14 THE AURORA AND THE AURORAL IONOSPHERE

H. C. Carlson, Jr., and A. Egeland

14.1 INTRODUCTION

NUMEROUS NATURALLY OCCURRING celestial phenomena have been observed and admired since the dawn of human history, but few have stirred human imagination, curiosity, and fear as much as the aurora. The aurora (also called the northern lights and polar lights) is certainly one of the most spectacular of nature's phenomena (Figure 14.1a–d).

When we search for records of the northern lights dating from more than 1,000 yr ago, we find that most come from the Mediterranean countries, that is, from low latitudes. Yet auroral displays are seen in that area only after unusually strong solar activity. The time lapse between such large auroral events can be 50–100 yr. Furthermore, an aurora seen at such low latitudes is significantly less dramatic and colorful than those at the higher latitudes of common auroral displays (see Section 14.3). Nonetheless, the ancient low-latitude events were dramatic enough to strike fear into the hearts of those who saw them.

At much higher latitudes, the aurora borealis (i.e., in the north) and aurora australis (in the south) routinely appear in the so-called auroral zones, far from most population centers. Even today, the southern auroral zone (roughly around Antarctica) is inhabited only intermittently. The northern auroral zone, which crosses Alaska, northern Canada, northern Scandinavia, and Siberia, has always been accessible to frontiersmen (hunters and fishermen) living in the polar region. More recently, the area under the northern lights has become permanently, although sparsely, populated.

In earliest historical times, inhabitants of Greenland and the Nordic countries interpreted the northern lights as omens from the gods portending disaster, as signs from deceased relatives, as signs of a battle among the gods, or as weather signs. From more recent and more scientific Scandinavian records (*The King's Mirror,* written about 1230 A.D.), it appears that the regions of auroral activity have shifted significantly during the past 1,000 yr. For those interested in the history of the aurora, monographs by Brekke and Egeland (1994) and Eather (1980) are available.

Those who appreciate the beauty of nature may find nothing compara-

FIG. 14.1a. Auroral forms seen from the ground during the polar night.

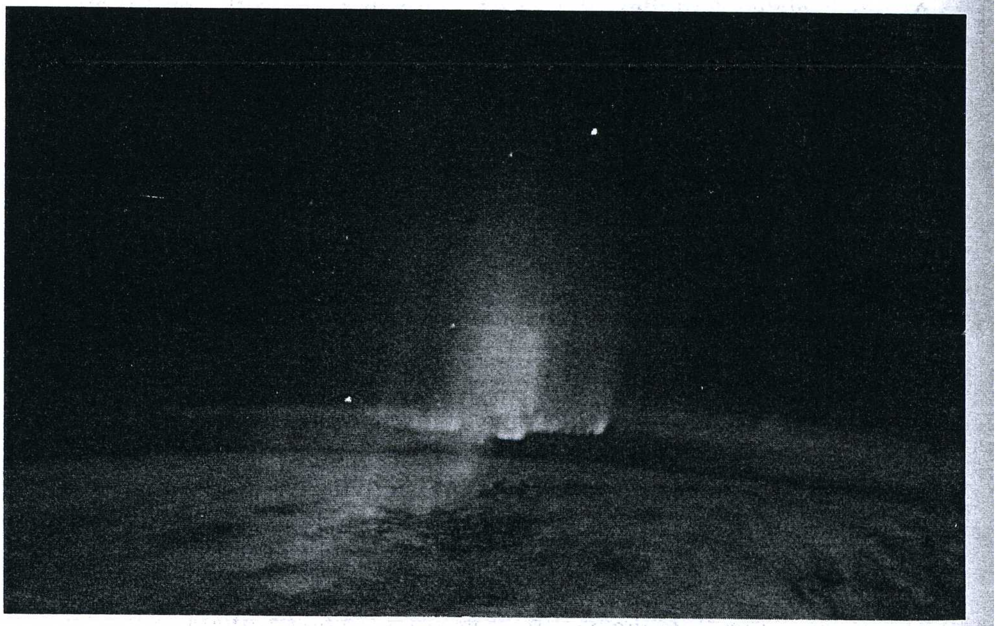

FIG. 14.1b. View from space of the aurora from low earth orbit (~250 km) of the Space Shuttle *Discovery*, 29 April 1991. (Courtesy of NASA.)

