13 MAGNETOSPHERIC DYNAMICS

R. L. McPherron

13.1 INTRODUCTION

THE MAGNETIC FIELD of the terrestrial magnetosphere is produced by superposition of magnetic fields from a variety of sources. At the earth's surface, the most important source is the main field produced by currents within the earth's liquid core. Remanent magnetization and magnetization induced by the main field are also important. At ionospheric heights of about 120 km, electric currents, including the solar dynamo, the equatorial electrojet, the convection electrojets, and the substorm electrojet, are other important sources of magnetic field. Field-aligned currents link these ionospheric currents to the outer parts of the earth's magnetosphere. These currents, which transmit stresses from the outer magnetosphere to the ionosphere, produce effects on the earth indistinguishable from those produced by current systems flowing solely in the ionosphere. At about $3–5R_E$, particles drifting in the Van Allen radiation belts produce a westward ring current that, depending on the total energy of particles, reduces the horizontal component of the surface field [see equation (10.21)]. On the dayside of the earth, at distances of $10–15R_E$ at the boundary between the solar wind and the main field, there is a thin sheet of current, the magnetopause current. The magnetopause current almost entirely confines the earth's field to the magnetosphere. This current increases the magnetic field everywhere inside the magnetopause. Tangential stress between the solar wind and the earth's field drags the earth's field lines and plasma antisunward, forming a long magnetic tail behind the earth. This process produces another current called the tail current, which flows across the midnight meridian in the same direction as the ring current. Its main effect at the surface of the earth is to reduce the total field.

All of the currents above the ionosphere are controlled by the solar wind. The two most important controlling parameters are dynamic pressure, which depends on the solar-wind velocity and density, and the dawn–dusk component of electric field, which depends on the velocity and the north–south magnetic field. When any of these variables change, there are corresponding changes in the strength, location, and distribution of currents. The changes in currents are reflected by changes in the magnetic field at the earth's surface. Such changes are called *geomag-*

netic activity. The field of study that examines the relation between the solar wind and geomagnetic activity (i.e, solar-wind–magnetosphere coupling) is called solar-terrestrial physics. Geomagnetic activity can affect us in many practical ways. Radio communications, radar observations, electrical utilities, long-distance pipelines, and synchronous spacecraft are among the systems affected when geomagnetic activity reaches peak values.

The study of solar-wind coupling has a long history, beginning with studies of the relationship of geomagnetic activity to events on the sun. These studies utilize measurements of the earth's field made by instruments called magnetometers. Various techniques for measuring magnetic fields have been developed; some of the most common are described in Appendix 13A. Originally, these depended on magnets and photographic recording. Today, proton precession and fluxgate magnetometers produce digital signals that are recorded on magnetic media such as computer tapes. The original records from many locations spread over the entire globe are too complex and too voluminous to be manipulated easily. Magnetic indices have been developed to replace these data with crude measures of the strength of the various sources. Most studies of solar-wind coupling are performed with these indices. Descriptions of the most commonly used indices can be found in Appendix 13B.

Magnetic-activity indices are well correlated with events on the sun. Magnetic activity rises and falls, as does the number of sunspots during the 11-yr cycle of sunspot activity. Magnetic activity is also modulated by the location of the earth in its orbit around the sun. Annual and semiannual variations can easily be seen in magnetic indices. The rotation of the sun about its axis approximately every 27 days (as seen from the moving earth) is also an important factor. Activity indices display a very strong tendency to repeat every solar rotation.

Although the sun is the original cause of almost all geomagnetic activity, the primary means by which it exerts its influence is the solar wind. As discussed in Chapter 4, at the earth the solar wind blows radially outward at velocities of 300–1,000 $km \cdot s^{-1}$, carrying particles at densities of 1–50 cm^{-3}. Embedded in the solar wind is a magnetic field that normally lies in the ecliptic plane, with a strength of 3–30 nT. It has been found that some activity indices increase as the square of the solar-wind velocity. However, almost no magnetic activity is seen unless the interplanetary magnetic field (IMF) embedded in the solar wind has a southward component, antiparallel to the earth's magnetic field near the subsolar point on the dayside magnetopause. This dependence is quantified in the half-wave rectifier model. According to this model, which is an approximation to the best available models, whenever the IMF has a southward component, activity is proportional to the cosine of the clock angle of the IMF around the earth–sun line. There is no activity predicted for northward IMF. Even when the IMF is along the

expected Archimedean spiral, a southward component can occur because of the orientation of the earth's magnetic axis. Solar flares, coronal mass ejections, and the perturbed solar wind found at the interfaces between high and low-speed streams, referred to as corotating interaction regions, can also generate substantial north–south components of the IMF.

