1. INTRODUCTORY OBSERVATIONS
I-1. AURORAL PHENOMENA

S.-I. Akasofu*

The great progress in auroral physics during the last few decades has created the awareness that most of what we call auroral phenomena are various manifestations of dissipation processes associated with the discharge of electrical power generated by the solar-wind/magnetosphere interaction. Here we review briefly the progress that has been made in understanding some of the basic processes in the solar-wind/magnetosphere/thermosphere/ionosphere interaction.

1. INTRODUCTION

As a star, the Sun is continuously emitting enormous amounts of energy into space. This energy emission takes several forms, the first of which is the familiar black-body radiation. The second mode of energy emission is the solar wind, which consists of protons with energies of about 1 kV and an equal number of electrons with energies of a few hundred electron volts. They stream out from the Sun at supersonic speeds. The solar wind tends to confine the Earth and its magnetic field into a comet-shaped cavity called the magnetosphere. As the solar wind interacts with the magnetosphere, as much as 10^6 MW of electrical power is generated, discharged, and subsequently dissipated, partly through that portion of the upper atmosphere called the polar ionosphere. Both the solar X-ray and ultraviolet radiations, the third mode of energy emission, are responsible for producing the ionosphere. Most of what we call auroral phenomena are various manifestations of dissipation processes associated with this discharge.

The discharge process produces, among many fascinating phenomena, visible emissions that we recognize as the aurora. In fact, of all the manifestations, the aurora is the only visible phenomenon. As described in the following chapters, a great variety of other manifestations occur and can be detected by specific instruments, such as magnetometers, ionosondes, and many satellite-borne instruments.

Three systems—the solar wind, the magnetosphere, and the ionosphere—interact, transmitting and transforming the solar-wind energy into energies of auroral phenomena, and eventually depositing most of it as heat energy in the ionosphere (Fig. 1). So far, the electrical connection between the magnetosphere and the upper atmosphere has been considered. Thus, its ionized component, the ionosphere, has been emphasized. However, the neutral component of the upper atmosphere also responds to the discharge process. Thus, the term thermosphere includes both the ionized and neutral component of the upper atmosphere above the mesosphere (to about 80 km). The importance of the thermosphere in auroral phenomena is emphasized in Chapter 2.

Sydney Chapman played the most important role in establishing the foundations of the scientific discipline to which all the participants of the International Conference on Auroral Physics belong. However, it is impossible to describe Chapman’s contributions to auroral science in a short paper. Thus, we confine ourselves to making a few remarks, and we limit the subject area to auroral science:

1. Chapman established the present concept of geomagnetic storms in terms of the initial phase and the main phase (1918).
2. He obtained the storm-time current system in terms of the equivalent (two-dimensional) currents (1918–35).
3. He published a theory of the night airglow and the formation of the ozone layer (1930).
4. With V. C. A. Ferraro, he published a theory of magnetosphere formation by proposing that solar-wind particles constitute a plasma and are not a cloud of individual particles (1931).
5. He published a theory of the formation of the ionosphere (1931).
6. With T. G. Cowling, he obtained the standard formulas for the ionospheric conductivities (1939) and published the classic treatise “The Mathematical Theory of Non-Uniform Gases” (1953).
7. With J. Bartels, he published the magnum opus “Geomagnetism” in 1940. It served as the basic reference and treatise until about 1970.

Each of these seven contributions can be regarded as fundamental. Even one such contribution may be considered to be sufficient by any single researcher in his entire scientific career. Those who are interested in Chapman’s contributions, not only in auroral physics but also in other fields, should refer to Sydney Chapman, Eighty, From his Friends, by Akasofu et al. [1968].

*Geophysical Institute, University of Alaska, Fairbanks, Alaska 99775-0800.

2. POWER GENERATION FOR AURORAL PHENOMENA

If the solar wind were a nonmagnetized plasma, we would not expect much more than the formation of the magnetospheric cavity. However, since the solar wind is a magnetized plasma (carrying the solar magnetic field, generally called the interplanetary magnetic field (IMF)), the interaction between the solar wind and the magnetosphere becomes very complex. Thus, progress has been slow in understanding this particular interaction process, which is called dayside magnetic reconnection.