FIG. 14.1c. Aurora seen from a Defense Meteorological Satellite Program (DMSP) satellite.

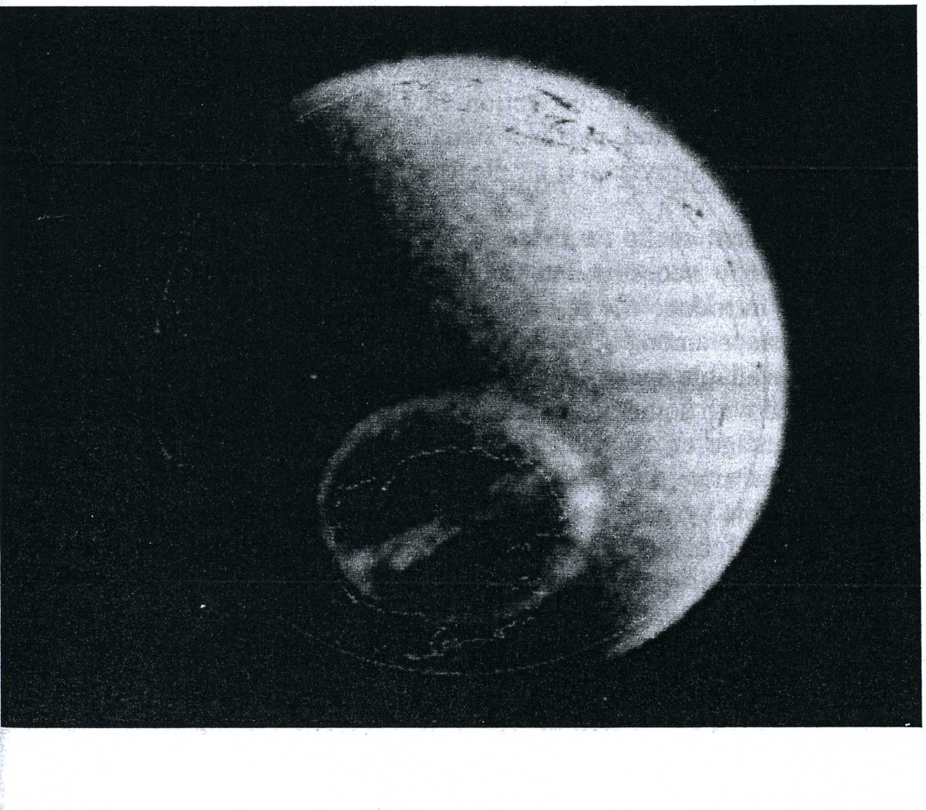

FIG. 14.1d. Aurora seen from the *Dynamics Explorer DE-1* in high earth orbit near $4R_E$ during an unusual event where the unusually circular oval has an extra transpolar sun-aligned crossbar, resembling the greek letter θ, giving this the name of "theta aurora." (From Frank et al., 1986.)

ble to a night with a magnificent auroral display. It is just as beautiful to watch today as it was in the earliest days of human history. This was given poetic expression by Tromholt (1885) in his book *Under the Rays of the Aurora Borealis:*

> Lovely celestial display! Before your fascinating mysterious play, in which enigmatic forces of Nature flood the heavens with light and color throughout the long Polar night, the golden sunsets of the Pacific Ocean, the gorgeous flora of the Tropics, the resplendent lustre of gems of Golconda, must pale. Lovely celestial display!

Less poetic, but fascinating in a different way, is the recognition that an aurora is the optical manifestation of auroral-particle precipitation and its interaction with atmospheric constituents (see Section 14.2). Auroral emissions are produced by particles, originating from the sun and the earth's atmosphere, that collide with the earth's atmosphere along streamlines modulated by electric and magnetic fields in the magnetosphere and ionosphere. The size and form of the aurora thereby reflect the forces acting on these auroral particles as they journey from their source to the earth's upper atmosphere (see Section 14.2). Auroral morphology, the study of the occurrence of the aurora in space and time, is described in Section 14.3, as is the electrodynamics of polar-cap arcs.

