Polar Ionospheric Research

VICTOR P. HESSLER

INTRODUCTION

The short wave electromagnetic radiation from the sun has sufficient energy to cause photo-ionization of the gases in the upper atmosphere. The resulting ionized region ranging in height from about 50 km. to 1000 to 2000 km. is called the ionosphere. The possibility of such a conducting layer was postulated in 1878 by Balfour Stewart to account for the daily variation of the geomagnetic field and was given additional credence in 1902 by Kennelly and Heaviside in explanation of the seemingly anomalous propagation of wireless signals across the Atlantic Ocean. The existence of the ionosphere could no longer be questioned after the classic pulsed radio experiment of Breit and Tuve in 1926.

The first indication of trapped particles in the earth's magnetic field, far above any trace of the atmosphere, was given by the inter-hemispheric propagation of lightning-induced electromagnetic wave propagation along magnetic field lines: whistlers. By the mid 1950's calculations based on the dispersion of whistlers had shown that electron densities of the order of 400 cm.⁻³ existed at an altitude of 12,000 km. in the region now called the magnetosphere. The magnetosphere may be defined as the region in which the geomagnetic field energy density is greater than the kinetic energy density of all charged particles and thus their motions are under magnetic field control. The magnetosphere extends from about 500 km. to several earth radii on the sunward side and to many earth radii on the nightside (Fig. 1).

FIG. 1. Present-day view in noon-midnight magnetic meridian plane of the geomagnetic field and tail as distorted by the supersonic solar wind. A detached bow shock forms upstream from the magnetosphere and leads to the development of the turbulent boundary layer, the magnetosheath (From King and Newman 1967).

Consultant, Geophysical Institute, University of Alaska, College, Alaska. Also guest worker, Environmental Science Services Administration, Boulder, Colorado, U.S.A.
From a practical point of view knowledge of the ionosphere and its variations is essential to all aspects of radio communications other than ground wave and line of sight transmissions. The Ballistic Missile Early Warning System radar signals must traverse most or all of the ionosphere twice and thus are subject to varying amounts of absorption and refraction in the highly active polar ionosphere. In addition spurious reflections from auroral-type activity may block extensive areas of the field of view. Electric currents in the ionosphere produce magnetic and telluric variations at the ground. Under disturbed conditions the variations may make it impossible to conduct geophysical prospecting involving magnetic and telluric techniques.

DESCRIPTION OF THE IONOSPHERE

Perhaps the most important ionospheric parameters are the height profiles of charge density, atmospheric pressure, and electron, ion, and neutral particle temperatures (Fig. 2a, b). These parameters determine the particle collision frequencies and the critical radio frequencies at the various levels of the ionosphere and

---

**FIG. 2.** Normal electron distributions at the extremes of the sunspot cycle: a. Daytime; b. Nighttime (From Johnson 1965).
thus the refraction and absorption of radio waves. Knowledge of the molecular and atomic constituents at the various levels is essential to an understanding of the absorption cross section of the solar electromagnetic radiation and the resulting ionization.

The electrical conductivity of the ionosphere together with the mechanical and electrical driving functions determine the magnitude and direction of the current density vector in the various regions and levels of the ionosphere. The guiding effect of the earth’s magnetic field upon the moving charges results in anisotropic conduction and thus a tensor conductivity. The two important components of this tensor from the standpoint of resulting ionospheric currents are the Hall and Pederson conductivities (Figs. 3 and 4).

**FIG. 3.** Hall conductivity $\sigma_2$ versus altitude (From Johnson 1965).

**FIG. 4.** Pederson conductivity $\sigma_1$ (reduced conductivity) versus altitude (From Johnson 1965).

**Ionospheric Regions**

The ionosphere divides naturally into three horizontal layers or regions based on distinct physical characteristics and ionization processes. The regions are termed the D-region, the E-region and the F-region from the bottom to the top of the ionosphere, with further subdivisions of the E- and F-regions.
THE QUIET D-REGION

During undisturbed solar conditions the daytime D-region extends from a height of about 60 km. to 85 or 90 km. (Fig. 1). The electron density of the quiet daytime D-region is of the order of $10^3$ cm.$^{-3}$ at 80 km. and decreases to 10 cm.$^{-3}$ at 60 km. The lower D-region ionization essentially disappears on a quiet night (Fig. 2b). Electron density values are of the order of $10^3$ cm.$^{-3}$ at 90 km. and 10 cm.$^{-3}$ at 80 km., down from the daytime level by 2 orders of magnitude. Owing to the high collision frequency in the D-region its conductivity is too low (Figs. 3 and 4) to permit appreciable electric currents to flow. Thus the D-region contributes very little to magnetic variations observed at the surface of the earth.