Attempts have been made to define precise solar-wind coupling parameters that can be used to predict the strength of magnetic activity. The procedures assume that the magnetosphere responds to the solar wind as if it were a deterministic system driven by the solar wind. Various techniques have been used to determine the best energy-coupling parameters and their most general linear relationships with geomagnetic activity. Such studies reveal that at time scales of less than about 3 h, only about half of the variance in magnetic-activity indices is predictable. The unpredictable residual is related to discrete events in which energy stored within the magnetosphere by the interaction of the solar wind with the magnetosphere is suddenly released. These events are known as magnetospheric substorms.

In the following sections of this chapter we shall discuss in greater detail the topics summarized here. In particular, we shall discuss the evidence for solar-wind control of the energy input that drives geomagnetic activity and identify the properties of the solar wind that are most closely coupled to the magnetosphere. A subsequent section extends the discussion to the topics of substorms and storms; this discussion focuses on the electric currents responsible for geomagnetic activity.

13.2 TYPES OF MAGNETIC ACTIVITY

13.2.1 Definitions

Electric and magnetic fields are produced by a fundamental property of matter: electric charge. Electric fields can be produced by charges at rest or moving relative to an observer, or by time-varying magnetic fields. Magnetic fields require moving charges. The two fields are different aspects of the electromagnetic force. The electric field at a point near a distribution of charges is defined as the force per unit charge when a positive test charge is placed at the observation point. It might seem that the magnetic induction **B** could be defined in a manner similar to **E**, as proportional to the force per unit pole strength when a test magnetic pole is brought close to a source of magnetization. However, no isolated magnetic pole has ever been observed. The consequence of this is that magnetic-field lines always close on themselves, rather than terminating on a pole. It is therefore common to define the magnetic field by the Lorentz-force equation, $\mathbf{F}=q(\mathbf{v}\times\mathbf{B})$, for the force on a particle of charge q with a velocity $\mathbf{v}$ [see equation (2.1)].

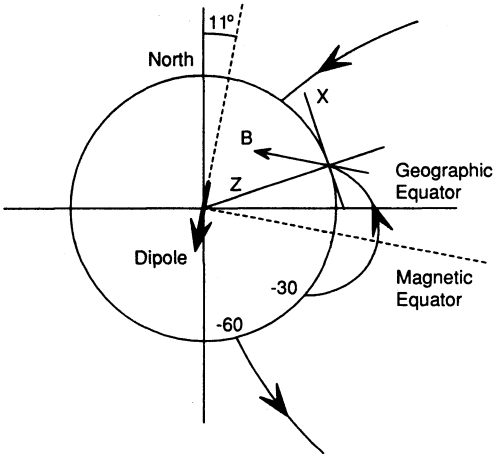

FIG. 13.1. Geometry of the earth's dipole magnetic field and the tangential plane with respect to which the earth's magnetic-field vector is measured.

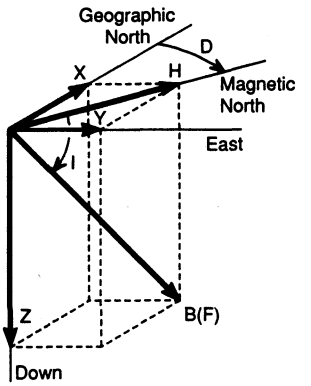

FIG. 13.2. Components of the geomagnetic field as measured in different coordinate systems.

The relationships of geomagnetic coordinate systems to the earth and its dipole magnetic field are illustrated in Figure 13.1. The figure shows an observatory located at 30° magnetic latitude on the geomagnetic prime meridian, which cuts through the eastern seaboard of the United States. Reference directions are provided by the horizontal plane tangential to the earth at that station and by a radial vector directed toward the center of the earth. In this meridian, the earth's dipole axis is tilted 11° toward the geographic equator.