The auroral-substorm concept is discussed in Section 14.4. Sections 14.5 and 14.6, respectively, discuss the auroral ionosphere and its effect on radio waves. The basics of thermal balance and energy balance, plasma convection, and thermospheric responses controlled by geomagnetic activity, the interplanetary magnetic field (IMF), and local magnetic time are presented in Section 14.7. The auroral boundaries, as defined by optical and particle signatures, as well as current- and plasma-convection-reversal (i.e., electric-field) signatures, are discussed in Section 14.8.

This treatment seeks to provide an introduction to the terminology and morphology necessary to read the extensive literature on these subjects, to introduce the relevant physical processes, and to describe the relationships among upper-atmospheric boundaries. Within the constraints of available space, we trace the development of some key concepts, spiced with some of the unsolved challenges of auroral physics.

The advantage of using the aurora as a monitor of those near-earth processes that arise through the link to the magnetosphere, rather than any conceivable system of in situ measurements, is that the size of the auroral oval differs in scale from the magnetosphere by perhaps a factor of 10^6. The high spatial and temporal resolution available through ground-based observations provides another advantage of studying the aurora from below (see Section 14.8).

The first International Polar Year (1882–3) can be regarded as marking the beginning of modern auroral research. The driving force behind the effort was Kristian Birkeland, the great auroral pioneer (Birkeland, 1908, 1913). In his day, only the simplest ground instruments were available for auroral investigations. Today, auroral research is conducted mainly through the use of sophisticated instruments on board rockets and satellites, as well as advanced balloon and ground-based equipment. Even artificial aurora have been produced in the earth's atmosphere (e.g., Winckler, 1980). Some of the mysteries of the northern lights have been solved, partly or fully, but new problems have appeared, and the study of the auroras continues to engross many scientists.

14.2 AURORAL-PARTICLE PRECIPITATION: THE AURORAL SPECTRUM

The optical spectrum of an aurora consists of a great number of spectral lines and bands – from ultraviolet to infrared wavelengths. The auroral radiation is emitted by atmospheric constituents that are excited by precipitating particles. Figure 14.2 shows parts of the optical spectrum of an aurora. These emissions are primarily due to a two-step process in which precipitating energetic auroral particles (electrons and ions) collide with the atoms and molecules of the earth's upper atmosphere, converting their kinetic energy, in part, into energy stored in the chemically excited states of atmospheric species; the chemically excited states relax, giving off photons of wavelengths determined by the energy transitions in the relaxation processes. We shall summarize some of the main characteristics of the energetic particles and the auroral optical emissions they produce.

14.2.1 Scattering and Absorption of Auroral Particles

The primary auroral particles, populations of electrons and ions with energies from less than 100 eV up to small multiples of 100 keV, can be measured directly by the use of instrumented rockets and satellites. Some of these will precipitate into the atmosphere, causing atmospheric excitation and ionization, as discussed in Chapter 7. Near the earth, such particles are found mainly above 55° magnetic latitude. At greater distances above the earth, they have their sources in the plasma sheet of the geomagnetic tail and in the polar-cusp region on the dayside of the magnetosphere, as described in Chapter 10. As the high-energy tail of the energy spectrum (>30 keV for electrons; >1 MeV for protons) is not important for the auroral emissions, it will not be discussed further here.

The rate of precipitation of auroral particles into the upper atmosphere is schematically illustrated in Figure 14.3a. The dots represent mainly the higher-energy (>20 keV) auroral particles, and the triangles

