difficult to achieve operationally for the scale involved and the difficulty is compounded by a need for selecting the most appropriate times. However, this is an objective worthy of pursuit that could be approached in a limited way using existing particle data relative to auroral and magnetic field determination of the instantaneous location of the discontinuity with careful attention to the limitations of these techniques. Simultaneous electric field (i.e., convection) mapping would be required for the ideal experiment.

References

Akasofu, S.-I., D. S. Kimball and C. I. Meng, 1964–1966: J. Atmosph. Terr. Phys., **26**, 205; **27**, 173, 189; **28**, 489, 497, 505, 627.

Axford, W. I. and C. O. Hines, 1961: Can. J. Phys., **39**, 1433.

Birkeland, Kr., 1908, 1913: 'The Norwegian Aurora Polaris Expedition, 1902–1903', Vols. 1 and 2.

Chapman, S. and J. Bartels, 1940: 'Geomagnetism', Oxford Univ. Press.

DeForest, S. E. and C. E. McIlwain, 1971: J. Geophys. Res., **76**, 3587.

Davis, T. N., 1962: J. Geophys. Res., **67**, 75.

Davis, T. N., 1971: J. Geophys. Res., **76**, 5978.

Föppl, H., G. Haerendel, L. Haser, R. Lüst, F. Melzner, B. Meyer, N. Neuss, H. H. Rabben, E. Rieger, J. Stöcker and W. Stoffregen, 1968: J. Geophys. Res., **73**, 21.

Grafe, A., 1969: Planet. Space Sci., **17**, 1395.

Haerendel, G. and R. Lüst, 1970: In 'Particles and Fields in the Magnetosphere' (Ed. B. M. McCormac), D. Reidel, Dordrecht, Holland, 213.

Harang, L., 1946: Terr. Mag. Atmos. Elec., **51**, 353; Geofys. Publ., **16**, No. 12.

Harang, L. and J. Tröim, 1960: Planet. Space Sci., **5**, 33.

Heppner, J. P., 1954: Thesis, Calif. Inst. of Tech.; 1958, Defence Research Board, Canada, Report No. DR 135.

Heppner, J. P., 1967: In 'Aurora and Airglow' (Ed. B. M. McCormac), Reinhold Publ. Corp., New York, 75.

Heppner, J. P., 1969: In 'Atmospheric Emissions' (Eds. B. M. McCormac and A. Omholt), Van Nostrand Reinhold, New York, 251.

Heppner, J. P., 1972: Planet. Space Sci. (to be published) Papers presented at Substorm Symposium, IUGG, Moscow, August 1971.

Heppner, J. P., J. D. Stolarik and E. M. Wescott, 1971: J. Geophys. Res., **76**, 6028.

Kaiser, T. R., 1958: Ann. Géophys., **14**, 76.

Leznaik, T. W. and J. R. Winckler, 1970: J. Geophys. Res., **75**, 7075.

Maynard, N. C., 1972: In 'Magnetosphere-Ionosphere Interactions' (Ed. K. Folkestad), Universitetsforlaget, Oslo, (in press).

Nagata, T., 1963: Planet. Space Sci., **11**, 1395.

Silsbee, H. B. and E. H. Vestine, 1942: Terr. Mag. Atmos. Elec., **47**, 195.

Vestine, E. H., I. Lange, L. Laporte and W. E. Scott, 1947: Carnegie Inst. of Washington, Report 580.

Wescott, E. M., J. D. Stolarik and J. P. Heppner, 1969: J. Geophys. Res., **74**, 3469.

Wescott, E. M., J. D. Stolarik and J. P. Heppner, 1970: In 'Particles and Fields in the Magnetosphere' (Ed. B. M. McCormac), D. Reidel, Dordrecht, Holland, 229.

Geofysiske Publikasjoner, Vol. 29, 1972

RECENT WORK ON IONOSPHERIC IRREGULARITIES AND DRIFTS

By B. H. Briggs
DEPARTMENT OF PHYSICS,
UNIVERSITY OF ADELAIDE, AUSTRALIA

Abstract

The use of large aerial arrays, and observations of phase as well as amplitude, have provided new information about ionospheric irregularities, and forced a reappraisal of some established methods of analysis of spaced-receiver records. Also, the height range has been extended by the use of weak partial reflections from the *D* region. These developments are reviewed, with particular reference to the validity of the various forms of analysis of drift records.

1. Introduction

When a radio wave is reflected from or transmitted through the ionosphere, temporal fluctuations of amplitude, phase, and direction of arrival are usually observed, due to the presence of irregularities of ionization. The understanding of these phenomena is important from a practical point of view, and also offers the possibility of studying the nature and movements of the ionospheric irregularities. Professor Harang was an active worker in this field, and made observations of ionospheric drifts using both reflected radio waves (Harang and Pedersen, 1957a and b) and radio star scintillations (Harang, 1963).

Deductions about ionospheric irregularities and drifts have usually been made by recording fading or scintillation at three spaced points on the ground, and then subjecting the records to various forms of analysis. Many papers have been published on methods of analysis of these 'spaced-receiver' records.

Recently, there have been some important developments in techniques of observation. Firstly, large arrays of aerials have been used to study the moving diffraction pattern formed over the ground when a radio wave is reflected from the ionosphere, and observations of phase as well as amplitude have been made. This work has given a clearer picture of the way the moving radio pattern is related to the ionospheric irregularities and movements. Secondly, the use of higher transmitter power has enabled ionospheric drifts to be measured using weak partial reflections from the *D* region in the height range 60–100 km. The height of scattering is in this case selected by a time gate and is therefore under the direct control of the experi-

menter. Height profiles of drift can be obtained without the need to use more than one radio frequency.

The object of the present paper is to review the present state of knowledge, to survey critically the various methods of analysis of spaced-receiver records in the light of this knowledge, and suggest some problems for future research. We will not discuss the actual results of ionospheric drift observations (see Kent and Wright (1968) for a recent review).

2. Observations with large aerial arrays

In the recent work with large aerial arrays, pulses on frequencies of a few MHz are reflected from the ionosphere near vertical incidence. The time-varying pattern formed over the ground by the reflected wave is observed with the array, and recorded in various ways.

In some cases, the number of aerials was increased above the basic minimum of three in order to improve the statistics and to study the effect of aerial spacing on the results (e.g. Kelleher, 1966; Harnischmacher, 1968; Beynon and Wright, 1969a). In other cases, the radio pattern was sampled at a sufficient number of points to enable it to be converted into a visible pattern for direct observation. Thus, MacDougall (1966) used a 4×4 matrix of aerials with a spacing of 100 m, Haubert and Doyen (1966) used a 6×6 matrix with a spacing of 30 m, and Briggs et al. (1969) used a circular array 1 km in diameter, containing 89 crossed dipoles on a rectangular matrix with a spacing of 91 m. The methods used to produce a visible picture are described in the references cited. Here we wish to emphasize certain features of the results, which have an important bearing on the problem of the relation of the diffraction pattern to the ionospheric irregularities.

The direct observations of the moving pattern show that it is often not random, but made up from bands or 'fringes'. Felgate and Golley (1971a) made a comprehensive study of the patterns observed with the 89 element array at Buckland Park near Adelaide, South Australia, using a frequency of 1.98 MHz, and found that they could be classified into seven types as follows:

Type 1. Patterns with a scale much larger than the array (i.e. much larger than 1 km).

Type 2. Fairly regular fringes moving across the field of view. The spacing varied from 400 m up to a distance greater than the size of the array.

Type 3. Crossing fringes, in which several sets of moving fringes appear to be superimposed.

Type 4. Random patterns moving with a steady velocity. Individual features may change as they move.

Type 5. Patterns usually random, as Type 4, but from time to time resolving themselves into one or more sets of fringes.

Type 6. Random patterns with rapidly changing direction of movement.

Type 7. No fading at all.

The relative frequencies of occurrence of the various types of pattern are given in Table I.

Table I.

Pattern Type	1	2	3	4	5	6	7
E	12%	10%	24%	17%	23%	12%	2%
E_s	19%	34%	20%	3%	14%	6%	4%
F	46%	18%	7%	3%	0	2%	24%

Patterns of Type 2 in which fringes with constant spacing and orientation cross the field of view produce periodic fading records on a single receiver. Several workers have noted the occurrence of periodic fading (e.g. Kushnerevsky and Zayarnaya, 1960; Monro, 1962; Essex, 1968), and Pfister (1971) has reported the presence of discrete frequencies in the power spectra of fading records. However, patterns of Types 3 and 5 would probably appear random on a single receiver, and the rapid changes which occur suggest that spectral analysis of a long record may often be misleading. When the whole pattern is made visible, rapid changes can be studied, and fringe structures consisting of as few as one or two bright or dark bands can be recognized. The proportion of records which contain fringe-like structures is therefore larger than would be estimated from the study of individual fading records. If we restrict attention to patterns which would produce 'usable' fading records (i.e. excluding Types 1 and 7), the proportions which contain fringes of one kind or another are as follows: E, 65%; E_s, 88%; F, 83%. We must conclude that the majority of drift results are influenced by the presence of fringes.

Direct observations of the pattern also enable rapid changes of velocity to be studied. Felgate and Golley (1971a) found that the velocity could sometimes double in magnitude, and/or reverse in direction, in a time of the order of 1 min. Such variations would be difficult to resolve with a three-receiver system, because the minimum length of record which can be usefully analysed is of the order of several minutes.

3. Models of the reflection process

The first models of the reflection process assumed that the ionosphere contained small irregularities and behaved like a random screen to which diffraction theory could be applied (Booker et al., 1950; Ratcliffe, 1956). If, in Fig. 1a, the irregularities of scale L have a mean drift velocity V, and also an r.m.s. velocity v_0 among themselves, it can be shown that the diffraction pattern will move over the ground with a velocity $2V$ and will also change in form as it moves. Application of the 'full-correlation analysis' (Sect. 5) to three-receiver records provides measures

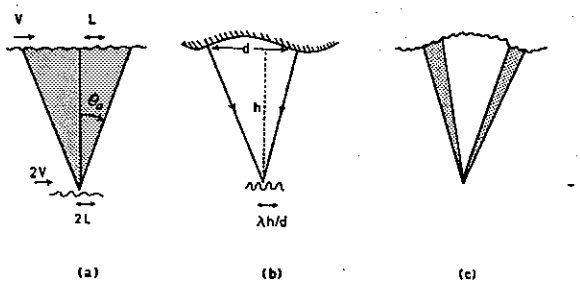


Figure 1. Various possible models for waves totally reflected from an irregular ionosphere. In (a), the random screen model, the angular distribution of downcoming energy takes the form of a continuous 'filled-in' cone. In (b), the specular point model, a small number of discrete downcoming waves occur. Model (c) is a combination of (a) and (b).

of V and, rather more indirectly, of v_0 (Kent, 1966). It has also often been assumed that the scale of the diffraction pattern will be $2L$. The factors of 2 arise from the so-called 'point-source' effect (Wright, 1968; Felgate, 1970). The angular distribution of the downcoming energy will take the form of a 'filled-in' cone centred on the zenith, which can be characterized by a half-width θ_0 (Briggs and Phillips, 1950).

Based on this model, the following typical values are obtained for cases of total reflection from the E, E_s, and F regions in the frequency range 2–6 MHz:

Drift velocities	$V \sim 50$ m s^{-1}.
R.M.S. velocities	$v_0 \sim$ few m s^{-1}.
Scales	$L \sim$ few hundred m.
Half-width	$\theta_0 \sim 5$ deg.

According to this model, the diffraction pattern should be random, and the occurrence of periodic fringes cannot be explained.

Several workers who have observed fringes or periodic fading have proposed the alternative model illustrated in Fig. 1b. Here the wave is reflected from an isoionic surface which is smooth enough to act locally as a specular reflector, but which contains large scale ripples or tilts. From any point in the ground a small number of normals can be drawn to this surface, and the feet of these normals will be specular points for propagation from a transmitter to a nearby receiver. Thus a small number of discrete downcoming rays will be present rather than a filled-in cone, and these rays will interfere to form the pattern over the ground. If there were just two such rays, parallel interference fringes would be formed. The scale of these fringes would depend on the horizontal separation of the reflection points, and their motion would be determined by the horizontal and vertical motions of the reflection points. If there are several specular points, each pair will contribute a set of fringes with a particular orientation and spacing. Computer simulation (Holmes, 1971) has shown that three downcoming waves produce a pattern in which crossing fringes are recognizable by eye, but more than three produce a pattern which looks random. Also, horizontal motion of the reflecting surface with velocity V produces a translation of the pattern with velocity $2V$. This model accounts very well for

the fact that patterns can look either random or fringe-like, and can change from one type to the other in a short space of time. Monro (1962) and Pfister (1971) have suggested a combined version of the two models, as shown in Fig. 1c.

To decide conclusively between the various models illustrated in Fig. 1, direct observations of the angular spectrum of downcoming waves are required, and these are possible with large aerial arrays by the use of image-forming or rapid beam-swinging techniques. It is expected that both methods will be brought into use in the near future in Australia (Briggs et al., 1969; Holmes, 1960; Mines, 1969). Valuable evidence can also be obtained from observations of phase as well as amplitude, and Pfister (1971) has deduced angles of arrival and Doppler shifts of discrete components in the angular spectrum, by the use of spectral and cross-spectral analysis.

If the specular point model is accepted, at least on those occasions when fringes are observed, the scale of the ionospheric irregularities can be estimated from the observed spacing of the fringes (see Fig. 1b). It is found that typical values for the horizontal separation of the reflection points are 15 km for *E* region reflections, and 50 km for *F* region reflections. Thus the ionospheric irregularities are very much larger than the observed scales of the patterns.

All the models of Fig. 1 are idealized, since they neglect scattering or angular deviation occurring below the reflection level itself.

4. Evidence from phase-path measurements, and the role of internal gravity waves

There is, of course, other evidence for the existence of large scale tilts and undulations in the effective reflecting surface. This comes from observations of group and phase height changes (e.g. Munro, 1958; Monro, 1962), from observations of directional fluctuations (e.g. Bramley, 1953), and from observations of Doppler shifts (e.g. Georges, 1968). However, most of these observations refer only to the *F* region. Table I shows that fringes are observed frequently for *E* region reflections as well as for *F* region, and it is therefore of interest to look for independent evidence of large scale undulations in the *E* region. The best technique for this purpose is the phase path method which enables very small changes in the height of reflection to be detected.

In fact as early as 1957, Landmark (1957) discussed the fading of radio waves reflected from the *E* region, and, on the basis of phase path measurements, suggested a model like that of Fig. 1b. The horizontal scales ranged from 10 to 40 km. He also pointed out that if the vertical amplitude became large, more than one discrete reflection point could occur, giving rapid amplitude fluctuations due to interference effects.

More recently, Vincent (1971) has recorded phase path variations of *E*-region echoes using three aerials of the Buckland Park array. The maximum available aerial separation of about 1 km was used. The phase was found to vary by 10 to 50 radians in a quasi-periodic manner and with periods of the order of 5 min. This corresponds to height fluctuations of the order of 120 to 600 m. Small time displacements were observable between the variations at the three aerials, from which the horizontal speed of the isoionic surface could be determined; the speeds were in the range 30 to 150 ms^{-1}. When combined with the observed periods,

these speeds enabled the horizontal scales to be determined; these were found to be in the range 10 to 60 km. The observations are quite consistent with the deductions made from the fringe patterns, in so far as velocities and horizontal scales are concerned.

If we accept that the actual ionospheric irregularities involved in drifts experiments have these large horizontal scales, the question arises as to the cause of such irregularities, and the significance to be attached to the observed drift velocities. Now the scales in both *E* and *F* regions are just the values expected for internal gravity waves (Hines, 1960). Additional evidence that gravity waves are involved comes from a further analysis carried out by Vincent (1971). The power spectrum of the phase height variations was calculated, and was found to show a small peak near to 5 min period, the expected Brunt-Vaisala frequency at *E*-region heights, followed by a rapid cut-off for shorter periods. (See also Vincent, 1969.)

It seems likely that the mean drift velocity obtained from close-spaced receivers and arrays will give the horizontal velocity of the isoionic surface, even if interference effects are present. However, the velocities obtained could be the horizontal phase velocities of gravity waves, and need bear no relation to the drift velocity of either the neutral air or the ionization. This conclusion must, however, be reconciled with the fact that the velocities observed at *E*-region heights do often show quite clear tidal effects and agree reasonably well with measurements of the drift of the neutral air determined by independent methods (Kent and Wright, 1968). The best solution to this apparent paradox seems to be the idea of critical layers proposed by Hines (1968) in which gravity waves travelling with the velocity of the neutral air are preferentially amplified, and therefore dominate the observations.

The significance of the *F*-region drifts is not at present clear, and more experiments are needed to compare the velocities of travelling ionospheric disturbances (T.I.D.'s) with velocities derived from closely spaced receivers. Some workers have reported reasonably close agreement (e.g. Jones et al., 1957; Gusev et al., 1960) but diurnal and seasonal variations differ. The presence of fringe patterns strongly suggests that T.I.D.'s of suitable scales may influence *F*-region drift observations, but that small scale irregularities also exist at certain times is indicated by the phenomena of 'spread-*F* echoes' and radio star scintillations. Other possible mechanisms by which T.I.D.'s might affect observations have been discussed by MacDougall (1966).

5. Implications for three-receiver record analysis

Many workers will continue to use the three-receiver method for the observation of ionospheric drifts, because of its simplicity and economy. It is therefore important to decide upon the best method of analysis of the records, in the light of the new evidence provided by the large arrays and by phase measurements.

Most workers would probably reject three-receiver records which were obviously periodic. However, it is now clear that this is not sufficient to guarantee that interference effects between discrete specular reflections are absent. Therefore, this possibility must be allowed for in devising methods of analysis, and in interpreting the results obtained.

We will consider first the method usually known as full 'correlation analysis' (Briggs et

al., 1950; Phillips and Spencer, 1955). In this method it is assumed that the pattern is random and moves with a constant velocity; it may also change in form as it moves. A possible elongation in a preferred direction is allowed for, and is described by an ellipse which is a contour at the 0.5 level of the two-dimensional spatial correlation function. The random changes are described by a parameter V_c of the dimensions of velocity. It can be shown that, on these assumptions, an 'apparent' velocity calculated in a naive way from the time displacements is always larger than the true velocity of drift. The method showed how the true velocity can be calculated in the presence of anisotropy and random changes.

