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.

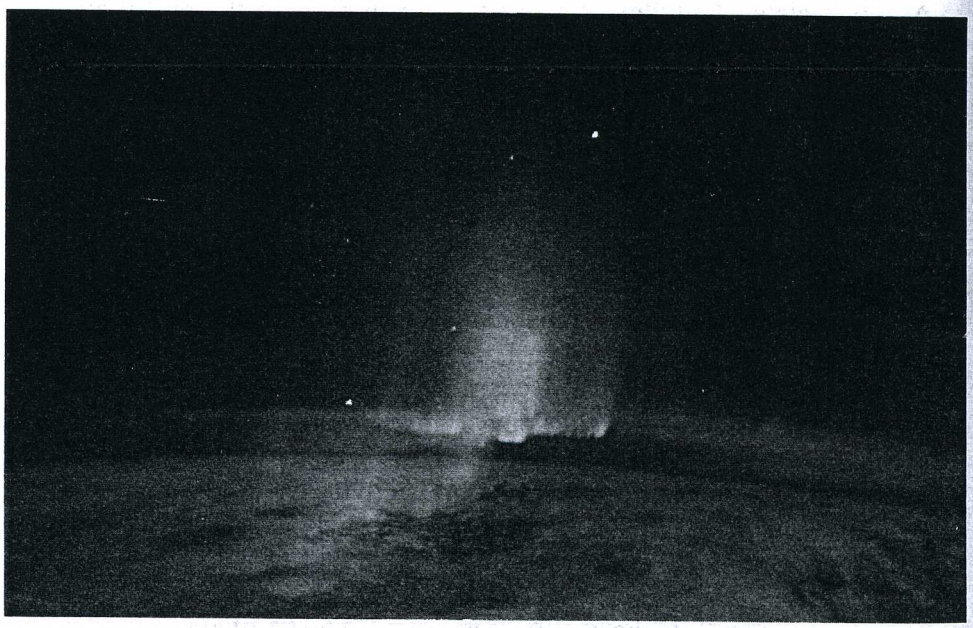


FIG. 14.1b. View from space of the aurora from low earth orbit (~250 km) of the Space Shuttle *Discovery,* 29 April 1991. (Courtesy of NASA.)

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

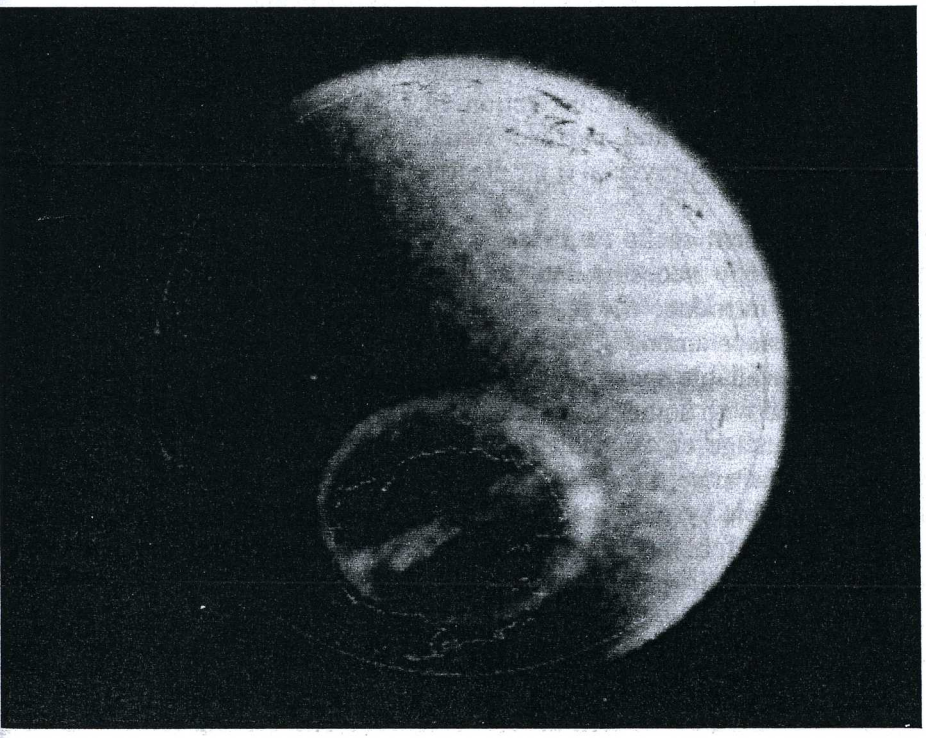


FIG. 14.1d. Aurora seen from the *Dynamics Explorer DE-1* in high earth orbit near $4R_E$ during an unusual event where the unusually circular oval has an extra transpolar sun-aligned crossbar, resembling the greek letter θ, giving this the name of "theta aurora." (From Frank et al., 1986.)

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

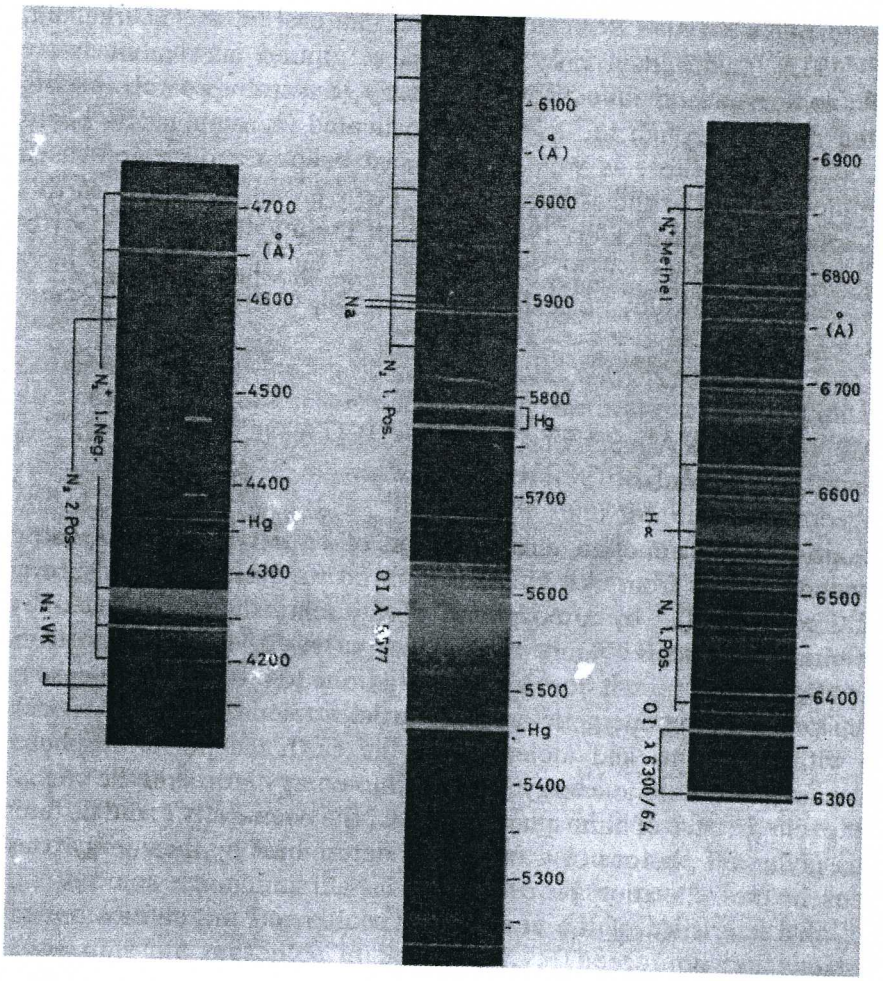
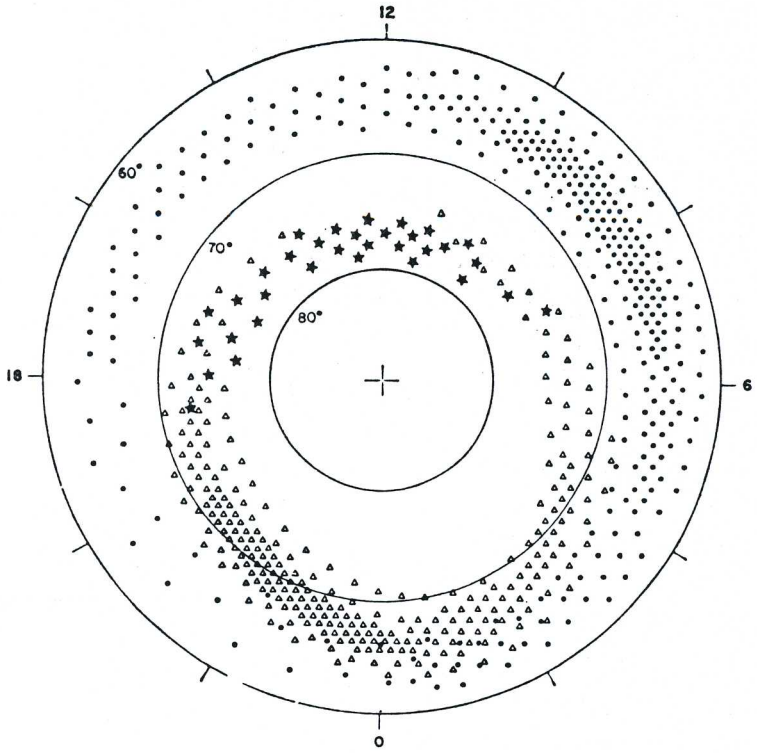
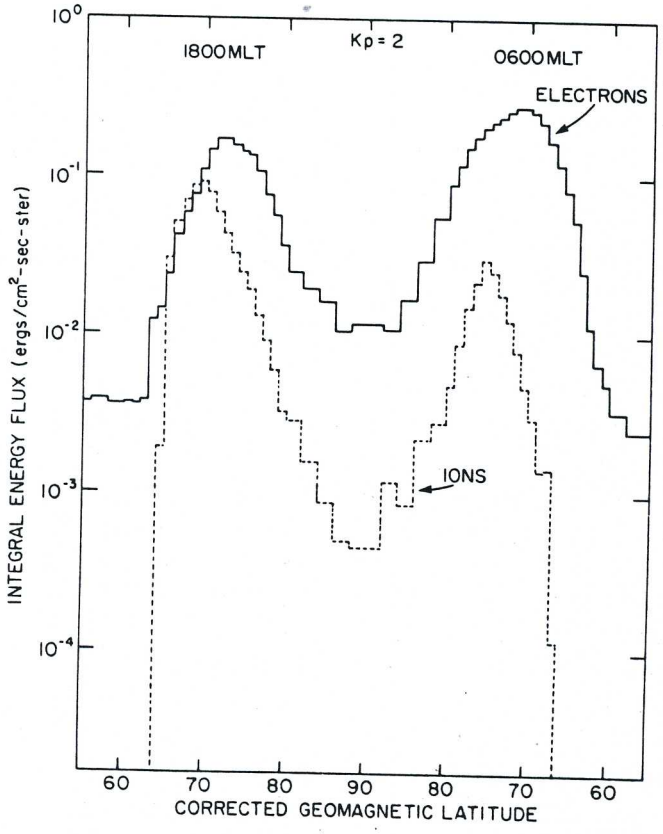


FIG. 14.2. Selected parts of the auroral spectrum in the visible range. (For details, see Vallance Jones, 1974.)

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.,



(a)



(b)

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:

$$X + e \rightarrow X^+ + e_n + (e - 36\ \text{eV}) \tag{14.2a}$$

$$X + \text{H}^+ \rightarrow X^+ + e_n + (\text{H}^+ - 36\ \text{eV}) \tag{14.2b}$$

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

spectroscopy, the reader is referred to Vallance Jones (1974) and Omholt (1971).

Photons may be emitted spontaneously as the excited species relax to lower energy levels and/or ground states, or further chemical reactions may take place as the energy cascades, with photons emitted along the way. The optical-emission wavelength λ, in nanometers, is related to the released energy E, in kiloelectron volts, by $E = 1.240/\lambda$.

Some energy reactions and transfer mechanisms important in auroral physics are the following:

Electron Impact

$$e + N \rightarrow N^* + e' \tag{14.3a}$$

Energy Transfer

$$X^* + N \rightarrow X + N^* \tag{14.3b}$$

Chemiluminescence Reaction

$$M + XN \rightarrow MX^* + N \tag{14.3c}$$

Cascading

$$N^{**} \rightarrow N^* + h\nu \tag{14.3d}$$

Here, N, X, and M are atmosphere neutrals, asterisks represent different levels of excitation, and $h\nu$ is an emitted photon. In each collision, the primary electron (e) is replaced by secondary electrons (e') of lower energy [equation (14.2)], which can be responsible for further excitation. The energy and altitude dependences of the absorption of auroral particles will not be discussed further here (see Chapter 7).

An additional class of energy-transfer processes involves ionization of neutrals. When ionization occurs, the ion may be left in an excited state and may radiate light immediately. The auroral spectrum can be used to detect the ionization process. This is the case with the first-negative-band system. These bands are excited directly from the ground state of N_2. Symbolically, the process involved is

$$N_2 + (e \text{ or } H^+) \rightarrow (N_2^+)^* + (e' \text{ or } H^{+\prime}) + e_n \tag{14.4a}$$

where $e' = e - 36$ eV, $H^{+\prime} = H^+ - 36$ eV. The asterisk indicates that the molecular ion is created in an excited state. This process is followed by radiation of the first negative bands:

$$(N_2^+)^* \rightarrow N_2^+ + (391.4\text{- and } 427.8\text{-nm}) \text{ aurora} \tag{14.4b}$$

The probability of process (14.4a) followed by (14.4b), compared with the probability of all possible ionization processes, is nearly independent

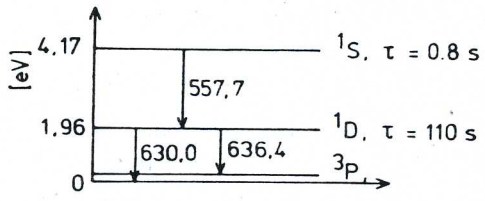


FIG. 14.4. Energy level of the oxygen atom. The terms as well as the radiative half-lives of the ^1S and ^1D levels are indicated, and the wavelengths of photons emitted in transitions between energy levels are shown.

of energy for electrons between 0.5 and 20 keV (not for electrons <0.5 keV). The intensity of these auroral bands can therefore be used to determine the net downward electron energy. Quantitative considerations show that about 25 ion pairs are produced for each photon emitted in the λ 391.4-nm band. The corresponding figure for the λ 427.8-nm band is 75 ion pairs per photon. For protons, the situation is somewhat more complicated (as will be discussed later).