The practical importance of the D-region lies in its high rate of absorption of radio waves. Under normal daytime conditions there is essentially no ionospheric reflection of 0.05 to 1 MHz waves; thus communications in this band are limited to line of sight or to ground waves: the well known day-night effect in the broadcast band.

THE QUIET E-REGION

The boundaries of the E-region, as stated in the various references, range from 85 to 90 km. at the bottom and from 140 to 160 km. at the top. The E-region electron density at the noontime, quiet sun condition is of the order of $10^5$ cm.$^{-3}$ (Fig. 2a). The density is about 50 per cent greater during sunspot maximum. The major variation from day to night is noted by comparing Figs. 2a and b, which show nighttime electron densities of approximately 2 to 3 orders of magnitude less than the daytime values. The maximum conductivity of the ionosphere in directions corresponding to existing driving functions, such as the dynamo and tidal actions, is located in the E-region. As noted above, the conductivity is low in

![Figure 5](image-url)
the D-region because of the high collision frequency. In the F-region the collision frequency is very low in comparison with the spiralling frequency perpendicular to the magnetic field lines, and thus any driving function across the magnetic field produces little net forward motion. In the E-region the collision frequency is of the same order of magnitude as the spiralling frequency, thus the spirals are broken up and the net forward motion increases. The auroral electrojet flows in this relatively high conductivity path.

The primary ionizing radiations of the E-region are solar X-rays (10 to 100 Å), the hydrogen Lyman $\beta$ line (1025.7 Å), the doubly ionized carbon, C$^{++}$, line (977 Å), and the hydrogen Lyman continuum (910 to 800 Å) (Fig. 5). The characteristics of the daytime E-region are very well known. These characteristics are such as to give a good radio reflection in the frequency range of the vertical incidence sounder. The delay time gives accurate height information and electron density is determined from the critical frequency. The E-region is monitored routinely by more than 100 ionosondes distributed throughout the world. The practical importance of the E-region lies in its relative low absorption and effective reflection of radio waves in the frequency band from about 1 to 3 MHz thus permitting long-distance radio communications in this frequency band. Daytime radio transmissions for distances up to 2000 km. generally take place via the E-layer.

THE QUIET F-REGION

The F-region consists of two layers, F1 and F2. The lower layer, F1, ranges in height from 140 to 200 km. (Fig. 2a) and is characterized by both the well-behaved E-layer below and the variable F2 layer above. Like the E-layer, it disappears at night and is related to solar zenith angle (Fig. 2b). Typical noontime electron densities range downward from $2.5 \times 10^{5}$ cm.$^{-1}$ at sunspot maximum. The principal ionizing agent of the F1 layer is the solar ultraviolet in the wavelength band from 200 Å to 900 Å. The ions are principally NO$^+$ and O$_2^+$ at the lower boundary and O$^+$ at the upper boundary. The deep valley above the residual nighttime E-region (Fig. 2b) cannot be recognized on the ionosonde records. Knowledge of this region is dependent on rocket-borne detectors. With the disappearance of the F1 layer at night the F2 layer becomes synonymous with the F-region.

The F2-layer ranges from 200 km. to 1000 to 2000 km., the lower limit being determined by electron density and the upper by ion composition considerations. Long distance radio communication between ground-based stations is possible only because of refraction in the F-layer. The various radio warning services publish predictions of MUF (maximum useable frequency) values which are based on measurements of electron densities and anticipated solar activity.