The method is only concerned with describing the motion of the pattern, and not with the way it is produced, and is therefore applicable in principle to patterns formed by interference effects. The difficulty is, however, that the necessary correlation functions can only be calculated from records of several minutes duration, and the direct observations of the pattern have shown that its velocity and type can change during this time. Velocity fluctuations, and random changes of the pattern as it moves, are probably due to horizontal and vertical motions of specular points, which cause related movements of the interference fringes, and so have no direct interpretation in terms of ionospheric movements. Similarly, the anisotropy may be only an average effect for the length of record analysed, and may be produced by one or two fringes occurring for a short time in a pattern which is at other times random in nature. In fact random patterns in which all features are consistently elongated in a preferred direction do not seem to occur; rather the patterns tend to be mixtures of random features with fringes of very large elongation. In any case, the spatial correlation ellipse will have a scale much smaller than the ionospheric irregularities themselves, as pointed out earlier, whenever interference effects are present.

In spite of the uncertain interpretation of the random velocity V_c and the pattern scale and anisotropy, it has been shown that, as far as the mean velocity of drift is concerned, the 'true' velocity of correlation analysis is the best available estimate, provided the aerial spacing is not too small relative to the scale of the pattern. If the aerial spacing is small the 'apparent' velocity calculated directly from the time displacements is preferable. This has been established as follows (Golley and Rossiter, 1970). By the use of the large aerial array at Buckland Park, velocities could be determined almost instantaneously from direct observations of the dis-

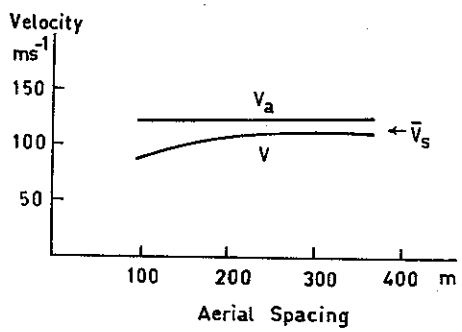


Figure 2. Comparison of different methods of determining the velocity of the pattern, using an array of 89 aerials. Values of V and V_a were determined by the standard correlation analysis, using three aerials at a time, with various separations. The velocity V_s was determined by the spatial correlation method, using all 89 aerials. $\overline{V_s}$ is the average value of V_s during the lengths of record used to determine V and V_a.

placement of the pattern in one second. We will call this method the 'spatial correlation' method (Briggs, 1968) and denote the velocity obtained by V_s. The velocity was also determined by applying the standard correlation analysis to fading records obtained from various sets of three aerials selected from the 89 aerials of the array. The 'apparent' and 'true' velocities of correlation analysis were compared with the average value of V_s during the time of the records. A typical result is shown in Fig. 2. It will be seen that the 'true' velocity V is less than the correct value V_s, but tends to this value as a limit as the aerial separation increases. The 'apparent' velocity V_a is independent of aerial spacing, but somewhat larger than the correct value. The departures from the value V_s are often larger than in the particular example shown.

Golley and Rossiter (1970) conclude that for total reflections from the E and F regions, the optimum arrangement of three aerials would be at the corners of an equilateral triangle of side 300 m, and that the best velocity estimate would be the 'true' velocity of correlation analysis.

There is at present some uncertainty as to the cause of the low values of 'true' velocity obtained by correlation analysis when the aerial separation is small, but it is generally agreed that the effect is instrumental in origin. Non-linearities and digitizing errors in the various channels (Golley and Rossiter, 1970), and interchannel coupling effects (Fedor and Plywaski, 1971) have been suggested. A related effect is a tendency for the spatial correlation ellipse to have its major axis along the hypotenuse of the aerial triangle, if small right-angled triangles are used (Beynon and Wright, 1969; Golley and Rossiter, 1970).

An important discussion of the relation of correlation analysis to different variants of the 'similar fades' type of analysis has been given by Sprenger (1969). The method of similar fades is, however, not very compatible with modern methods of digital recording and computer analysis, and the use of correlation functions is to be preferred.

We consider next the 'three-dimensional' form of correlation analysis in which an attempt is made to measure vertical as well as horizontal velocities by the use of two or more radio frequencies reflected from different heights (Gusev and Mirkotan, 1960; Fedor, 1967; Wright and Fedor, 1969). The principle of this method is shown in Fig. 3. Suppose we had four probes O, X, Y, Z situated in the ionosphere itself, each capable of recording the temporal fluctuations of electron density. We could then apply a three-dimensional form of correlation analysis in order to determine; (*i*) the mean drift vector V, which could be at any angle to the (x, y, z) coordinate system, (*ii*) a random velocity V_c, and (*iii*) an ellipsoid which would be a contour surface at the 0.5 level of the three-dimensional spatial correlation function of the electron density fluctuations. This ellipsoid has been called a 'correloid' (Wright and Fedor, 1969). It has been proposed that a system equivalent to that of Fig. 3 can be obtained by the use of receivers on the ground, provided two radio frequencies f_1 and f_2 are used. Thus receivers O, X, Y operate at f_1 and an additional receiver at O operates at f_2, for which the reflection level is higher than for f_1 by the desired amount.

Unfortunately, this proposed equivalence is open to criticism on several grounds. In the first place, we have already seen that the horizontal scale of the ionospheric irregularities may be much larger than that of the diffraction pattern over the ground, so that even the three

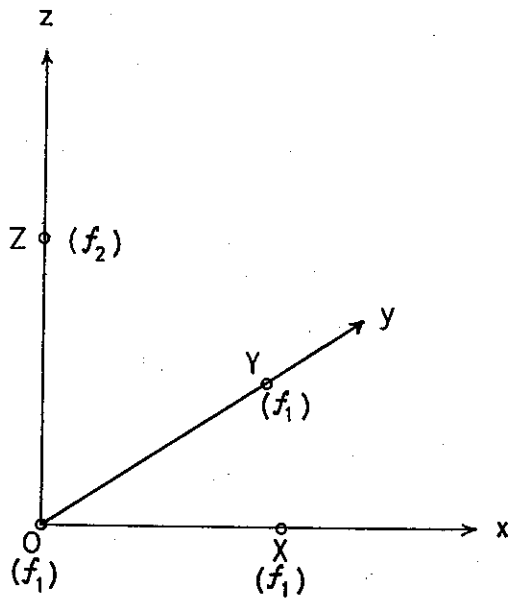


Figure 3. Arrangement of probes in the ionosphere which could be used to determine three-dimensional drifts.

receivers on the frequency f_1 are not equivalent to three probes in the ionosphere itself. This difficulty cannot be overcome by the use of complex correlation, because, although the complex correlation function is invariant (Ratcliffe, 1956), its scale is smaller than that of the diffracting screen when the phase deviation is large. Secondly, the use of two frequencies to obtain an effective vertical separation is open to the objection that the diffraction pattern for a frequency f_2 would be different from that for f_1, even if the ionospheric structure were independent of height. Indeed, several workers who have investigated 'frequency diversity' effects have shown that they can be explained by the path differences between waves arriving from different directions (Briggs, 1951; Essex and Hibberd, 1968). Because of the neglect of these effects, the length of the vertical axis of the 'correloid' will be underestimated. If the 'correloid' is incorrectly determined, so will all the derived quantities, including the drift velocity. Also the vertical motion would have to be of a rather peculiar type in which irregularities moved relative to reflection levels which remained at fixed heights. We know from phase path observations that the reflection heights are, in fact, continually changing by amounts larger than the required vertical separation *OZ* in Fig. 3. Thus, the three-dimensional form of correlation analysis seems quite incompatible with the specular point model of Fig. 1b.

It is interesting that fading records obtained on two frequencies sometimes *do* show time displacements which suggest, at first sight, a vertical component of velocity. However, several alternative explanations must be excluded before this is accepted. On the specular point model such displacements would be expected whenever one set of fringes dominates the records, because the fringes on neighbouring frequencies will be displaced relative to each other. A consistent sense of time displacement would not be likely to be maintained for more than a few minutes, however. Essex and Hibberd (1968) have given an alternative explanation based on ionospheric tilts. If consistent time displacements are maintained for times of the order of

hours, as apparently reported by Drobzhev (1966), some kind of vertical motion would seem to be indicated. Further investigation of such effects is desirable.

Observations on several frequencies are, of course, very valuable in determining the height variation of the horizontal velocity of ionospheric drift.

The last method of analysis to be discussed is often called 'dispersion' analysis. In this method, the same Fourier component is selected from each of the three amplitude records, and the phase differences are used to deduce the velocity and direction of motion of a plane wave of the given frequency past the observing system (Yerg, 1956; Jones and Maude, 1965, 1968). If the velocity is independent of frequency, the motion may be called 'non-dispersive'. If the velocity is found to increase with frequency the dispersion is 'positive', and if it decreases with frequency, it is 'negative'.

Jones and Maude (1965) found marked positive dispersion for some records of waves totally reflected from the *E* region. They regarded this as evidence for the propagation of a form of surface gravity wave through the ionosphere, because the observed dispersion matched that expected theoretically for this type of wave.

This interpretation rests heavily on the assumption that a given spatial Fourier component in the ground pattern is produced by a wave of equal wavelength in the ionosphere. However, when interference between discrete downcoming waves occurs, this is not the case. In fact, a particular spatial Fourier component (i.e. set of fringes) corresponds rather to a particular pair of specular points, and its motion depends upon the horizontal and vertical motions of these specular points. The wavelengths in the ionosphere are very much larger than those in the pattern, and of the order expected for internal gravity waves.

The frequent occurence of positive dispersion, and absence of negative dispersion, is an interesting experimental fact which must be explained. There are several secondary effects which can produce an apparent or spurious positive dispersion; in the present case, velocity variations during the length of record analysed seem a possible cause (Felgate and Golley, 1971b).

6. Drift measurements using partial reflections from the D region

It is known that the *D* region contains small irregularities which are able to back-scatter waves on frequencies in the region of 2–4 MHz, and which also cause 'forward-scatter' propagation on higher frequencies. Recently, the back-scattering at vertical incidence has been used to observe drifts by the spaced receiver technique (Fraser, 1965, 1968; Fraser and Kochanski, 1970; Haug and Pettersen, 1970; Rossiter, 1971; Golley and Rossiter, 1971). This has the advantage of extending the height range to lower levels and, because the reflections are partial rather than total, a range of heights may be investigated on a single radio frequency. Energy returned from the desired height is selected by a time gate set at the required time delay after the emission of the transmitted pulse. There is the possibility of recording height profiles of drift over the height range 60–100 km by manual or automatic scanning of the gate.

The occurrence of back-scatter on frequencies of the order of 2 MHz must indicate the

presence of irregularities with scales as small as 75 m in the vertical direction, and the process seems to be one of random scattering from small irregularities, perhaps of turbulent origin.

Some observations of the nature of the pattern produced by *D* region echoes have been made with the Buckland Park aerial array, but so far are limited to observations of the strongest echoes, which come from heights near 90 km. They have shown that the patterns are random and isotropic, and do not have rapid velocity variations. Aerial separations for a correlation of 0.5 are of the order of 100 m, compared with 300 m for total reflections. The angular spectrum is correspondingly wider; with the model of Fig. 1a, the value of θ_0 would be of the order of 16° (Golley and Rossiter, 1971).

These facts suggest that full correlation analysis would be applicable, and we might be encouraged to deduce drifts, scale sizes, and random velocities. Dispersion analysis has also been applied (Haug and Pettersen, 1970; Golley and Rossiter, 1971). However, we must be cautious once again in interpreting the results, because there is a different complication in this case; because of the finite height resolution, a small range of heights of the order of 5 km is sampled for any one position of the time gate. Now sharp gradients of wind are known to exist in this region of the atmosphere, and we cannot assume that the velocity will be constant over this height range. Thus patterns moving with different velocities may be superimposed. This can affect the value of the random velocity V_c of correlation analysis, and introduce spurious positive dispersion, as discussed by Haug and Petersen (1970). It can be shown that the velocity gradients are of the right order of magnitude to explain the effects observed.

In spite of these complications, several workers have shown that the mean drift velocity determined by the spaced receiver technique is in reasonable agreement with the neutral air velocity at *D*-region heights (e.g. Fraser and Kochanski, 1970). This is also true for low frequency waves *totally* reflected from the *D* region at oblique incidence (Sprenger and Schminder, 1968). The method is therefore a promising one for synoptic observations in the height range 70–100 km.

It would be interesting to see whether vertical drifts could be measured in the *D* region by comparing the fading of waves returned from different heights. This technique would not be open to the objections mentioned in connection with the spaced frequency experiment.

7. Suggestions for future work

1. Further observations should be made of phase as well as amplitude.
2. Direct observations of the form of the angular spectrum of downcoming waves are needed, both for *E*- and *F*-region reflections, and *D*-region partial reflections.
3. The reason for the dependence of 'true' velocity on aerial separation requires further investigation.
4. Multi-frequency observations are desirable to study the height variation of drift, and relations between movements at different heights.

5. The possible occurrence of consistent time displacements between fading records on spaced frequencies should be further investigated.
6. Observations of *D*-region drifts at more places are needed.
7. The possibility of detecting vertical motions in the *D* region should be explored.
8. In the *F* region, more comparisons are needed of the velocities of T.I.D.'s and the velocities obtained from fading records on closely spaced receivers.
9. Every opportunity should be taken to compare ionospheric drifts with other observations such as drifts of meteor trails and motions of rocket launched vapour trails.

Acknowledgements

I am greatly indebted to my colleagues for valuable discussions, and especially to Dr. D. G. Felgate, Dr. M. G. Golley, Dr. R. A. Vincent and Mr. N. Holmes, for making their results available prior to publication.

References

Beynon, W. J. G. and J. C. Wright, 1969a: J. Atmosph. Terr. Phys., **31**, 119.
Beynon, W. J. G. and J. C. Wright, 1969b: J. Atmosph. Terr. Phys., **31**, 593.
Booker, H. G., J. A. Ratcliffe and D. H. Shinn, 1950: Phil. Trans. Roy. Soc., **242**, 579.
Bramley, E. N., 1953: Proc. Roy. Soc., **A220**, 39.
Briggs, B. H., 1951: Proc. Phys. Soc., **B64**, 255.
Briggs, B. H., 1968: J. Atmosph. Terr. Phys., **30**, 1777.
Briggs, B. H. and G. J. Phillips, 1950: Proc. Phys. Soc., **B63**, 907.
Briggs, B. H., G. J. Phillips and D. H. Shinn, 1950: Proc. Phys. Soc., **B63**, 106.
Briggs, B. H., W. G. Elford, D. G. Felgate, M. G. Golley, D. E. Rossiter and J. W. Smith, 1969: Nature, **223**, 1321.
Drobzhev, V. I., 1966: Geomag. Aeronomy (English trans.), **6**, 878.
Essex, E. A., 1968: J. Atmosph. Terr. Phys., **30**, 1441.
Essex, E. A. and F. H. Hibberd, 1968: J. Atmosph. Terr. Phys., **30**, 1019.
Fedor, L. S., 1967: J. Geophys. Res., **72**, 5401.
Fedor, L. S. and W. Plywaski, 1971: Private communication.
Felgate, D. G., 1970: J. Atmosph. Terr. Phys., **32**, 241.
Felgate, D. G. and M. G. Golley, 1971a: J. Atmosph. Terr. Phys., **33**, 1353.
Felgate, D. G. and M. G. Golley, 1971b: Aust. J. Phys., **24**, 397.
Fraser, G. J., 1965: J. Atmosph. Sci., **22**, 217.
Fraser, G. J., 1968: J. Atmosph. Terr. Phys., **30**, 707.
Fraser, G. J. and A. Kochanski, 1970: Ann. Géophys., **26**, 675.
Georges, T. M., 1968: J. Atmosph. Terr. Phys., **30**, 735.
Golley, M. G. and D. E. Rossiter, 1970: J. Atmosph. Terr. Phys., **32**, 1215.
Golley, M. G. and D. E. Rossiter, 1971: J. Atmosph. Terr. Phys., **33**, 701.
Gusev, V. D. and S. F. Mirkotan, 1960: 'Issled neodrorednostei v ionosfere' Moscow, Isdatel'stro Akad Nauk SSSR, **4**, 7.
Gusev, V. D., J. W. Kushnerevsky and S. F. Mikotan, 1960: 'Some ionospheric results obtained during the IGY', Elsevier Pub. Co., Amsterdam, 322.
Harang, L., 1963: J. Atmosph. Terr. Phys., **25**, 109.
Harang, L. and K. Pedersen, 1957a: J. Atmosph. Terr. Phys., **10**, 44.
Harang, L. and K. Pedersen, 1957b: J. Geophys. Res., **62**, 183.
Harnischmacher, E., 1968: In 'Winds and Turbulence in Stratosphere, Mesosphere and Ionosphere' (Ed. K. Rawer), North Holland Pub. Co. Amsterdam, 277.
Haubert, A. and G. Doyen, 1966: Ann. Géophys., **22**, 338.
Haug, A. and H. Pettersen, 1970: J. Atmosph. Terr. Phys., **32**, 397.
Hines, C. O., 1960: Can. J. Phys., **38**, 1441.
Hines, C. O., 1968: J. Atmosph. Terr. Phys., **30**, 837.
Holmes, N. 1970: University of Adelaide, Dept. of Physics, Report No. ADP 95.
Holmes, N. 1971: Private communication.
Jones, D. and A. D. Maude, 1965: Nature, **206**, 177.
Jones, D. and A. D. Maude, 1968: J. Atmosph. Terr. Phys., **30**, 1487.
Jones, I. L., B. Landmark and C. S. G. K. Setty, 1957: J. Atmosph. Terr. Phys., **10**, 296.
Kelleher, R. F., 1966: J. Atmosph. Terr. Phys., **28**, 213.
Kent, G. S., 1966: Ann. Géophys., **22**, 427.
Kent, G. S. and R. W. H. Wright, 1968, J. Atmosph. Terr. Phys., **30**, 657.
Kushnerevsky, Yu. V. and Ye. S. Zayarnaya, 1960: NASA Technical Translation T.T.F.-20.
Landmark, B., 1957: J. Atmosph. Terr. Phys., **10**, 288.
MacDougall, J. W., 1966: J. Atmosph. Terr. Phys., **28**, 1093.