The nomenclature used for the various components of the vector field in geomagnetism is summarized in Figure 13.2. **B** is the vector magnetic field; F is the magnitude or length of **B**, originally referred to as the total force; X, Y, and Z are the three cartesian components of the field, measured with respect to a geographic coordinate system; X is northward, Y is eastward, and, completing a right-handed system, Z is vertically down toward the center of the earth. The magnitude of the field projected in the horizontal plane is called H. This projection makes an angle D (for declination) measured positive from north to the east. The dip angle I (for inclination) is the angle that the total-field vector makes

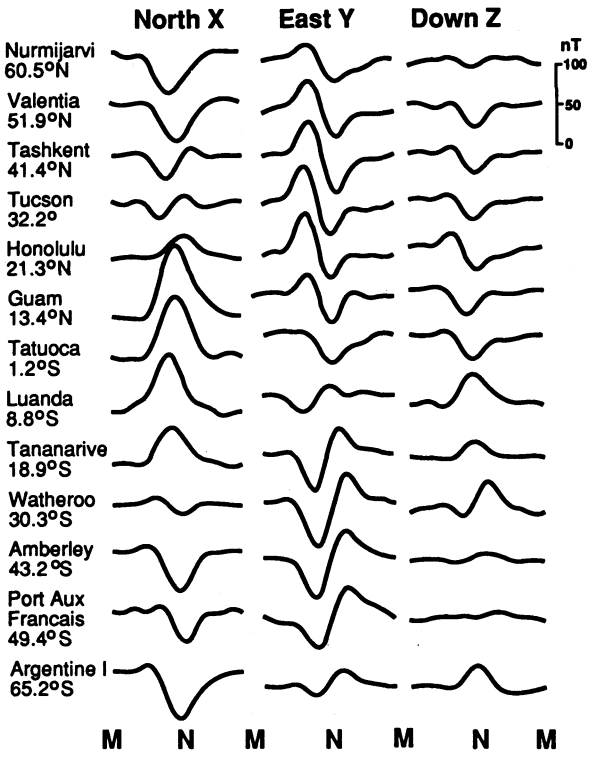

FIG. 13.3. Average quiet-day traces of the geographic north (*X*), east (*Y*), and vertical (*Z*) components of the surface magnetic field at stations distributed from north to south latitudes and plotted versus time. The times when the stations cross midnight (M) and noon (N) are indicated. (From Parkinson, 1983.)

with respect to the horizontal plane and is positive for vectors below the plane. It is the complement of the usual polar angle of spherical coordinates.

13.2.2 The Solar-Quiet Variation

The most thoroughly studied and best understood of the various types of geomagnetic activity is the diurnal variation S_q. On quiet days, every midlatitude observatory records a systematic variation in each field component (*X, Y, Z*), as illustrated in Figure 13.3. Stations at the same magnetic latitude, but separated in longitude, record similar patterns, but delayed in time by the earth's rotation. The pattern has considerable symmetry with respect to the magnetic equator and local noon, suggesting that the diurnal variation is produced by an ionospheric current system fixed with respect to the sun, under which the stations rotate. The form of this current system at equinox is shown in Figure 13.4. Two cells of current circulate around foci located at about 30° magnetic latitude. At the equator, both currents flow from dawn to dusk.

Fourier analysis of the diurnal variation at a single station reveals that it is primarily solar diurnal, with a smaller semidiurnal component. In addition, there is a weak, lunar semidiurnal (12 h 25 min) component. The primary cause of the diurnal variation is a dynamo created by