The Disturbed Sun

The solar electromagnetic radiation ranges from radio wavelengths through the visible, ultraviolet, to X-rays of less than 1 Å. The continuum radiation is relative constant even in the ultraviolet, but the shorter wavelengths increase materially at sunspot maximum and the very short wavelengths show striking in-
creases with solar flare activity. For example, studies of solar electromagnetic energy flux show ratios of maximum to minimum solar intensity of 7:1 for wavelengths $<200 \text{ Å}$, 60:1 for wavelengths $<20 \text{ Å}$, and 600:1 for wavelengths $<8 \text{ Å}$. The hard X-rays are very flare sensitive. One measurement of 2 to 8 Å X-ray flux during a class 2+ flare shows an increase 770,000 times the sunspot minimum, quiet sun value. In contrast such measurements normally show only several percent increase in Lyman $\alpha$ (1216 Å). The solar wind also increases appreciably with sunspot activity. Under quiet solar conditions typical values are 5 protons cm.$^{-3}$ at a velocity of 500 km. sec.$^{-1}$ ranging to 10 protons cm.$^{-3}$ at 1500 km. sec.$^{-1}$ for disturbed conditions.

The Disturbed Ionosphere

The advent of a solar disturbance, particularly a class 2 or 3 flare, results in major changes in the characteristics of the several ionospheric regions.

The Disturbed D-region

There are three significant types of D-region disturbances associated with increased solar activity, particularly with solar flares. They are termed “sudden ionospheric disturbances” (SID), “auroral absorption events”, and “polar cap absorption events” (PCA). An SID is characterized by a sudden increase in absorption of MHz-frequency radio waves in the sunlit hemisphere with maximum attenuation at the subsolar point and decreasing monotonically to the twilight region. Its onset is normally precipitate and simultaneous with the visual appearance of a major flare. The attenuation gradually decreases to the pre-flare level over a period of 60 to 90 minutes. The increased low-level ionization responsible for the attenuation is produced by the 1.0 to 10 Å X-rays associated with the flare (Fig. 5).

Auroral absorption, an auroral zone effect, is considered to result from the precipitation of energetic electrons into the upper D-region. It is predominantly a nighttime effect associated with active auroras and magnetic disturbances. The concept of the PCA event dates from the IGY period. “The very heavy radio absorption which began less than an hour after the great west-limb solar flare of 23 February 1956 and whose onset occurred during the most intense outburst of solar cosmic rays yet recorded, was the first generally recognized case of high-latitude absorption that could not be considered as merely a special case of auroral absorption.” (Bailey 1964). It is now known that these PCA events are caused by low energy solar cosmic rays — protons with energies from 300 to approximately 5 Mev. (The most pronounced PCA event in several years occurred at the close of the Naval Arctic Research Laboratory Symposium completely disrupting radio communications from NARL to the north.)

The Disturbed E-region

Variations in the E-region due to solar disturbances generally appear in the form of thin enhanced layers of ionization called Sporadic E. The source of the Sporadic E ionization is not well understood. However, in the case of the aurorally-associated Sporadic E the agency must be energetic particle bombardment, but
how the particles can produce such thin layers is not at all clear. The role of Sporadic E in radio communications is quite complicated. The layers serve as reflectors with a wide range of possible ray paths, involving multiple reflections at the earth and from the E and F regions. The reflection from a Sporadic E layer may occur either at the top or bottom of the layer.

THE DISTURBED F-REGION

The first observation of an F-region disturbance was made at Slough, England, in 1935. The disturbances are characterized by a marked decrease in the F2 layer critical frequency and electron density at mid and high latitudes and by a corresponding increase at low latitudes.

IONOSPHERIC RESEARCH TECHNIQUES

Until the advent of high altitude rockets and of satellites, all ionospheric investigations were of necessity indirect. Most of the techniques are based on analysis of radio wave reflection or refraction and attenuation in the frequency range from about 10 to 50 MHz. The whistler techniques extend the useful spectrum down to about 300 Hz and below this the magnetic and telluric field techniques utilize the band in the range from about 50 to 0.001 Hz plus diurnal variation effects. At the other end of the spectrum, the spectral lines, intensity, and morphology of the visual aurora provide significant information. With the advent of rockets and satellites direct measurements of the various ionospheric parameters became possible as extensions of the radio techniques. Only some representative techniques will be mentioned.