The simplest situation occurs when the IMF is directed southward, so that it is antiparallel to the Earth’s magnetic field near the nose of the magnetosphere. It is a recent finding by Russell and Elphic [1979] that this interaction is not a steady process. Indeed, a more recent computer-simulation study confirms that it is basically a nonsteady process (Lee and Fu [1985]).

It is through dayside magnetic reconnection that some magnetic field lines from the magnetosphere are connected to solar-wind magnetic field lines across the magnetopause. We are indebted to Dungey [1961] for the present concept of the so called open magnetosphere. It is understood that solar-wind particles flow along the magnetopause, crossing the newly connected magnetic field lines, although the details have not yet been fully understood. This process is basically the same as that of a magnetohydrodynamic (MHD) generator. Thus, the entire magnetopause constitutes a gigantic natural generator that we call the solar-wind/magnetosphere generator. Individual solar-wind ions lose only a very small fraction of their kinetic energy by this interaction, but it is through this process that more than $10^6$ MW of power is generated. This amount is estimated on the basis of the total energy-deposition rate in the polar ionosphere. The total potential drop generated, about 100 kV, is estimated from the potential difference between the dawnside and duskside of the auroral oval.

3. MAGNETOSPHERE/IONOSPHERE INTERACTION

The dynamo process described above would have little significance if the ionosphere did not exist. Without the ionosphere, the dynamo has to power an open circuit. Many auroral phenomena occur because the power is transmitted to the ionosphere from the magnetosphere. Furthermore, the ionosphere is not simply a passive load, and the magnetosphere and the ionosphere constitute a complex feedback system. As the magnetosphere is filled with a very rarefied plasma, which is permeated by the Earth’s magnetic field, electric currents tend to flow along the magnetic field lines. Called Birkeland currents, these field-aligned currents (FAC) transmit the power generated by the dynamo to the ionosphere. Actually, the current system that connects the magnetosphere and the ionosphere is very complex, consisting of the primary (region 1) and secondary (region 2) currents [Iijima and Potemra, 1976]. The aurora is the result of this discharge process.

When the aurora can be seen from well above the northern polar region, it appears as a ring of luminosity around the geomagnetic pole, the auroral oval. The
auroral oval delineates approximately the area called the polar cap; the geomagnetic field lines that anchor to the polar cap are connected to the IMF lines across the magnetopause, and such field lines are termed open field lines. On the other hand, other geomagnetic field lines connect two points (one in each hemisphere) across the equatorial plane; such field lines are called closed field lines. The auroral oval consists of two parts: the first is the oval of discrete (curtain-like) aurora, and the second is the oval of the diffuse aurora (Fig. 2). In spite of the great progress made in auroral physics in recent years, one of the long-standing unsolved problems is how the region of discrete aurora and the diffuse aurora are connected by geomagnetic field lines from different plasma regimes in the tail of the magnetosphere. This is a subject of Chapter 4.

4. AURORAL POTENTIAL STRUCTURE

The magnetosphere and the ionosphere together constitute a complex interactive system. Alfvén waves and the FAC carry information between the magnetosphere and the ionosphere. However, in conducting the upward electric current from the ionosphere, the magnetic field lines have only a limited capacity because, like all charged particles in the Van Allen radiation belt, the current-carrying electrons have a helical motion along geomagnetic field lines. As the electrons approach the Earth (move into a region of stronger field), the pitch of the helical motion increases. As a result, the electron motion becomes completely circular at a certain height. At this mirror point, the electrons are reflected back and start to move upward, so that they cannot reach the ionosphere. However, it has been suggested that when the generator power and the FAC density become high enough, an interesting potential structure develops in this rarefied plasma environment. An electron moving upward along the center of the structure is accelerated toward the Earth (ionosphere) by the upward-directed electric field, increasing its velocity component along the magnetic field line. Thus, its pitch decreases, allowing the electron to reach the ionosphere. The potential drop in the structure is estimated to be a few kilovolts, so that the electron has acquired a few kiloelectron volts of energy by the time it emerges from the bottom of the structure. The energy spectral characteristics of auroral electrons have been studied extensively by rocket- and satellite-borne instruments in the past [Burch, 1988].