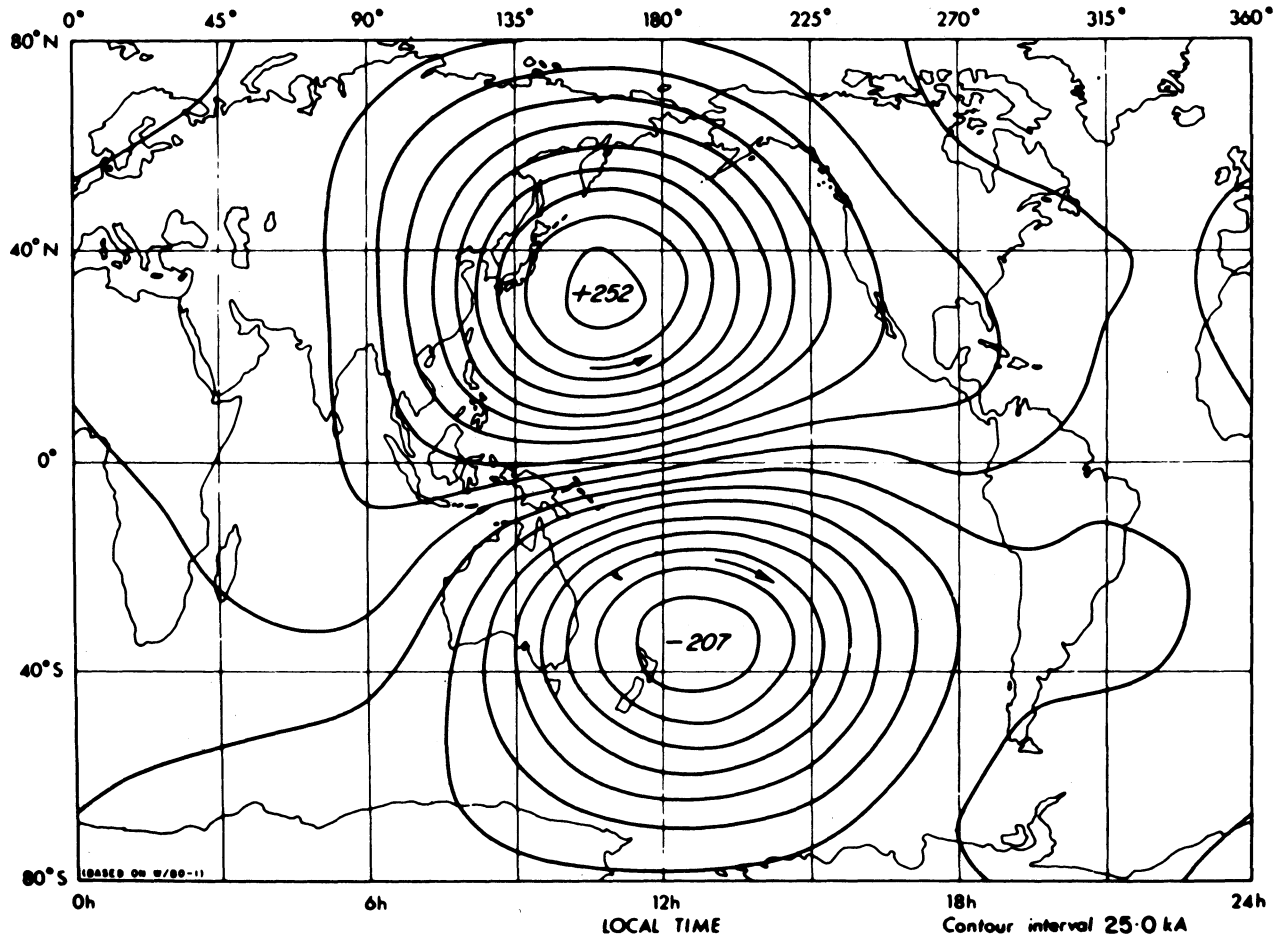

FIG. 13.4. Ionospheric current system responsible for the magnetic variations known as S_q. (From Parkinson, 1983.)

motion of electric charges in the ionosphere across the earth's magnetic-field lines. This motion is driven by winds in the ionosphere. These winds are driven by solar heating and lunar and solar tides. The dominance of the diurnal component suggests that solar heating is the primary force driving the dynamo.

13.2.3 Magnetospheric Substorms

The most frequent type of geomagnetic activity is referred to as a *magnetospheric substorm*. A substorm is the ordered sequence of events that occurs in the magnetosphere and ionosphere when the IMF turns southward and increased energy flows from the solar wind into the magnetosphere (McPherron, 1979, 1991; Akasofu, 1979; Rostoker et al., 1980). The most obvious manifestation of a substorm is the aurora, discussed in detail in Chapter 14. During a substorm, quiet auroral arcs suddenly explode into brilliance. They become intensely active and