Optical techniques

THE ALL-SKY CAMERA

The all-sky camera provides a minute by minute record of the auroral morphology out to a radius of about 800 km from the camera. Since auroral luminosity is closely related to electric charge precipitation into the lower ionosphere (because ionospheric currents flow along auroral arcs, and because brilliant and active auroral displays occur at the site of major ionospheric disturbances) a net of all-sky cameras may provide significant information about ionospheric activity over a vast area. The camera also has the advantage of any visual technique in showing the exact location and contour of the disturbance. In a recent experiment all-sky cameras carried on jet aircraft flying along magneto conjugate paths in the northern and southern hemisphere showed a remarkable similarity in the quiet time auroral forms.

AURORAL SPECTROSCOPY

Spectrographic techniques together with photometry and triangulation provide a wealth of information concerning ionospheric processes. With the aid of the diverse techniques available, deductions may be made concerning many aspects of the aurora. For instance, details such as the energy and flux density of the
precipitating electrons and protons responsible for auroral excitation, the type of neutral or ionized atom or molecule being excited, and the height at which the excitation occurs may be deduced.

Radio Propagation Techniques

THE IONOSONDE

The "pulse method" of ionospheric exploration, first introduced by Breit and Tuve is by far the most commonly used. The instrument, called an ionosonde, is essentially a pulsed radar (of about 10 kw. peak power) in which the exploring frequency is varied, normally over a range from 1 to 25 MHz. The wave is directed vertically and the echoes are usually recorded on a panoramic display from an oscilloscope with virtual height on the Y-axis and frequency displayed logarithmically on the X-axis. The frequency is normally swept through the 24 MHz in about 15 sec. More than 100 ionosondes are in continuous operation throughout the world.

THE RIOMETER

The riometer, or relative ionospheric opacity meter, measures the absorption of cosmic radio noise in its passage through the ionosphere. The instrument basically consists of a calibrated sensitive radio receiver with an output circuit capable of driving a strip chart recorder. In the frequency range of 25 to 50 MHz the riometer is insensitive to normal D-region absorption but quite sensitive to the increased D-region absorption which accompany SID's, auroral events, and PCA's. At the auroral zone it may be difficult at times to separate auroral and PCA events, particularly at night when electron attachment greatly reduces the PCA effect. However, the incidence of the accompanying auroral event is often delayed from several hours to two days after the onset of the PCA. The riometer provides excellent PCA records during such periods.

VHF FORWARD SCATTER

A forward scatter system consists of a transmitting and receiving station separated by distances of from 1000 to 2000 km. and usually operating in the frequency range from 30 to 40 MHz. The relatively high gain antennas are directed to the lower ionosphere at the midpoint between the sending and receiving stations. Bailey (1964) has made extensive use of the technique for the study of PCA's in the Arctic. The several systems used had midpoints ranging from 60° to 83°N. geomagnetic latitude. The range in latitude corresponded to calculated vertical cutoff energies for protons of from 500 down to 0.09 Mev.

AURORAL RADAR

Auroral radar echoes are observed with oblique incidence sounders operating in the frequency range from 20 to 800 MHz. At lower frequencies the signal may be attenuated too much and at higher frequencies the back-scattering effect is small. As the term implies the scattering electrons in auroral radar are associated with visual auroral phenomena, but the correlation is far from one to one. The
scattering irregularities are highly aspect sensitive. Radar reflections at College, Alaska, come from the quadrant bisected by magnetic north and in ranges from 400 to 1000 km. At Barrow where most of the visual aurora is seen to the south, the predominant direction for radar reflection is at a low angle in the general northerly direction. The auroral radar technique seems to have produced only limited results. However, knowledge of its characteristics is, of course, critical to the BMEWS operation.

WHISTLERS

Whistlers are electromagnetic signals in the audio frequency range. They are often preceded by a click followed by a sliding tone commonly ranging from 5000 down to 1000 Hz in about a second. It is now known that the click is propagated directly to the listener from a lightning stroke. Some of the broad-band electromagnetic wave energy from the stroke is transmitted to the ionosphere where it couples to a tube of magnetic flux along which it is guided by the free electrons in the magnetosphere. The energy packet is reflected at the base of the tube in the opposite hemisphere. Since the high frequency components travel faster than the lower frequencies the wave is dispersed and produces the whistler. Whistler analysis has shown that during sunspot maximum the electron density is of the order of 100 cm.$^{-3}$ at 5 earth radii. During a magnetic storm the electron densities in the magnetosphere may drop to one-tenth of the normal quiet day value.