The brightest visible feature of the aurora, the "green line" at 557.7 nm, is due to the transition of an electron from the ^1S excited state to the ^1D state of atomic oxygen, as illustrated in Figure 14.4. Another commonly observed line – particularly in the polar cusp and cap – is the "red line" at 630 nm as the ^1D state relaxes to the ground state. (Fine structures of the electron shells in the ground state allow 636.4-nm emission as well.) If the O(^1S)-state electron gives up its full 4 eV in a single step, instead of two nominal 2-eV steps (^1S to ^1D, and then ^1D to ^3P), it emits a photon at 297.2 nm, that is, about half the wavelength of the emissions from the smaller energy steps (Figure 14.4).

The forbidden oxygen line at 557.7 nm and the red doublet at 630 and 636.4 nm can be excited by the following process (where e' has less energy than e):

$$O(^3P) + e \rightarrow O(^1S) + e' \tag{14.5a}$$

followed by

$$O(^1S) \rightarrow O(^1D) + h\nu \qquad (557.7 \text{ nm}) \tag{14.5b}$$

or

$$O(^1S) \rightarrow O(^3P) + h\nu \qquad (297.2 \text{ nm})$$

For the red doublet, we have

$$O(^3P) + e \rightarrow O(^1D) + e' \tag{14.5c}$$

followed by

$$O(^1D) \rightarrow O(^3P) + h\nu \qquad (630/636.4 \text{ nm}) \tag{14.5d}$$

where ^1S and ^1D have total electron spin $s=0$, while ^3P has spin $s=1$ (Vallance Jones, 1974). These excitations have a small probability for higher energies. There are also great numbers of permitted oxygen and nitrogen lines from higher excited states. Most of the permitted transitions observed in O and N have excitation potentials of about 10–13 eV,

and in O^+ and N^+, 20–30 eV above the ground state of the neutrals. The sodium doublet at 589.0 and 589.6 nm (^2S–^2P) is occasionally observed in auroras. Also, the helium line at 587.6 nm (^3P–^3D) has been observed on rare occasions.

In sunlit auroras (Størmer, 1955), the spectral distribution is somewhat different from that of ordinary auroras, with enhanced intensity of the first negative bands. This is due to resonance scattering of sunlight by N_2^+ produced by the primary particles.

The statistical residence time in an excited state before emission is determined by the *Einstein transition probability*. Governed by quantum-mechanical selection rules, *allowed* or *permitted transitions* occur very rapidly (in times on the order of 10^{-7} s). Transitions that violate these selection rules are called *forbidden transitions*. They do occur, but only after a much longer time; for example, the transition for $O(^1S)$ shown in equation (14.5b) occurs in about 0.8 s, and that for $O(^1D)$ shown in equation (14.5d) requires about 110 s. The latter is so slow that much below 200 km, an $O(^1D)$ atom is likely to suffer a collision that will knock it out of the $O(^1D)$ state before it has a chance to emit. Thus, the 630-nm (OI, or neutral oxygen) emission is expected to peak above 200 km, even though the excitation to the $O(^1D)$ state is expected to peak near an altitude of 100 km. (A roman numeral following an atomic symbol refers to the ionization state, with I designating un-ionized, II designating singly ionized, etc.) This *quenching* by collision with other atmospheric constituents substantially reduces the number of 630-nm photons emitted below the number of $O(^1D)$ states excited. The 1.96 eV used by the oxygen atom in exciting the $O(^1D)$ state goes into vibrationally exciting the local atmospheric gases, instead of dissipating into the atmosphere (Figure 14.4).

Molecules also can store energy in the form of vibrational energy (along the molecular axis) and/or rotational energy (along a transverse axis). Because of the close spacing of vibrational-energy levels, auroral emissions from molecules have bandwidths of nanometers, whereas atomic-line bandwidths are on the order of 0.1 nm or less.

A few definitions and categorizations should help to put these ideas of auroral emission processes and rates in perspective. The photon-emission intensity I ($cm^{-3} \cdot s^{-1}$) is

$$I = N^* A$$

where N^* (cm^{-3}) is the density of the excited-state emitting molecules and A (s^{-1}) is the Einstein coefficient. The density of excited molecules is the ratio P/L: the excitation rate per unit volume P ($cm^{-3} \cdot s^{-1}$) divided by the loss rate L (s^{-1}). L must include all collisional deactivation (or quenching) processes of the excited state by each quenching species, weighted by its rate coefficient ($molecule^{-1} \cdot cm^{-3} \cdot s^{-1}$).

An important part of the optical spectrum in the visible region is shown in Figure 14.2, and the average intensities of the most prominent

TABLE 14.1. Typical Relative Auroral Intensities

Molecule	Atomic Emission (nm)	Relative Intensity	
		Night	Day
O	^1D–^1S 557.7	1	1
	^3P–^1D 630/636.4	0.1–0.5	1–100
N	^4S–^2D 519.9	0.01	
	^2D–^2P 1,040	1	
O_2	b-X atmospheric (0–0) bands	2	
	a-X IR (0–0) bands	10^2–10^3	
O_2^+	B-A, first negative bands	0.4–1	
N_2	B-A, first positive bands	5–20	
	A-X, Vegard-Kaplan bands	1	
	a-X, Lyman-Birge-Hopfield band	0.5–1	
N_2^+	B-X, first negative bands	0.5–1	0.3
	A-X, Meinel bands	7–20	

Note: Several bands of O_2, O_2^+, N_2, and N_2^+ radiation, each band extending over a range of wavelengths and containing emissions at discrete wavelengths, contribute to the auroral emission spectrum. These bands are listed in standard auroral spectroscopy notation in column 2 of this table. For an explanation of the nomenclature and other details, see Chamberlain (1961).

auroral emissions relative to the green line are listed in Table 14.1 (Omholt, 1971; Vallance Jones, 1974). The auroral intensities are highly variable, and the values listed in the table are estimated averages. Because of the difficulties inherent in accurate optical-intensity measurements, the intensities quoted have relatively large uncertainties.

Some weak but important hydrogen lines, first discovered by Vegard in 1939, exist in the auroral spectrum (Vegard, 1939). The emissions H_α at 656.3 nm and H_β at 486.1 nm result from excited hydrogen atoms that are produced when energetic protons (H^+) bombard the atmosphere. The excitation mechanism, illustrated in Figure 14.5, can be written

$$X + H^+ \rightarrow X^+ + H^* \tag{14.6a}$$

followed by the auroral emission

$$H^* \rightarrow H + h\nu \qquad \text{(hydrogen/proton aurora)} \tag{14.6b}$$

The hydrogen atom again collides:

$$H + X \rightarrow H^+ + X + e_n \tag{14.6c}$$

and process (14.6a) can start again.

The hydrogen atom formed in process (14.6a) has almost the same velocity and direction as the original proton. The fast atom – after process (14.6b) – collides with an atmospheric particle (X) and may be reionized (14.6c) or excited. The latter is more likely to occur at low particle energies. An average particle goes through a great number of processes of electron capture and loss before it has lost its energy and is brought to rest in the upper atmosphere.

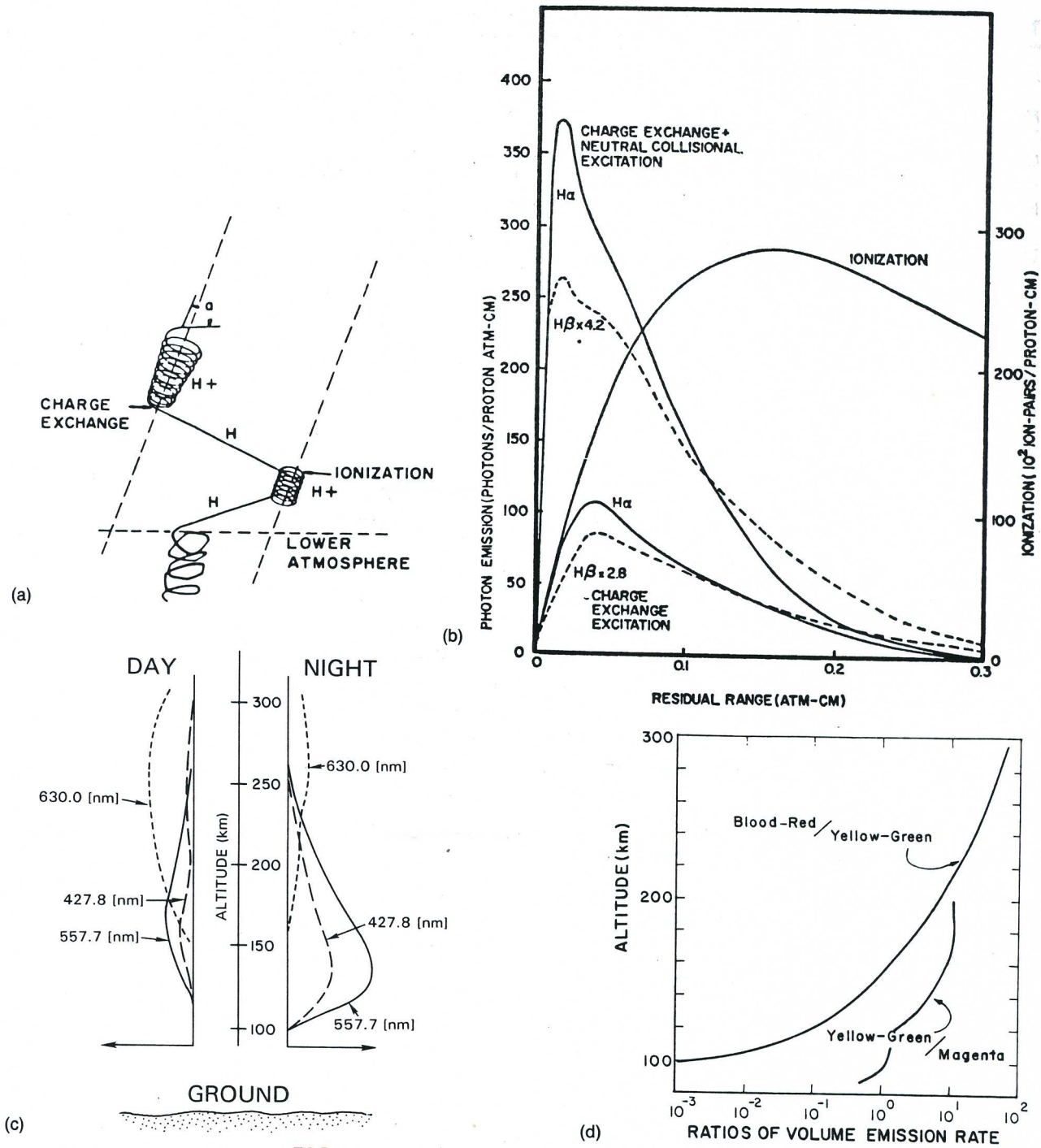


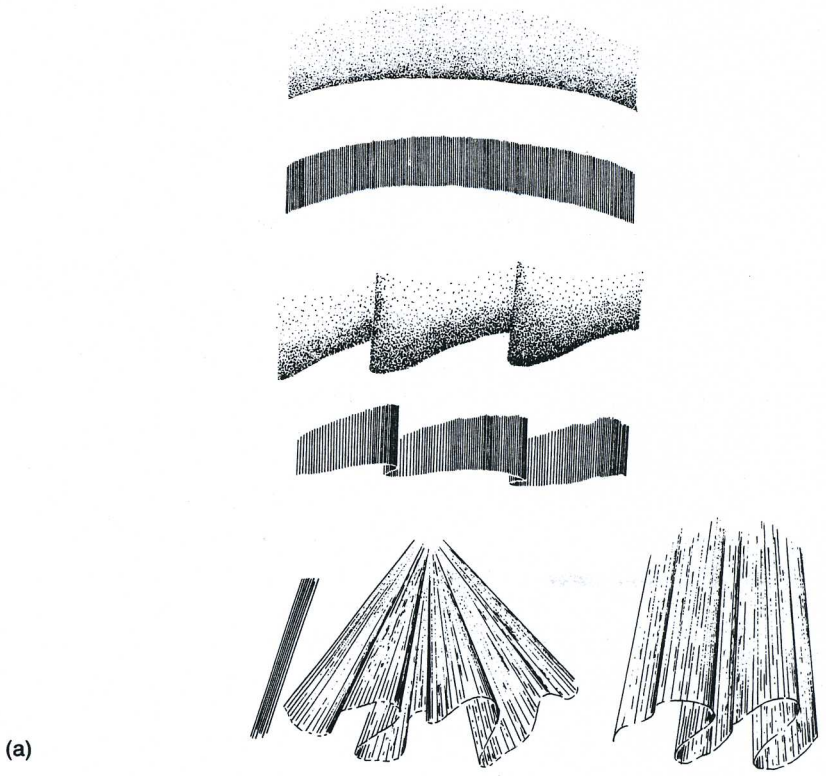
FIG. 14.5. (a) Typical path of protons entering the atmosphere. The figure is schematic and not to scale. (b) Photon-emission curves for H_α and H_β and ionization curves for protons in the air. (From Omholt, 1971.) (c) Main differences between dayside and nightside auroras versus altitude. (d) Ratios of emission intensities in various spectral ranges versus altitude. On the basis of observations of spectral ratios, the average auroral height can be determined.

