

TOR HAGFORS* AND KRISTIAN SCHLEGEL*

Earth's ionosphere

1 INTRODUCTION

In this chapter we will deal with the ionosphere, a part of the terrestrial atmosphere ranging from about 80 km to more than 1000 km height. Its lower boundary lies in the mesosphere, at large heights a transition into the magnetosphere occurs. In altitude the ionosphere approximately coincides with the thermosphere a part of the neutral atmosphere (Figure 1). The term thermosphere is used when the properties of the neutral gas are under consideration, the term ionosphere when the electric properties and processes are regarded. Although the charged particle concentration in the ionosphere is at most around 1/1000 of the total particle density (around the F-layer peak, see Figure 1), the ions and electrons enable fundamentally different processes to take place in the ionosphere as compared to those occurring in the neutral atmosphere. Electric currents, electric fields, particle drifts in these fields and the reflection of radio waves are among these phenomena.

In Section 2 we shall first follow the exploration of ionospheric structure in its historical context. We will deal with the different layers (see Figure 1), will discuss peculiarities thereof and geographic as well as temporal anomalies. A separate subsection is devoted to ionospheric irregularities and some consequences of their existence.

Section 3 describes the development of our understanding of the fundamental physical processes leading to the formation and spatial and temporal behavior of the ionosphere. We shall deal with production and loss processes, chemistry, dynamics and transport, winds and waves in the ionosphere and thermosphere. Influences from above

(of the magnetosphere) as well as from below (of the lower atmosphere) will be discussed as well. A brief summary of the current status of our knowledge of ionospheric processes completes this section.

The techniques for probing the ionosphere from the ground and *in situ* are the subjects of Section 4. Experiments by Appleton and Barnett (1925) and by Breit and Tuve (1926) designed to determine the height of reflection of radio waves from the ionospheric layers signaled the beginning of ionospheric research as a science. They in many ways can be said to represent the start of modern space science. They also, as it turned out, had an enormous influence on other fields. The development of the magnetionic theory to describe the propagation of radio waves in magnetized plasmas (Lassen 1927; Appleton 1932) marked a large step forward in plasma physics which before primarily had been confined to studies of gas discharges. The later studies of reflection and scattering from and transmission through irregular ionospheric layers and their interpretation had a profound influence on radio and radar astronomy (Ratcliffe 1948; Booker and Clemmow 1950; Booker *et al.* 1950). They helped to establish relations between the behavior of radio signals transmitted through the interplanetary and the interstellar medium, and led to increased understanding of the properties of plasmas in our solar system, in our galaxy and in other galaxies. Studies designed to detect small radio sources by observing scintillation caused by interplanetary plasma irregularities led to the discovery of pulsars (Hewish *et al.* 1968). They contributed to the interpretation of radar echoes from planets, asteroids and comets in terms of their degree of surface irregularity. The studies of the ionosphere even brought with them important technological advances in that all modern radar systems trace their roots to Breit