Mechanisms for accelerating energetic charged particles in natural conditions have greatly concerned astrophysicists, solar physicists, and auroral (magnetospheric) physicists, because such particles are common in cosmic, solar, and magnetospheric environments. It has been widely believed that it was impossible to maintain a significant electric field along magnetic field lines in a very rarefied plasma, making it impossible to accelerate charged particles by an electric field along the field line. Thus, as alternatives, a variety of MHD processes has been conceived. It seems, however, that, at least in the magnetosphere, an electric field parallel to the geomagnetic field lines can be produced and maintained in a limited region when an intense electric current flows along the field lines, as first suggested by Alfvén [1950]. There is no doubt that there are other processes involved in the acceleration of auroral particles. This important subject is discussed in Chapter 3.

The auroral potential structure and other acceleration processes are crucial in producing the aurora and associated auroral phenomena. A large number of current-carrying electrons reach the top of the ionosphere after being accelerated to a few kiloelectron volts. As a result, the electrons are capable of ionizing and dissociating atoms and molecules in the polar upper atmosphere. The potential structure also accelerates positive ions upward, producing upward-streaming ions. Complex plasma-wave/particle interaction processes also occur in the auroral potential structure, generating intense radio emissions in the kilometric range (0.6–1.5 MHz). These subjects are discussed particularly in Chapter 4, and also in all other chapters.

5. AURORAL PHENOMENA AT THE IONOSPHERIC LEVEL

As the electrons penetrate downward, they collide with atmospheric atoms and molecules, losing about
Figure 3—Flow chart for magnetosphere/ionosphere coupling and production of the aurora.
30 eV at each collision. The kinetic energy of a few kiloelectron volts is needed for auroral electrons to penetrate to an altitude of 100 km, where the atmospheric density is high enough that the optical emissions from the excited atoms and molecules can be detected by the naked eye (see Fig. 3). By colliding with atmospheric atoms and molecules, the precipitating electrons ionize and excite them, and dissociate molecules. Subsequently, a complex chain of chemical reactions takes place, including excitation by secondary electrons. The excited ions, atoms, and molecules are responsible for electromagnetic radiation over a wide spectral range extending from the extreme ultraviolet (EUV) to the infrared (IR). The most common emission from the aurora is a whitish-green light (5577 Å) that is emitted by oxygen atoms, excited in part by colliding secondary electrons. A red emission near the bottom of the auroral curtain partly from excited nitrogen molecules. The EUV (from excited oxygen and hydrogen atoms, ionized oxygen atoms, and others), UV (from ionized nitrogen molecules and others), and IR (from oxygen, nitrogen molecules and ionized nitrogen molecules, and others) emissions cannot be observed on the ground because they are absorbed by the intervening atmosphere. These topics are dealt with in Chapters 2 and 7.

The voltage produced by the dynamo process varies between 25 and 150 kV. This potential drop is transmitted to the ionosphere by FACs. The resulting electric field in the ionosphere and the energetic electrons produced by the auroral potential structure are responsible for most of the phenomena associated with the aurora.

The aurora has a thin, curtain-like form; its thickness (north-south) is about 1 km or less, while its lateral extent (east-west) is thousands of kilometers. The reason for this particular auroral form is simply that the accelerated electrons are confined to a thin sheet. How­ever, it is not known at present why the FACs tend to develop such a thin sheet. Multiple curtains (two or more) can form, but the cause is not well understood. The bottom of the curtain is at about 100 km altitude, because most of the precipitating electrons lose their penetrating power at that height as the atmospheric density increases rapidly downward. The thin-electron sheet exhibits a variety of instabilities, including a series of vortices and curls, as well as large-scale structures (such as westward traveling surges). This subject is discussed in Chapter 6.