As a fast particle (proton/atom) penetrates the atmosphere and slows down, it spends a large fraction of its time as a neutral atom. Therefore, its probability for emission of light increases until its energy drops below the excitation level. Figure 14.5b shows the photon emission in H_α and H_β per proton as a function of residual range. When a photon is emitted, it has a Doppler displacement that depends on the velocity of the emitting hydrogen atom and the angle between the velocity vector and the direction of the photon (Omholt, 1971). The first realistic estimates of the auroral-particle energies (made before the space age) were based on such Doppler profiles.

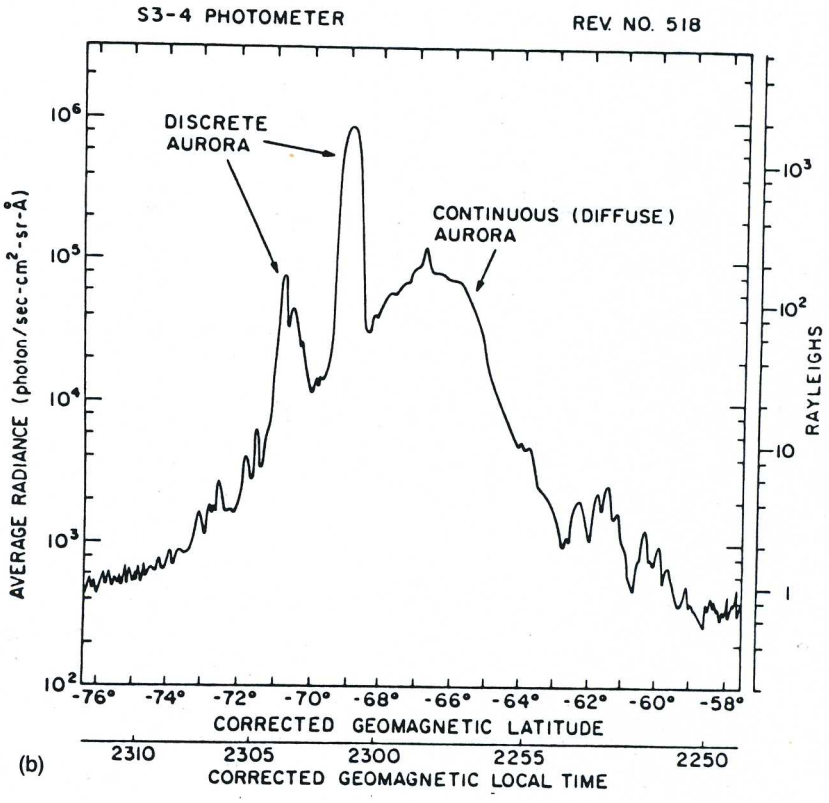
The protons arrive in helical paths (Figures 14.5a), spiraling around the magnetic lines of force with a given pitch angle. With a known distribution of an ensemble of protons of initial energies and pitch angles, we can estimate the total emission from hydrogen as a function of height, as well as the Doppler profile of the hydrogen light for any direction of observation. These computations are not difficult in principle, but somewhat cumbersome in practice (Omholt, 1971). As a result of charge exchange, the proton auroras are more defocused (i.e., diffuse) than the incident proton precipitation. These auroras occur in an oval displaced duskward of the electron oval (Figure 14.3b) and have a response time different from that of electron auroras (Section 14.4).

The spectral characteristics listed earlier are more generally described in terms of the colors observed in auroral forms. The 630-nm (OI) emission, which is created by auroral "soft" primaries, is seen as the red aurora; this forms the diffuse background radiation in which the discrete arcs are embedded (Figure 14.6b). "Blood-red" auroras are produced by low-energy electrons ($\ll 1$ keV). Reports of "red lower borders" (usually fast-moving) indicate the presence of particles with energies greater than 10 keV. The majority of the auroras are yellow-green, but sometimes appear gray. The gray appearance refers to observations below the color threshold of the eye and corresponds to low particle flux. Figure 14.5d shows how the ratios of the main visual emissions change relative to one another as a function of altitude. Specifically, blood-red auroras dominate in the altitude region above 200 km, whereas a magenta color predominates below approximately 100 km. Blood-red is the (OI) emission at 630–636.4 nm, yellow-green is the (OI) emission at 557.7 nm, and magenta is a combination of N^2 and O_2^+ emissions near 600 nm and N_2^+ first-negative-band emissions in the blue end of the spectrum.

The 630-nm (OI) emission can also be excited by a process called thermal excitation. The ambient electron gas will have a thermal or Maxwellian distribution of energies, with a population decreasing exponentially with increasing energy. The fraction of the electron population above a fixed energy is obviously strongly dependent on the electron temperature. For electron temperatures much above 3,000K, there may be enough electrons in the high-energy *tail* of the thermal distribution to excite detectable emission from O atoms. This follows because the



(a)



(b)

FIG. 14.6. (a) Illustrations of typical forms of the northern lights. Top to bottom: homogeneous arc; arc with ray structure; homogeneous band; band with ray structure. The three lower forms are (left to right) rays, corona, and draperies. (b) Downward-looking photometric measurements of discrete and diffuse (continuous) aurora performed by the S^3-4 satellite. Intensity plotted versus satellite position in latitude and local time.

$O(^1D)$ level is only 1.96 eV above the ground state, as shown in Figure 14.4. Thus, thermal electrons can produce excitation of 630-nm (OI) auroras. Stable auroral red (SAR) arcs are formed on the equatorward edge of the auroral oval by this excitation process, provided that increased electron-gas heating from above and/or decreased electron-gas cooling by the ionosphere below allow the electron temperatures in the F region to rise to well over 4,000K (e.g., Kozyra et al., 1990).

14.2.3 Auroral Intensities

An aurora appears as a luminous cloud, having an apparent surface brightness. Absorption within the visible spectrum is negligible. Hence, the apparent surface brightness is proportional to the integrated emission per unit volume along the line of sight. The surface brightness is used to define the intensity of an aurora.

If the surface brightness I is measured in photons per square centimeter per second per steradian, then $4\pi I$ represents the total emission in photons per square centimeter per second integrated along the line of sight. This is defined as the intensity of an aurora. The unit adopted for $4\pi I$ is the rayleigh (R). One rayleigh is equal to an integrated emission rate of 10^6 photons per square centimeter per column per second (inclusion of "column" in the units refers to the unknown height of the column above the apparent source; it is included to show that this is a volume emission, not a true surface emission). The observed intensity of a particular auroral form depends on the direction of observation. A thin auroral layer covering a large part of the sky is most intense when viewed at low elevation angles.

The auroral intensity in rayleighs at a particular wavelength λ from an incoming electron/proton beam, with isotropic pitch-angle distributions assumed at all energies, can be evaluated from the following expression:

$$I(X(\lambda))[\mathrm{R}] = \frac{\pi}{10^6}\int_{E=0}^{\infty} j(E)\cdot P(E)\, dE \tag{14.7a}$$

where $j(E)$ is the differential electron/proton energy spectrum, and $P(E)$ is the total number of photons for gas X at wavelengths λ (i.e., for N_2^+ at 427.8 nm or for H_β at 486.1 nm). For protons, $P(E)$ depends strongly on the energy spectrum (see Problem 14.2), and H^+ above 20 keV contributes significantly to the output light. For the N_2^+ first negative bands, Rees and Roble (1986) have given the following formula:

$$4\pi I(427.8\ \mathrm{nm})[\mathrm{R}] = 213\left(\frac{E_{\mathrm{ave}}}{2}\right)^{0.0735}\cdot \epsilon \frac{\mathrm{erg}}{\mathrm{cm}^2\cdot \mathrm{s}} \tag{14.7b}$$

where ϵ is the energy flux, and E_{ave} is the average energy of incident auroral particles.

When classifying auroral intensities, the line used for reference is the

green oxygen line at 557.7 nm, which is dominant in the wavelength region near the maximum sensitivity of the human eye. Typical intensities of nightside auroral arcs and bands vary from one to a few tens of kilorayleighs. During an active period with a bright display, auroral intensity in or near the zenith may be several hundred kilorayleighs. Thus, as 1 kR near 550 nm is the visibility threshold for the dark-adapted naked eye, the optical aurora is a relatively weak but very dynamic optical phenomenon. The large variations in intensity are closely correlated with the net downward particle energy. In order for the northern lights to be visible to the eye, the particle energy input to the atmosphere must be about 1 erg·cm^{-2}·s^{-1} or approximately 10^{-3} W·m^{-2}. For a medium-strong northern light about 10 km wide and 1,000 km long, approximately 10^6 kW are needed, which is comparable to the power capacity of a large power plant. Because only about 1 percent of the particle input to the atmosphere is used to produce visible light, it is clear that there is an enormous quantity of energy deposited in the upper strata of the atmosphere during each auroral night.

Auroral Photography. On a clear, dark winter night, the northern lights may look bright to the dark-adapted naked eye. Still, it is not easy to take good pictures of the aurora with short exposures, though with a modern 35-mm camera with a small-focal-ratio lens, it is possible. With fast color film (e.g., 400 ASA), an exposure time of 1 s to a few seconds is normally needed, and therefore it is advisable to use a tripod for mounting the camera.

14.3 AURORAL DISTRIBUTION IN SPACE AND TIME

14.3.1 Auroral Forms

Looking at an auroral display, one normally sees a bewildering number of auroral forms and situations (Størmer, 1955). Each instantaneous auroral situation may be considered to be composed of various superimposed elementary auroral forms or structures that vary in space and time. For practical purposes, we have to consider only four such elementary forms, as illustrated in Figure 14.6: (1) the quiet homogeneous arc and band stretching along the magnetic east–west direction across the sky in a straight or curved line; (2) auroral rays and combinations of rays, which may vary considerably in length; (3) diffuse or irregular auroral clouds; (4) spirals and curls. Their intensities may vary over several orders of magnitude. For a more detailed description of the different auroral forms, the reader is referred to Størmer's excellent book *The Polar Aurora* (1955).

Another scientifically useful method of categorizing auroral forms is to identify them as discrete or diffuse forms (Figure 14.6b). However,

even diffuse auroras may contain some weak discrete structures. This has been shown by the use of very sensitive optical instruments. The *diffuse aurora* is so named because its weak, small striated structures were not observable in the satellite data where it was first documented as a separate form. It is associated in the evening with the region of greatest proton energy flux and maps magnetically just inside the boundary of stable trapping (see Chapter 10). This explains its more or less circular form around the magnetic pole. A diffuse aurora in the E region is produced by higher-energy particles bouncing from one hemisphere to another and losing energy to the atmosphere as they diffuse into the *loss cone* while drifting in longitude, electrons to the east and protons to the west.

14.3.2 Height Distribution of Auroras

Generally, the light from an aurora is proportional to the deposition of energy into the atmosphere by the primary particles. As discussed in Section 7.2.2, the height distribution is related to the energy and pitch-angle distribution of the precipitating particles, as well as the atmospheric composition. The energy distribution and flux of the particles can be studied directly only by rocket and satellite techniques. Still, auroral height measurements, as well as spectral ratios (Figure 14.5d), may provide useful information on representative energies and systematic variations in the energy spectrum.

Figure 14.7 shows the distribution of 12,330 auroral points measured by Størmer (1955). Størmer's measurements demonstrated that auroral arcs and bands, homogeneous as well as rayed, lie predominantly within the height interval 95–150 km, whereas isolated rays or bundles of rays can lie significantly higher. Long rays may stretch several hundred kilometers up into the atmosphere. As Figure 14.5c illustrates, the average height of dayside auroras is significantly higher (by 100–200 km) than that for the nighttime auroras.

The detailed height distributions for individual forms have been measured by both rockets and ground-based techniques. Auroral height and the height distribution for each form are related to the average energy and the energy distribution of the precipitating particles.

14.3.3 Auroral Locations, Auroral Morphology

In the literature up to the 1960s, auroral location was described in terms of the auroral zones, which represent the regions in which nighttime auroras occur in an average statistical sense. Within the auroral zones, which are centered roughly 23° from the geomagnetic poles and are approximately 10° wide, auroras are observed by the naked eye on more than 50 percent of clear nights, even during years of low solar

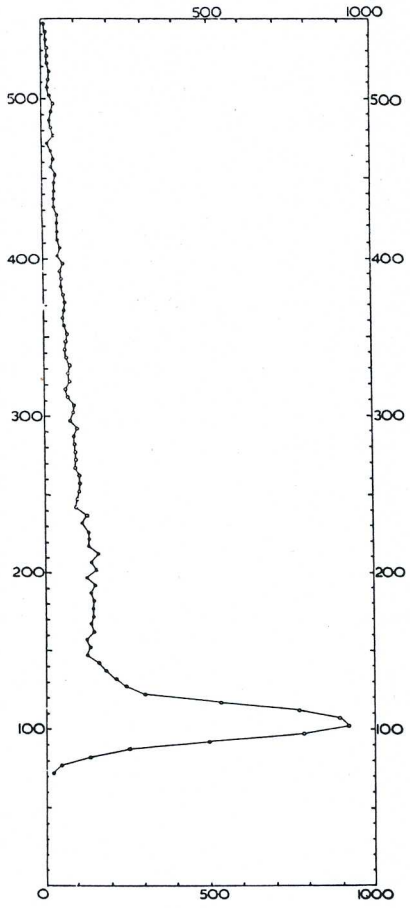


FIG. 14.7. Statistical distribution of 12,330 height measurements of the northern lights derived by Størmer and his colleagues. The vertical scale gives the height in kilometers, and the horizontal scale gives the number of measurements. Most of the northern lights are found between 90 and 150 km.

activity. The frequency of occurrence and the intensities of auroras are clearly correlated with the activity of the sun. This is borne out by the 27-day recurrence tendency, which correlates with the rotational period of the sun, as well as by the 11-yr variation that correlates with the sunspot cycle. However, as first pointed out by Størmer (1955), the auroral activity peaks 1–2 yr after a sunspot maximum. The auroral zones follow L = constant to a good first approximation. The instantaneous distribution of maximum auroral activity versus magnetic time and latitude was found by Feldstein and Starkov (1967) to be given by an oval-shaped belt called the auroral oval (Figure 14.8).