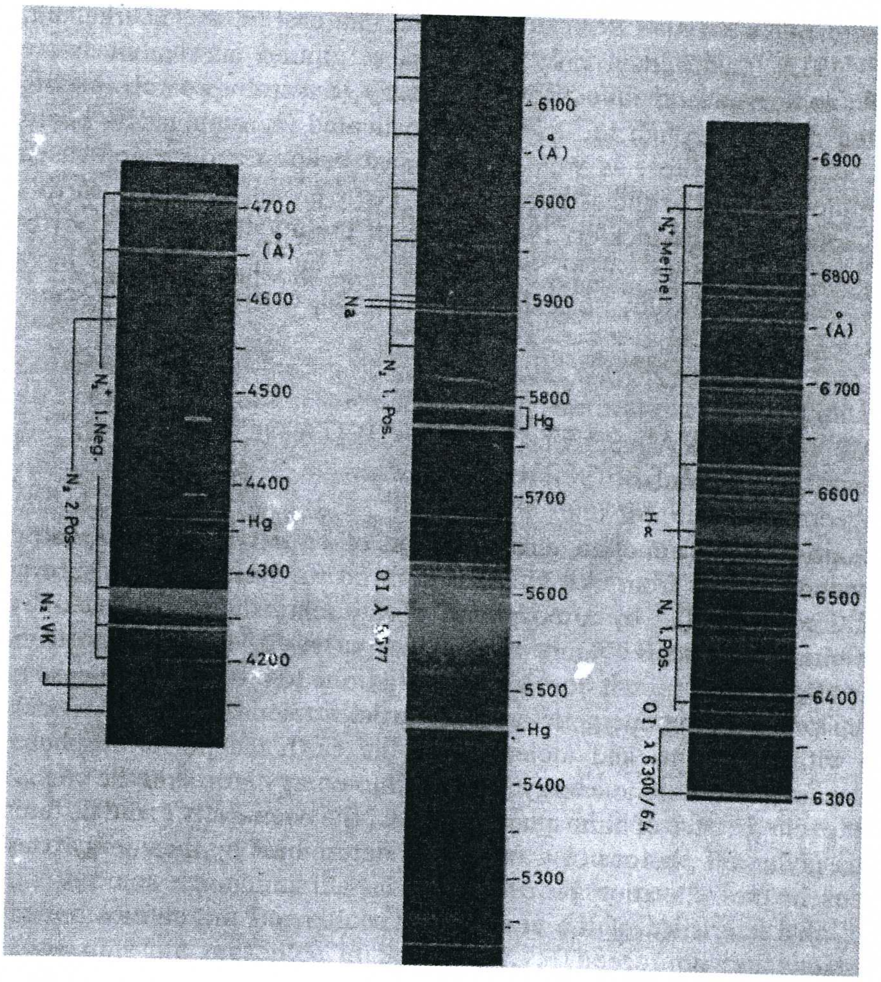

FIG. 14.2. Selected parts of the auroral spectrum in the visible range. (For details, see Vallance Jones, 1974.)

represent the medium-energy (~0.5–20 keV) particles responsible for the visual aurora. The stars mark the particles (<1 keV) that enter the magnetosphere through the polar cusp, causing the dayside oval aurora. Interpretation of Figure 14.3a, dating from 1971, illustrates the way in which straightforward ideas can advance our understanding of auroras. The most energetic particles lie on a circle of constant latitude, as would trapped particles leaking out of a loss cone as they drifted in a longitudinal ring current (see Chapter 10). The medium-energy particles lie on a circle tipped back away from the sun, as would particles accelerated downward along the earth's higher-latitude magnetic-field lines draped upward and downwind from the solar source of the solar wind. The low-energy particles are confined to the footprint of those midday magnetic-field lines that plausibly could funnel solar-wind particles directly into the earth's upper atmosphere with minimal acceleration. Far more extensive observations show overlapping of principal zones, a more gradual transition from one to another, and a much more sophisticated framework for relating particle populations to boundaries (e.g.,