*Max-Planck-Institut für Aeronomie, Katlenburg-Lindau, Germany

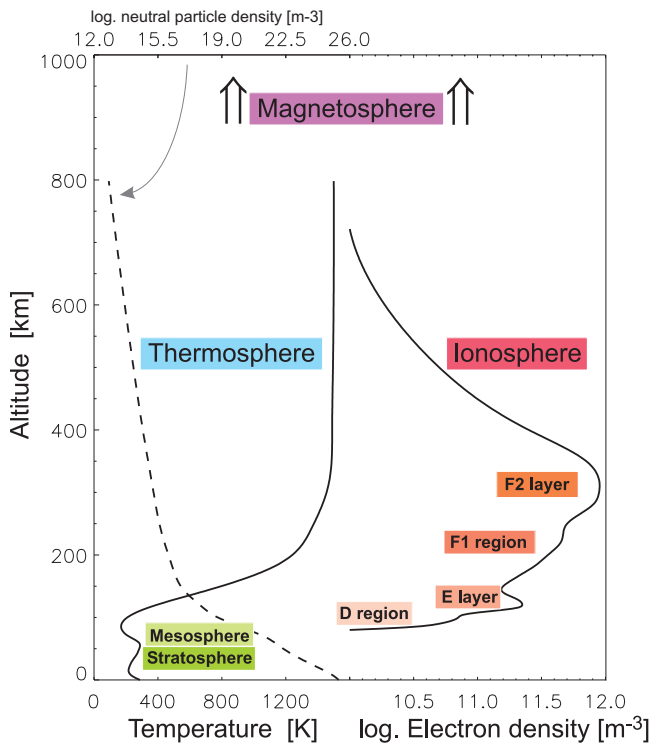


Figure 1 Schematic diagram of the ionosphere, one of the “spheres” of terrestrial atmosphere, between the mesosphere and the magnetosphere. It is characterized by its electron density height profile, subdivided into different layers and regions as described in the following sections (right hand part of the figure). The ionosphere covers approximately the same altitude range as the neutral thermosphere (left hand part of the figure), characterized by the neutral gas temperature. The neutral particle number density is plotted as a dashed line, its scale is on the top of the figure.

and Tuve’s experiments with pulsed radar to determine the height of the ionosphere from the time delay of the echoes.

Finally, in Section 5 we briefly describe the role of the ionosphere and of ionospheric research in our modern society. Impacts on communication, geodesy and space technology are discussed in the context of solar-terrestrial relations and space weather.

The literature on the ionosphere is very extensive comprising several tens of thousands of papers, and it will be impossible to refer to them all in a book on the ionosphere, and let alone in such a brief chapter on the subject. The choice of references in this chapter is highly subjective and to some extent reflects our own interests and areas of research. We apologize for the omission of many important contributions to the field. These omissions must be ascribed to our ignorance of important contributions rather than to ill intent. We sincerely apologize for this limitation in acknowledging all those who have contributed to the field.

Speaking about important contributions, the most outstanding scientist in our field was certainly Professor Edward V. Appleton (1892–1965) as will become evident in the following sections. He received the Nobel Prize for his achievements in 1947, the only scientist thus honored for ionospheric research.

2 EXPLORATION OF IONOSPHERIC STRUCTURE

2.1 Early ideas

The first ideas of an electrically conducting layer in the atmosphere were advanced by Gauss who in 1839 argued that small daily variations of the geomagnetic field could be explained by electric currents flowing in such a layer (see Kaiser 1962). Later Stewart (1882) suggested that these currents may be produced by the dynamo actions of winds blowing across the geomagnetic field. The basis of the electric conductivity however, remained unclear before the discovery of the electron in 1897 and ions around 1900.

After the successful reception of a radio signal transmitted from Poldhu/Cornwall at St. Johns/Newfoundland by Marconi on 12 December 1901, it became obvious that these waves did not propagate on a rectilinear path. Electromagnetic waves discovered by Hertz in 1887 were expected to travel along straight rays like visible light in vacuum, but could also undergo reflection and diffraction. Either one of these processes had to be involved to explain the propagation over such long distance. The first clear hypotheses of an electrical conducting layer and of a reflection of the signal by this layer were published independently by Kennelly (1902) and Heaviside (1902). The diffraction of radio waves around a conducting Earth was treated by MacDonald (1903), Rayleigh (1903), Poincare (1903), Zenneck (1907) and Sommerfeld (1909) and was discarded as possible explanation of Marconi’s results.

Although radio transmissions were more and more extensively used for broadcasting and information transfer in the following years, it took more than two decades before the height of the layer could be established. Appleton’s famous wavelength change experiment was performed in December 1924 (Appleton and Barnett 1925). It proved the existence of a ground wave (direct path) and a sky-wave (reflected at the layer) which are subject to interference at the reception point, causing a fading of the signal. From the fading characteristics the authors were able to derive a first height estimate of the conducting layer of about 90 km. From our present knowledge (see Figure 1) we know that this very low height had two causes. The sounding frequency of 0.75 MHz used was very low and therefore the reflection height was well below the E-region

peak and, the measurements were made during night (because the BBC transmitter used was available for this research purpose only after the end of the normal broadcasting at midnight). About a year later the so-called Peterborough experiment, gave the first height variation: during the time one hour before sunrise to sunrise the layer descended from about 115 to 95 km. The interference pattern formed on the ground between transmitter and receiver was also utilized in another experiment to estimate the layer height. Hollingworth (1926) measured the variation of signal strength of a transmitter near Paris at various receiving locations between London and Edinburgh from which he deduced the height of reflection. At the same time, as mentioned in the introduction, Tuve and Breit (1925) used a transmitting method still applied today, the pulse technique. It gave similar results, namely around 100 km for the reflection height.