In the polar upper atmosphere surrounded by the auroral oval, the electric field lies in the dawn-dusk direction, causing the ionospheric plasma to have an \( \mathbf{E} \times \mathbf{B} \) drift motion from the dayside to the nightside. Just outside the oval, the drift motion reverses direction (namely, from the nightside to the dayside in both the dawn and dusk hemispheres), resulting in two large-scale vortex motions that are, as a whole, called convection. There are several methods of observing the \( \mathbf{E} \times \mathbf{B} \) drift motion, including direct measurement by satellite-borne instruments. Barium-ion clouds released from a rocket at an altitude of a few hundred kilometers participate in the same \( \mathbf{E} \times \mathbf{B} \) drift motion and provide a method that has been used extensively to map the electric field in the auroral upper atmosphere [Heppner and Maynard, 1987; Heels, 1988; Fälthammar, 1989]. An incoherent scatter radar is a powerful ground-based observing device for detecting ion motions. The drifting ions impart their momentum to the neutral particles in the upper ionosphere, causing them to be dragged in the same direction, i.e., from the dayside to the nightside. The resulting motion is another type of atmospheric wind [Killeen and Roble, 1988].

In the lower ionosphere, only electrons can participate in the convection motion. As a result, two large-scale vortex currents occur in the polar ionosphere. They are particularly concentrated along the oval and are called the westward and eastward electrojets. Intense Joule heat is produced along the auroral oval. This heating is another cause of large-scale winds in the upper atmosphere. A number of researchers have made a detailed study of the atmospheric motions resulting from such heating [e.g., Rees and Fuller-Rowell, 1987]. The hot secondary-electron gas resulting from the ionization heats the atmospheric atoms and molecules, and the interaction becomes an important cause of large-scale winds in the polar upper atmosphere. As the energetic electrons are decelerated by the collisions, X-rays are generated that can be detected by a balloon-borne X-ray detector at an altitude of 30 km and by satellite-borne detectors from above. For details of this subject, see new textbooks by Akasofu and Kamide [1987], Kamide [1988], and Rees [1989].
Akasofu—Auroral Phenomena

A large-scale wavy motion is generated near the western end of the poleward motion. This wavy motion, called the westward traveling surge, propagates westward (toward the dusk-sunset line) with a speed of about 1 km s\(^{-1}\). In the morning sector, auroral curvatures appear to disintegrate into many patches. The poleward motion in the midnight sector lasts typically for about 30 min to 1 hour. After this poleward advancing curtain reaches its highest latitude, auroral activity begins to subside; however, the westward traveling surge often continues to propagate along the day-side part of the oval. During the last decade, excellent auroral images have been taken from satellites, with the result that many global features of the auroral substorm have been clarified. These findings are reported in Chapter 5.

The auroral substorm is the only visible manifestation of what we call the magnetospheric substorm. There are many other different manifestations of the magnetospheric substorm [Akasofu, 1977]. The electrojets are greatly intensified during the auroral substorm, causing intense geomagnetic disturbances. This phenomenon is called the polar magnetic substorm. The flow patterns of the electrojets have been studied extensively by Kamide [1988], Baumjohann et al. [1981], and others.

The causes of the magnetospheric substorm have been one of the major topics among magnetospheric physicists during the last two decades. Many theorists have speculated that magnetic reconnection in the magnetotail is responsible for the energy supply. In the magnetotail, the magnetic field is directed toward Earth in the northern half and away from Earth in the southern half; the magnetotail can be considered to consist of two solenoids, producing antiparallel magnetic fields. They have speculated that such an antiparallel magnetic-field system is intrinsically unstable and that the fields can spontaneously and explosively annihilate themselves. Thus, they have hypothesized that the magnetic energy accumulated and stored in the magnetotail would be suddenly converted into energy for the magnetospheric substorm by a process that is intrinsic to the magnetosphere. In fact, it has long been said that the magnetotail contains enough energy for many intense substorms and that a search should be made for internal processes that could trigger magnetic reconnection explosively.