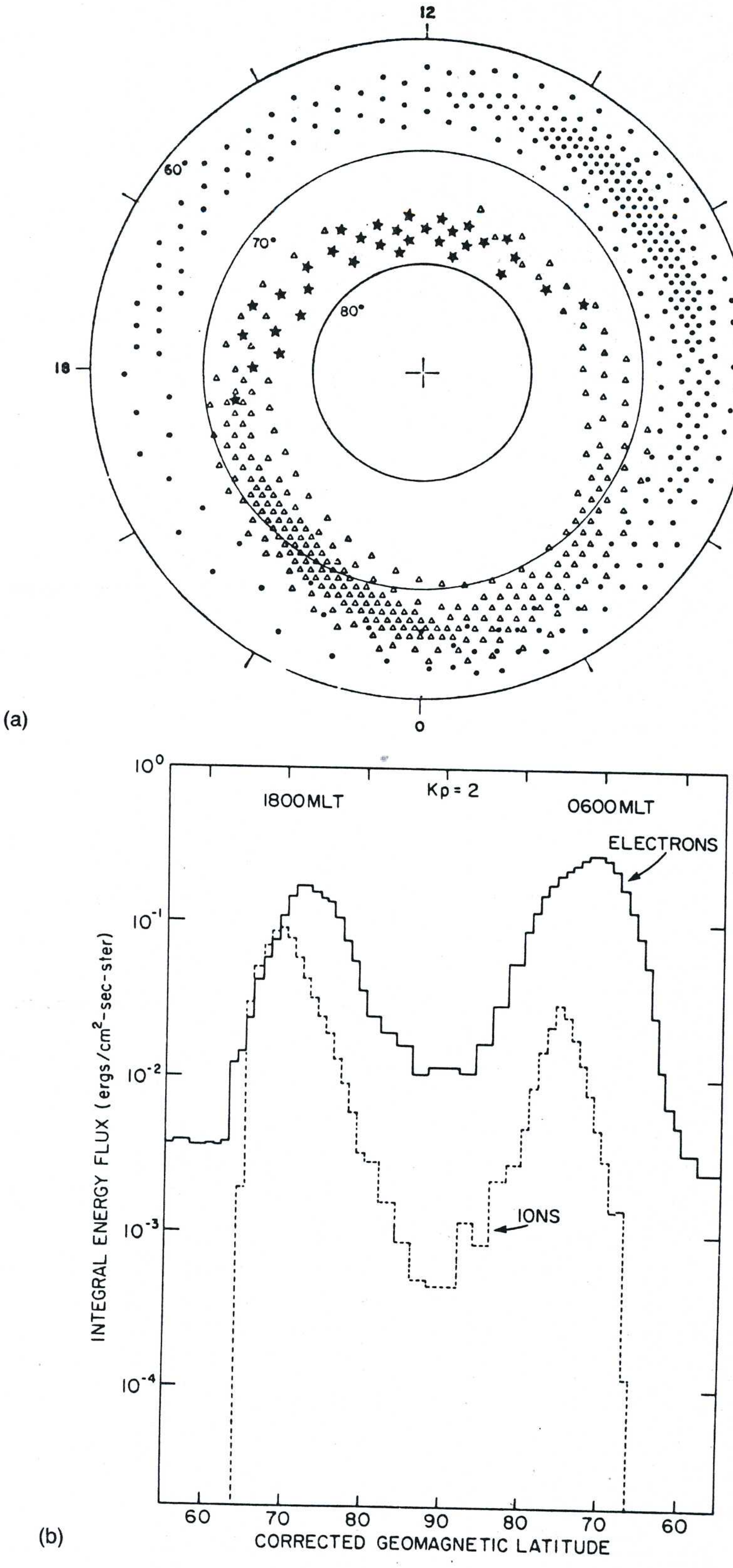

FIG. 14.3. (a) Idealized representation of a three-zone auroral-particle precipitation pattern. The auroral-oval (medium-energy) precipitation (splash type) is represented by the triangles, the auroral-zone (high-energy) precipitation (drizzle type) by the dots, and the polar-cusp (low-energy) precipitation 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. (From Hartz, 1971.) (b) Integrated energy flux into the auroral ionosphere across the dawn–dusk plane as a function of geomagnetic latitude for electrons and protons.

Figure 9.18). The strong asymmetry in the location of the aurora on the dayside and nightside of the earth gives the impression of an oval-shaped band around the polar regions. This band, within which auroras are common, is referred to as the auroral oval and is discussed fully in Section 14.3.3. Precipitating ions producing auroras show a dawn–dusk asymmetry, displaced toward dusk with respect to auroral electrons, as illustrated in Figure 14.3b.

In the high-latitude region on the dayside (i.e., between 70° and 80° Λ, where Λ is magnetic latitude), the characteristics of the energetic particles are similar to those of the magnetosheath; that is, the average energies are well below 1 keV. Both electrons and protons from the magnetosheath penetrate down to the atmosphere in the cusp/cleft region, where they produce dayside auroras. The precipitation occurs in a narrow region at about 78° invariant latitude stretching from late morning to early evening magnetic local time (Section 14.3.3), referred to as a cusp or a cleft.