2.2 The ionospheric layers

The name of the layer was at that time still associated with researchers in the field, i.e. 'Heaviside layer' or 'Appleton layer'. After the British National Committee for Radio-Telegraphy in 1926 had suggested that a non-personal name should be found, the term 'ionosphere' was introduced by R.A. Watson-Watt and apparently also independently by Appleton. Appleton used the letter 'E' first to denote the electric field of signals reflected from the layer. Only later the letter was used to indicate the layer itself.

A thorough interpretation of these early radio experiments was only possible after a suitable theory of radio wave propagation was developed. Larmour (1924) was the first to realize that a plasma has a refractive index less than unity causing incoming waves to be bent away from the normal to the layer and consequently reflect radio waves. He derived the formula for the index of refraction

$$n^2 = 1 - \left(\frac{\omega_p}{\omega} \right)^2 = 1 - X \quad (2.1)$$

with the plasma frequency

$$\omega_p = \sqrt{\frac{Ne^2}{m_e \epsilon_0}}$$

where N and m_e are the electron density and electron mass, respectively, e the elementary charge, ϵ_0 the dielectric constant of free space, and ω the angular frequency of the transmitted signal. Appleton realized that the influence of the geomagnetic field on the motion of electrons and ions had to be taken into account. This finally led to the derivation of the so-called Appleton-Hartree formula (eqn (4.1)), (Appleton 1932; Hartree 1931) for the index of refraction. Lassen had

derived the same formula even a few years earlier (Lassen 1926), but the name of the formula remained.

On the basis of eqn (2.1) it was obvious that the waves do not propagate with the velocity of light through the plasma and that measurements using the signal travel time only give a 'virtual' height h' . In the general case the virtual height is related to the true height h by

$$h' = \int_0^h \frac{c}{V_g} dz = \int_0^h \left(n + \omega \frac{\partial n}{\partial \omega} \right) dz \quad (2.2)$$

where V_g is the group velocity of the wave and n the index of refraction (the simplified form is given in eqn (2.1)) depending on the height profile of the electron density $N(h)$ and the sounding frequency ω .

Experiments with vertically reflected waves on a higher frequency led Appleton (1927) to the conclusion that another layer existed above the E layer. He called it F layer and found later that during day time it splits into an F1 and an F2 layer. Both layers (cf. Figure 1) were systematically studied in the following years making use of ionosondes developed at the beginning of the thirties (see Section 4). It was found that the ionization followed the solar zenith angle (Appleton and Naismith 1932) suggesting that the ionizing radiation came from the sun. The prevalence of the F layer ionization during night remained a puzzle for many years (see Section 3). Information about the layer thickness was first obtained in 1937, when Appleton showed that the E and F layer had a thickness ratio of 1:4.

A layer below the E layer escaped detection for quite a long time, because of measuring difficulties. The early ionosondes were not able to measure it, because of the low electron densities in this layer and the consequently necessary long wavelength to probe it (cf. eqn (1)). First indication of this 'D layer' came from fadeouts accompanying solar flares (Mögel 1931), later also from the study of the interference pattern between ground wave and sky wave near long-wavelength transmitters. Best and coworkers (1936) were able to derive reflection heights of 74 km during day and 92 km at night. More precise data about the electron density distribution were not obtained before the 1950s when the partial reflection (Gardner and Pawsey 1953) and the cross modulation techniques (Fejer 1955) were introduced (cf. Section 4).

The F1 and the D layers are not really layers with a well-defined peak as is the case for the E and the F2 layers (Figure 1). They form a ledge of enhanced ionization just below the E and the F2 layer, respectively. During the last twenty years, however, it became clear that they are really different from the layers above in terms of the physics involved (ionization, recombination, dynamics, see Section 3). Today we speak of D and F1 regions, rather than using the term layer.

It should be remembered that ionosondes do not give electron density data above the peak of a layer. This led to a gap in the measurements between the E and the F layer only closed after incoherent scatter facilities and rocket and satellite measurements became available. The electron density above the F2-layer peak was first observed with topside sounders from satellites in 1962 and also later with the incoherent scatter technique. The latter provides the most accurate measurements from the ground between about 80 km and several hundreds of km altitude (Figure 2 and Section 4).