The auroral ovals (one for each hemisphere) are continuous bands centered near 67° magnetic latitude at magnetic midnight and near about 77° at magnetic noon during quiet periods and periods of moderate activity (Akasofu, 1968). (Examples of auroras, as seen from both space and ground, are shown in Figure 14.1.) The latitude extremes of the oval are the average boundaries of auroral luminosity, as seen by all-sky cameras during the International Geophysical Year (IGY) 1957–8. The locations of the auroral oval for quiet, moderate, and active conditions

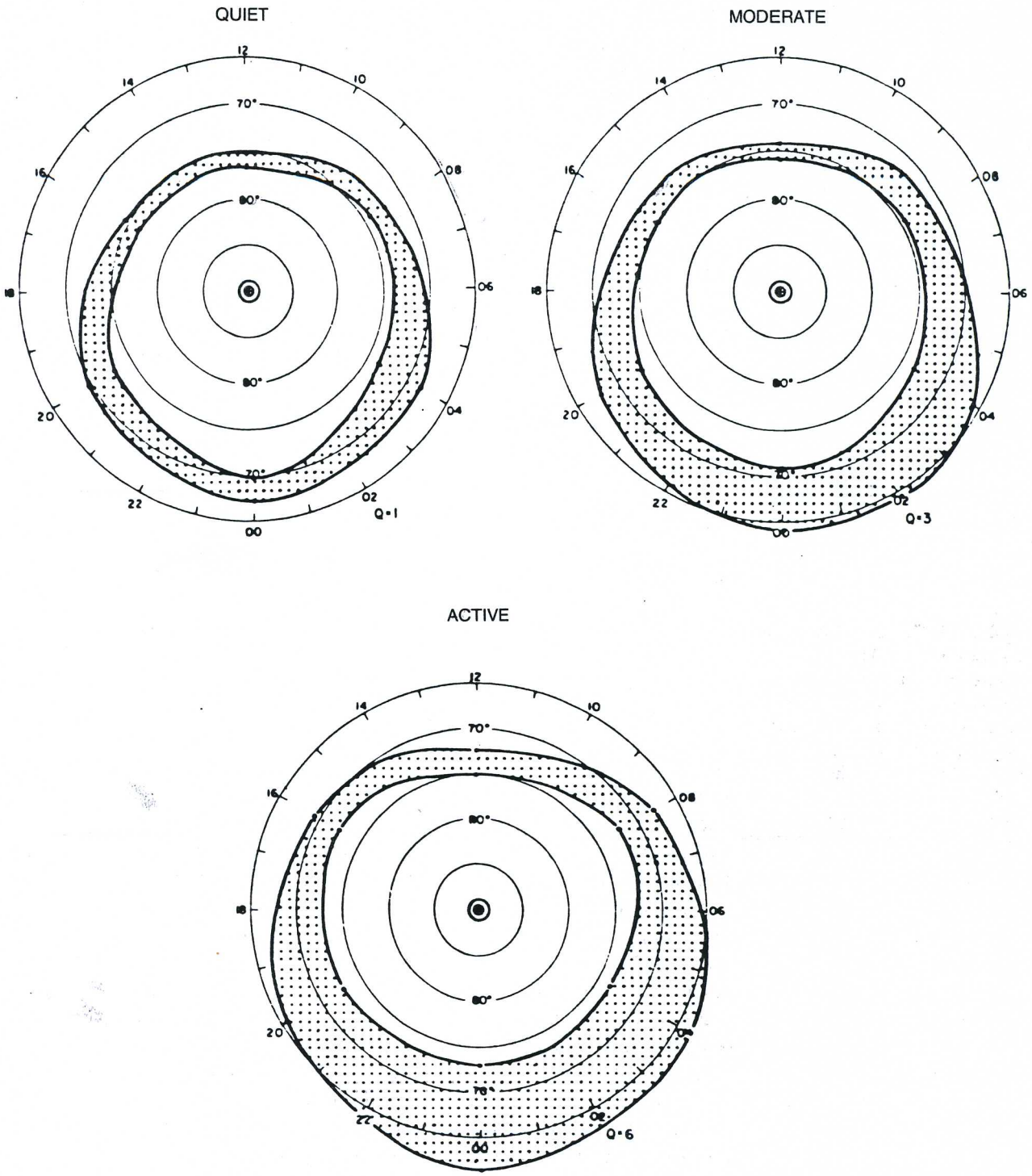


FIG. 14.8. Variation in the size of the auroral oval with activity. The shaded area represents the distribution of maximum auroral activity in the northern hemisphere. Coordinate system is corrected geomagnetic (CG) latitude and CG local time, and noon is at the top. (Adapted from Feldstein and Starkov, 1967.)

are shown in Figure 14.8 as functions of geomagnetic local time and geomagnetic latitude. Evidently, the location of an aurora is connected with disturbances in the geomagnetic field, the aurora moving mainly equatorward with increasing magnetic activity.

Auroras also appear regularly poleward of these ovals, but with less frequency and intensity. Equatorward, the occurrence rate falls rapidly. The occurrence of auroras at low latitudes is associated with enhanced magnetic activity and magnetic storms.

As an observation station rotates with the earth, its proximity to the oval (of a particular activity level) fixed in the earth–sun frame of reference, or magnetic local time (MLT), changes with a 24-h periodicity. Because the magnetic pole is displaced from the earth's rotation axis (roughly an 11° difference in the northern hemisphere, tipped toward the U.S. east coast), there is also a longitude dependence for many auroral effects. These are best ordered by universal time (UT), that is, the time counted positive westward from Greenwich with 15° per hour.

By means of careful comparisons among ground-based, rocket-borne, and satellite observations, it has been found that nightside auroras are primarily due to electrons between 1 and 15 keV, whereas dayside auroras are primarily due to electrons with energies less than 1 keV (Figure 14.3). This explains well the large differences between both the emission altitudes and spectra of nightside and dayside auroras.

Unstructured auroras, down to about 95 km in a circular band of roughly constant magnetic latitude nearly equal to that of the equatorward edge of the midnight auroral oval, are produced by electrons of approximately 5–20 keV. Diffuse and structured auroras are produced at intermediate altitudes by particles of intermediate energy (on the order of 1–10 keV) from the magnetospheric plasma sheet (see Section 14.8 and Chapter 10).

In contrast to the morphological and statistical descriptions of the auroral zones developed up to the 1960s, and the auroral oval (Figure 14.8) described thereafter, a more physical view of an instantaneous auroral state and location controlled by the solar wind, or by the interplanetary magnetic field (IMF), emerged after the 1970s. For a northward IMF, the solar wind couples poorly to the magnetosphere; the auroral oval becomes quite contracted, and the auroral intensity becomes weak. For a southward IMF, the coupling is much stronger and the oval much larger; its size expands with increasing solar-wind pressure, and the auroral intensity increases. Ground-based measurements show poleward motion of the poleward boundary of the midday aurora as the IMF changes from southward to northward, with monotonic progression of the contraction over extended periods.

Conjugacy of nightside auroral forms (mirror images, as it were) between the northern and southern hemispheres has been seen and taken as strong support for the argument that these auroras are on closed field lines.

A complication in describing and studying auroras arises from the fact that the sun usually is not in the geomagnetic equatorial plane. Because of the inclination of the earth's axis of rotation with respect to the earth's orbital plane (23°) and the angle between the dipole axis and the axis of rotation (~11°), the direction to the sun may deviate as much as ±34° from the geomagnetic equatorial plane during a year. These effects give rise to both diurnal and seasonal variations in auroral occurrences. The most important variation is probably the seasonal one, due to the angle between the geomagnetic axis and the earth–sun direction. There is a pronounced equinoctial maximum in auroral occurrences.

14.3.4 Dayside Cusp/Cleft Auroras

The increasing interest in dayside ionospheric phenomena in recent years has been motivated by general questions of the solar-wind–magnetosphere–ionosphere interactions. In this section we shall concentrate on optical ionospheric signatures on the dayside (i.e., ~±6 h of magnetic noon) in the cusp/cleft region, which corresponds to the auroral oval between about 70° and 80° Λ.

Some characteristic differences between dayside and nightside auroras regarding emission wavelengths and height distributions are seen in Figures 14.5c and 14.5d. Notice also that the dayside portion of the oval is located closer to the magnetic pole (i.e., centered at a magnetic latitude between 75° and 78° Λ) than are the nightside auroras, and the width of the dayside oval is approximately half of that at night (Figure 14.8).

Cusp auroras form in the daytime polar thermosphere, where atomic species are abundant (see Chapter 7). Consequently, the intensity of atomic lines in these auroras is significantly greater than the intensity of molecular bands, leading to a simplification of the dayside auroral spectrum. Even atomic lines that are not observed in nighttime auroras, partly because their wavelengths coincide with those of the more intense molecular bands, are detected in midday auroras. The relative weakness of the auroral molecular bands in midday cusp auroras also permits daytime observation of the chemiluminescent airglow OH emissions from the polar mesosphere.

The dominant dayside cusp aurora is, by definition, the diffuse band where emissions with I (630 nm) $\geqslant I$ (557.7 nm) are normally observed in the region 11–13 MLT. The red-line emission above 200 km is caused by the soft fluxes of magnetosheath-particle precipitation (Figure 14.4). Its intensity during quiet conditions is normally below 1 kR; that is, the dayside cusp aurora is subvisual, and the average height is 250 km or higher. The cusp aurora is diffuse in character.

The first direct evidence that the dayside aurora is related to plasma entering from the magnetosheath into the polar F region dates back no further than the 1970s. Ground-based dayside auroral measurements

TABLE 14.2. Cusp/Cleft-Aurora Electron Precipitation

Auroral Forms	*I* (630 nm)	*I* (557.7 nm)	Electron Energy	Energy Flux ($erg \cdot cm^{-2} \cdot s^{-1} = 10^{-3}\ W \cdot m^{-2}$)
Quiet cusp (midday gap)	<1 kR	<0.3 kR	<0.2 keV	<0.5
Active cusp (i.e., transient events)	Typically 0.5–5 kR	Typically 0.5–10 kR	<2 keV	Typically 0.5–5

were largely neglected until 1980, partly because of inaccessibility of ground sites with the necessary proximity to both the geomagnetic and geographic poles. Lack of equipment with the sensitivity necessary to study these weak, subvisual auroras was an additional factor.

Recently (e.g., Sandholt and Egeland, 1989) it has been realized that these dayside ionospheric effects are as important to solar-terrestrial research as are the nightside processes. Research related to the dayside cusp/cleft regions (based on coordinated ground, rocket, and satellite observations) has increased significantly, and important findings are being reported. The islands of Svalbard, Franz Josef Land, and part of Greenland, located near 75° Λ, are the only accessible sites in the northern hemisphere where midday auroras can be observed by optical methods (which can be used only if the sun is at a minimum of 10° below horizon near magnetic noon). Recent developments of more sensitive all-sky TV cameras and multichannel scanning photometers (sensitive primarily to 630- and 557.7-nm wavelengths) have provided high-resolution data (both in time and space) that can be studied together with coordinated particle and field measurements from polar satellites.

The region of cusp auroras is also called the "midday gap" because of the lack of discrete structures. Some typical parameters, including particle characteristics, are given in Table 14.2. Discrete, active auroral forms (often multiple, long-lived auroral bands) are observed before and after the midday gap. Transient discrete forms (often with a recurrence period of ~10 min) occur between 09 and 16 MLT (even at 12 MLT). The discrete arcs are produced by particles of higher energy than those producing the 630-nm region, but the average energy is more variable than on the nightside. Altitudes range from 150 to 200 km, and the intensities and occurrences of the arcs do not vary directly as the magnitude of the auroral substorm on the nightside. Discrete arcs are also found in the polar cap. They are also short-lived arcs of the same spectral type as the dayside arcs. At times, nightside and dayside auroras are simultaneously active. The most striking feature of simultaneous nightside and dayside auroras is the poleward expansion on the nightside, coinciding with equatorward expansion on the dayside. However, active cleft auroras are often observed without any simultaneous

changes in the nightside, and vice versa. The large-scale dynamics of the cusp/cleft auroras are mainly controlled by the IMF B_z component. During IMF $B_z < 0$ conditions, the auroral intensity is particularly sensitive to the solar-wind activity, which controls the efficiency of plasma transfer. When the IMF vector turns due north, the auroral intensity decreases, and the aurora contracts poleward, indicating reduced plasma transport into the cusp region.

One type of active event (breakup arcs, poleward expansion, longitudinal motion depending on IMF B_y) can be superposed on a quiet, gaplike background arc. The majority of these midday transients, less than 10 min in duration, with intensities of a few kilorayleighs (10–20-kR events are rare), are not likely to be observed by satellite imagery, because of inadequate sensitivity and temporal/spatial resolution.