Mines, D. F., 1969: Aust. Telecom. Res., **3**, 28.
Monro, P. E., 1962: Aust. J. Phys., **15**, 387.
Munro, G. H., 1958: Aust. J. Phys., **11**, 91.
Pfister, W., 1971: J. Atmosph. Terr. Phys., **33**, 999.
Phillips, G. J. and M. Spencer, 1955: Proc. Phys. Soc., **B68**, 481.
Ratcliffe, J. A., 1956: Rep. Prog. Phys., **19**, 188.
Rossiter, D. E., 1971: Aust. J. Phys., **24**, 103.
Sprenger, K., 1969: J. Atmosph. Terr. Phys., **31**, 1085.
Sprenger, K. and R. Schminder, 1968: J. Atmosph. Terr. Phys., **30**, 693.
Vincent, R. A., 1969: J. Atmosph. Terr. Phys., **31**, 607.
Vincent, R. A., 1971: Private communication.
Wright, J. W., 1968: J. Atmosph. Terr. Phys., **30**, 919.
Wright, J. W. and L. S. Fedor, 1969: J. Atmosph. Terr. Phys., **31**, 925.
Yerg, D. G., 1956: J. Atmosph. Terr. Phys., **8**, 247.

THE OCCURRENCE OF RADIO AURORA AT HIGH LATITUDES: THE IGY PERIOD, 1957-1959

BY A. G. MCNAMARA

ASTROPHYSICS BRANCH,
NATIONAL RESEARCH COUNCIL OF CANADA,
OTTAWA, CANADA

Abstract

Two years of data resulting from the operation of an auroral radar network in the Canadian Arctic have been analyzed to determine the factors controlling the occurrence of radio aurora. The radars were located at Ottawa, Saskatoon, Baker Lake, and Resolute, at magnetic latitudes of 58.9, 61.5, 75.1, and 84.3°N. It was found that magnetic aspect control is not very strong at 48.5 MHz, and that echoes were obtained at all azimuth angles even where the aspect was 19° away from perpendicularity. The latitude of most frequent echo occurrence coincided with maximum of the visual aurora zone. Midnight maxima were observed in the diurnal occurrences at Ottawa, Saskatoon, and Baker Lake, but the maximum occurrence at Resolute was near local noon.

1. Introduction

Since VHF radar echoes were first obtained from aurora (Harang and Stoffregen 1940; Aspinal and Hawkins, 1950; Currie et al., 1953), observations have been made from many locations. A good review of radio aurora observations has been published by Hultqvist and Egeland (1964) and by Leadabrand et al. (1965). For geographical reasons virtually all of these observations were made from locations either well outside the auroral zone or near the low latitude edge. These observations led to conclusions that the polar diagram of the scattering process was very narrow, and was strongly controlled by the magnetic aspect angle defined by the intersection of the radar line-of-sight with the Earth's magnetic field at the reflection point. In general, the exact magnetic normality condition at auroral scattering heights can only be obtained by radars looking toward the pole from magnetic latitudes of about 55° to 60°. Simple aspect theory would lead one to the (erroneous) conclusion that auroral echoes could not be obtained within or inside the auroral zone.

The IGY offered an opportunity to study radar aurora at large aspect angles from inside the auroral zone in Canada. Earlier research in Canada had shown the occurrence of radar echoes at aspect angles up to 10° away from perpendicularity for radars located at Saskatoon and Ottawa. It was inferred that echoes probably could be obtained at much higher latitudes

where the aspect geometry was much more unfavorable. Newly-developed radars were installed in 1956 at Ottawa, Saskatoon, Baker Lake, and Resolute, and were operated continuously for two years, from June 1957 to May 1959. These stations spanned the auroral zone and the polar cap. Their geophysical parameters are listed in Table I. Throughout this paper, the mnemonics OT, SA, BL, and RB are used to denote the four stations Ottawa, Saskatoon, Baker Lake, and Resolute respectively.

Table I. *Characteristics of the auroral radar stations*

	OT	SA	BL	RB
Geographic latitude	45.4 N	52.1 N	64.3 N	74.7 N
Geographic longitude	75.8 W	106.7 W	96.1 W	94.9 W
Geomagnetic latitude	58.9 N	61.5 N	75.1 N	84.3 N
Dip angle	74.8	77.2	86.2	89.2
Declination	—13	+18	+5	—82
L-value	3.6	4.3	14.7	94.2

It may be noted that RB is located between the geomagnetic dipole north pole and the dip pole. The great circle distance from SA to BL is 1485 km, and the distance from BL to RB is 1155 km.

The radar data presented were obtained during the IGY period, June 1957 to May 1959, and represent approximately 34,000 half-hour intervals or 10^6 minutes of operation at each station. Of this time, an average of 3% of the data were lost from all causes.

The auroral echoes discussed in this paper were obtained by direct backscatter, with ranges from 200 to 1200 km. The most frequent occurrence of echoes was in the interval 500 to 800 km.

Aurorally-associated echoes propagated by 2 *Es* ground backscatter, and non-auroral ground backscatter propagated by the 2*F* mode, were also observed on the radar network. These echoes are easily recognized on the recorded display and they are not included in this paper.

2. The Canadian IGY auroral radar network

Identical auroral radars were installed at the four locations tabulated in Table I. The radars were continuously operating automatic systems of the type described by McNamara (1958), differing mainly in a new antenna system which was employed. Table II gives the radar parameters. The principal novel feature of this network was the orientation and synchronization of the rotating beam patterns to provide simultaneous overlapping coverage of large regions by pairs of radars, with the object of studying the aspect mechanism.

Table II. *Radar system parameters*

Frequency	48.5	MHz
Wavelength	6.19	m
Peak power	1	kW
Pulse length	300	μs
Pulse repetition frequency	70	s^{-1}
Horizontal beam width (1/2 power)	56	degrees
Antenna gain (two way)	12.5	db
Receiver sensitivity	$2 \cdot 10^{-16}$	W

Figure 1 shows the beam pattern coverage of the network with the beam pattern scaled to correspond to 1140 km range, the horizon intersection with the 100 km level. The antenna consists of an array of four corner reflectors with the beam stepped electronically through the four increments of 90° each. The beam dwells for 30 s in each position thus taking two minutes for a complete 'rotation'.

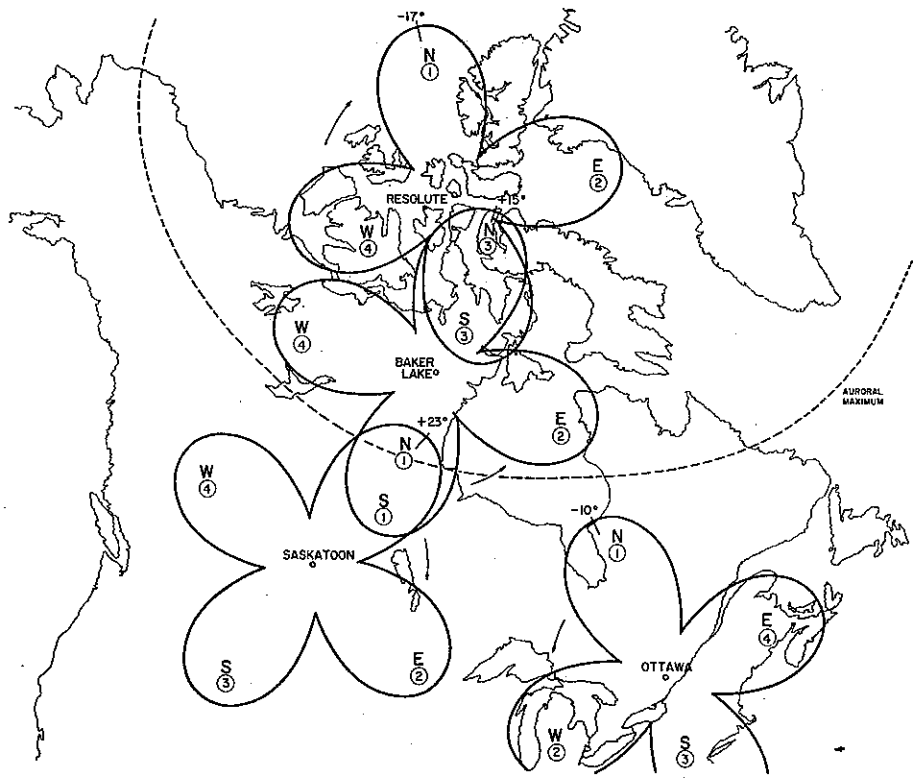


Figure 1. The locations of the auroral radars in the Canadian IGY network. The cloverleaf patterns represent the simultaneous, successive antenna beam positions of the synchronized network. The beam patterns are scaled to represent the approximate geographical coverage, out to 1140 km range.

The switchings of the beams at the four stations were such that at any instant all four radars were in beam position 1, or 2, or 3, or 4. Synchronism was maintained at all times by precision electronic oscillators and clocks which were checked daily against WWV

or CHV standard time signals. These clocks also provided the time information on the data records. Accumulated time errors were not allowed to exceed 2 s at any time.

The transmitted frequencies were in the interval 48.2 to 48.8 MHz, and were spaced approximately 200 kHz from each other. Thus, any possibility of mutual interference was avoided and the frequencies were close enough that the data could be interpreted in terms of a single frequency.

The 'clover-leaf' beam patterns were offset from exact alignment with the magnetic meridian in order to obtain the best overlapping coverage. The offsets from true geographic north are indicated on the north beams in Fig. 1. Designation of the beam positions are with reference to geographic directions.

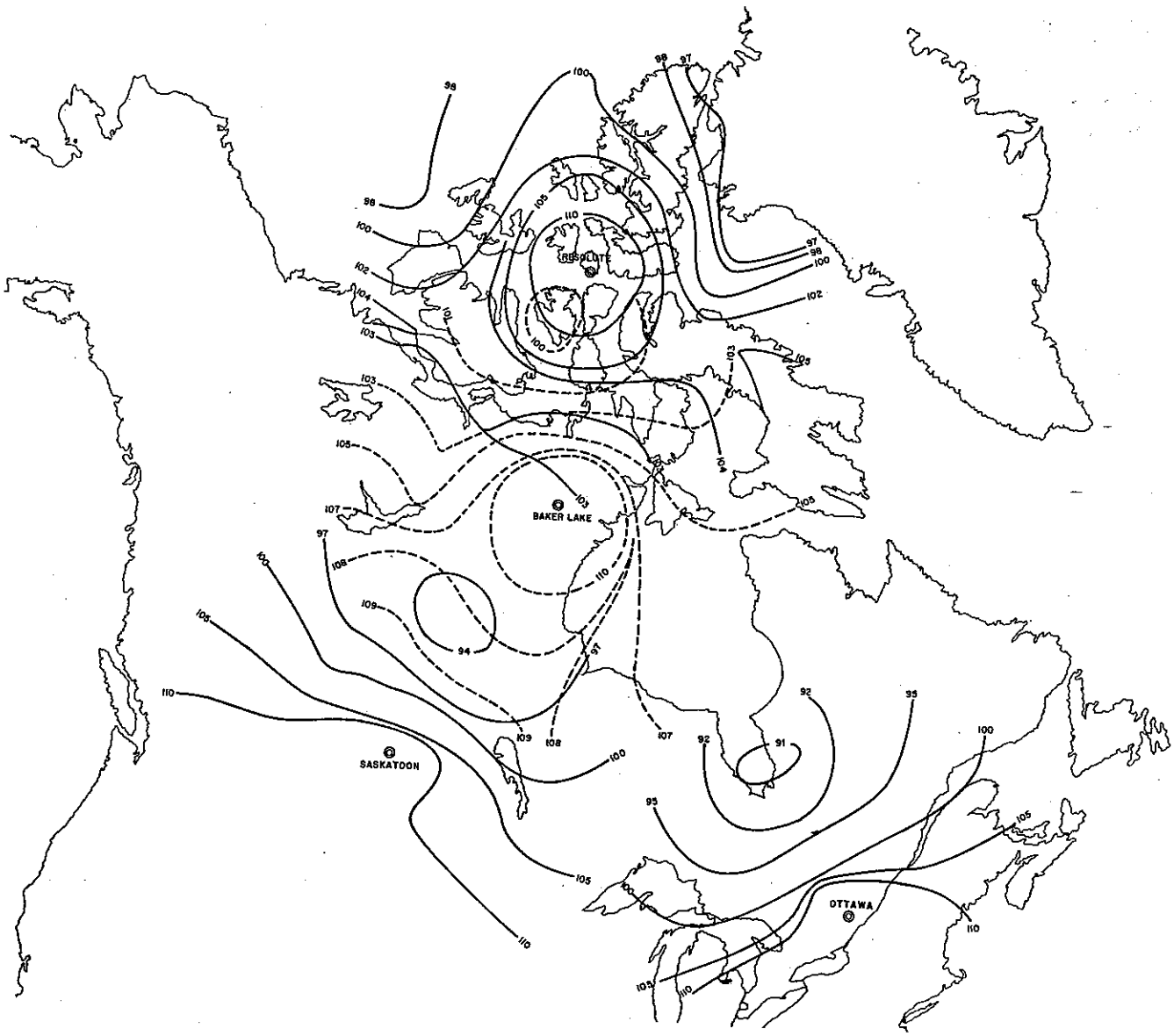


Figure 2. Magnetic aspect contours for the four radars. The contour numbers are the angles between the radar line of sight and the Earth's magnetic field. A reflection height of 100 km is assumed.

The antenna radiation was directed horizontally and was vertically polarized. Accurate antenna pattern and gain measurements were obtained by scale model tests. In the vertical plane, the antenna pattern was controlled primarily by the height of the antenna above ground. The antenna height was set such that no nulls occurred in the radiation pattern between 0 and 15° elevation.

3. Magnetic aspect angles

Figure 2 shows the magnetic aspect angle contours for the radar network. The aspect angles are derived from the intersection of the radar line-of-sight with the Earth's magnetic field at an altitude of 100 km. Since the actual magnetic field at this level controls the plasma configuration, the measured surface values of the Earth's main geomagnetic field (smoothed to remove local anomalies) were used in the calculations rather than an approximate dipole model. The values of inclination and declination were obtained from 1955 magnetic survey data of Canada's Department of Mines and Technical Surveys. At that epoch, the north magnetic dip pole was located at 73.7° N, 101.0° W. The best aspect angles at Ottawa, Saskatoon, Baker Lake, and Resolute were 1°, 4°, 8°, and 7°, respectively, away from perpendicularity. The regions of best aspect may be seen in Fig. 2, and in all cases they lie generally to the north of the station.

On a few occasions, strong echoes were obtained right in to the limit of readability, at about 200 km range. At this range, the vertical angle of arrival indicates off-perpendicular scattering angles of up to 30°.

4. Observed azimuth distributions

The data in Figs. 3 and 4 show conclusively that magnetic aspect control of 48 MHz radar auroral signals within the auroral zone is not nearly so strong as commonly believed. In these locations the geographical regions of occurrence of the scattering structures determine the occurrence frequencies more strongly than do the aspect angles at which they are observed. The same effect has also been observed by Bullough (1961) working within the Antarctic auroral zone.

Figure 3 shows an example of auroral echoes occurring to the east, south, and west at Baker Lake where the aspect angles were 19° away from perpendicularity, but none were occurring to the north at an aspect of 10°. Figure 4 shows the azimuth distributions of the entire two years of data for the four stations. At Ottawa and Saskatoon where the directions of maximum auroral occurrence and best aspect both occur to the north, the aspect control is more evident and shows a centering on magnetic north. (See for example the Saskatoon azimuth data of Currie et al. (1953), where the preferred echo azimuth is approximately 20° east of north.) At Baker Lake and Resolute, the occurrence distributions obviously dominate the aspect function.

Analysis of a portion of the IGY data from the Saskatoon-Baker Lake overlap region was used by McDiarmid and McNamara (1967) to derive the aspect attenuation rate. Using the

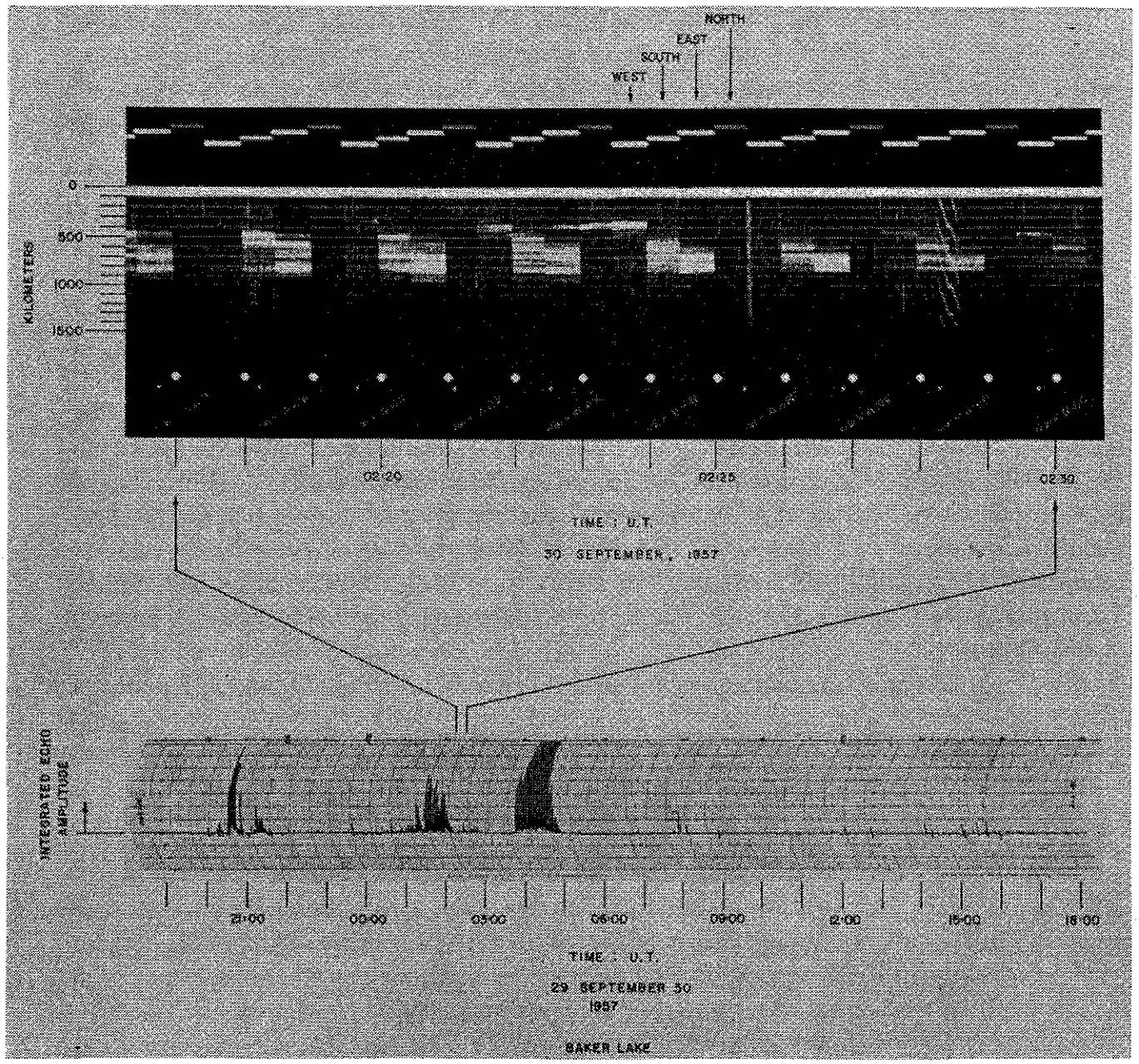


Figure 3. An example of radio aurora at Baker Lake, with echoes occurring to the east, south, and west.