Balloon and Rocket Techniques

Balloons, having a normal ceiling in the order of 30 km, permit the study of only the more penetrating radiations. They have found their most extensive use in the recording of bremsstrahlung X-rays produced when precipitated energetic electrons are brought to rest in the upper atmosphere of the auroral zone. In situ atmospheric and ionospheric measurements in the region between balloon heights and the usual satellite perigees, roughly 30 to 200 km., are made with rockets. In contrast with the several hours duration of balloon experiments, rocket experiment durations are of the order of many seconds to several minutes. Among the many instrumentations and experiments carried aloft by rockets are magnetometers, energetic particle detectors, auroral spectrophotometers, barium cloud ionospheric wind-detecting experiments, and temperature detectors.

The first magnetometer probes of the ionosphere were made in 1948-49 when three triaxial fluxgate units were flown to about 110 km. aboard Aerobee rockets by the Naval Ordnance Laboratory. The next flight, made in 1956, carried a proton precession magnetometer. All succeeding rocket borne magnetometers through, at least, 1964 were either proton precession or rubidium vapor magnetometers. Such flights have determined the height of the equatorial electrojet, and some mid-latitude phenomena; they are now being used to explore the relationship between the auroral electrojet and the visible aurora.
MAGNETIC AND TELLURIC DISTURBANCES

From the standpoint of ionospheric research, magnetic and telluric records taken at the earth's surface are used to a large extent as indices of ionospheric activity. Ionospheric disturbances, particularly in the region of the auroral zone involve marked changes in the ionospheric conductivity and in electric forces. The resulting ionospheric current systems produce magnetic changes on the ground at the auroral zone of more than 2500 γ (5 per cent of the total field). Magnetic observatories and their international associations publish a group of indices which correlates well with various ionospheric parameters. Real time magnetic and telluric records are used to predict radio propagation conditions and, for example, to control release times for balloon and rocket experiments.

Geomagnetic micropulsation studies, while still in the natural history stage of study of the phenomena themselves, begin to provide insight into other magnetospheric and ionospheric phenomena. For example, there is a prominent band of pulsations in the period range from 1 to 5 seconds, with amplitudes from several milligammas to several gammas which are generated in the magnetosphere on closed field lines and propagate horizontally in a waveguide centred on the $F_2$ peak. Recent studies of the incidence of these pulsations at auroral-zone stations and at the north geomagnetic pole show major seasonal variations in the propagation characteristics of the polar cap ionosphere.

THE AURORAL OVAL

Analysis of the IGY and succeeding all-sky camera data, and other auroral type activity has shown that the precipitation of auroral primary particles occurs in an oval displaced about 3° toward the dark hemisphere. As a first approximation the earth rotates under the oval once a day. The oval may be termed the instantaneous auroral zone, whereas the auroral zone is now considered to be the locus of the midnight sector of the auroral oval. The auroral oval expands and contracts with variations in the intensity of magnetic activity.

POLAR IONOSPHERIC RESEARCH PROBLEMS

An extensive discussion and analysis of the current polar ionospheric research problems is contained in the forthcoming report by the Panel on Upper Atmosphere Physics (1969). The following sections reflect a number of the views and recommendations of the above report.

The Magnetospheric Model

Many aspects of the magnetospheric model are not clearly defined; in particular we lack understanding of the high latitude lines which are swept away from the sun by the solar wind to form the magnetospheric tail.

Auroral Theory

A unifying theory is needed to follow the solar electrons and protons from their incidence and capture at the edge of the magnetosphere, through the magneto-
sphere to their final precipitation in the polar regions, with the accompanying auroral type phenomena, visual, radio, thermal, magnetic, etc.