At very large heights (above 1000 km) the ionospheric ion composition changes from O^+ to H^+ and He^+ ions, therefore this outer part of the ionosphere is sometimes called protonosphere. The whole near-Earth space filled with thermal plasma was named plasmasphere. This 'sphere' is corotating with the Earth and the thermal plasma within is transported upward and downward along magnetic field lines, even from one hemisphere to the other (Lemaire and Gringauz 1998). The plasmasphere exhibits a relatively sharp outer boundary, a steep drop in plasma density from about 10^9 to about $10^7 m^{-3}$ at 3–5 Earth radii. This plasmopause was detected almost simultaneously from early satellite measurements (Gringauz *et al.* 1961) and from ground-based whistler observations (Carpenter and Park 1963). The plasmopause can also be regarded as the ultimate border of the ionosphere.

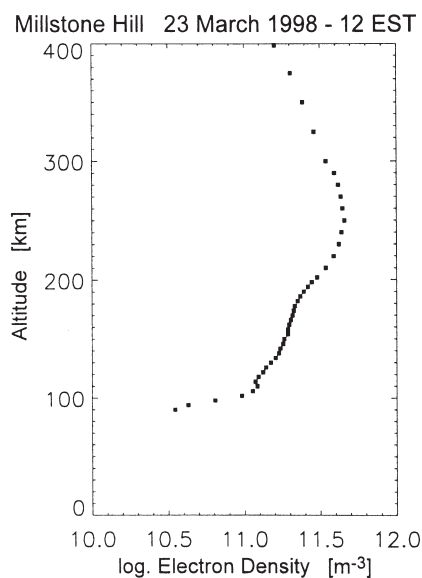


Figure 2 Example of an incoherent scatter measurement of the ionospheric electron density at mid latitudes with the Millstone Hill radar. It covers most of the ionosphere from the D layer to the topside above the F2-layer peak. The F1 layer is visible as well. With the help of advanced transmitter pulse schemes the height resolution can be kept variable, i.e. high in regions with small scale details (D, E, F1 region) and low in the F2 region and above. (J. Holt, private communication.)

2.3 Special effects and geographic peculiarities

The study of diurnal and seasonal variations of the F2 layer revealed several anomalies. The diurnal anomaly addresses the fact that the highest electron density does occur not at local noon (smallest solar zenith angle) but in the afternoon. The winter anomaly refers to the higher ionization in winter than in summer at mid latitudes (Berkner and Wells 1938). The former can be explained when dynamic effects are included in the physical description of the layers rather than only ionization, production and recombination (see Section 3), the latter by seasonal changes of the O/N_2 concentration ratio.

Already in the thirties a special form of the E layer was detected and later named 'sporadic E', or E_s (Appleton and Naismith 1940). It showed high electron densities with a peak often below the normal E-region peak, it was very narrow, and it occurred only sporadically. At mid latitudes observations showed a very pronounced maximum in the probability of occurrence during summer, whereas no clear seasonal variation was found at low and high latitudes (Smith 1957). Later, through measurements with sounding rockets, it was shown that E_s layers contain a considerable content of metallic ions, i.e. Fe^+ and Mg^+ besides O_2^+ and NO^+ , the usual dominant ions at E-region height (Narcisi *et al.* 1967). In the sixties a wind-shear theory of sporadic E was developed by several authors (e.g. Whitehead 1961; Axford 1963). An altitude variation of the wind speed acts in the presence of the geomagnetic field to compress the ionization. This wind-shear theory is generally accepted today as one cause of mid latitude sporadic E.

A special form of E_s with peak electron densities well exceeding those of the F2 layer was observed at high latitudes (Appleton *et al.* 1933) and extensively studied particularly by Scandinavian scientists in the following years. After rocket and satellite measurements became available in the sixties, it was shown that these very pronounced layers (Figure 3) are created by electrons with energies of several keV precipitating from the magnetosphere along geomagnetic field lines. The particle energy determines the height of these layers, the particle flux its peak density (see Section 3). In recent years it became clear that the precipitation of these electrons (and sometimes protons) is controlled by the solar wind. The morphology of these layers is determined by magnetospheric processes also associated with geomagnetic disturbances. During these events significant changes in the F layer were also observed: so-called positive and negative storm effects, denoting an increase or decrease of the F-layer ionization, respectively (see Section 3).

Enhanced electron densities in the high latitude D layer were inferred from radio wave propagation disturbances (Mögel 1931; Dellinger 1937). These 'sudden ionospheric



<http://www.springer.com/978-0-7923-7196-0>

The Century of Space Science

Bleeker, J.A.; Geiss, J.; Huber, M. (Eds.)

2001, XLIX, 1846 p., Hardcover

ISBN: 978-0-7923-7196-0