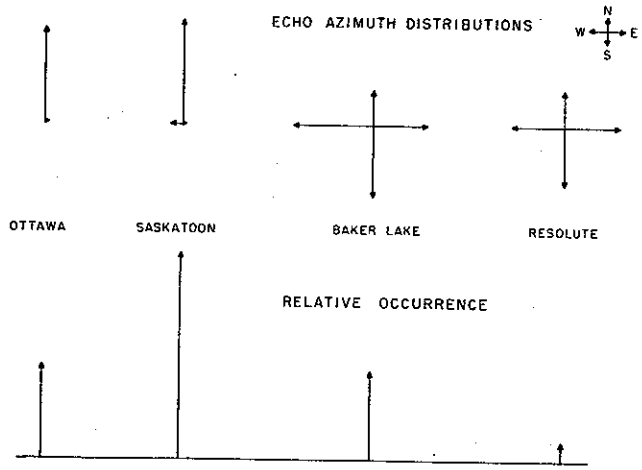


Figure 4. Upper part: Echo azimuth distributions observed at Ottawa, Saskatoon, Baker Lake, and Resolute in the directions indicated by Figure 1. The four plots are not to the same scale. Lower part: The relative total occurrence of echoes, independent of direction but plotted to the same scale.

simultaneous echoes from these two stations, a mean aspect attenuation of 1.1 db per degree was found for the interval from 5° to 20° off-perpendicularity.

This result is also confirmed statistically using the observational data of McNamara (1960) on the cross-section probability distributions of the auroral echoes. In that paper it was shown that the cross-section occurrence above a limiting value of σ_L is proportional to $1/\sigma_L{}^x$, where x is between 1.5 and 1.67. Using the observed ratio of 6.9 for the total minutes of occurrence of echoes on Saskatoon-north compared to Baker Lake-south, a value of 1.2 to 0.9 db per degree is obtained.

5. Seasonal and latitudinal variations

Most of the data presented in this paper are based upon the occurrence of radio aurora in each of the 48 half-hour intervals throughout the day. The figures show the occurrence on this basis or as a percentage of the total number of half-hour intervals. Over the two years of the IGY operation, the average percentage occurrence was as shown in Column 1 of Table III. The data were also read on a finer time scale in terms of the actual number of minutes of occurrence; these data are listed in Column 2. The differences in the two measures of occurrence reflect the more transient nature of many of the echoes observed at the higher latitudes. The half-hourly intervals were used in most of the analysis because they were more suitable for computer analysis and because other phenomena could be more readily scaled and correlated on this basis.

Table III. *Average occurrence of radio aurora*

Station	% of half-hour intervals	% of minutes	Ratio $\frac{\%\text{ intervals}}{\%\text{ minutes}}$
OT	9.8	5.6	1.8
SA	21.6	12.1	1.8
BL	9.3	4.1	2.3
RB	1.8	0.78	2.3

Figure 5 shows the monthly occurrence at each station throughout the two year period. The occurrence rates at Ottawa and Saskatoon, outside the auroral zone, correlate well. However, the most notable feature is the tendency toward anti-correlation within the polar cap. Baker Lake occurrence tends to be a maximum when the lower latitude stations are minimum. Resolute is somewhat independent in its occurrence characteristic. This is particularly evident when one examines the day-to-day occurrence of echoes.

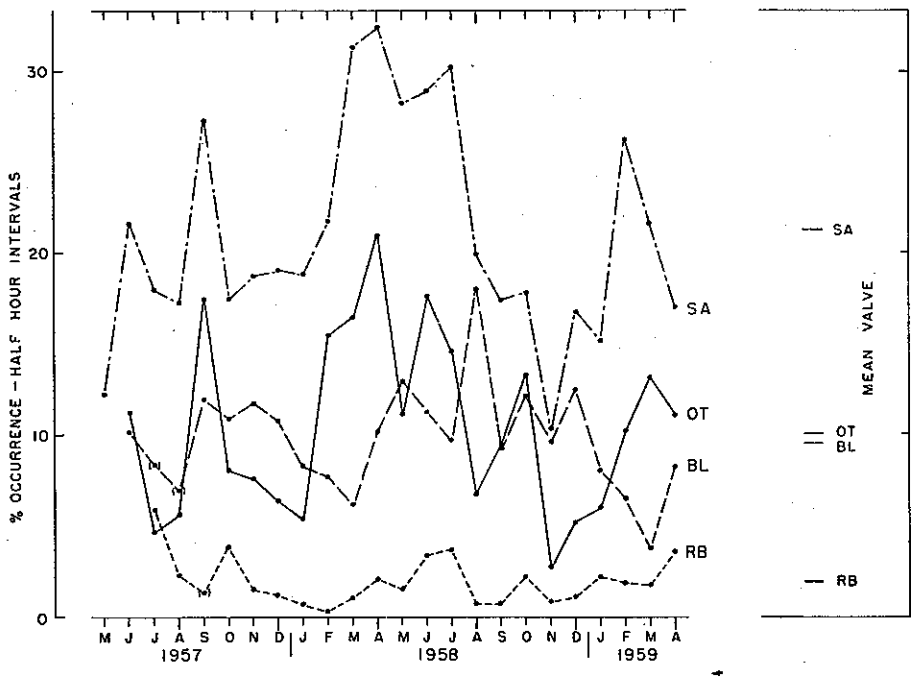


Figure 5. The monthly occurrence values of radio aurora throughout the IGY interval.

The azimuth distributions for Baker Lake and Resolute have also been examined for seasonal effects. The data were divided into four quarters, centred on the equinoxes and the solstices. No significant seasonal variation was found for Baker Lake. In the Resolute data there appeared to be some preference for eastward echoes in the summer and a preference for westward echoes in the winter.

6. Diurnal variations

6.1 *All echoes.* Selected directions of the diurnal distributions are shown in Fig. 6. The data are from the full two years of continuous observation. Ottawa has a diurnal maximum occurring near local midnight. However, the 48 MHz data taken over shorter periods have shown a doublepeaked night-time maximum with a small dip which varies around 0400–0530 UT, near local midnight. The time of the first maximum varies from 0100 to 0230 UT and the second maximum is at 0700 to 0800 UT. (This characteristic also appears in the 488 MHz data of Blevis et al., 1963.) Saskatoon has a midnight maximum and a small morning peak on the side at 1300 UT. This characteristic is always present on the Saskatoon data, and was also observed in 1952 by Currie et al. (1953).

For better comparison with the other stations, only the north and south distributions are plotted in Fig. 6 for Baker Lake. The southward occurrence has a single nighttime maximum at 0430 UT, but the northward occurrence has maxima at 0430 and at 1730 UT. The latter activity is seen to coincide with the activity at the pole station, Resolute.

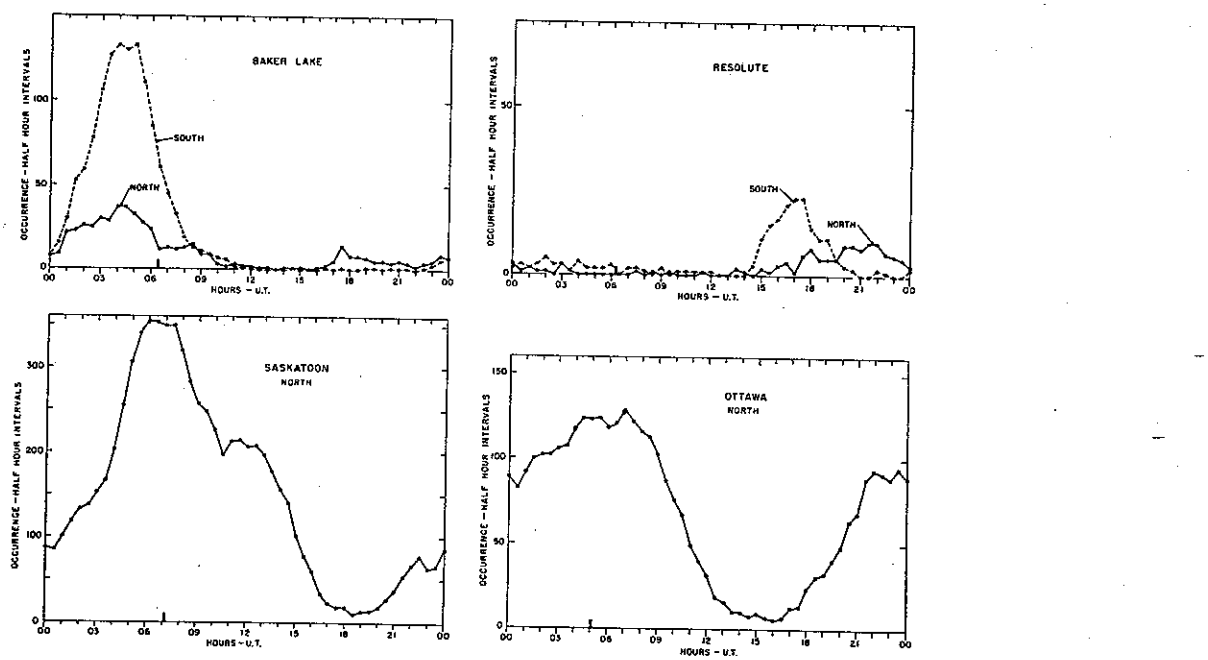


Figure 6. Diurnal occurrence variations for selected directions at Ottawa, Saskatoon, Baker Lake, and Resolute. Data are from two years of continuous operation. The black bars on the abscissae mark the time of local midnight at the station.

Diurnal plots of Baker Lake activity for the east and west directions (not reproduced here), are similar to those for Baker Lake-south but displaced in time. The eastern maximum occurs at 0300 UT, the southern at 0430 UT, and the western at 0500 UT.

6.2 *Diffuse and discrete radio aurora.* Radio aurora has frequently been categorized as diffuse or discrete, and attempts have been made to relate the two types to different scattering mechanisms. Leadabrand (1961) has ascribed diffuse and discrete echoes to horizontal layer structure and to vertical ray structure respectively, and Unwin (1968) has associated diffuse

echoes with the two-stream instability. The IAGA Working Group on Radio Auroral Nomenclature has attempted to further divide the types of echoes according to range spread, time, duration rate, and height of reflection (Unwin, 1967). In general, the classification of echoes as diffuse or discrete is somewhat arbitrary, and is very dependent upon the parameters of the radar. With the pulse length, beam width, antenna elevation, and antenna scan employed in the Canadian network, it was only possible to recognize the two basic categories, based upon the appearance of the echoes on the range vs. time display.

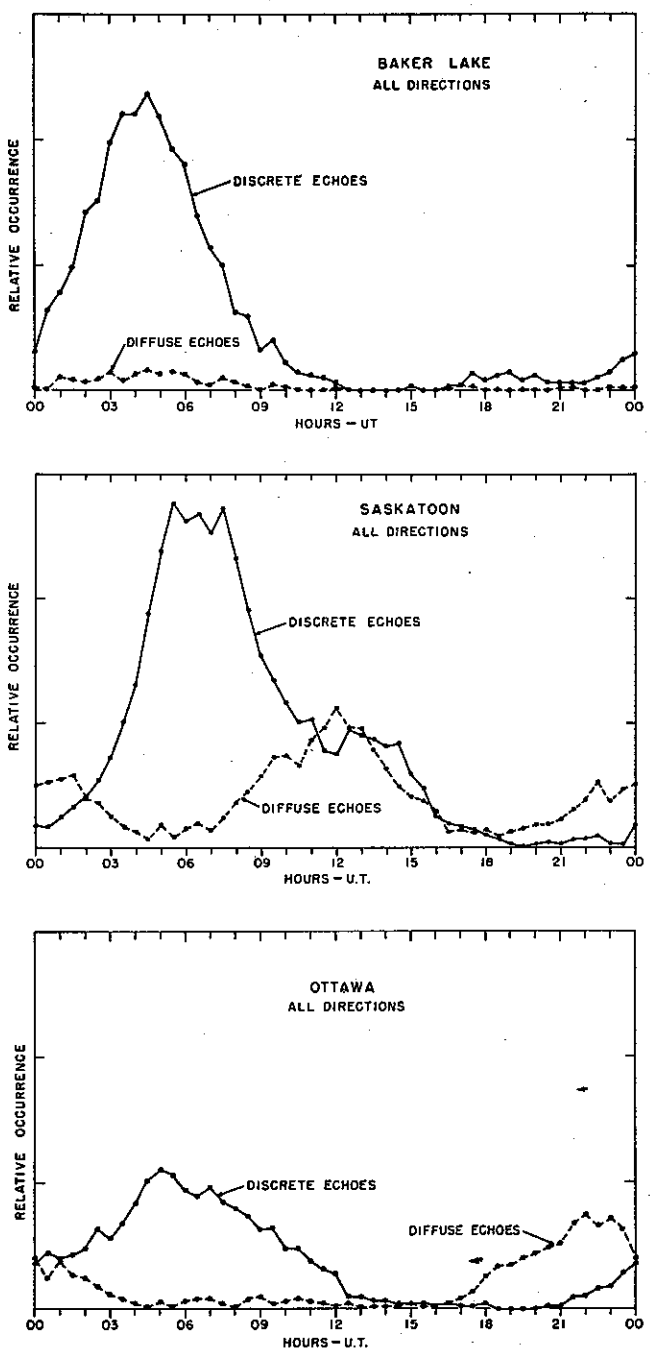


Figure 7. Diurnal variations of the diffuse and discrete types of radio aurora.

Figure 7 shows the diurnal distributions of the diffuse and discrete types of echoes. In general, the discrete type tends to be characteristic of the midnight period, while the diffuse type is an afternoon and morning phenomenon.

The diffuse echo is mainly a type seen at the lower latitude stations. This relative occurrence is quantized by taking the ratio of the number of minutes of discrete echo to the number of minutes of diffuse echo. The ratios are 1.7, 2.1, 21, and 21 for OT, SA, BL, and RB, respectively.

In the paper of McDiarmid and McNamara (1967), the data were grouped in various discrete and diffuse classification combinations to see if any difference in aspect sensitivity could be found. If a difference were found, then one might infer a difference in the scattering mechanism. However, the analysis revealed no significant difference in the aspect functions for the two types.

7. Comparison with optical aurora

Comparable statistical properties of the optical aurora in the same region of Canada where the radars were located were derived by reading 35 mm all-sky camera film from nine stations in Canada. The stations are listed in Table IV. In this comparison of the optical and radio aurora, interval-by-interval correlations have not been made. Only the total occurrence statistics are compared.

Analyses were performed on data derived by noting the presence or absence of any aurora on the pictures, taken at one-minute intervals, during each half-hour interval. The latitudinal location of aurora was also noted by recording the appropriate combination of three zones in which the aurora was observed: the zenith zone, defined as a band normal to the geographic meridian, with borders 1/2° of latitude north and south of the station; the north zone, the entire area north of the zenith zone; the south zone, the entire area to the south of the zenith zone.

Observing and operating conditions were noted for each half-hour, and assigned to one of four categories:

I. Unlimited visibility	=	stars visible, no cloud above 15° elevation.
II. Partially cloudy or imperfect record	=	some cloud above 15°, partially fogged mirror, moonlight, uncertain time determination.
III. Film run but no record	=	completely clouded or fogged, covered with snow.
IV. No film	=	daylight, complete mechanical failure.

For the purpose of this analysis, records bearing the quality indices I and II were defined as 'usable time', and hence used as valid samples. Table IV lists the total number of usable half-hour periods obtained at each station in the 12 months of data utilized. The station coordinates are for a dipole field.

Table IV. *The Canadian 35 mm all-sky camera stations*

Name	Symbol	Geographic		Geomagnetic		No. of usable Intervals
		Lat.	Long.	Lat.	Long.	
Alert	AL	82.6	62.0	85.5	195.0	1305
Resolute Bay	RB	74.7	94.9	83.0	73.1	2541
Baker Lake	BL	64.3	96.0	73.7	44.3	3122
Churchill	CH	58.8	94.1	68.7	37.2	2862
Yellowknife	YE	62.4	114.4	68.8	65.4	994
Meanook	ME	54.6	113.3	61.9	59.0	1992
Saskatoon	SA	52.1	106.6	60.5	49.5	3479
Ottawa	OT	45.4	75.9	56.8	8.7	3161
Victoria	VI	48.4	123.4	54.2	67.2	4888

Figure 8 shows the distribution of the auroral occurrence defined in terms of 'percent usable time' where the basic unit of time is the half-hour. The dashed curve is based on all aurora visible from each station, i.e. the complete hemisphere seen by each camera, whereas the solid curve is restricted to the aurora visible within the 1°-wide band of latitude passing through the camera zenith. Restriction of the data to this zenith zone reduced the auroral occurrence only to 70% of that obtained utilizing the entire field of the camera. The severe limitations on visibility imposed by atmospheric extinction and geometrical compression is evident when it is noted that the geographical area covered by the zenith zone is only 6.4% of the total potentially available area down to the horizon.