The Polar Substorm

Perhaps the most significant progress in polar ionospheric research in the past decade has been the rapid development of the polar substorm concept, the aspects of which are reviewed in detail by Akasofu (1968). Studies of all-sky camera records for several years show that the quiet auroral forms which characteristically lie along the auroral oval are activated intermittently in a disturbance extending all along it. The disturbance originates in the midnight sector and spreads rapidly and violently in all directions. The substorm lasts from one to three hours and is usually repeated every few hours during a magnetic storm.

The auroral substorm is always accompanied by an intense auroral electrojet which is responsible for the polar magnetic substorm. The equivalent current system of the magnetic substorm is assumed to flow in the ionospheric E-region and is inferred from magnetic perturbation vectors, or more commonly from the H or X component at polar magnetic stations. Many of the aspects of the polar substorm are concomitants of the magnetospheric substorm. Studies to date suggest that the intense plasma cloud ejected during a major solar flare generates a shock wave in the interplanetary plasma. If the flare occurs within say a 20° radius around the centre of the solar disk as viewed from the earth the shock wave may produce a sudden compression of the magnetosphere sufficient to trigger explosive processes in the magnetosphere which are responsible for the polar substorm phenomena described above. The magnetospheric-polar substorm complex represents one of the most important of the current polar ionospheric research problems.

RESEARCH PROGRAMS

Routine recording

Our knowledge of the polar ionosphere has progressed to the point where we may perform many special experiments to test new or old hypotheses. Since funding is always a problem and the experiments are expensive there is a temptation to curtail the older programs. However, the value of the experiment often depends materially upon synoptic data from the various observatories. Every effort should be made to keep the polar net of ionosondes, magnetic observatories, forward scatter systems, all-sky cameras, and riometers in operation. For example, for two years, there has been no ionosonde in operation at Thule, the north geomagnetic pole, and now the data are much needed to support the micropulsation analysis.

A polar meridian net

The most urgent need in the continued study of the eccentric auroral oval is a meridian net of magnetic, riometer and all-sky camera stations extending from the geomagnetic pole to say 62°N. geomagnetic. In addition to further delineation of the auroral oval, such a net would provide a frame of reference for polar orbiting satellite data taken in its spatially and temporally-variable coordinate system.
Satellites and rockets

Experiments utilizing satellites and rockets are needed for studies of the D and lower E regions, Sporadic E, distribution in latitude and altitude of ionospheric minor constituents such as atomic oxygen, ozone, water vapor and nitric oxide, and determination of the positive and negative ionic composition.

The importance of ground-based support for satellite research such as magnetometers and all-sky cameras at the base of geostationary satellite L-shells must not be overlooked. Fig. 6 shows that such experiments in the arctic fall within the principal Naval Arctic Research Laboratory support area.

FIG. 6. Locus of the base of L-shells corresponding to a range of geostationary satellite positions over the Pacific Ocean.

LOGISTICS AND FACILITIES

Logistics will presumably always be one of the major problems of polar ionospheric research. In Antarctica the Navy has the capability of placing a research team anywhere on the continent at almost any time but, of course, the expense precludes any but the most urgent programs being situated away from the established bases. In the Arctic NARL has the same type of capability as evidenced by their monumental and courageous support of Arctic drifting station ARLIS II. Their establishment and support of ARLIS III and ARLIS IV exemplifies their capability of supporting individual programs and at a moderate cost.

Future logistics of polar research will surely involve unmanned automatic observatories with telemetry analogous to the current satellite techniques. The Panel on Upper Atmospherics report (1969) states, “The unmanned automatic observatory with real-time communication link has nearly all the observational capabilities of a manned station, as has been demonstrated by various observatory satellites. The experience gained from manned stations has provided enough information to enable the observatories and sensors to be interfaced to the local
environment. In addition, an unmanned automatic observatory in the polar regions can provide vastly improved data acquisition capabilities. Data can be transmitted from the field site to the experimenter’s laboratory in real time using synchronous satellites.”

That NARL will assist in the accomplishment of future polar ionospheric research cannot be doubted. Its record to date, its geographical location, and the interest evidenced by the Department of the Navy all indicate the continuation of its valuable contribution.

ACKNOWLEDGEMENTS

I should like to express my sincere appreciation to the Office of Naval Research and to the Naval Arctic Research Laboratory for the support of my ionospheric research program at Point Barrow and on the arctic drifting stations.

REFERENCES