The proton and electron motions in near-earth space are governed by the three adiabatic invariants introduced in Chapters 2 and 10. A fraction of these particles will have their mirror points in the atmosphere, below about 200 km in altitude. Particles penetrating the atmosphere collide with atmospheric atoms and molecules and gradually lose their energy to the neutrals. The energy loss rate for a subrelativistic electron is given by the formula

$$-\frac{dW_e}{dx} = -\frac{dW_e}{Q\,ds} = \frac{2\pi e^4 Z A_0}{W_e A}\,\ell n(W_e/I) \tag{14.1}$$

where dx ($=Q\,ds$) is the atmospheric depth, given in grams per square centimeter, A_0 is Avogadro's number, Z is the average atomic number of atmospheric atoms of atomic weight A, I is the average energy loss per ionization, Q is the mass density of scattering atoms, and ds is a differential distance along the electron trajectory. In fact, the main sink for fast, charged particles in the magnetosphere is the atmosphere (Rees, 1989).

Precipitated charged particles in the ionosphere are subject to inelastic and elastic collisions with the atmospheric constituents. They lose their energy gradually by (1) ionizing and exciting the upper atmosphere, (2) dissociating atmospheric molecules, (3) heating the upper atmosphere, and (4) producing bremsstrahlung x-rays. (This latter process, which is discussed in Chapter 7, is negligible for low-energy particles and will not be discussed further here.)

Thus, energy deposited in the upper atmosphere by precipitating particles is, in part, used to produce optical emissions (i.e., the aurora). As discussed in Chapter 7, a downcoming beam of monoenergetic particles entering the atmosphere will penetrate to about the altitude of "unity optical depth" for the particles. Most of the absorption will be within a neutral scale height of this altitude. In a realistic situation, one must sum the collision cross section s_{ij} over the different absorption processes available for each of the j atmospheric constituents present. These are weighted by the relative cross sections for the i processes and the relative number densities of the j constituents present. The cross section has an important energy dependence, with higher-energy particles penetrating more deeply. The approximate penetration depths for various proton and electron energies are shown in Figure 7.4. Because the particle penetration is governed by statistical processes, the actual penetration depths are not identical even for two particles with identical initial conditions. The values given in Figure 7.4 should therefore be considered as the average height where most of the energy is absorbed for vertical incidence. A detailed discussion of the problems of particle scattering and absorption is given in the monograph by Rees (1989).

Experimental data show that fast electrons and protons produce about one ion pair (ion–electron) per 36 eV of their initial energy. This can be written symbolically for electrons and protons, respectively, in the following equations:

$$X+e \rightarrow X^{+}+e_{n}+(e-36\ \mathrm{eV}) \tag{14.2a}$$

$$X+\mathrm{H}^{+} \rightarrow X^{+}+e_{n}+(\mathrm{H}^{+}-36\ \mathrm{eV}) \tag{14.2b}$$

where X is an atmospheric constituent, and e_n is a thermal electron in the ambient electron gas, rather than an energetic auroral electron. Because the ionization potential of the atoms and molecules, on average, is about 15 eV, about 40 percent of the energy goes into ionization, whereas about 60 percent goes into the motion of the product electron, which subsequently thermalizes.

14.2.2 The Auroral Spectrum

As discussed in Section 14.2.1, the kinetic energy of the auroral particles can be deposited – through collisions – into the translational, vibrational, and rotational energies of atoms and molecules, expended in impact-excitation of bound electrons from their ground state to a higher level, or spent in electron ionization by impact (Vallance Jones, 1974). The distribution of energy among these initial options sets the stage for the energy subsequently liberated in the ultraviolet (UV), visible, and infrared (IR) emissions of auroras. Thus, the auroral emissions contain atomic lines and molecular-band spectra of the primary constituents of the upper atmosphere (Figure 14.2), plus some important emissions from minor species (e.g., NO, He, and CO_2, which are efficient in the cooling of the thermosphere via strong IR emissions). The auroral emissions can therefore be considered as the "fingerprints" of the atmospheric constituents.

This section provides a brief description of the mechanisms that account for the auroral spectrum. For more thorough reviews of auroral