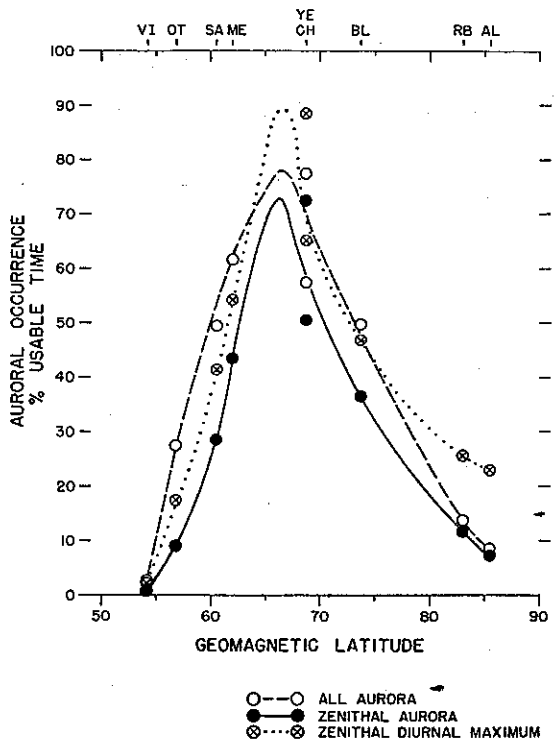


Figure 8. Occurrence of visual aurora determined from all-sky camera photographs. The camera stations are identified in Table IV. The magnetic coordinates are based on a simple dipole field.

The period of data chosen for this analysis was one year, 1 November 1957 to 31 October 1958, to help minimize possible effects of night-to-day ratios and seasonal differences. In addition, separate analysis of the data divided into the four seasons of the year was carried out. This analysis revealed some shifts in the distributions across the zone but these effects created relatively minor variations in the over-all shape of the zone. Hence, only the yearly average is presented in the figure.

To search for possible effects of different diurnal distributions in the zenithal auroral occurrence among the stations, diurnal curves at half-hourly intervals were plotted for each season and each station, using the zenithal zone data only. The diurnal peak was located from these plots and the occurrence rate at the peak noted and summed for the year. The resulting curve of maximal occurrence of zenithal aurora is shown by dotted curve of Fig. 8.

The average radio auroral occurrences on a half-hour interval basis as noted previously were 9.8, 21.6, 9.3 and 1.8% for OT, SA, BL, and RB, respectively. The observational statistics of the optical occurrences are weighted differently because 'usable time' is a factor. A more realistic comparison with the optical occurrence can be obtained by taking the radio auroral occurrence at the diurnal maximum for comparison with the similar optical rates. When this is done, the maximum radio auroral occurrences observed at the four stations are 18, 49, 19, and 3.2% in the radar beam positions OT(N), SA(N), BL(S), and RB(S). From the zenithal aurora diurnal maximum curve of Fig. 8, the optical occurrences at the latitudes corresponding to the maximum echo regions of the four selected radar beams are found to be 63, 85, 68 and 35%. The radar auroral zone is more sharply peaked in latitude than the optical, and appears to maximize at the same latitude. At the radar sensitivity employed here, the radio aurora occurrence at the diurnal maximum is somewhat less frequent than the visual aurora, by a factor of about 2 to 3.

8. Discussion

The VHF radar network operated during the IGY at high latitudes clearly demonstrates that under the intense excitation in the auroral disturbance, the conditions for strong aspect sensitivity are not valid. Based upon several independent analyses of these data, the aspect attenuation, at least at angles greater than 5° from perpendicularity, has a slope of about 1 db per degree. Moreover, the relative occurrence seen from Ottawa, where the aspect angle is within 1° of perpendicularity, compared with the radio and optical aurora seen at the other stations in the network suggests that this aspect attenuation rate must be essentially the same down to 1°. If any marked enhancement occurs at exact perpendicularity, then it must be relatively small and extremely narrow.

Numerous direct measurements of densities in aurora (McNamara, 1969) have shown that critically-dense electron concentrations ($3 \cdot 10^7$ cm^{-3} at 48.5 MHz) probably never occur even in the most intense auroral forms, and certainly not outside the forms. It is generally accepted now that the radar aurora at these and higher frequencies is due to under-dense back scattering from very small perturbations in the plasma, with dimensions of $\lambda/2$ metres (where λ is the radio wavelength).

The source of the micro-scale irregularities (3 metres to produce back scatter of 48.5 MHz radar waves) (McNamara, 1970), is undoubtedly due to some form of plasma instability. A particular theoretical model which is the most developed at the present time is the two-stream instability which generates ion-acoustic waves (Buneman, 1963; Farley, 1963). While some characteristics of the radio aurora do coincide with the predictions of this linear theory, others do not. For example, echoes are obtained at angles far off-perpendicular and Doppler spectra are frequently broad and shifted by amounts that do not always coincide with the ion-acoustic model (Hofstee and Forsyth, 1969).

Since plasma instabilities generally require some particular combination of magnetic field, electron density, electron temperature, electron gradient, conductivity, and electric field and current flow in both magnitude and direction, it is not surprising that detailed correlation is not usually found between the occurrence or the position of particular visual auroral forms and the occurrence of radar echoes.

Since the data presented in this paper are derived from many observations over a period of two years, the observational fact that radio aurora can be seen at 48.5 MHz without difficulty over a considerable range of aspect angles must be accepted. Any realistic model of the radio aurora process must recognize this fact.

Acknowledgements

The author gratefully acknowledges the work of L. A. Petch, B. E. Bourne, J. B. L. Heney, and W. M. Roy, in the construction of the equipment and the operation of the Ottawa radar. The dedicated work of L. R. McNarry and R. Wlochowicz at Resolute, E. E. Budzinski and J. Bleackley at Baker Lake, and D. Glass at Saskatoon is sincerely appreciated.

REFERENCES

Aspinall, A. and G. S. Hawkins, 1950: J. Brit. Astr. Assoc., **60**, 130.
Blevis, B. C., J. W. B. Day and O. S. Roscoe, 1963: Can. J. Phys., **41**, 1359.
Bullough, K., 1961: Ann. Géophys., **17**, 195.
Buneman, O., 1963: Phys. Rev. Letters, **10**, 285.
Currie, B. W., P. A. Forsyth and F. E. Vawter, 1953: J. Geophys. Res., **58**, 179.
Farley, D. T., Jr., 1963: J. Geophys. Res., **68**, 6083.
Harang, L. and W. Stoffregen, 1938: Nature, **142**, 832.
Hofstee, J. and P. A. Forsyth, 1969: Can. J. Phys., **47**, 2797.
Hultqvist, B. and A. Egeland, 1964: Space Sci. Rev., **3**, 27.
Leadabrand, R. L., 1961: J. Geophys, Res., **66**, 421.
Leadabrand, R. L., J. C. Schlobohm and M. J. Baron, 1965: J. Geophys. Res., **70**, 4235.
Lovell, A. C. B., J. A. Clegg and E. D. Ellyett, 1947: Nature, **160**, 372.
McDiarmid, D. R. and A. G. McNamara, 1967: Can. J. Phys., **45**, 3009.
McNamara, A. G., 1958: Can. J. Phys., **36**, 1.
McNamara, A. G., 1960: Can. J. Phys., **38**, 425.
McNamara, A. G., 1969: Can. J. Phys., **47**, 1913.
McNamara, A. G., 1970: In "The Radiating Atmosphere" (Ed. B. M. McCormac), D. Reidel Publ. Co., Dordrecht-Holland, 301.
Unwin, R. S., 1959: Ann. Géophys., **15**, 377.
Unwin, R. S., 1967: J. Atmosph. Terr. Phys., **29**, 1581.
Unwin, R. S., 1968: Ann. Géophys., **24**, 201.

Geofysiske Publikasjoner, Vol. 29, 1972

ELECTRON-DENSITY INCREASE IN THE E LAYER BELOW AN ARTIFICIAL BARIUM CLOUD

By W. Stoffregen
UPPSALA IONOSPHERIC OBSERVATORY,
RESEARCH INSTITUTE
OF NATIONAL DEFENCE, SWEDEN

Abstract

During the first two minutes after the release of an artificial barium cloud at a height of 187 km, a local enhancement of the electron density in the *E* layer geomagnetically below the cloud was observed. This observation was made by using a vertical ionospheric sounder tuned to 2.5 MHz. A similar effect had previously been studied indirectly by means of spectral observations at 5577 Å and 4278 Å. An increased emission was observed when the instrument was pointed towards the area below the cloud in the *E* layer.

1. Introduction

Barium clouds have been successfully used mainly for the study of electric fields in the ionosphere at heights above 150 km (see e.g. Haerendel et al., 1967). When an artificial barium cloud is released, an interaction between the cloud and the ambient atmosphere will take place. The geomagnetic field will guide electrons from one part of the cloud down to the *E* layer, where the high conductivity of this layer will facilitate the return transport of electrons up to the opposite part of the cloud. This has been proposed by Haerendel et al. (1967), who concluded that a local increase of the electron and ion density must thus be expected.

A detailed theoretical study of the electron density change in the *E* layer associated with the artificial barium cloud has been made by Papamastorakis (1968). From these two papers, it is concluded that within the area of the *E* layer, which is connected with the cloud by the geomagnetic field-lines, a zone of increased electron density, corresponding to the rear side of the cloud, will be created. Another zone of decreased electron density will appear, corresponding to the end of the cloud which points in the direction of its motion.

Recently, during a rocket campaign at Esrange (lat. 68.1° N, long. 21.0° E), some evidence was found that the barium cloud, which constitutes irregularities in the *F* layer, may trigger an increased particle precipitation. During the flight of the ESRO rocket S66 on 14 February 1970, it was observed that, in connection with one of the releases, the electron flux increased sharply by a factor of 2 and remained above the steady background for 10–15 seconds. The

pitch-angle distribution of the electrons also changed to a direction more perpendicular to the geomagnetic field, according to Köhn et al. (1971).

Attempts have been made on several occasions to obtain experimental verification of the electron- and ion-density increases in the area of the *E* region below the cloud (the *E*-down area). As it was considered possible that an increased flux of electrons would give rise to a slightly higher emission on OI (5577 Å) and N_2^+ (4278 Å), a spectrophotometer was directed towards the *E*-down area. It soon became clear that a measurable effect could only be obtained when the background emission was rather small and stable. Hence an observable in-

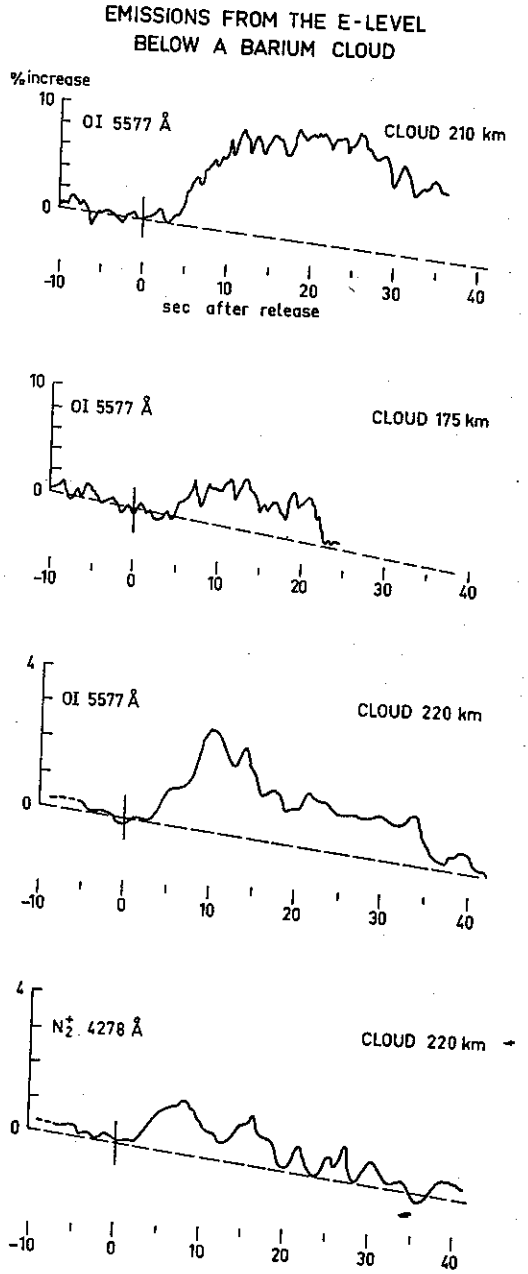


Figure 1. Enhancement of emissions OI (5577 Å) and N_2^+ (4278 Å) from an area in the *E*-layer geomagnetically below an artificial barium cloud.

dication of an enhanced emission was only expected when the release took place during undisturbed conditions. On two occasions (23 March, 1968 and 15 March, 1969), an increase of the emissions of at least 2–10% was observed. One typical feature of the optical records was that the increase started about 5 s after the release. The probable reason for this delay is that the ionization of the cloud is produced with a corresponding time constant. In Fig. 1 four examples of an observed increase of emissions from the *E*-down area are shown, according to Stoffregen (1970).

2. Direct observation of the electron-density increase

Any direct measurement of an *E*-down area effect is of importance for the understanding of the behaviour of an ion cloud. Such measurements can be made by means of an ionospheric

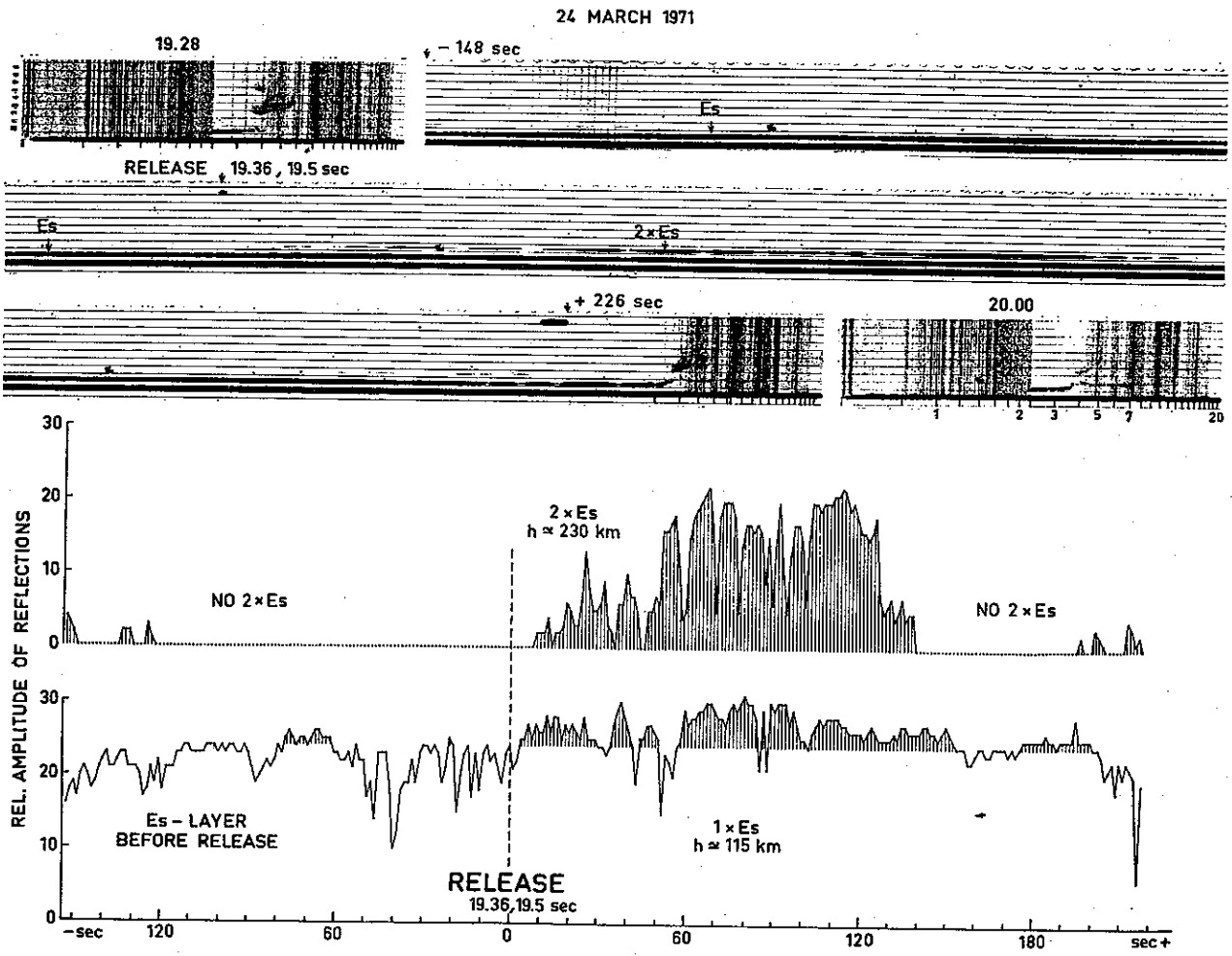


Figure 2. Upper part: Record of an electron-density increase in the area of the *Es* layer situated geomagnetically below a barium cloud. Multiple reflections (2 × *Es*) appear a few seconds after release. Lower part: Relative amplitudes scaled from the record above. The ionospheric sounder operated at a fixed frequency of 2.5 MHz.

sounder situated as close as possible below the *E*-down area. The sounding station at Esrange used for the observations reported here is located ≈30 km to the north of the launching site, approximately directly below the apogee of the estimated rocket trajectories. During the actual ESRO experiment (S87, March 1971), the trajectory of the Skylark rocket was such that the *E*-down area was at an elevation of ≈75° with respect to the ionospheric sounder. This is well within the radiation diagram of the sounder's antenna. As the normal operation of the sounder only permits complete sweep-frequency records in one minute, which is too slow for this purpose, the sounder was adjusted to a fixed frequency. During the first barium release on 15 March, 1971, a frequency of 0.5 MHz was used. Because of high absorption, no ionospheric reflections were recorded at all. A frequency of 2.5 MHz, about 1 MHz above the critical frequency of the *E* layer at this time of day, was chosen during the next experiment on 24 March, 1971. The record of this event is shown in Fig. 2. Before and after the fix frequency sequence, the recorder scanned over the frequency range of 0.25–20 MHz. A rather strong sporadic *E* layer with a critical frequency varying between 4 and 5 MHz was present during the actual period. The normal *E* layer could not be identified but is known from records on the days before and after to have had a critical frequency, *foE* between 1 and 1.5 MHz at 1930 h.