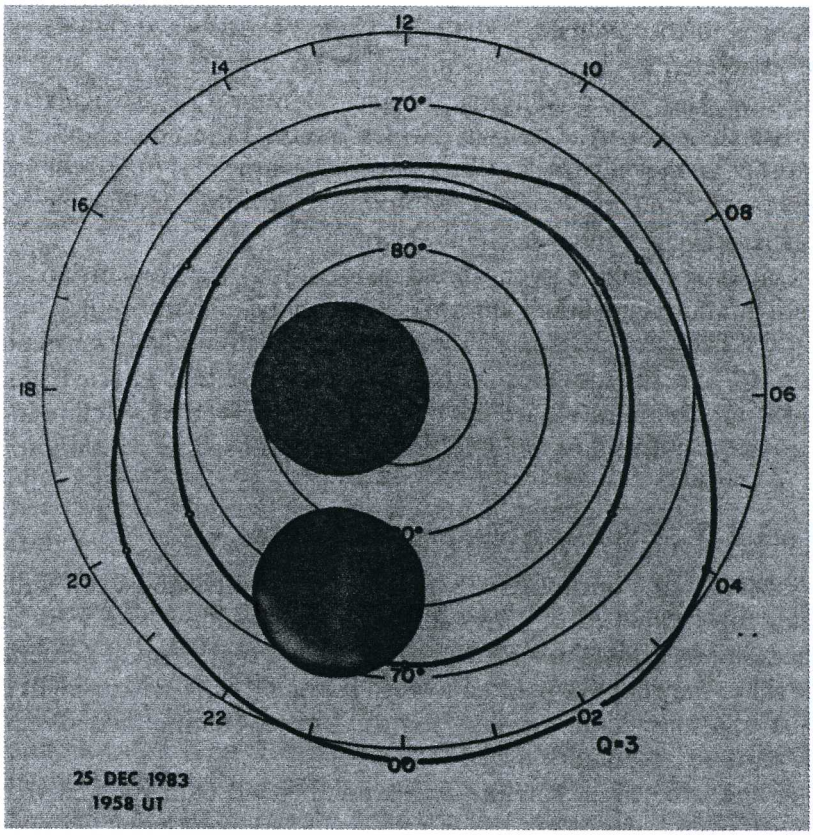
14.3.5 Polar-Cap Sun-aligned Auroras

Observations during IGY 1957–8 demonstrated that deep within the polar cap (poleward of ~75°–80° magnetic latitude) there were auroras oriented along the sun–earth direction. Seen only a small percentage of the time, they always pointed toward the sun; in the winter-night central polar cap, one could virtually tell the time by noting their direction as the earth rotated under them.

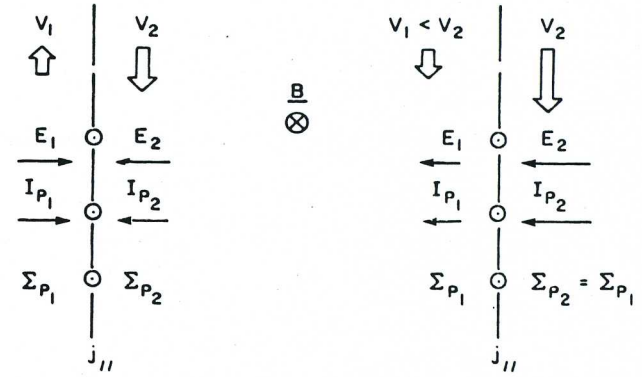
In the late 1970s, satellite and ground-based data showed that such occurrences correlated with northward IMF conditions. However, it was not until the 1980s that the improved sensitivity of intensified-imaging photometers showed these arcs (an example is shown in Figure 14.9a) to be present about half of the time, when the IMF was northward (or near zero). Those images, when combined with ground-based optical and radar measurements and satellite observations, showed that such arcs were of electrodynamic origin and determined their thermal and energetic characteristics.

A simple situation for which electrodynamic effects lead to an optical-arc signature along a boundary line is illustrated in Figure 14.9b. Consider flow near a boundary, and assume a vertical magnetic field and initially uniform conductivity across the boundary. Suppose the plasma velocity reverses (left-hand side) or that there is a velocity gradient (right-hand side) across the boundary line. This means that the electric field is discontinuous across the boundary. The horizontal electric-field differential will produce a horizontal Pedersen-current convergence at the boundary in the absence of a vertical current. Thus, a vertical (actually, magnetic-field-aligned) current flows with the magnitude required to maintain a divergence-free current state.

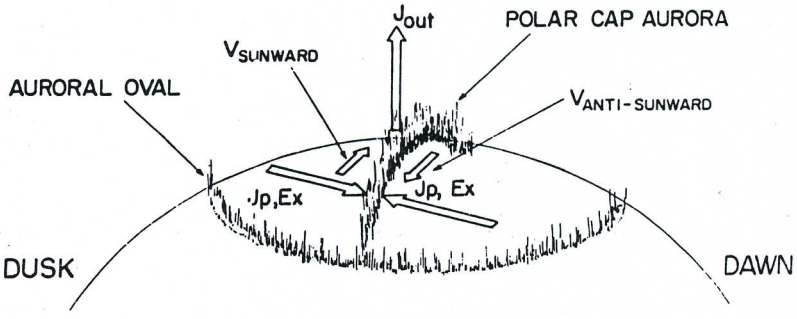
The difference between the left-hand-shear reversal and the right-hand-shear differential flow is merely the velocity of the frame from which the flow is viewed. It is the rest frame of the neutral gas that determines the currents. The special significance of the neutral-gas rest



(a)



(b)



(c)

FIG. 14.9. (a) Example of a sun-aligned arc, connected to the premidnight auroral oval, extending over 2000 km sunward, as seen in a pair of 630-nm all-sky imaging photometers. (From Carlson, 1990.) (b) Two examples of a simple electrodynamic situation that would produce a sun-aligned arc. (c) Cartoon representation of the simple electrodynamics characteristic of stronger sun-aligned arcs. The plasma flow on the duskside may more generally be (for a range of arc intensities) sunward, stagnant, or antisunward, but slower than the antisunward flow on the dawnside.

frame may be understood in part by recalling that it is collisions between the charged and neutral-gas particles, particularly different collision frequencies (mobilities) of the ions and electrons with the neutral-gas particles, that produce a finite conductivity and result in Pedersen current perpendicular to the velocity-shear boundary.

The sense of the velocity differential across the boundary determines whether it drives a horizontal convergence or divergence of current, and thus requires a vertically upward or downward current to maintain a divergence-free state. In Figures 14.9b and 14.9c, the sense leads to an upward current, presumably carried into the ionosphere by a flux of downcoming suprathermal electrons.

Note that if there is a series of velocity differentials across a portion of the ionosphere, with plasma alternately speeding up and slowing down, the upward current sheets will be interspersed between downward current sheets along velocity-difference boundary lines. The downward currents presumably will be carried by upgoing thermal electrons.

Combined ground-based images and satellite data have shown that the simple arc electrodynamics are in fact the electrodynamics that pertain to sun-aligned arcs that are stable in time and extensive in sunward direction (Carlson et al., 1988). We should note that a change in conductivity across a boundary in the ionosphere can drive current sheets out of the ionosphere even in the absence of a velocity shear. We have found no examples of this effect alone creating sun-aligned arcs.

Because the current carriers of the vertical (actually, magnetic-field-aligned) current are suprathermal electrons with energies of tens to hundreds of electron volts, they excite the optical signature of the arc. By the same token, they will be expected to produce impact ionization, thereby enhancing the conductivity within the arc and modifying the distribution of currents that flow within the arc itself. This feedback effect must be allowed for in detailed self-consistent treatments of the current patterns related to the arc.

Here, our main point is that the sun-aligned arcs are visual markers of velocity-shear lines, lines of sharp velocity differentials, of a specific sense: greater antisunward velocity on the dawnside than on the duskside of a polar-cap sun-aligned arc. The shear line points toward the sun because the antisunward flow is driven by coupling to the solar wind blowing past the earth and away from the sun (as described in Chapter 10). Because sun-aligned arcs are signatures of sharp slowing (or even reversal) of an antisunward ionospheric plasma flow, they are thereby a most valuable tool for discovery and definition of the character of polar-cap convection for northward IMF conditions (i.e., the half of the time for which convection is most poorly understood). These sun-aligned arcs are also of importance to the thermal balance and energetics of the polar-cap ionosphere and thermosphere (Valladares and Carlson, 1991).

14.4 THE AURORAL SUBSTORM

An auroral substorm, which can have many different auroral forms, large variations in both color and intensity, and rapid motions, may look completely disordered. However, an experienced observer will soon notice that such large auroras follow a fixed pattern, almost as in a classical play or symphony, with four scenes or movements (Akasofu, 1968).

For a nighttime observer near the oval, the display usually starts with one or more quiet, homogeneous arcs of fairly low intensity (1–10 kR) elongated approximately in the geomagnetic east–west direction. After some time, possibly hours, the aurora starts to move equatorward, increases in intensity, and may develop ray structure and take the form of less regular bands. Then, suddenly, the whole sky explodes, and the aurora spreads over the entire sky. Simultaneously, the aurora moves rapidly, with changes in form and intensity, at times increasing to several hundred kilorayleighs, and the individual structures in the aurora may show apparent speeds in an easterly or westerly direction of several tens of kilometers per second. After a few minutes, the aurora becomes weaker and rather diffuse. This marks the beginning of what is called the recovery phase.

Substorms must be considered within a fixed earth–sun frame of reference, as is true of all auroral phenomena. A ground-based observer thus must remove the effects of the earth's rotation in tracing auroral motion in this reference frame. Whereas that seems straightforward in principle, the offset of the geomagnetic pole from the geographic pole introduces more complexity than at first meets the eye.

Each (four-scene) active period is called an auroral substorm. An illustration of the development of an auroral substorm in the dipole local-time coordinate system is shown in Figure 14.10a. An example of an auroral "breakup" is shown in Figure 14.10b, a sequence of auroral pictures taken with an imager on board the Dynamics Explorer *DE-1* spacecraft at roughly 20,000 km from the earth. The proton aurora (Section 14.2.2) moves poleward of the trapping boundary some time after the poleward expansion of the discrete arcs associated with the substorm. The most spectacular part of the display usually lasts only a few minutes (<10 min), the whole active period being perhaps 0.5–1 h. Afterward, the sky is more or less covered with a homogeneous surface of weak auroral light. The intensity is then usually between 1 and 5 kR, but this aurora may be subvisual. Because of the lack of contrast, the intensity of such weak, widespread emission surfaces, compared with more discrete and distinct forms, is often underestimated by the eye. The duration and intensity of the substorm may vary greatly. The usual time interval between two substorms in one night ranges from 0.5 to 3 h. For disturbed periods, several auroral breakups may occur in one night.

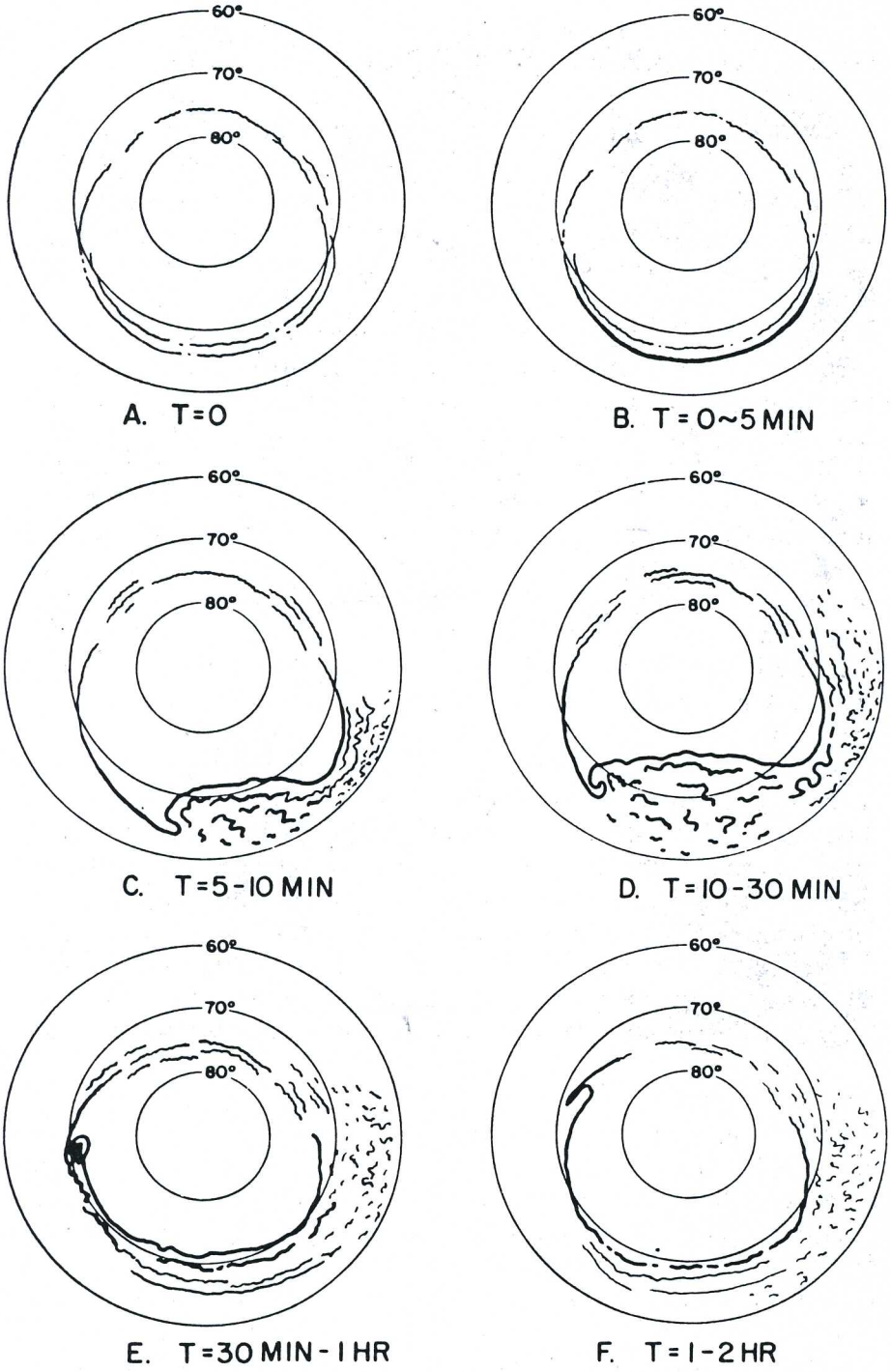


FIG. 14.10. (a) An auroral breakup event is illustrated by this sequence of drawings. The sequence of the pictures is indicated by the time labels given from $T=0$. (From Akasofu, 1968.) *(Continued)*