However, it has become increasingly clear that the occurrence of magnetospheric substorms is at least partially controlled by the solar wind and the IMF. In other words, the growth and decay of magnetospheric substorms are controlled by the rise and fall of the power generated by the solar-wind/magnetosphere generator, which is a function of at least the speed \(V\) of the solar wind, the magnitude \(B\) of the magnetic field, and the polar angle \(\theta\) of the magnetic field vector. Time variations of the power equation are similar to those of the rate of total energy dissipation in the inner magnetosphere (which includes the ring-current injection rate, the Joule heat production rate, the kinetic energy injection rate of auroral electrons, etc.). The exact dependence of the power on these quantities is, however, a matter of great controversy. One empirical formula suggested by Perreault and Akasofu [1978] is given by

\[
\text{Power (MW)} = 20 \ V \ (\text{km s}^{-1}) \times B^2 \ (\mu\text{T}) \times \sin^2 (\theta/2)
\]

where \(\theta\) is approximately the polar angle \(\theta = 0^\circ\) for a northward-directed field, and \(\theta = 180^\circ\) for a southward-directed field. The equation has been theoretically confirmed by Pudovkin et al. [1986], Reiff et al. [1981] have also shown that the total potential drop across the polar cap is closely related to the power given in the equation above.

Among the solar-wind quantities that control the power of the solar-wind/magnetosphere generator, the angle \(\theta\) is, on the average, the most variable. Consider a simple situation in which circularly polarized Alfvén waves propagate along the IMF (which lie approximately in the equatorial plane); in this situation the angle \(\theta\) is most effective in modulating the power (since \(V\) and \(B\) do not vary in this situation). Obviously, the power will be highest when \(\theta = 180^\circ\), namely, when the IMF is directed southward. Therefore, the occurrence of magnetospheric substorms is most often associated with the southward turning of the IMF vector. When \(\theta \approx 180^\circ\), the intensity of substorms depends on the magnitude of the magnetic field, the greater the magnitude, the more intense is the magnetospheric substorm.

At present, we are still far from a firm understanding of processes that lead to the onset of auroral substorms. If magnetic reconnection is involved in substorm processes, it is not certain whether or not it is the primary cause [Hones, 1984] or an effect [Kan et al., 1988; Akasofu, 1989]. It is important to note that the onset is signaled by the sudden brightening of an auroral curtain in the midnight sector. Thus, one of the most interesting problems in this regard is how an intensified FAC along a narrow strip in the midnight sector can arise after the magnetospheric convection becomes enhanced. Several ideas on onset processes are presented in Chapter 5.

The size of the auroral oval is a function of the power (or the north–south component of the IMF). In the midnight sector, the latitude of the oval is about 67° or above when the power is less than 107 MW. As the
power increases to $10^6$ MW, the oval expands to 65° or lower. Figure 4 shows schematically how the auroral distribution varies as the north-south component $B_{NS}$ of the IMF varies from $+5$ nT to a larger negative value ($-7$ nT) and then to a very large positive value ($+20$ nT). At the beginning, the oval is small and much of the polar cap is covered by a weak glow (mostly subvisual) except in the vicinity of the geomagnetic pole. As $B_{NS}$ becomes negative (namely, as the power of the auroral generator increases), the oval expands rapidly, and the subvisual glow disappears except for a narrow belt just to the poleward side of the oval.

Then, a substorm begins. Bright and active auroral curtains advance toward higher latitudes. As a result, the area enclosed by the auroral oval contracts. Frank and Craven [1988] suggest that this is an indication that magnetic energy accumulated in the magnetotail (during the initial expansion of the oval) is released. When $B_{NS}$ becomes positive again, the substorm begins to subside and the oval begins to contract poleward. At the same time, a subvisual glow starts to fill the polar cap.

Many other interesting phenomena also occur in the magnetosphere during magnetospheric substorms. Par-
particles with energies of a few hundred kilovolts are produced in the magnetotail and stream along the magnetic field lines [Williams et al., 1985]. The distributions of the electrons in the magnetosphere undergo drastic changes. Together with magnetotail reconnection, these changes are expected to be directly or indirectly related to auroral dynamics during auroral substorms.