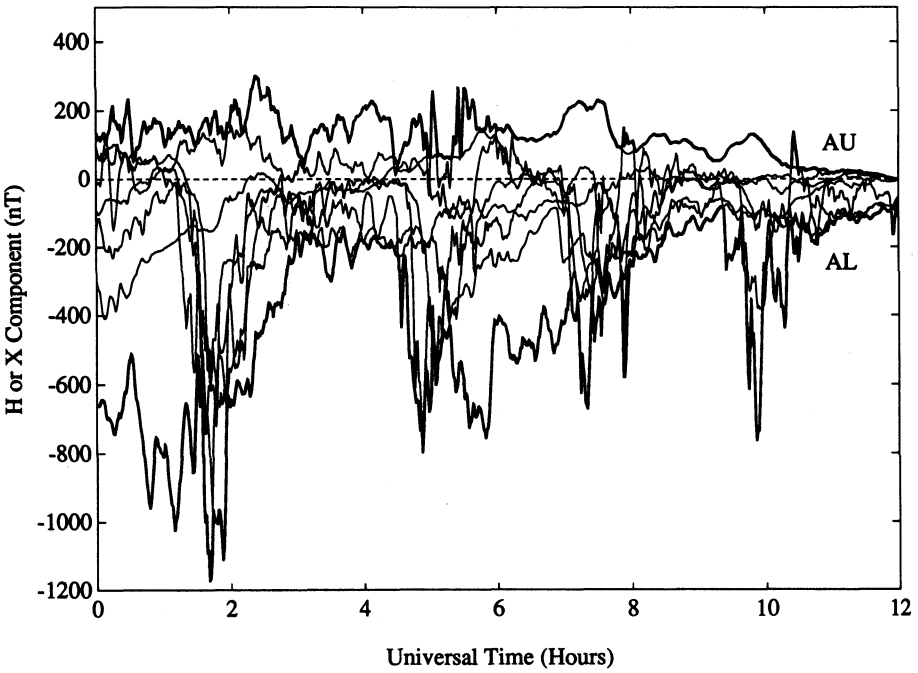

FIG. 13.5. Magnetic perturbations in the *H* component observed by auroral-zone observatories during a sequence of substorms on 3 May 1986. Positive perturbations are produced by a concentrated current or electrojet flowing eastward. Negative perturbations are produced by a westward electrojet.

colored. Over a period of about an hour, they develop through an ordered sequence that depends on time and location. Magnetic disturbances also accompany the aurora. On the surface beneath the aurora, a magnetometer will record intense disturbances caused by electric currents in the ionosphere. Figure 13.5 illustrates the form of these magnetic disturbances as a superposition of the perturbations recorded in the *H* component at a number of auroral-zone observatories. Stations located in the afternoon-to-evening sector record positive disturbances, whereas stations near and past midnight record negative disturbances relative to the field measured on quiet days. Applying the right-hand rule to currents assumed to be overhead leads to the conclusion that the currents are respectively eastward and westward toward midnight. These currents, later to be discussed further, are called electrojets, because the currents flow in concentrated channels of high conductivity produced at 120 km by the particles that generate the auroral light. Typical disturbances have amplitudes in the range of 200–2,000 nT, and durations of 1–3 h.

13.2.4 Magnetic Storms

When the coupling of the solar wind to the magnetosphere becomes strong and prolonged and geomagnetic activity becomes intense, as illustrated in Figure 13.5, a magnetic storm will develop. During a magnetic storm, auroral currents are almost continuously disturbed, such as seen on 3 May 1986 in Figure 13.5. The development of the storm is