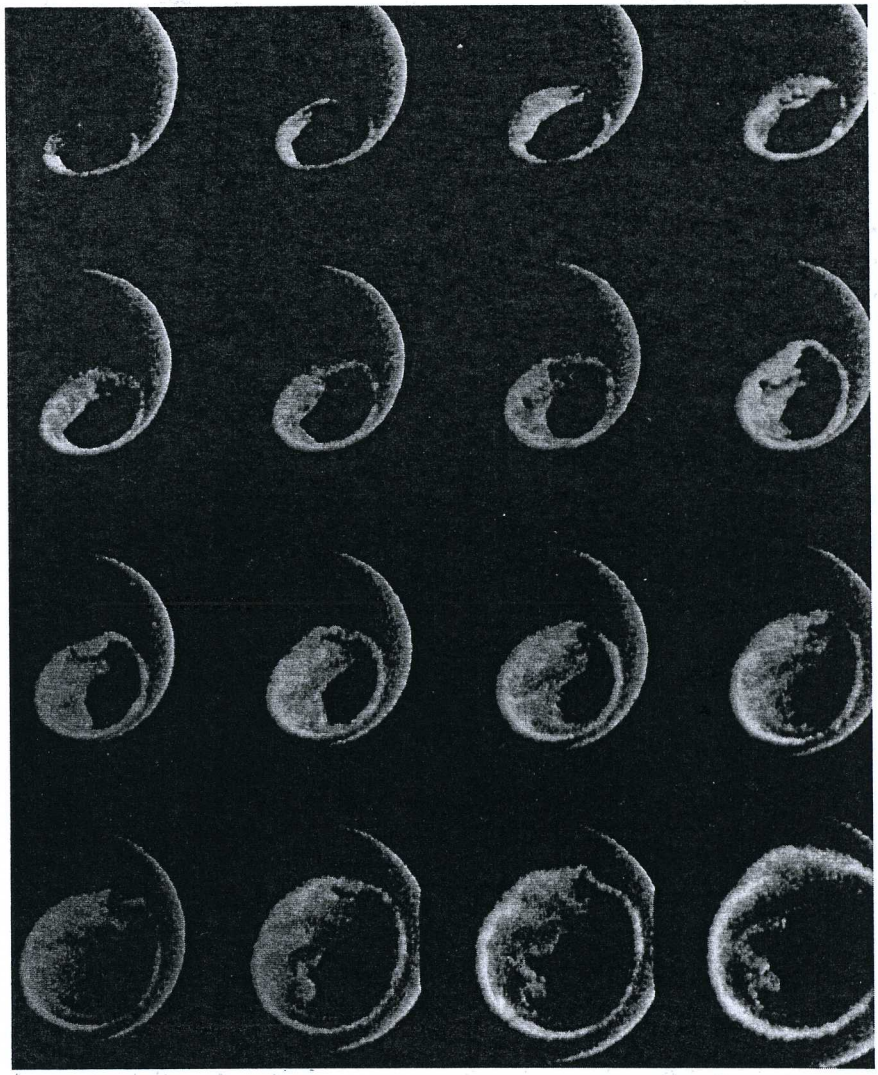


FIG. 14.10 *(cont.)* (*b*) Sequence of auroral pictures from *DE-1* during an auroral substorm. Time increases from left to right and from top to bottom. (From Frank and Craven, 1988.)

During an auroral display, several forms may appear simultaneously, partly overlapping or embedded in each other. Auroral arcs and bands frequently show ray structure. Each form usually moves and varies in intensity with time. For homogeneous, quiet forms, the velocities are small ($\approx$100 m$\cdot$s^{-1}), whereas particular ray structures may show rapid motion, with velocities of up to 50 km$\cdot$s^{-1}. Simultaneously, there are intensity and color variations in the moving forms. Large-scale intensity variations are combined with the movements. The form of the aurora may vary with the MLT. A few hours after geomagnetic midnight, the auroral pattern is predominantly diffused, but the forms often show rapid fluctuations in intensity and pulsations. The substorm is associated with disturbance current systems, giving rise to characteristic magnetic disturbances (Akasofu, 1968; Akasofu and Chapman, 1972).

Because an aurora is the result of charged particles interacting with the earth's upper atmosphere, and the injection of these particles is due to interaction with the solar wind, the aurora depends both on the plasma streaming away from the sun and on the earth's magnetic field. The appearance of auroras in space and time depends not only on the properties of the solar wind and the geomagnetic field but also on the composition of the atmosphere. The occurrences and spatial distributions of auroras depend also on local time and the substorm phase. The physical processes that control the different phases of an auroral substorm are not yet well understood.

14.5 THE AURORAL IONOSPHERE

Before addressing the electrodynamics of the ionosphere, we shall summarize some characteristics of the ionized portion of the radiating medium: the auroral ionosphere (see also Chapter 7). The auroral ionosphere has properties significantly different from those of the ionosphere at subauroral latitudes (Rishbeth and Garriott, 1969). For the midlatitude ionosphere, the dominant ionizing source is solar radiation [extreme-ultraviolet (EUV) and UV rays and x-rays], the dominant heating source is likewise solar radiation, and transport is usually dominated by neutral winds and downward diffusion due to gravity. In disturbed auroral regions, the dominant ionizing source is particle precipitation, the dominant heating source is particle precipitation (for electrons), and neutral particles, Poynting flux, or Joule heating (for ions), and the dominant transport is driven by electric fields.

14.5.1 Ionization

Below 180–200 km in the ionosphere, primary ions in the auroral ionosphere (as at midlatitudes) are molecular ions. They are relatively short-lived, recombining rapidly with thermal electrons (dissociative recombination at a rate of 10^{-6}–10^{-7} $cm^{-3} \cdot s^{-1}$). Thus, any significant (10^4 cm^{-3}) electron density seen in the nighttime auroral region much below 200 km in altitude is a measure of production of ionization at the time and place it is seen. Because the recombination rate is the product of the number of electrons per cubic centimeter (n_e) times the number of ions per cubic centimeter (n_i) available for recombination, and because $n_e \simeq n_i$, n_e^2 is a measure of precipitating auroral particles. For a nominal altitude of maximum ionization in the auroral E region (near 100 km), n_e^{max} is a measure of auroral-particle energy deposition (see Section 14.7.1), as the characteristic energy of particles that are able to penetrate to that altitude is known. From the ground, an ionosonde can measure the maximum plasma frequency of the E region, f_0E, giving the nominal relationship $(f_0E)^2 \propto n_e^{max}$.

Molecular ions dominate in the E region; in contrast, atomic ions dominate the F region. Recombination of atomic ions with electrons is slower than recombination of molecular ions, by a factor of about 10^4. These ions are generally lost through the following two-step process:

$$O^+ + N_2 \rightarrow NO^+ + N \tag{14.8a}$$

or

$$O^+ + O_2 \rightarrow O_2^+ + O \tag{14.8b}$$

The limiting process (14.8), a chemical reaction with neutral molecules to form a molecular ion (by charge exchange), is followed rapidly by dissociative recombination (see Section 7.3). Reaction (14.8a) explains the dominance of NO^+ below the F region.

14.5.2 Motion

The earth's magnetic field greatly influences the motion of plasma in the ionosphere (this is why many ionospheric properties are better described in geomagnetic coordinates than in geographic coordinates). If collisions with neutral particles do not become very frequent, electrons and ions will move together in the presence of an electric field **E**, frozen to magnetic-field lines **B**, in the direction perpendicular to both the electric and magnetic fields at a velocity

$$\mathbf{v}_e = \mathbf{v}_i = (\mathbf{E} \times \mathbf{B})/B^2 \tag{14.9}$$

where the subscripts refer to electrons (e) and ions (i).

Because of the long lifetimes of ions at F-region heights, the transport described by equation (14.9) becomes important. Ionized particles in the F region may survive for many hours before chemically recombining. At auroral and polar latitudes, strong auroral electric fields may move ionospheric plasma at velocities of the order 1 $km \cdot s^{-1}$. Ionization may have occurred thousands of kilometers away from the location where ions are observed. As polar ionospheric convection (Heelis, 1988) is always observed in the high-latitude ionosphere around the auroral oval, the displacement of ions from their source location in the F region is a continuing process.

14.5.3 Currents

Equation (14.9) applies to the motion of the F-region ions, but does not apply to the E region. In the E region, collisions of charged particles with neutrals occur so frequently that no transport of bulk ion-density profiles can occur.

In Figure 14.11a, positive-ion trajectories are shown for different ratios between the collision frequency and the ion-cyclotron frequency Ω_i. Above about 200 km, the charged-particle trajectories are described

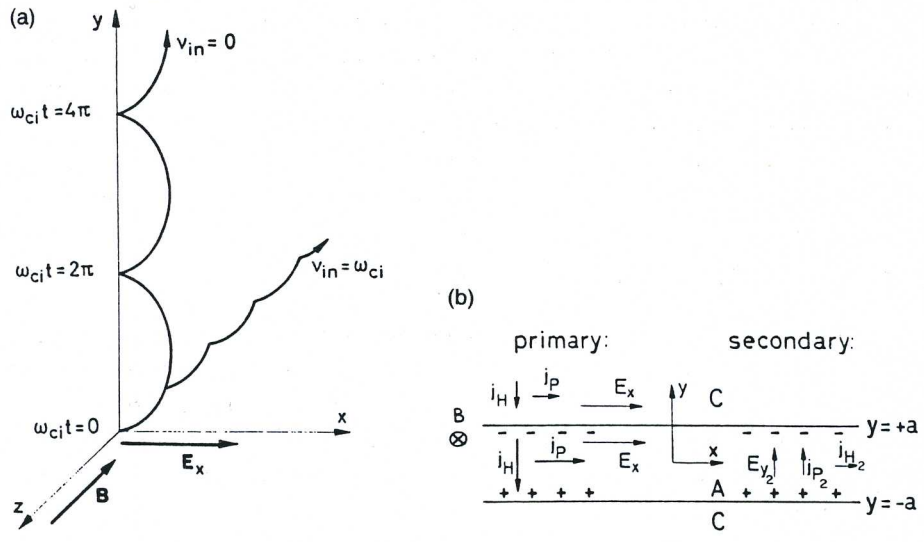


FIG. 14.11. (a) Motion of a positive ion in the absence of collisions ($\nu_{in}=0$) and when the collision frequency and gyrofrequency are equal ($\nu_{in}=\Omega_i$). The orbit is shown for a case when there is no electric-field component E_z along the magnetic field. (b) Simple model of an auroral arc, including both primary and secondary currents. (From Egeland et al., 1973.)

in terms of bulk plasma-drift velocity $[(\mathbf{E}\times\mathbf{B})/B^2]$. However, at around 140 km, the average ion will complete only part of a gyrocycle before it collides with a neutral particle. It then will resume its motion under the influence of **E** and start on another gyroarc, until the next collision occurs. When the ion–neutral collision frequency ν_{in} just equals Ω_i, the ions will drift, on average, at a 45° angle to $\mathbf{E}\times\mathbf{B}$. The subscript n refers to neutrals. When $\Omega_i \gg \nu_{in}$, the ions will never have a chance to move in the $\mathbf{E}\times\mathbf{B}$ direction. They will simply diffuse through the neutrals in the direction of **E** (see Problem 14.5). Because ν_{in} is so much greater than the electron–neutral collision frequency ν_{en}, electrons are bound to the geomagnetic-field line ($\omega_e \gg \nu_{en}$), and equation (14.9) applies to them down to an altitude of about 100 km. Eventually, even the electrons suffer significant collisions with the neutrals ($\nu_{en} \sim \omega_e$ at 80 km). Ion motion relative to electrons arising from the different effects of neutral collisions produces an electric current in the ionosphere. The mathematical treatment given in Section 7.7 applies in general.

Auroral currents (j = current density, $\mathrm{A}\cdot\mathrm{m}^{-2}$)

$$j=\sigma E=n_i e(v_i-v_e) \tag{14.10}$$

flow, where n_i provides current carriers, primarily above roughly 90 km; ions drift relative to electrons primarily below 140 km. Here, E and n_i lead to a current with components perpendicular to **B** and parallel to **E** (Pedersen current), perpendicular to both **B** and **E** (Hall current), and parallel to **B** (Birkeland current).

14.5.4 Secondary Currents

Auroral electrojets, intense localized currents, mainly flow from east to west, following long, narrow enhanced-conductivity channels near

the midnight portion of the auroral oval during substorms (see Chapter 13).This characteristic geometry produces distinctive current patterns. In order to understand them, it is necessary to become familiar with the concept of secondary-polarization electric fields.