The record is not perfect, in that the recording oscillograph tube was not correctly adjusted, so that because of overloading by the signal the image of a strong echo was broadened. For this reason, the height of the reflections had to be measured when they were weak. This technical imperfection, however, was of benefit, as it enabled us to scale the broadening as an arbitrary measure of the amplitude of the reflections. This is plotted in the lower part of Fig. 2.

3. Discussion

In discussing these observations, the primary question is whether the measured enhancement is a real *E*-down effect or not. The height of the cloud was measured as 187 km. A direct reflection from the cloud cannot be expected because of screening due to the presence of a strong *Es* layer. The second reflections have to be regarded as a multiple reflection ($2 \times Es$) from the sporadic *E* layer. This is confirmed by the fact that the $1 \times Es$ reflection is enhanced too, as shown in the diagram. Finally, there remains one possible objection, and this is that the observed enhancement is purely accidental and due to ionospheric activity. In that case, however, one would expect a corresponding change of the geomagnetic records as well as of the auroral activity. No evidence has been found for this.

On the other hand, it is striking that the start of the observed increase of electron density was delayed by almost exactly the same time as was observed earlier by other techniques (Fig. 1), i.e. by approximately 5 seconds. The duration of the increase is of the order of 2 minutes. Compared with three of the optical observations (Fig. 1), the duration is somewhat longer. Because of instrumental noise, the optical records are extremely difficult to deduce, especially when the increase is only a few per cent. However, the upper curve in Fig. 1, which unfortunately is interrupted because of the observation of a second cloud, shows an increase

of the emission to about 10%. If the observation had continued, this emission would probably have decreased to the background level at a time comparable with the observation in Fig. 2. Both before and after the record of the rather strong multiple reflection $2 \times Es$, only the $1 \times Es$ was present. It seems rather unlikely that an accidental increase would appear and disappear in this manner. To obtain further confirmation of this observation, the increase discussed here should be further investigated.

4. Concluding remarks

It is not yet understood in detail how an electron transport from the barium cloud towards the E layer is started. At an early stage it was assumed that *electrons from the cloud* were guided along the geomagnetic field lines to the E layer below because of high conductivity along the line of force. The probability that *energetic electrons* are involved in this process is still being discussed. Evidence has recently been found for a sudden increase of the energetic particle flux immediately after release. In any case, when electrons from a higher altitude penetrate into the E layer, this will cause a local density gradient. If this gradient, $\Delta N/N$, is very pronounced to begin with, it will give rise to radio reflections and scattering from this area. Diffusion and recombination will smooth out such gradients fairly rapidly and the radio reflections will disappear. In the case reported here the E-down area was slightly off the position of the ionospheric recorder (elevation $\approx 75°$). Hence, if after some time the gradient tends to be horizontally orientated, reflections from this area may disappear quite abruptly, as the observations have shown.

Acknowledgements

The author's thanks are due to Prof. R. Lüst and Dr. G. Haerendel for giving him the opportunity to take part in the observations of barium clouds. The making of the special records with the ionospheric sounder by the personnel responsible for the down-range station at Esrange is gratefully acknowledged. This work has partly been sponsored by the Swedish Natural Science Council.

References

Haerendel, G., R. Lüst and E. Rieger, 1967: Planet. Space Sci., **15**, 1.

Köhn, D., D. E. Page and M. L. Shaw, 1971: Paper presented at the Spring Meeting. German Physical Society, Heidelberg.

Papamastorakis, J., 1968: 'Jonosphärische Inhomogenitäten', Ph. D. Thesis, Technische Hochschule, München.

Stoffregen, W., 1970: J. Atmosph. Terr. Phys., **32**, 171.

Geofysiske Publikasjoner, Vol. 29, 1972

THE GENERATION AND PROPAGATION OF VLF EMISSIONS

By R. E. Barrington
COMMUNICATIONS RESEARCH CENTRE,
SHIRLEY BAY, OTTAWA,
CANADA

Abstract

A survey of the established features of VLF emissions is given. The characteristics of the propagation of emissions in the magnetosphere and ionosphere are reviewed with emphasis on those topics that are important in determining the regions in which emissions are generated. Both single particle and collective processes that have been studied as possible sources of VLF emissions are discussed. A number of recent experiments involving rockets and satellite-borne instruments have provided new clues to the generation process and these are briefly summarized.

1. Introduction

A wide variety of naturally occurring emissions (Helliwell, 1965) is found in and around the VLF band of the radio spectrum. Ground based observations of these emissions have shown that they be relatively steady for several minutes or even hours, or they may occur in discrete bursts lasting for as little as a fraction of a second. Generally the emissions are confined to welldefined and often quite narrow frequency bands, with bandwidths ranging from a few tens of hertz to twenty or thirty kilohertz. Several frequency bands may sometimes be observed simultaneously, but they are not normally harmonically related. Regular diurnal variations in the occurrence of VLF emissions have been found as well as systematic variations with geomagnetic latitude. They are observed most commonly at middle and high latitudes, are generally localized in geographic extent, and in some cases are closely associated with whistlers. Emissions observed at geomagnetically conjugate locations are often closely related both in form and in time of occurrence. While many unanswered questions about their generation remain, the available evidence indicates that the emissions seen on the ground propagate to the earth in the whistler mode from a region of generation in the ionosphere or magnetosphere.

The advent of satellite observations of VLF emissions (Barrington and Belrose, 1963) has led to a great inpouring of new information, but as yet this increase in experimental knowledge has not been matched by a corresponding increase in theoretical understanding. In part this is because the satellite-borne sensors, in addition to the types of VLF emissions observed

by ground based instruments, have also found a number of new types. Moreover, for satellite observations many of the assumptions that are valid on the ground (Cornwall, 1970) no longer apply. In a plasma such as the ionosphere and magnetosphere, both electromagnetic and electrostatic VLF waves can propagate, but only electromagnetic waves can escape and reach the Earth's surface. For whistler mode waves to escape from the ionosphere they must normally be incident on its lower regions at a small angle to the local zenith, whereas a satellite sensor can usually detect waves propagating in an arbitrary direction. In view of such factors it is not surprising that during the first decade of space observations much of the effort has been spent in relating satellite and ground observations of emissions.

In this paper recent observations of VLF emissions, particularly by satellites, are reviewed. These results are discussed in the light of current theories of the generation of VLF emissions and recent advances in the field of plasma physics. In order to provide a meaningful framework in which to discuss these current theories and observations certain features of magnetospheric propagation must be considered. These are particularly important in attempts to deduce the regions in which VLF emissions are generated. To simplify the situation only frequencies above the maximum proton gyrofrequency will be considered. This restriction excludes a number of intense and interesting emissions that occur at frequencies of a few tens or hundreds of hertz, but these are normally restricted to the ELF band. There is also increasing evidence that many such emissions are not closely related in their generation to those found at VLF frequencies (Barrington, 1971).

2. Propagation of VLF emissions

While VLF emissions are generally found to propagate in the right-hand circularly polarized whistler mode, two significantly different types of propagation are encountered. As was originally pointed out by Storey (1953), the refractive index surfaces for the whistler mode are highly anisotropic, leading to the guiding of whistler waves in the direction of the Earth's magnetic field. In the terrestrial ionosphere, however, with its relatively small gradient in the refractive index transverse to the field direction, the wave normal direction does not change as rapidly as the direction of the field. As a result the ray may deviate progressively from the direction of the Earth's fields. For frequencies below hybrid resonance frequency, propagation transverse to the field is permissible, and hence an increase in the angle between the ray and the field direction may lead to reflection. The magnetospherically reflected (MR) whistlers observed by satellites (Smith and Angerami, 1968) demonstrate that such reflection does in fact occur. Frequently, whistler mode wave packets have been observed to bounce back and forth several times between reflection points. Such propagation has been termed non-ducted and has the following characteristics: the signals can be observed by satellites but not on the ground, the propagation path depends on the waves launch angle and frequency as well as on the properties of the medium, and the wave normal angles with respect to the field direction are usually large.

A second, and in some ways more complicated type of propagation, has been termed the ducted mode. Here the waves are trapped or confined by field-aligned columns or ducts of

enhanced ionization. As in a waveguide they are constrained to travel primarily along the axis of the duct. Thus the path of propagation is determined by the duct and is independent of frequency and wave normal direction. Usually, ducted propagation is assumed to be longitudinal since the dispersion of whistlers travelling in such ducts is fairly accurately described on the basis of this assumption, and the physics of duct formation suggests that they should be field aligned. Ducting is important in the analysis of VLF emissions since most of the emissions seen on the ground have propagated in ducts. In addition, whistlers travelling in ducts often appear to trigger emissions. Only occasionally are ducted emissions seen by satellites, presumably when the satellite happens to be in a duct at the same time as an emission. Whistler evidence suggests that ducted propagation occurs primarily within the plasmasphere, i.e. the innermost region of the magnetosphere that is characterized by relatively high electron density and is in diffusive equilibrium with the ionosphere (Carpenter, 1966). Satellite observations of ducted whistlers over a broad range of latitudes suggests that there is often considerable leakage of whistler mode energy from the ducts.

The question of ducting or non-ducting of VLF emissions is important in deducing the region in which emissions originate. Of similar importance is the direction of propagation of an emission and in particular whether it is upgoing or downgoing. The ability of the VLF experiment onboard the Injun V satellite (Mosier and Gurnett, 1969) to determine if the Poynting flux is directed up or down the geomagnetic field represents an important development in this area. For discrete VLF emissions, where only waves propagating in a specific direction are encountered, the sign of the Poynting flux indicates whether the signals are propagating up or down the geomagnetic field. When many waves with different wave normal angles are present the sign of the average Poynting flux of all the waves can only be determined in special circumstances, i.e. near any gyrofrequency or for waves whose normals are uniformly distributed in the plane perpendicular to the direction of the magnetic field (Mosier and Gurnett, 1971).

Gurnett et al. (1971) have employed the Injun V Poynting flux observations to study a number of types of VLF emissions. Their results indicate that generally chorus is downgoing although occasionally at invariant latitudes of less than 60° upgoing chorus is found. In such cases the transition from upgoing to downgoing propagation usually occurs near the boundary between the plasmapause and the light ion trough with the upgoing waves observed inside the plasmapause. Of the several types of VLF hiss observed by satellites, all are found to consist of a mixture of upgoing and downgoing waves. Since much of the hiss is observed to be downgoing and is at frequencies above the LHR frequency it must be generated in the opposite hemisphere or at higher altitudes than the satellite. At present it has not been determined whether the upgoing hiss is generated below the satellite or is due to the reflection of downgoing waves.

Another question related to the propagation of VLF emissions is whether any of the emissions observed by satellites are due to electrostatic waves. In the vicinity of the lower hybrid resonance both theory and experiment indicate that waves propagating transversely to the field are essentially electrostatic (Gurnett et al., 1969). These authors also report the observation by Injun V of bands of noise that have an electric field component but no magnetic field.

Such noise bands are not related to the local LHR frequency and could be due to electrostatic waves. Injun V also observes effects at harmonics of the proton gyrofrequency. These are attenuation bands and although most clearly defined in the electric field spectrum, they are also observed in the magnetic field spectrum. As a result the waves involved in the gyrofrequency harmonic interaction are probably electromagnetic. OGO 5 (Kennel et al., 1970) has observed magnetospheric electric field emission, as well as the sporadic narrowbanded electric field bursts observed in interplanetary space. These latter are probably electron plasma oscillations. Scarf et al. (1969) have also suggested that some of the emissions seen by OGO 5 in interplanetary space may be due to ion acoustic waves. Undoubtedly, electrostatic as well as electromagnetic waves are generated in the warm magnetized plasmas of space. In many instances, however, the separation of these waves into pure electromagnetic or pure electrostatic waves is an over-simplification of the types of waves that are encountered.

3. Theories of the generation of VLF emissions

There is fairly general agreement that VLF emissions result from wave-particle interactions occurring within the terrestrial ionosphere or magnetosphere. Such interactions are of prime importance in the developing field of plasma physics and considerable progress has been made in the past decade in their study. In the magnetosphere, plasma conditions that cannot be duplicated in laboratories exist, and hence the wave-particle interactions that generate VLF emissions are potential new tools for the study of plasma physics. As yet, however, there is little general agreement on the details of the interactions that lead to particular types of VLF emissions.

Early theories of VLF emissions concentrated on well-known processes by which charged particles radiate electromagnetic waves. In the ionosphere and the magnetosphere, energetic particles move along helical trajectories, and the radiation resulting from such motion has some unusual properties due to the dispersion and anisotropy of the plasma in these regions. When the local phase velocity is less than the particle velocity the polarization induced in the medium by the charged particle leads to Cerenkov radiation. The acceleration associated with the gyration of particles about the field direction produces cyclotron radiation which is appreciable in the low-order harmonics at Doppler-shifted frequencies. In both cases the energy is radiated along cones in the magnetic field direction due to the symmetry of the medium. Cerenkov radiation is always emitted in the foreward hemisphere with respect to the motion of the particle's guiding centre. For the whistler mode, cyclotron radiation from electrons is emitted entirely in the backward hemisphere.

Liemohn (1965) has carried out an extensive analysis of the radiation from electrons in a cold collisionless magneto-plasma. He has evaluated the complicated analytic formulas for the radiated power for several special cases. For the whistler mode most of the energy is found to radiate along wave normals at large angles to the magnetic field direction and to occur at frequencies other than the rectilinearly Doppler-shifted fundamental cyclotron harmonic. Using reasonable approximations to eliminate the resonance singularity of the whistler mode refractive index for a cold collisionless plasma, the total power radiated into this mode is found to be a

slowly varying function of frequency and electron energy and has an average value of 10^{-30} watts per hertz per electron. Liemohn concludes that for incoherent radiation this power level per particle is inadequate to explain the observed intensities of VLF emissions.

Jørgensen (1968), using satellite and ground based observations of VLF hiss intensity and space observations of the flux of electrons, has proposed that the incoherent Čerenkov process could explain the intensities of auroral hiss. The differences between Jørgensen's and Liemohn's calculations lie in their estimates of hiss intensity, the flux density of suitable electrons, and the L value of the tubes of force on which the mechanism is operative. In view of Jørgensen's work, it is not possible to rule out incoherent processes as a source of some types of VLF emissions. It is clear, however, that mid-latitude hiss and many forms of discrete emissions are of such intensity and so located that they cannot be explained by an incoherent process.

In studying emission processes involving coherence, it is usual to work with dispersion relations, assumed particle distribution functions and to look for the existence of instabilities and their associated growth rates. Such an approach is not only a powerful tool for studying VLF emissions but it can be extended to show the impact of VLF wave fields on the precipitation of particles. This has brought into focus the fact that not only can energetic particles lose energy wave fields, but also VLF waves can have a significant impact on particle precipitation. The question of which way the energy may flow depends critically on the detail of the particle distribution function and on the amplitude and wave normal angles of the waves. Here, however, consideration is limited to the production of VLF emissions and hence only cases where the particles lose energy to waves are considered.

While a number of instabilities may occur in the ionospheric and magnetospheric plasma, the transverse cyclotron instability has undoubtedly attracted the most attention as a candidate for explaining VLF emissions. This instability may occur with a proton or electron beam and requires that the Doppler-shifted wave frequency seen by the particle equals the particle gyrofrequency and has the same sense of rotation. Since whistler mode waves are right circularly polarized, their fields rotate about the Earth's field in the same sense as an electron and in the opposite sense to that of protons. Thus, an anomalous Doppler-shift (i.e. the frequency must be shifted through zero so that a reversal of polarization occurs) is required to obtain a transverse instability involving whistler-mode waves and protons, whereas for an electron beam the Doppler shift is normal. As Brice (1964) has pointed out, when an electron beam produces VLF waves as a result of this instability, the longitudinal energy of the beam increases and the transverse energy decreases, leading to a reduction in the average pitch angle of the beam particles. Thus an electron beam in which the particles have no transverse energy cannot generate whistler mode waves by this process. For a proton beam, if this instability generates whistler mode waves, the longitudinal energy of the beam decreases, the transverse energy increases, and the average pitch angle increases. Thus a proton beam that contains no transverse energy can still generate whistler mode waves by this mechanism. Brice has proposed that this mechanism operating in the equatorial plane can account for the intense VLF emissions that are observed, since such an instability can involve feedback between the waves and the electrons and hence can be self-sustaining over a relatively long time. Bell and Buneman (1964) have studied this possibility in some detail.

For instabilities involving a longitudinal resonance between whistler mode waves and particles the longitudinal velocity of the particle must match the wave phase velocity and the wave electric field must have a longitudinal component. The observed flux densities of the particles that can satisfy this resonance condition are such that only electrons are likely to produce VLF emissions by this mechanism. The whistler mode waves are generated at the expense of the longitudinal energy of the electron beam with no effect on the transverse energy. Thus, particle precipitation as a result of this instability occurs only when whistler mode waves give up some of their energy to electrons.

Many proposals (Kimura, 1967) involving one or another of the mechanisms that have been described have been made to explain the discrete types of VLF emissions seen both in satellites and on the ground. None of them, however, adequately explain the narrow bandwidth, the rapidly varying frequency, and the long duration of these emissions. Since discrete emissions and those triggered by whistlers or VLF transmitters are very similar, it seems possible that they both may result from very similar processes. If this is correct then two features of artificially stimulated emissions provide valuable clues to the type of mechanism required to explain them. One is that there is a delay of about 100 ms between the onset time of the triggering signal and the time the triggered emission begins. This time tends to be longer than the dot and shorter than the dash of a coded VLF transmission. Secondly, since triggering is often observed when no spontaneous emissions are present, a small signal instability may not be required to produce the observed emissions.