7. POLAR CAP DURING PERIODS OF NORTHWARD IMF

One of the important questions among magnetospheric physicists today is: what happens to the aurora when $\theta$ becomes almost 0° for an extended period? As the power of the solar-wind/magnetosphere generator decreases, the aurora becomes dim and the auroral electrojets become weak. However, an unexpectedly interesting auroral phenomenon takes place in this situation. There appear a number of auroral curtains and subvisual patches across the auroral oval that are parallel to the noon−midnight meridian. Such auroras are called polar-cap auroras and have recently been studied extensively by Frank and Craven [1988], Meng and Lundin [1986], and others. This phenomenon and many others associated with it (field-aligned currents, convection, etc.) cannot be simply understood in terms of the east−west component of the IMF plays an important role in determining the distribution of the aurora when the angle $\theta$ becomes small. The convection pattern also becomes significantly asymmetric with respect to the noon−midnight meridian.

8. NEED FOR A GLOBAL IMAGING OF THE MAGNETOSPHERE

High-time resolution global imaging of the aurora has made a major contribution in advancing magnetospheric physics during the last decade. It provides a "visible frame of reference" in studying individual substorms and in dealing with very complex magnetospheric phenomena, as well as with a number of vital quantitative parameters. In the past, despite the fact that the concept of the auroral oval has been useful as a natural frame of reference (rather than geomagnetic latitude) in sorting out a great variety of ground-based and satellite observations, auroral imaging had not necessarily been the top-priority project, compared with observations of "invisible" physical quantities, such as electric fields, magnetic fields, particle fluxes, etc., and even "visible" components, including auroral spectra. The present success in imaging the aurora by spacecraft has made it clear that imaging has been established as a truly vital tool in auroral and magnetospheric physics.

It may not be an exaggeration to say that imaging of the entire magnetosphere is needed for the future advancement of magnetospheric substorm studies. The complexity of magnetospheric phenomena and their three-dimensional aspects cannot easily be studied completely by observations made at single points by a few satellites. Although past satellite observations have suggested many interesting phenomena, proving their validity is not an easy task. There have been many two- and three-dimensional simulation studies to explain these phenomena. However, we must be cautious in inferring magnetospheric phenomena on the basis of two- or three-dimensional computer simulation studies alone. The simulation studies must be tested by theoretical analyses and by global observations.

In this context, like auroral imaging, it is expected that the global observation of the magnetosphere could make a major advance in magnetospheric substorm studies. It must be stressed that such global imaging does not reduce the importance of in situ observations of the magnetosphere by satellites and from ground-based stations. On the contrary, these observations will be truly complementary, just as satellite and ground-based observations are. Such a complementary effort was dramatically demonstrated by both in situ spacecraft (ICE) and ground-based imaging observation of P/Giacobini-Zinner comet. Figure 5 shows an anticipated view of the magnetosphere at a distance of 0.5 Earth radii ($R_e$). The figure was made by computing the line-of-sight densities, along different view directions, of sunlight resonantly scattered by oxygen ions.

9. CONCLUDING REMARKS

Perhaps, the research for truth follows the geometry of a polyhedron. Often, a researcher stands on one surface of it, while another researcher stands on another surface. For the first, everything on his surface appears to be consistent with his own model (a paradigm), while the same holds true for the second on his surface. Unfortunately, the first does not understand why the second attempts to understand a phenomena differently from his, or vice versa. Such a difference becomes a cause of controversy. Eventually, however, both will come to the realization that each has been studying only one aspect of a multifaceted phenomenon and that the phenomenon in which they are interested has at least two surfaces. Often, this task has been left to a new generation. This is what the present generation has done; it is why we have made such good progress in this field. Often, it is in the polyhedron that understanding of a natural phenomenon deepens, by revealing one new surface after another, which collectively constitute the truth. Alfvén, Birkeland, Chapman, Dungey, and many others have established