Consider a slab of aurorally enhanced ionospheric plasma A immersed in a background ionosphere C of lower electron density and thus conductivities as illustrated in Figure 14.11b. A primary electric field in the x-direction, E_x, is applied in both the A and C regions; it must be continuous at the interface at a consequence of $\nabla \times \mathbf{E} = 0$. The geomagnetic field is into the figure. The primary currents (J = height-integrated, $\mathrm{A \cdot m^{-1}}$, see Section 14.7) driven by E_x include both the Pedersen current J_P along $\hat{x}$ and the Hall current J_H in the $\hat{y}$-direction. (J_P is greater in A than in C simply because of the higher conductivity in A.) However, the y current component must, in the steady state, be continuous across the A–C interface. This causes a transient accumulation of excess positive and negative charges at the A–C interface, as shown in the figure. The secondary-polarization field E_y built up in A by this process drives a secondary Pedersen current J_P that cancels the excess primary J_H across the boundary. However, this secondary polarization E_y also drives a secondary Hall current (J_H) along the arc A in the $+x$-direction. This adds to the primary J_P, making the current even more intense along the arc.

Quantitatively, the equalities just noted lead to (Egeland, Holter, and Omholt, 1973)

$$J_x = \Sigma_P E_x + \Sigma_H E_y = E_x(\Sigma_P + \Sigma_H(\Sigma_H - \Sigma_H^C)/\Sigma_P) \tag{14.11a}$$

$$= E_x(\Sigma_P + (\Sigma_H)^2/\Sigma_P) \qquad (\text{for } \Sigma_H \gg \Sigma_H^C) \tag{14.11b}$$

$$J_y = -\Sigma_H^C E_x \tag{14.11c}$$

where Σ is the height-integrated conductivity, and all terms refer to the auroral arc, except Σ_H^C, which refers to background and E-region ionospheric conductivity.

This simplified argument neglects Birkeland currents ($\mathbf{J} \| \mathbf{B}$) at the arc boundaries, and when applied to auroral arcs, it gives an upper limit for the degree to which the current is enhanced. In the actual ionosphere, Birkeland currents along $\mathbf{B}$ reduce the polarization charges and thus reduce E_y. This discussion of secondary-polarization fields was applied to the auroral electrojet, but in general it applies to any enhanced-conductivity slab. [Note: Equation (14.11a) shows that the effective conductivity along the slab is enhanced. In the limiting case of a very high conductivity ratio between regions A and C, we find that the effective conductivity approaches $\Sigma_P + (\Sigma_H)^2/\Sigma_P$, as in equation (14.11b). This expression is similar in structure to the Cowling conductivity.]

14.6 AURORAL EFFECTS ON RADIO WAVES

Although it is beyond the scope of this treatment to include an exposition of the ways in which an aurora affects the radio-frequency spectrum, it would be a serious omission not to note that its effects are significant. We mention a few to convey a sense of those effects.

The large auroral and polar-cap electric fields, currents, and plasma drifts, relative to the neutral atmospheric velocity, lead to a variety of plasma instabilities. These instabilities at F-region altitudes can be reduced to two major classes. One, known for 20 yr or more, is an $\mathbf{E} \times \mathbf{B}$ or gradient-drift class of instability, somewhat modified by strong field-aligned currents. The other, discovered to be important in 1988, is an inertial or Kelvin-Helmholtz instability that develops in strong-velocity-shear regions, especially in auroral latitudes (Basu et al., 1990). These plasma instabilities cause otherwise relatively smooth plasma to develop inhomogeneous structures, **B**-aligned, over scale sizes ranging from 0.1 to 10 km, through which strong radio scintillations are produced. These auroral scintillations lead to strong amplitude fading and phase fluctuations up to gigahertz (GHz) frequencies, disrupting VHF (30–300 MHz), UHF (0.3–3 GHz), and GHz communications, and even navigation systems at lower frequencies.

Magnetic-field-aligned irregularities in the auroral ionospheric plasma, driven by instability processes, also scatter radio waves, especially at HF (3–30 MHz) and VHF ranges, much as glass rods reflect light where the geometry is similar. This auroral scatter or *clutter* can blind radar tracking, disrupt or improve HF communication, and serve as an important geophysical diagnostic tool. Its existence has been known since early in the 1930s, when amateur radio operators discovered that during an aurora it was possible to receive transmissions from an operator located to the south by directing their antennas toward the north.

Enhanced ionization at D-region altitudes leads to strong radio absorption (proportional to f^{-2}). A wave propagating through an ionized medium exchanges energy between the wave and the particles. The AC electric field alternately accelerates the electron for half a cycle and decelerates it on its next half cycle. If the electron suffers a collision with a neutral particle before it can reradiate, this ordered energy becomes disordered, heating the electron at the expense of dissipating the energy in the radio wave. Increased D-region electron densities in the auroral region (auroral absorption and/or polar-cap absorption events) seriously attenuate HF communication from earth. Measurements of attenuation of extraterrestrial radio waves passing through the attenuating D region can serve as a diagnostic for D-region ionizing sources (Rosenberg et al., 1991).

Wave–particle interactions at auroral latitudes also lead to a variety of electromagnetic emissions. The particle energy is converted to electromagnetic energy through wave–particle plasma processes, not all of which are understood (Helliwell, 1988).

14.7 ENERGY TRANSFER TO THE IONOSPHERE

Particle energy, momentum, and mass are transferred to the magnetosphere from the solar wind. Consequently, significant energy is conveyed to the ionosphere in the form of particle energy and/or as electromagnetic energy.

We have addressed the role that the auroral particles play in direct excitation of auroral optical emissions (Section 14.2), as well as in ionization of neutrals. Both mechanisms imply that energy has been deposited. However, another form of energy deposition often exceeds the rate of energy input to the auroral and polar regions by particles. The downward Poynting flux ($\propto \mathbf{E} \times \mathbf{H}$) from the magnetosphere can be dissipated in the ionosphere–thermosphere and can do mechanical work against $\mathbf{j} \times \mathbf{B}$ forces there.

We can think of this type of energy input to the ionosphere as starting with mechanical energy in the solar-wind generator that gets converted to electromagnetic energy. That, in turn, is conveyed down the geomagnetic-field lines (as Poynting flux) into the ionosphere, where it is dissipated as Joule heating. Joule heating in the ionosphere is $\mathbf{j} \cdot \mathbf{E}$ in the ionospheric rest frame. Thus, for an applied electric field, energy is dissipated by the current component parallel to $\mathbf{E}$, that is, the Pedersen current; Hall current (perpendicular to $\mathbf{E}$) is nondissipative. That leads to a Joule-heating dependence on characteristic energy, because higher-energy particles penetrate more deeply into the atmosphere, where $\Sigma_H > \Sigma_P$, and the currents tend to be nondissipative. Joule heating is then less important. Less energetic particles produce ionization at higher altitudes where Pedersen conductivities are greater than Hall conductivities and Joule heating is relatively more important. Because the current density $j_P = \sigma_P E$, where σ_P is the Pedersen conductivity, height-integrated Joule heating can be expressed as $\Sigma_P E^2$, where $\Sigma_P = \int \sigma_P \, dh$. For limited times and locations, this heating rate may be many ergs per square centimeter per second, while the energy deposited in sunlit regions by solar UV is about 0.5 $\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$ above roughly 110 km. The latter is a valuable reference quantity to use as a measure of the significance of upper-atmospheric energy input. It should be compared with the integral (over altitude) of the rate of Joule heating ($\text{erg} \cdot \text{cm}^{-2} \cdot \text{s}^{-1}$) expressed in the rest frame of the neutral atmosphere as

$$\mathbf{j} = \sigma(\mathbf{E} + \mathbf{v}_n \times \mathbf{B}) \tag{14.12}$$

where $\mathbf{v}_n$ is the neutral-wind velocity.

The total heating of the lower thermosphere by auroral energy dissipation, with heating rates transiently exceeding that of UV from an overhead sun, will change the temperature, density, composition, and winds of the local thermosphere. In fact, the global-scale thermospheric and ionospheric responses to such variable heating are subjects of ongoing research.

14.7.1 Particle-Energy Deposition

Arriving at an estimate of the particle heating affords an opportunity to bring out several important points. At E-region altitudes, recombination is proportional to a rate multiplied by the number of electrons and ions or n_e^2 (see Section 7.3). The production rate of ionization can be estimated by the energy of the particle divided by 35 eV, as discussed in Section 14.2. In a quasi-steady state, the production rate just balances the recombination rate. This leads to a quantitative estimate of Q_p, the rate of ionospheric heating by precipitating particles, as

$$Q_P = 5.6 \times 10^{-6} \int \alpha(h) n_e^2(h)\, \mathrm{dh}\ (\mathrm{erg \cdot cm^{-2} \cdot s^{-1}}) \qquad (14.13)$$

where α is the effective-altitude recombination coefficient (roughly $2–3 \times 10^{-7}$ $\mathrm{cm^3 \cdot s^{-1}}$ around 120 km), n_e is the electron density ($\mathrm{cm^{-3}}$), and h is the altitude (km). Note that the particle heating rate is also proportional to n_e^2. A 3-MHz E-region plasma frequency equates roughly to an electron density of 10^5 $\mathrm{cm^{-3}}$, a Q_P of 1 $\mathrm{erg \cdot cm^{-2} \cdot s^{-1}}$, and 0.5 kR of 391.4-nm emission for a typical auroral layer. This is representative of diffuse auroras, which are produced by loss-cone particle precipitation, which is also responsible for an extensive, highly uniform, quasi-equilibrium layer of E-region ionization.

14.7.2 Joule Heating and Energy Deposition

The earlier discussion of Joule heating was formulated as a collective-interaction description (the ionosphere as a resistive load). It is often instructive to take both a collective view and a particle view of interaction processes. Thus, this section examines energy deposition from the viewpoint of particles. Auroral and transpolar electric fields drive plasma at a velocity $v_i = (\mathbf{E} \times \mathbf{B})/B^2$ at high latitudes. The plasma is electrically neutral, with $n_e = n_i$. An individual ion, upon collision with a neutral atmospheric particle, will exchange energy and be deflected in some new direction. Thus, the ion's ordered motion in the plasma-drift direction becomes disordered motion, that is, heat. From this particle view, it is thus easy to see that the ion-gas temperature heats up rapidly over a few ion–neutral collision times. [A useful number to recall for collision frequencies (Chapter 7, Figure 7.8) is that at about 140 km, the ion–neutral collision frequency roughly equals the ion gyrofrequency, about 200 $\mathrm{s^{-1}}$]. The ν_{in} decreases exponentially with altitude, in propor-

tion to the neutral density. At F-region altitudes (~300km), the time for the ions to heat up by this process is only on the order of seconds.

The ion temperature increases by this ν_{in} process at a rate (Banks and Kockarts, 1973)

$$\frac{\partial}{\partial t}T_e = \nu_{in}\left(T_n - T_i + \frac{m_n}{3k}(v_i - v_n)^2\right) \tag{14.14a}$$

so that for a quasi-steady state, ion temperature T_i exceeds T_n by

$$T_i - T_n = \frac{m_n}{3k}(v_i - v_n)^2 \tag{14.14b}$$

where m_n is the mean mass of the neutral atmospheric gas, and k is Boltzmann's constant.

Note that temperature enhancement varies as the square of the plasma velocity in the neutral rest frame and is negligible for midlatitude velocities of the order of 100 $m \cdot s^{-1}$ or less, but exceeds 1,000K for auroral and polar-cap velocities much greater than 1 $km \cdot s^{-1}$.

14.7.3 Thermospheric Heating and Momentum Transfer

In sharp contrast to the ions, the neutral-particle gas takes much longer to respond and responds dramatically by a momentum change, rather than a thermal change. Heat into ions exposed to an electric field and an $\mathbf{E} \times \mathbf{B}$ force (which physically may be thought of as resistive Joule heating in a conductor, randomization of ordered ion-particle motion, or "frictional-drag heating" of the ion gas dragged through the neutral-particle gas) is ultimately lost to the neutral gas. However, the thermospheric temperature increases much less dramatically than the ionospheric temperature previously discussed, because of its greater particle and mass densities, which lead to far greater heat capacity.

For a steady electric field $\mathbf{E}$ moving the ions along an ordered trajectory, we noted that ν_{in} causes the ordered motion of individual ions to become randomized in direction, resulting in heating. However, for an upper atmosphere at rest, the momentum transfer from the ions to the neutral particles is systematic in that it always has a component downstream in the direction of plasma flow. Whereas (Figure 7.8) an average F-region ion collides with a neutral particle on the order of once per second, an average neutral particle collides with an ion much less often. Thus, the thermosphere responds much more slowly. This is because, at F-region altitudes, there are more than 1,000 times as many neutrals as ions (~10^9 cm^{-3} neutral particles versus typically 10^6–10^5 cm^{-3} ions and electrons). Thus, the average ion has to experience 10^3–10^4 ion–neutral collisions, which occur with a frequency ν_{in}, before the average neutral particle experiences one neutral–ion collision, occurring with a frequency ν_{in}. For large ion densities, the thermosphere

may eventually be brought up to the speed of the ions. Quantitatively, the "ion-drag" force on the neutrals acts over a time scale given by the neutral-gas momentum equation (neglecting gradient terms):

$$\frac{\partial v_n}{\partial t} = \frac{\rho_i}{\rho_n} \nu_{in}(v_i - v_n) \tag{14.15}$$

where we assume that the F-region neutral wind varies negligibly with altitude, and where ρ is the gas density (number density times mass) for the ions and neutrals, as indicated by the i and n subscripts.

The time for the thermosphere to respond is then given by $\rho_n \rho_i \nu_{in}$. Using typical values for n_n, m_n, and ν_{in} for atomic oxygen ions, we find that the response time for typical F-region altitudes and thermospheric temperatures/densities (250–450 km, 750–1,500K) is $0.23 \times 10^{10}/n_i$. Thus, for a typical daytime ionospheric density of $n_i = 10^6$ cm^{-3}, the thermosphere is dragged up to the ionosphere plasma drift velocity (typically $\sim$1 km$\cdot$s^{-1} over the polar cap for a southward IMF) in about 30 min, whereas for ionospheric ion and electron densities of 10^5 cm^{-3}, it takes 6 h. In 6 h, the earth has rotated 90°, and the thermosphere is exposed to a very different part of the ion convection pattern, which is fixed in the earth–sun reference frame for a fixed IMF condition. Thus, for ionospheric densities near and above 10^6 cm^{-3}, the upper-atmospheric winds are tightly coupled to the plasma drift (with roughly 30 min smoothing or averaging). For ionospheric densities near and below 10^5 cm^{-3}, this wind coupling is almost negligible.