To explain these features of discrete emission Helliwell (1967) has developed a phenomenological theory that is predicated on the cyclotron resonance between whistler mode waves and energetic electrons. In this theory a feed-back between waves, and electrons connected with the oscillation of the electrons in the magnetic 'potential well' of the wave is invoked. This feed-back occurs over a period of about 10^{-1} s and a path length that is long compared to the wavelength. In essence the feed-back or interaction region is determined by the path length over which the phase angle between the transverse velocity of the electron and the magnetic field of the wave lies within $\pm \pi$ radians. Typically the interaction region is 1000 km, and the variation of gyrofrequency and the spread in parallel velocities of the resonant electrons of this region would lead to emission bandwidths of about 100 Hz.

The narrow band emission comes from the current associated with the resonant electrons that are bunched in phase with the output wave. As electrons move through the interaction region they are acted on by a longitudinal force due to the magnetic field of the wave. This force causes the electrons to execute an oscillatory motion about their position in the absence of the wave and leads to phase bunching, with a bunching time dependent on the amplitude of the wave magnetic field. An increase in this field beyond a critical value leads to a decrease in the physical length of the bunching region without a corresponding increase in the associated transverse current that produces radiation. Thus the amplitude of the emission is limited by a process that is more or less independent of the strength of the triggering wave and hence could explain the almost constant intensity of discrete or triggered emissions. The bunching time is typically about 30 ms and this coupled with the time for waves to traverse the interaction region leads to a time delay between the emission and the triggering pulse of about

60 ms which is not too different from the observed 100 ms delay.

A rapid change of frequency with time, as is often characteristic of VLF emissions, results from the fact that the radiation from each electron changes in frequency as the electron moves through the interaction region due to the changing parameters of the medium. If it is assumed that energy is conserved in the interaction region then this region must drift. For drift across the equator in the streaming direction a hook in the frequency time spectrum results. Drift in the opposite direction inverts the hook, while successive crossings of the equator yield an oscillating tone. Thus a wide variety of spectral forms can be produced by this mechanism. The drift velocity of the interaction region does not affect the rate of change of frequency with time for a specific emission, since this is determined solely by the distance of the interaction region from the equatorial plane.

Das (1968) has developed independently a theory of VLF emissions that invokes the same basic process as Helliwell's. There are, however, significant differences in their conclusion. In the Helliwell model a cyclotron resonance instability is not necessary since all the energy for the emission comes from the triggering wave. In Das's work non-linear effects such as trapping are shown to lead to enhancement of the growth rate of this instability for particular narrow frequency intervals, and hence the energy of the emission can be obtained from the electrons. Both workers conclude that the required triggering power is proportional to the fourth power of the trigger pulse length, but neither approach at present explains the preference for triggering at one half the equatorial gyrofrequency.

Instabilities involving electrostatic waves have received far less attention in the contex of VLF emission theory since little observational evidence exists to show that such waves occur in the ionosphere or the magnetosphere. Some calculations have been made and these show that the plasma density must exceed some critical value for instability, but this criterion is usually met in the magnetosphere. There are also indications that the instabilities tend to occur at even or odd multiples of half the electron gyrofrequency. As indicated in the next section there is growing evidence to support the existence of electrostatic waves in the magnetosphere and this will no doubt attract attention to the theory of the instabilities in which they are important.

4. Satellite observations of VLF emissions

Satellite observations of VLF emissions have yielded considerable information on the spatial and temporal distribution of VLF emissions. There are a number of problems encountered in determining such distribution. When narrow-band receivers are used, it is easy to determine the intensity of the noise seen at a given frequency, but it is not usually possible to differentiate between different types of noise. Broadband receivers permit the identification of different noise types, but in many cases it is difficult to determine the intensity of a particular type accurately. Observations by both the Alouette (Barrington et al., 1971) and Injun (Gurnett, 1966) satellites indicate that the most intense noise is found in the ELF part of the spectrum and is primarily a daytime phenomenon, with a maximum occurrence around local noon at

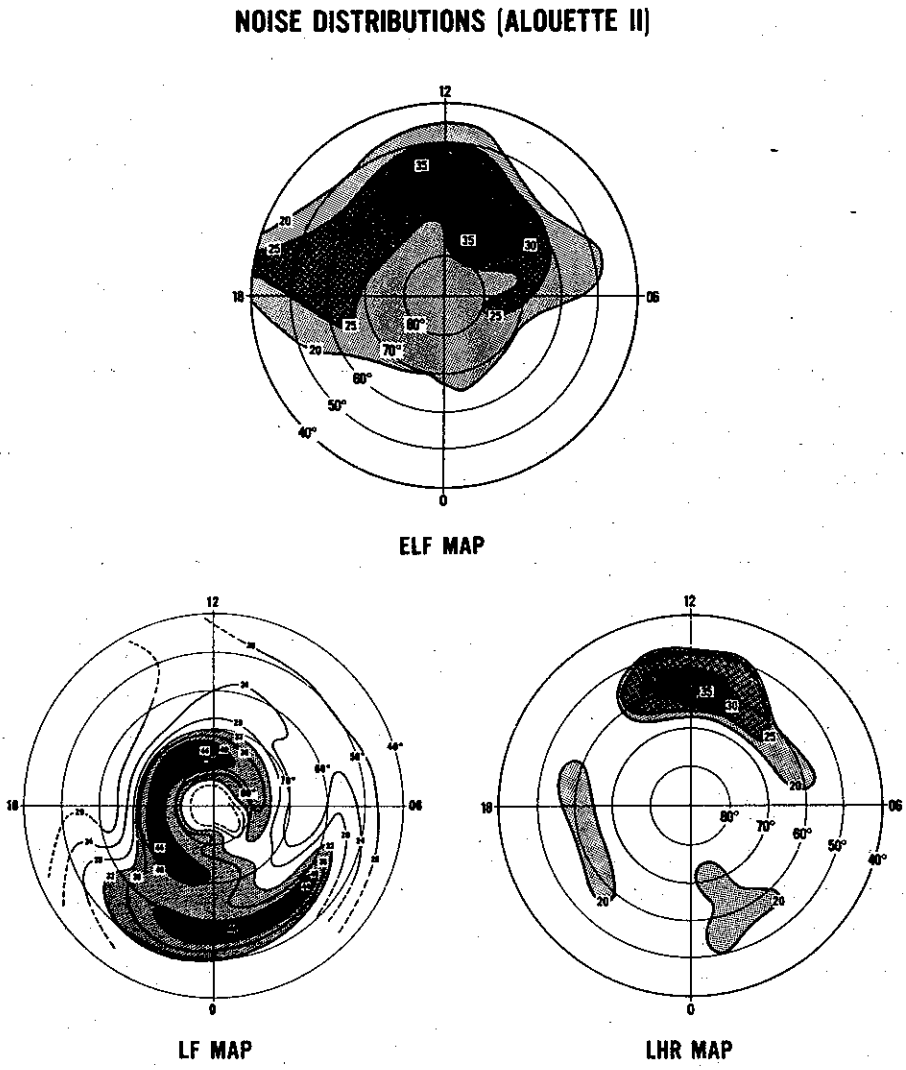


Figure 1. The variation in local time and latitude of three types of noise observed by the Alouette and ISIS satellites. The noise contributing to the ELF map always has a peak in its intensity at frequencies of a few hundred Hz. The LF map gives the mean intensity of noise observed at a fixed frequency of 200 kHz. The LHR map plots the average of the maximum intensities of noise bands that are cut-off at the LHR frequency.

an invariant latitude of about 60°. These features are illustrated by the contour plots shown in Fig. 1.

The commonest form of emission in the VLF part of the spectrum is auroral hiss. These emissions are extremely broadband (Laaspere et al., 1971) extending usually from approximately the local lower hybrid resonance frequency to frequencies of the order, but always less than, the local electron gyrofrequency. The latitude of maximum occurrence of auroral hiss varies throughout the day in much the same way as the auroral oval of Feldstein (cf. this volume), but lies slightly on the poleward side of this region. Maxima in the occurrence rate of this hiss occur near noon and midnight, but the intensity of the hiss is somewhat greater on the noon side of the Earth as shown in Fig. 1.

Dunckel and Helliwell (1969) have observed VLF emissions using OGO 1, a magnetospheric satellite. These observations differ from those made with low altitude satellites such as Alouette and Injun, in that they include downward propagating emissions that are reflected or absorbed at high altitudes. In the magnetosphere, emissions are found at all local times except for the midnight to dawn sector, at L values exceeding 5. The intensity of these emissions peaks at L values of 4 and 9 and at a local time of 1000. The upper frequency limit of most of the emissions is proportional to the minimum electron gyrofrequency along the magnetic field line on which the satellite is situated. No emissions were observed beyond the estimated position of the shock boundary.

Satellites, in addition to providing new insight into the occurrence and characteristics of the VLF emissions seen on the ground, have discovered new types as well. Burtis and Helliwell (1969) have reported that the OGO 1 and OGO 3 satellites observe a new type of VLF radiation in the magnetosphere that they call 'banded chorus'. The frequency, f, of this chorus depends on the equatorial electron gyrofrequency, f_{HO}, for the field line of observation, and ratios f/f_{HO} of 0.2–0.5 are typical. This ratio is also found to be latitude-dependent with the lower values of the ratio occurring at higher latitudes. This chorus has been observed at all local times but is most common during the morning hours outside the plasmapause. It is claimed that these emissions are produced near the equator at a fraction of the equatorial electron gyro-frequency by the electron cyclotron resonance generation mechanism.

The OGO 5 satellite in addition to detecting banded chorus has also yielded interesting data on what appear to be electrostatic waves (Kennel et al., 1970). These emissions are localized near the geomagnetic equator on auroral lines of force and occur at frequencies above the local electron cyclotron frequency f_{HO}. They are typically narrow band emissions and the most commonly occurring frequency is $3/2\, f_H$. The electric field intensities of these waves are in the 1 to 10 millivolt per meter range, implying that these waves could be effective sources of pitch angle diffusion and energization for electrons. OGO 5, unfortunately, lacks appropriate wave magnetic field diagnostics and hence the identification of these waves as largely

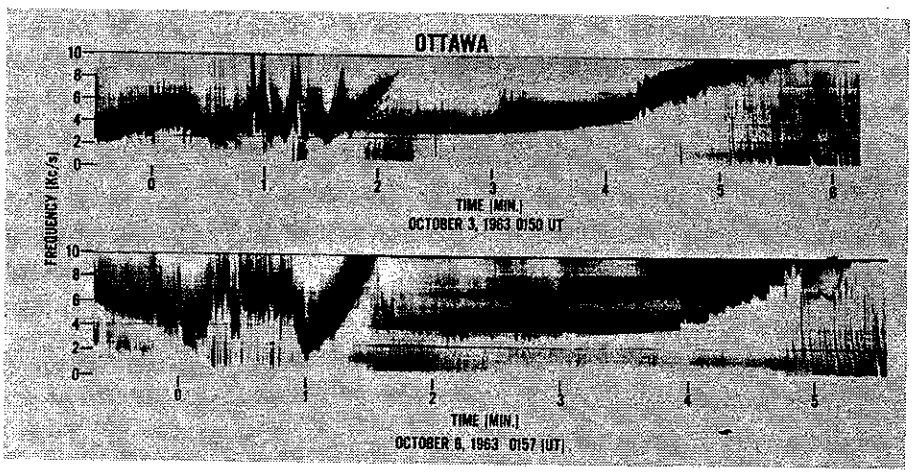


Figure 2. Saucer-shaped emissions observed by the Alouette II satellite. Two events of a few seconds in duration can be seen around the 1 minute mark of the upper trace. These occur in conjunction with other saucer-shaped events with durations of about 20 seconds.

electrostatic must be considered tentative. There are, however, several indirect arguments supporting this identification.

The Alouette satellites were the first to detect a type of VLF emission that has been termed by Gurnett et al. (1971) a saucer-shaped emission. Figure 2 shows an example of such an emission. An analysis of these events using the Poynting flux technique of Gurnett shows clearly that all frequency components of such events propagate up the geomagnetic field. At present this is the only type of emission whose origin is definitely known to be at low magnetospheric altitudes, i.e. below a few thousand kilometers. These emissions have only been observed by satellites, and, particularly for the short duration events, their hyperbolic form is often quite symetric. Gurnett et al. (1971) has reported that the same saucer event has been seen on as many as three successive orbits of Injun 5, indicating that the frequency time spectrum is primarily a spatial rather than a temporal effect. It has been proposed (Mosier and Gurnett, 1969), that the frequency time spectrum is due to a frequency dependent limiting ray angle for whistler mode propagation from the source to the satellite. At a given frequency the observed intensity variations are postulated to be due to the motion of the satellite through the array of possible ray paths from the source to the satellite. Smith (1969) has suggested that the source of these emissions is an auroral beam that radiates over a limited region (perhaps at the altitude of break-up of the beam). He has also found evidence in saucer events that extend to low frequencies that he interprets as due to the presence of electrostatic beam modes. In fact he suggests that the minimum frequency of a saucer event is an increasing function of the minimum distance from the satellite path to the source beam.

Since all forms of VLF emissions with the possible exception of those triggered by whistlers or Earth-based transmitters are thought to derive their energy from energetic particles, there have been many attempts recently to correlate emission observations with energetic particle data. Oliven and Gurnett (1968) have reported that the Injun 3 satellite has observed that microbursts of precipitating 40 keV electrons are always accompanied by a group of VLF chorus emissions, but that chorus is not necessarily accompanied by microbursts. In this study it was not generally possible to find a one to one correspondence between individual bursts of chorus and particular microbursts.

Using rockets, Gendrin et al. (1970) have observed VLF emissions and particle fluxes simultaneously. During chorus events an increase in the flux of newly injected electrons, both trapped and precipitated, was detected. When periodic hiss emissions were observed, a hydromagnetic wave of the same period and fluctuations in the intensities of trapped electrons were present. These observations are interpreted as evidence of pitch-angle scattering of electrons associated with VLF emissions.

Rosenberg et al. (1971) have reported on simultaneous X-ray observations and VLF observations made from balloons over Siple Station in the Antarctic. The X-rays arise from the precipitation of energetic electrons. During a period of enhanced activity at Siple Station a one to one correlation was found between short X-ray bursts and discrete VLF emissions with a centre frequency near 2.5 kHz.

A significant step in the study of wave particle interactions that generate VLF emissions has been taken by Cartwright and Kellogg (1971). These experimenters used a rocket-borne

electron gun to inject bursts of electrons of 35–45 keV energy into the ionosphere over a range of pitch angles from 65° to 115°. Receivers installed in a capsule that was ejected from the main part of the payload observed the radio spectrum from 16 Hz to 10 MHz. It was found that the electron beam generated waves at frequencies around the local plasma frequency and over the full whistler range from the low frequency limit of the receiver. The authors consider that this radiation is generated by a Cerenkov-like emission from the electrons of the beam. In addition to this radiation a 50 eV argon ion beam, used to keep the accelerator payload neutral, generated a hiss band centred close to the lower hybrid resonance frequency. The amplitude of this hiss was modulated at the spin rate of the accelerator payload and the minimum intensity occurred when the plasma jet was directed away from the nose cone containing the receivers.

Summary

Both theory and experiment over the past decade have revealed that VLF emissions are an integral part of the wave-particle interactions that occur in the ionosphere and the magnetosphere. As such they provide a valuable tool for detecting and studying such interactions, since they can be observed and recorded at locations remote from the interaction region. Thus, it is not unrealistic to anticipate that with time such emissions will become a convenient monitor of the behaviour of the magnetosphere and its interactions with the solar wind and with the interplanetary plasma.

References

Barrington, R. E., 1972: In 'Magnetosphere-Ionosphere Interactions' (Ed. K. Folkestad), Universitetsforlaget, Oslo.
Barrington, R. E. and J. S. Belrose, 1963: Nature, **198**, 651.
Barrington, R. E., T. R. Hartz and R. W. Harwey, 1971: J. Geophys. Res., **76**, 5278.
Bell, T. F. and O. Buneman, 1964: Phys. Rev., **133**, 1300.
Brice, N., 1964: J. Geophys. Res., **69**, 5415.
Burtis, W. J. and R. A. Helliwell, 1969: J. Geophys. Res., **74**, 3002.
Carpenter, D. L., 1966: J. Geophys. Res., **71**, 693.
Cartwright, D. G. and P. J. Kellogg, 1971: Nature, **231**, 11.
Cornwall, J. M., 1970: In 'Progress in Radio Science 1966–69' (Eds. G. M. Brown, N. D. Clarence and M. J. Rycroft), URSI, Brussels.
Das, A. C., 1968: J. Geophys. Res., **73**, 7457.
Dunkel, N. and R. A. Helliwell, 1969: J. Geophys. Res., **74**, 6371.
Gendrin, R. J., J. Etcheto and B. de la Porte des Vaux, 1970: J. Geophys. Res., **75**, 1970.
Gurnett, D. A., 1966: J. Geophys. Res., **71**, 5599.
Gurnett, D. A., G. W. Pfeiffer, R. R. Anderson, S. R. Mosier and D. P. Caufman, 1969: J. Geophys. Res., **74**, 4631.
Gurnett, D. A., S. R. Mosier and R. R. Anderson, 1971: J. Geophys. Res., **76**, 3022.
Helliwell, R. A., 1965: In 'Whistlers and related ionospheric Phenomena'. Stanford University Press, Stanford, California, 203.
Helliwell, R. A., 1967: J. Geophys. Res., **72**, 4773.
Jørgensen, T. S., 1968: J. Geophys. Res., **73**, 1055.
Kennel, C. F., F. L. Scarf, R. W. Fredericks, J. H. McGehee and F. V. Coroniti, 1970: J. Geophys. Res., **75**, 6136.
Kimura, I., 1967: Planet. Space Sci., **15**, 1427.
Laaspere, T., W. C. Johnson and L. C. Semprebon, 1971: J. Geophys. Res., **76**, 4477.
Liemohn, H. B., 1965: Radio Science, **69D**, 741.
Mosier, S. R. and D. A. Gurnett, 1969: J. Geophys. Res., **74**, 5675.
Mosier, S. R. and D. A. Gurnett, 1971: J. Geophys. Res., **76**.
Oliven, M. N. and D. A. Gurnett, 1968: J. Geophys. Res., **73**, 2355.
Rosenberg, T. J., R. A. Helliwell and J. P. Katsufrakis, 1971. Techn. Note BH-713. University of Maryland.
Scarf, F. L., G. M. Crook, R. W. Fredericks, I. M. Green and C. F. Kennel, 1969: In 'Plasma Waves in Space and Laboratory'. (Eds. J. O. Thomas and B. J. Landmark), Edinburgh University Press, Edinburgh, 379.
Smith, R. L., 1969: Nature, **224**, 351.
Smith, R. L. and J. J. Angerami, 1968: J. Geophys. Res., **71**, 711.
Storey, L. R. O., 1953: Phil. Trans. Roy. Soc., **A246**, 113.