14.8 RELATION TO BOUNDARIES AND PHYSICAL PROCESSES IN THE MAGNETOSPHERE–IONOSPHERE–THERMOSPHERE

The era of satellites has provided a sophisticated framework within which to order our developing understanding of the transfer of energy, momentum, and mass from the solar wind through the magnetosphere and into the ionosphere and atmosphere.

Figure 9.18 identifies magnetospheric regions, distinguished by particle population characteristics, from which particles precipitate into the upper atmosphere to produce auroras. The figure illustrates the topology of the magnetosphere (x points toward the sun, and z to the north, in the right-hand x, y, z coordinate system) and also the projection of the plasma source regions onto the northern polar upper atmosphere, with the sun again to the left.

Currents flow between the dawn and dusk regions of the magnetosphere across the tail current sheet, separating the tail lobes. Most of them flow straight across the magnetotail, but some are diverted along field lines into and across the auroral ionosphere. These currents are carried principally by electrons (which are more mobile than ions). The

primary current to the auroral region enters the dawnside, crosses the polar cap, and exits the duskside. Currents into the ionosphere are carried by upgoing thermal electrons, which do not excite auroral radiation. Currents out of the ionosphere are usually carried by precipitating-electron current carriers, which do excite auroras. Hence, the primary currents produce auroras mainly premidnight.

As the conductivity is enhanced in the auroral oval (see Section 14.5), currents also flow from the inner (poleward) edge to the outer (equatorward) edge of the oval, leading to secondary currents. The outward return flow (secondary current) on the dawnside then also produces auroral emissions. The dynamo effect converts the kinetic energy of the solar wind, which is the ultimate source of the magnetospheric current system, to electrical energy. The net power involved is of the order of 10^6 MW when integrated over the entire magnetosphere.

The potential drop across the polar cap (typically 50 kV) also applies an $\mathbf{E} \times \mathbf{B}$ force driving ionospheric plasma antisunward across the polar cap. These plasma velocities typically are 1 $km \cdot s^{-1}$ to within a factor of 2. This convection dominates the character of the polar ionosphere, and often even that of the polar thermosphere (the part of the upper atmosphere above 120 km). The flow lines of antisunward ionospheric convection close on themselves by returning toward the sun equatorward of the auroral region.

The degree to which this solar-wind-driven convection occurs smoothly, versus a series of spurts, and the implications of that issue for the underlying physical processes and boundary relationships are important topics (Cowley and Lockwood, 1992). Improved time-resolution measurements, over large areas of the polar regions, are needed to help resolve these issues.

14.9 STABLE SUN-ALIGNED ARC: ENERGETICS AND THERMAL BALANCE

We have discussed the optical character and electrical character of sun-aligned arcs (Figures 14.1d and 14.9). Discussion of the energetics and thermal balance of such arcs at this point will serve not only to present the properties of a common class of arcs but also to review the material of Sections 14.5 and 14.7 and consolidate it with the material introduced in the earlier sections of this chapter.

We can anticipate much about the physical characteristics of a sun-aligned arc. Recall the right-hand side of Figure 14.9b. As the conductivity changes more slowly than $\mathbf{E}$ across the sun-aligned element of an arc, the decrease in $\mathbf{E}$ from the dawnside to the duskside means that the horizontal-current component decreases across the arc. This can happen only if current flows out of the plane of the horizontal-current slab. In the ionosphere, this happens through upward magnetic-field-aligned currents. We can say that these arcs satisfy a simple Ohm's-law relation-

ship between **E** and the current it drives through the conducting ionosphere. Electrons are the current carriers. Because of their low mass, they are more mobile than ions. The electrons must be drawn into the ionosphere from the magnetosphere. The ultimate driver of all this is a solar-wind-driven mechanical force, generating an electric field in the magnetosphere that maps down magnetic-field lines to the ionosphere to drag plasma across the polar cap. Now, the presence of this sun-aligned sheet of incoming electrons causes the temperature of the electron gas to be higher above an arc (double or more) than elsewhere in the surrounding ionosphere. The actual increase depends on the incident flux of suprathermal electrons from the magnetosphere and on how well the lower ionosphere is thermally coupled to its thermospheric heat sink. For low electron densities, and thus low electron-gas cooling rates, the electron temperature increase may exceed 3,000K, leading the heated ambient electrons to enhance the 630-nm emissions.

As the cross-arc gradient of antisunward plasma flow velocity (cross-arc gradient of **E**) increases, the upward current out of the arc must increase to carry off the enhanced horizontal current converging on the arc. Field-aligned-current densities can grow by increasing the number and/or the speed of the current carriers. The latter effect causes precipitating-electron fluxes to become harder, as is observed. The fluxes of precipitating hard electrons increase E-region ionization. It then follows that intense arcs will have underlying E-region ionization, whereas weak arcs will not. It also follows that the particle energy flowing into intense, stable sun-aligned arcs can be estimated from measurements of either molecular optical emission (which can exceed kilorayleighs) or E-region electron densities (which can exceed 2×10^5 cm^{-3}). Recall that the global mean thermospheric EUV heating rate is 0.5 $erg \cdot cm^{-2} \cdot s^{-1}$. Electron-particle energy deposition within an intense arc exceeds this reference solar-radiation rate!

Plasma on the dawn edge of the arc is found to flow antisunward at about 1 $km \cdot s^{-1}$ and to decrease duskward across the arc. If thermospheric winds are light, the velocity difference between the ions and the neutrals can be used to calculate the Joule heating rate. It follows from the square-law dependence of the velocity difference that the heating will be concentrated near the high-plasma-velocity (dawn) edge of the arc. Measured ion and neutral velocities across an intense arc imply Joule heating rates of several ergs per square centimeter per second, concentrated near the dawn edge, with net energy into the arc exceeding that from the incident-electron source. This Joule heating rate has been confirmed, based on calculating the rate of ion heating required to maintain the approximately 1,000K by which ion temperatures have been observed to exceed those of the neutral gas. This energy that appears in the ionosphere in the form of heat is conveyed into the ionosphere in the form of electromagnetic energy (Poynting flux). Measurement of the Poynting flux simultaneously with Joule heating rates has demonstrated that the foregoing calculation is consistent with observation (Valladares

and Carlson, 1991). These physical arguments are not confined to sun-aligned arcs.

Polar-cap convection can take different forms, depending on whether the solar-wind magnetic field is nearly parallel or nearly antiparallel to the subsolar field of the earth. When these fields are parallel (i.e., the IMF is northward), sun-aligned arcs are seen to occur. Near the central polar cap, the sun-aligned boundaries of the arc drift toward dawn (dusk) for negative (positive) IMF B_y in the northern hemisphere. The connection of the arcs and the relation of their drifts to the coupling of the magnetosphere and the solar wind are challenging topics, but beyond the scope of this treatment.

ADDITIONAL READING

Akasofu, S.-I., and S. Chapman. 1972. *Solar-Terrestrial Physics*. Oxford: Clarendon Press.

Alfvén, H., and C.-G. Fälthammar. 1963. *Cosmical Electrodynamics, Fundamental Principles*. Oxford University Press.

Banks, P. M., and G. Kockarts. 1973. *Aeronomy*. New York: Academic Press.

Brekke, A., and A. Egeland. 1994. *The Northern Lights, Their Heritage and Science*. Dreyer: Grøndahl.

Chamberlain, J. W. 1961. *Physics of the Aurora and Airglow*. New York: Academic Press.

Eather, R. H. 1980. *Majestic Lights*. Washington, DC: American Geophysical Union.

Egeland, A., O. Holter, and A. Omholt. 1973. *Cosmical Geophysics*. Oslo: Universitetsforlaget.

Jursa, A. S. (ed.). 1985. *Handbook of Geophysics and Space Environment*. Springfield, VA: National Technical Information Service.

Kelley, M. C. 1989. *The Earth's Ionosphere*. San Diego: Academic Press.

Meng, C. I., M. J. Rycroft, and L. A. Frank (eds.). 1991. *Auroral Physics*. Cambridge University Press.

Rees, M. H. 1989. *Physics and Chemistry of the Upper Atmosphere*. Cambridge University Press.

Sandholt, P. E., and A. Egeland (eds.). 1989. *Electromagnetic Coupling in the Polar Clefts and Caps*. Dordrecht: Kluwer.

Størmer, C. 1955. *The Polar Aurora*. Oxford: Clarendon Press.

Vallance Jones, A. 1974. *Aurora*. Dordrecht: Reidel.

PROBLEMS

14.1 Assume that monoenergetic electrons at 3 keV precipitate parallel to the geomagnetic field and produce an auroral arc of 10^3 km in east–west extent and north–south width of 10 km, with a homogeneous surface brightness of 5 kR in the 391.4-nm band.

(a) Estimate the flux needed to produce this arc.

(b) Calculate the corresponding net downward energy in watts per square meter for this electron flux.

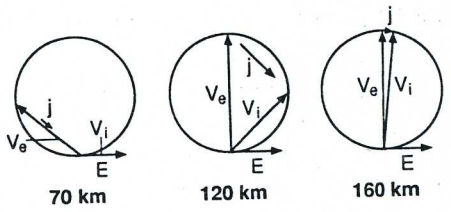


FIG. 14.12. Electric current j and the direction and magnitude of electron and ion drift ($\mathbf{v}_e$ and $\mathbf{v}_i$) relative to an ionospheric electric field $\mathbf{E}$ parallel to the x-axis (while $\mathbf{B}_E$ is along the z-axis) for three different heights.

14.2

(a) Discuss the hydrogen lines (H_α and H_β) in the auroral spectrum and explain why this aurora is diffuse in character.

(b) Discuss how the Doppler shift (see Figure 14.5b) can be used to estimate the average H^+ energy. Assume that a remote observer at an auroral latitude makes observations of the spectrum both parallel and perpendicular to the earth's magnetic field. A comparison of the spectra reveals a 1-nm shift in the wavelength of the peak intensity. Estimate the average velocity of the downflowing H^+ assuming 0° pitch angles.

(c) Assume that the H_α production versus H^+ energy can be estimated in the 0.4–100-keV energy range by a straight line on log-log coordinates, defined by $x = H_\alpha$ photon production/proton; $y =$ energy (keV); ($x = 1$, $y = 1$), ($x = 10$, $y = 15$). Calculate the flux needed to produce 100 R at H_α, for 100, 10, and 1 keV H^+-particle energies.

(d) Discuss why the hydrogen lines are so weak in the dayside cusp and cap.

14.3 Calculate the L values for the following auroral observatories (see Section 6.2.1):

Tromsø (geomagnetic coordinates 117.54°E, 66.96°N)
Ny Ålesund (geomagnetic coordinates 131.24°E, 75.31°N)
Nord (geomagnetic coordinates 112.0°E, 80.9°N)
Qaanaaq (geomagnetic coordinates 39.9°E, 86.8°N)

14.4 Discuss how the direction of the geomagnetic-dipole axis relative to the sun varies with time (diurnal and seasonal) and how that variation will influence the auroral occurrences at the observatories listed in Problem 14.3.

14.5 The electric current j and the direction and magnitude of electron and ion drift ($\mathbf{v}_e$ and $\mathbf{v}_i$) relative to an ionospheric electric field $\mathbf{E}$ parallel to the x-axis (while $\mathbf{B}_E$ is along the z-axis) are plotted in the Figure 14.12 for three different heights. The angle a between electrons and ions relative to $\mathbf{E}$ is given, for a charged particle k, by $a_k = \tan^{-1}(\Omega_{ck}/\nu_{kn})$ (Ω and ν are gyrofrequency and collision frequency). For a height of about 180 km, $\nu_{kn} \ll (\Omega_{ck})$; electrons and ions move in the same direction, with a velocity $\mathbf{E} \times \mathbf{B}/B^2$. This drift produces no net current. Discuss the importance of collisions for two cases. In case a, assume that the electron-neutral collision frequency greatly exceeds the electron gyrofrequency (as, for example, near 70 km). In case b, assume that the electron gyrofrequency greatly exceeds the electron-neutral collision frequency (as, for example, near 180 km).

14.6 Discuss how the Biot-Savard law can be used to estimate the magnetic disturbance at the ground (in nT) for a line current at 120 km. Assume different values for the conductivity and the electric field.

14.7 Current continuity at the equatorward boundary of the arc in Figure 14.11b can be written

$$E_x^A = \frac{\Sigma_P^C}{\Sigma_P^A} E_x^C + \frac{\Sigma_H^A - \Sigma_H^C}{\Sigma_P^A} E_y^C + \frac{J_\parallel}{\Sigma_P^A}$$

where the values inside and outside the arc are indicated by A and C.

(a) Discuss the different terms.

(b) We assume that conductivity gradients play a minor role when soft particles dominate (e.g., dayside cusp aurora). The equation then reduces to

$$E_x^A = E_x^C + \frac{J_\parallel}{\Sigma_P^A}$$

where E_x^C is the polarization electric field, and $J_\parallel$ is the Birkeland current (often given approximately by $\partial E_x/\partial x \cdot \Sigma_p$, if E_x^C plays a minor role).

Discuss the importance of these different terms in the E and F regions for nightside and dayside auroras.

14.8 Explain how auroral spectrometer measurements could be used to determine the composition and height distribution of the upper atmosphere between 95 and 300 km.