PUBLICATIONS BY LEIV HARANG

(Collected and arranged by J. Holtet)

HARANG, L. (1931): Über das Auftreten eines besonderen Nordlichtbogens am 26. Januar 1931. *Zs. Geophys.*, **7**, 271.

HARANG, L. (1931): Filteraufnahmen von Polarlicht. *Zs. Geophys.*, **7**, 324.

HARANG, L. und W. BAUER (1932): Über einen Nordlichbogen in weniger als 80 km Höhe über der Erde. *Gerl. Beitr. Geophys.*, **37**, 109.

HARANG, L. (1932): Observations of micropulsations in the magnetic records at Tromsö. *Terr. Mag. Atmos. Elec.*, **37**, 57.

HARANG, L. (1932): An apparatus for registration of the aurora borealis. *Terr. Mag. Atmos. Elec.*, **37**, 167.

HARANG, L. and H. TØNSBERG (1932): Investigations of the aurora borealis at Nordlys Observatoriet. *Geofys. Publ.*, **9**, No. 5.

VEGARD, L. and L. HARANG (1933): The auroral spectrum in the region of long waves. *Geofys. Publ.*, **10**, No. 5.

HARANG, L. (1933): Eine Untersuchung der Polarisation des Nordlichtes. *Zs. Geophys.*, **9**, 162.

HARANG, L. (1934): Filteraufnahmen von Polarlicht. *Geofys. Publ.*, **10**, No. 8.

VEGARD, L. and L. HARANG (1934): The wavelength of the green auroral line determined by an interferometer method. *Geofys. Publ.*, **11**, No. 1.

HARANG, L. (1935): Filteraufnahmen von rot gefärbten Nordlichtern. *Gerl. Beitr. Geophys.*, **48**, 229.

HARANG, L. and L. VEGARD (1935): Interferometer measurements of the red auroral line 6300. *Nature*, **135**, 542.

HARANG, L. (1936): Änderungen der Ionisation der höchsten Atmosphärenschichten während der Nordlichter und erdmagnetischer Störungen. *Gerl. Beitr. Geophys.*, **46**, 438.

HARANG, L. (1936): Höhenbestimmungen und Spektralaufnahmen von sonnenbelichteten und rot gefärbten Nordlichtern. *Gerl. Beitr. Geophys.*, **48**, 1.

HARANG, L. (1936): Vertical movements of the air in the upper atmosphere. *Terr. Mag. Atmos. Elec.*, **41**, 143.

HARANG, L. (1936): Oscillations and vibrations in magnetic records at high-latitude stations. *Terr. Mag. Atmos. Elec.*, **41**, 329.

HARANG, L. (1937): Further studies on the vertical movements of the air in the upper atmosphere. *Terr. Mag. Atmos. Elec.*, **42**, 55.

VEGARD, L. and L. HARANG (1937): Continued investigations on the auroral spectrum. *Geofys. Publ.*, **11**, No. 15.
(In "Recent results regarding the spectral analysis of the auroral luminescence.")

HARANG, L. and L. VEGARD (1937): Interferometer measurements of the red auroral line 6300. *Geofys. Publ.*, No. 15.
(In "Recent results regarding the spectral analysis of the auroral luminescence.")

VEGARD, L. and L. HARANG (1937): Interference measurements of the red auroral line 6300 carried out at Oslo. *Geofys. Publ.*, **11**, No. 15.
(In "Recent results regarding the spectral analysis of the auroral luminescence.")

HARANG, L. (1937): Änderungen der Ionisation der höchsten Atmosphärenschichten während der Nordlichter und erdmagnetischen Störungen. *Geofys. Publ.*, **11**, No. 17.

HARANG, L. (1937): Height measurements of selected auroral forms. *Geofys. Publ.*, **12**, No. 1.

HARANG, L. (1938): Annual variation of the critical frequencies of the ionized layers at Tromsö during 1937. *Terr. Mag. Atmos. Elec.*, **43**, 41.

HARANG, L. and W. STOFFREGEN (1938): Scattered reflections of radio waves from a height of more than 1000 km. *Nature*, **142**, 832.

HARANG, L. (1939): Annual variation of the critical frequencies of the ionized layers at Tromsö during 1938. *Terr. Mag. Atmos. Elec.*, **44**, 15.

HARANG, L. (1939): Pulsations in an ionized region at height of 650–800 km during the appearance of giant pulsations in the geomagnetic records. *Terr. Mag. Atmos. Elec.*, **44**, 17.

HARANG, L. (1939): Höhenändrungen des unteren Randes der Nordlichten beim Übergang von der dunkeln zu der sonnenbelichteten Atmosphäre. *Gerl. Beitr. Geophys.*, **54**, 81.

HARANG, L. und W. STOFFREGEN (1939): Der Polarisationzustand der Radiowellen bei der Reflexion an Schichten, die während erdmagnetischer Störungen und Nordlichter gebildet werden. *Hochfreq. u. Elektroak.*, **53**, 181.

HARANG, L. und W. STOFFREGEN (1940): Echoversuche auf Ultrakurzwellen. *Hochfreq. u. Elektroak.*, **55**, 105.

HARANG, L. (1941): Maximalwerte der Erdstromspannungen in der Nähe der Nordlichtzone während sehr intensive erdmagnetischen Störungen. *Gerl. Beitr. Geophys.*, **57**, 310.

HARANG, L. (1941): Polarization-studies of echoes reflected from the abnormal *E*-layer formed during geomagnetic storms. *Terr. Mag. Atmos. Elec.*, **46**, 279.

HARANG, L. (1942): Pulsations in the terrestrial magnetic records at high latitude stations. *Geofys. Publ.*, **13**, No. 3.

HARANG, L. (1942): Experimental studies on the reflections of radio waves from the ionized regions. *Geofys. Publ.*, **13**, No. 4.

HARANG, L. (1944): A study of the auroral arcs and draperies. *Geofys. Publ.*, **13**, No. 14.

HARANG, L. (1945): The luminosity curve of the aurorae. *Geofys. Publ.*, **16**, No. 6.
Also 1946 in: *Terr. Mag. Atmos. Elec.*, **51**, 381.

HARANG, L. (1945): Scattering of radio waves from great virtual distances. *Terr. Mag. Atmos. Elec.*, **50**, 287.

HARANG, L. (1945): A study of auroral arcs and draperies. *Terr. Mag. Atmos. Elec.*, **50**, 297.

HARANG, L. (1945): Radio echo observations at Tromsö during the solar eclipse on July 9, 1945. *Terr. Mag. Atmos. Elec.*, **50**, 307.

HARANG, L. (1945): Note on methods of determining the directions of auroral arcs from a single station. *Terr. Mag. Atmos. Elec.*, **50**, 311.

HARANG, L. (1946): The mean field of disturbance of polar geomagnetic storms. *Terr. Mag. Atmos. Elec.*, **51**, 353.

HARANG, L. (1946): Radio echo observations at Tromsö during the solar eclipse on July 9, 1945. *Geofys. Publ.*, **16**, No. 11.

HARANG, L. (1946): The mean field disturbance of the polar earthmagnetic storm. *Geofys. Publ.*, **16**, No. 12.

HARANG, L. (1951): Aurora and magnetic storms. In "Compendium of Meteorology" (Ed. F. K. Thomas), *Ann. Met. Soc.*, Boston, 347.

HARANG, L. and B. LANDMARK (1953): Radio echoes observed during aurorae and terrestrial magnetic storms using 35 Mc/s and 74 Mc/s waves simultaneously. *Nature*, **171**, 1017.

HARANG, L. and B. LANDMARK (1954): Radio echoes observed during aurorae and geomagnetic storms using 35 and 74 Mc/s waves simultaneously. *J. Atmosph. Terr. Phys.*, **4**, 322.

OMHOLT, A. and L. HARANG (1955): Measurements of the mean lifetime of the metastable *S*-state of the oxygen atom in the upper atmosphere during auroral displays. *J. Atmosph. Terr. Phys.*, **7**, 247.

HARANG, L. (1956): Height distribution of auroral emissions. *J. Atmosph. Terr. Phys.*, **9**, 157.

HARANG, L. and K. PEDERSEN (1956): Drift measurements of the *E*-layer. *Geofys. Publ.*, **19**, No. 10.

HARANG, L. and K. PEDERSEN (1957): Drift measurements of the *E*-layer during the solar eclipse 30 June, 1954. *J. Atmosph. Terr. Phys.*, **10**, 44.

HARANG, L. and K. PEDERSEN (1957): Drift measurements of the *E*-layer. *J. Geophys. Res.*, **62**, 183.

HARANG, L. (1957): Results of drift measurements in Norway. In "Polar Atmosphere Symposium, Part II Ionospheric Section" (Ed. K. Weeks), Pergamon Press, London, 26.

LOVELL, A. C. B., P. A. FORSYTH and L. HARANG (1957): Radio reflections from aurora. *Ann. IGY*, **3**, Part 4, Pergamon Press, 337.

HARANG, L. (1958): Height distribution of the red auroral line in polar aurorae. *Geofys. Publ.*, **20**, No. 5.

HARANG, L. and J. TRØIM (1959): Determination of the angle of arrival of auroral echoes. *J. Atmosph. Terr. Phys.*, **14**, 107.

HARANG, L. and J. TRØIM (1959): An example of heavy absorption in the VHF-band in the arctic ionosphere. *Planet. Space Sci.*, **1**, 102.

HARANG, L. and K. MALMJORD (1960): Drift measurements of the *E*-layer at Kjeller and Tromsö during the international geophysical year 1957–58. *Geofys. Publ.*, **22**, No. 1.

HARANG, L. and A. OMHOLT (1960): Luminosity curves of high aurorae. *Geofys. Publ.*, **22**, No. 2.

HARANG, L. and J. TRØIM (1961): Studies of auroral echoes – I. *Planet. Space Sci.*, **5**, 33.

HARANG, L. and J. TRØIM, (1961): Investigation of auroral echoes – II. *Planet. Space Sci.*, **5**, 105.

HARANG, L. and K. MALMJORD (1961): Determination of drift movements of the ionosphere at high latitudes from radio star scintillations. *Geofys. Publ.*, **23**, No. 3.

HARANG, L. (1962): A study of the auroral OI feature at 8446 Å. *Planet. Space Sci.*, **9**, 383.

HARANG, L. (1963): Drift of the ionosphere at high latitude determined from radio star scintillations. *J. Atmosph. Terr. Phys.*, **25**, 109.

HARANG, L. (1964): Electromagnetic radiation in the 8 kc/s band observed near the auroral zone. *Astrophys. Norv.*, **9**, 136.

HARANG, L. and R. LARSEN (1964): Observations of enhanced radiation in the kc/s band (5 and 8 kc/s) during auroral and geomagnetic disturbances. In "Natural Electromagnetic Phenomena below 30 kc/s" (Ed. D. F. Bleil), Plenum Press, New York, 143.

HARANG, L. and R. LARSEN (1965): Radio wave emissions in the v.l.f.-band observed near the auroral zone I. Occurrence of emissions during disturbances. *J. Atmosph. Terr. Phys.*, 27, 481.

HARANG, L. and K. N. HAUGE (1965): Radio wave emissions in the v.l.f.-band observed near the auroral zone II. The physical properties of the emissions. *J. Atmosph. Terr. Phys.*, **27**, 499.

HARANG, L., R. LARSEN and J. SKOGTVEDT (1965): The distribution of intensities in v.l.f.-emissions. *J. Atmosph. Terr. Phys.*, **27**, 1147.

HARANG, L. (1967): VLF emissions observed near and south of the auroral zone. In "Aurora and Airglow" (Ed. B. M. McCormac), Reinhold Publ. Corp., New York, 607.

HARANG, L., R. LARSEN and J. SKOGTVEDT (1967): VLF-emissions in the kc/s band observed at stations close to the auroral zone and at stations on lower latitudes. *Phys. Norv.*, **2**, 271.

HARANG, L. (1968): VLF-emissions observed at stations close to the auroral zone and at stations on lower latitudes. *J. Atmosph. Terr. Phys.*, **30**, 1143.

HARANG, L. (1968): Emission of VLF during the great disturbance of 25–26 May 1967. *Planet. Space Sci.*, **16**, 1081.

HARANG, L. (1968): The Forbush-decrease in cosmic rays and the transit time of the modulating cloud. *Planet. Space Sci.*, **16**, 1095–1101.

ERIKSEN G. and L. HARANG (1969): Radio noise from the ionosphere on 225 MHz during a great ionospheric disturbance. *Phys. Norv.*, **4**, 1.

HARANG, L. (1969): Radio noise emissions in the MHz-band during aurorae and ionospheric disturbances. *Phys. Norv.*, **4**, 5.

HARANG, L. (1969): Radio noise from Aurora. *Planet. Space Sci.*, **17**, 869.

HARANG, L. (1969): Electromagnetic emissions in the kHz band during aurora. In "Atmospheric Emissions" (Eds. B. M. McCormac and A. Omholt), Van Nostrand Reinhold Comp., New York, 153.

HARANG, L. (1969): Bursts of VLF-emission simultaneous with sharp dip of absorption and SC in geomagnetic *H*. *Planet. Space Sci.*, **17**, 1565.

HARANG, L. and S.-I. AKASOFU (1969): Low latitude VLF emissions and polar substorms. *J. Atmosph. Terr. Phys.*, **31**, 1445.

HARANG, L. (1970): Observations of VLF emissions along a N-S chain of stations crossing the auroral zone. In "Plasma Waves in Space and in the Laboratory, Vol. 2" (Eds. J. O. Thomas and B. J. Landmark), Edinburgh University Press, Edinburgh, 389.

HARANG, L. and J. TRØIM (1970): Polarization of auroral echoes. *Planet Space Sci.*, **18**, 655.

HARANG, L. (1970): World-wide VLF-emissions during the great disturbance of 25–26 May, 1967. *Planet. Space Sci.*, **18**, 944.

MONOGRAPHS:

HARANG, L. (1940): "Das Polarlicht und die Probleme der höchsten Atmosphärenschichten." Probleme der Kosmischen Physik, Band XX. Akad. Verlag Becker & Erbr., Leipzig.

HARANG, L. (1951): "The Aurora". International Astrophysics Series, Vol. 1. Chapman & Hall, London.

Papers published in *Geofysiske Publikasjoner* may be obtained from: Universitetsforlaget, Blindern, Oslo 3, Norway.

Vol. XXIV.

In memory of Vilhelm Bjerknes on the 100th anniversary of his birth. 1962.

Vol. XXV.

No. 1. Kaare Pedersen: On the quantitative precipitation forecasting with a quasi-geostrophic model. 1963
– 2. Peter Thrane: Perturbations in a baroclinic model atmosphere. 1963.
– 3. Eigil Hesstvedt: On the water vapor content in the high atmosphere. 1964.
– 4. Torbjørn Ellingsen: On periodic motions of an ideal fluid with an elastic boundary. 1964.
– 5. Jonas Ekman Fjeldstad: Internal waves of tidal origin. 1964.
– 6. A. Eftestøl and A. Omholt: Studies on the excitation of N_2 and N_2+ bands in aurora. 1965.

Vol. XXVI.

No. 1. Eigil Hesstvedt: Some characteristics of the oxygen-hydrogen atmosphere. 1965.
– 2. William Blumen: A random model of momentum flux by mountain waves. 1965.
– 3. K. M. Storetvedt: Remanent magnetization of some dolerite intrusions in the Egersund Area, Southern Norway. 1966.
– 4. Martin Mork: The generation of surfa e waves by wind and their propagation from a storm area. 1966.
– 5. Jack Nordø: The vertical structure of the atmosphere. 1965.
– 6. Alv Egeland and Anders Omholt: Carl Størmer's height measurements of aurora. 1966.
– 7. Gunnvald Bøyum: The energy exchange between sea and atmosphere at ocean weather stations M, I and A. 1966.
– 8. Torbjørn Ellingsen and Enok Palm: The energy transfer from submarine seismic waves to the ocean. 1966.
– 9. Torkild Carstens: Experiments with supercooling and ice formation in flowing water. 1966.
– 10. Jørgen Holmboe: On the instability of stratified shear flow. 1966.
– 11. Lawrence H. Larsen: Flow over obstacles of finite amplitude. 1966.

Vol. XXVII.

No. 1. Arne Grammeltvedt: On the nonlinear computational instability on the equations of one-dimensional flow. 1967.
– 2. Jørgen Holmboe: Instability of three-layer models in the atmosphere. 1968.
– 3. Einar Høiland and Eyvind Riis: On the stability of shear flow of a stratified fluid. 1968.
– 4. Eigil Hesstvedt: On the effect of vertical eddy transport on atmospheric composition in the mesosphere and lower thermosphere. 1968.
– 5. Eigil Hesstvedt: On the photochemistry of ozone in the ozone layer. 1968.
– 6. Arnt Eliassen: On meso-scale mountain waves on the rotating earth. 1968.
– 7. Kaare Pedersen and Knut Erik Grønskei: A method of initialization for dynamical weather forecasting, and a balanced model. 1969.

Vol. XXVIII.

No. 1. Haakon Mosby: Atlantic Water in the Norwegian Sea. 1970.
– 2. Kaare Pedersen: Balanced Systems of Equations for the Atmospheric Motion. 1971.
– 3. Arnt Eliassen and Jan-Edgar Rekustad: A Numerical of Mesco-Scale Mountain Waves. 1971
– 4. Per B. Storebø: Steady State Aerosols. 1972.
– 5. Thomas A. McClimans: An Approximate Theory for the Structure of Strong Pycnoclines. 1972
– 6. Aksel Wiin-Nielsen: Simulations of the Annual Variations of the Zonally-Averaged State of the Atmosphere. 1972. In Press.

FYENS STIFTSBOGTRYKKERI