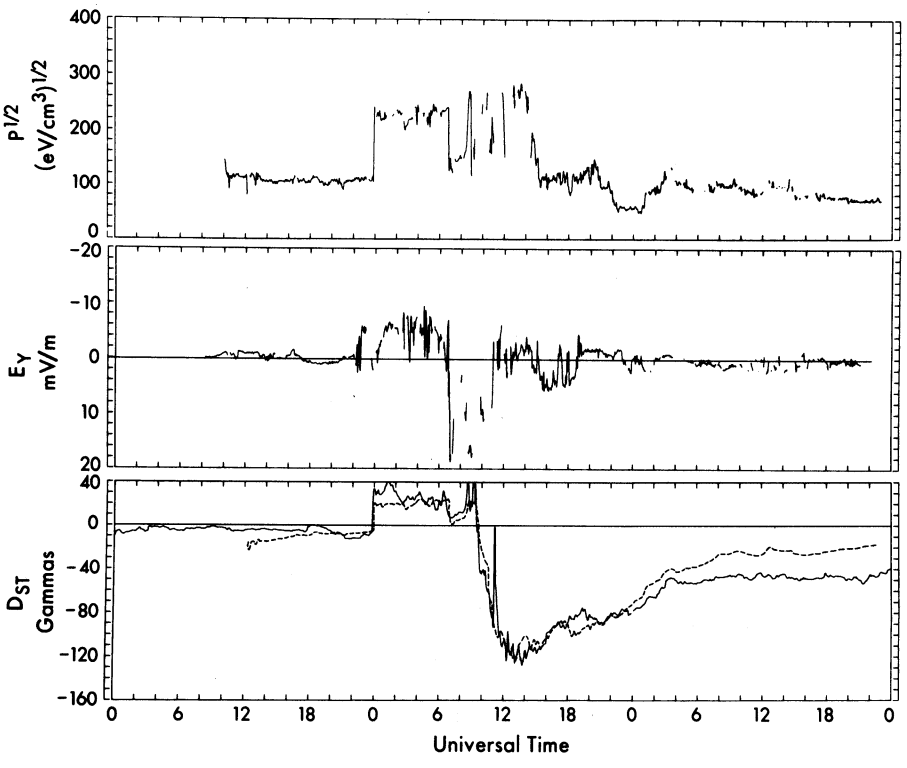

FIG. 13.6. Measurements of solar wind and magnetic field on the surface of the earth on 15–17 February 1967. Top: Solar-wind dynamic pressure. Middle: Dusk-to-dawn component of solar-wind electric field ($E_y = -uB \cos \theta$). Note that negative values of E_y are plotted upwards. Bottom: Effects of a magnetic storm as recorded in the D_{st} index (see Figure 13B.4). (From Burton et al., 1975.)

best identified at midlatitudes in the D_{st} index. D_{st} is defined as the instantaneous worldwide average of the equatorial H disturbance (see discussion in Appendix 13B). Figure 13.6 shows a classic storm signature (Burton, McPherron, and Russell, 1975) as recorded in D_{st}. A storm often begins with a sudden increase in magnetic field that may last for many hours. This initial phase is followed by a rapid and sometimes highly disturbed decrease in D_{st}. This is the storm main phase. Subsequently, D_{st} begins a rapid recovery, the first stage of the recovery phase. Eventually, a stage of long, slow recovery ensues. Typical storm durations are 1–5 days. The initial phase may be of any length, from zero to more than 25 h. Main phases last about 1 day. The recovery phase lasts many days. The distribution of storm magnitudes obeys a power law. Storms with D_{st} of order 50–150 nT occur almost every month. Several times per year the disturbance may reach 150–300 nT. Only a few times per solar cycle does one exceed 500 nT.

Studies of the spatial distribution of the storm disturbance field during the main phase indicate that it is nearly uniform over the entire earth and is directed parallel to the earth's dipole axis (i.e., southward). Such a disturbance would be produced by a current encircling the earth called a *ring current*. This ring current is created by particles drifting in the Van Allen radiation belts at distances of $3–5R_E$. The initial phase is produced by the magnetopause current. An increase in solar-wind dynamic pressure forces this current closer to the earth and increases its

strength. The perturbation magnetic field at the surface is northward. Recovery is caused by loss of particles from the radiation belts, leading to a decrease in the intensity of the ring current. In Figure 13.6, the correlation between dynamic pressure (top panel) and the initial phase of elevated D_{st} is obvious. Similarly, it can be seen that the main-phase decrease is a delayed response to an interval of strongly southward IMF (negative E_y).

13.2.5 ULF Waves

Magnetic activity can also take the form of periodic disturbances of the magnetic field, with periods of 1–1,000 s. These disturbances are called magnetic pulsations. They were originally discovered through microscopic examination of the minute fluctuations of the tip of a large compass needle; hence they are also referred to as magnetic micropulsations. Spacecraft observations have led to an understanding of their causal mechanisms. The subject of magnetic pulsations is discussed in Chapter 11, and examples are shown there.

13.3 MEASURES OF MAGNETIC ACTIVITY: GEOMAGNETIC INDICES

Magnetic activity at the earth's surface is produced by electric currents in the magnetosphere and ionosphere. Magnetic measurements at many locations provide one means of remotely sensing these currents and recording how they change with time. Because it is relatively inexpensive to make magnetic measurements, and because they are unaffected by weather, they provide an ideal way of routinely monitoring these currents. Over the past two centuries, a large network of more than 200 permanent magnetic observatories has been established. Data from these observatories and from temporary stations are frequently used to study magnetospheric phenomena, often in conjunction with in situ observations by spacecraft. However, the interpretation of raw magnetograms requires experience, and it is essential to accumulate data from many stations for a study. To simplify the task, several organizations routinely generate indices of magnetic activity. Ideally, an index is simple to generate, but useful, as it varies monotonically with some meaningful physical quantity, such as the total current flowing in a system.

Originally, magnetic indices were defined without much understanding of the processes that caused the disturbances. As understanding developed, new indices were defined, and older ones abandoned. Some indices are continued simply for historical reasons. They provide an ever-lengthening sequence of measurements that can be used for long-

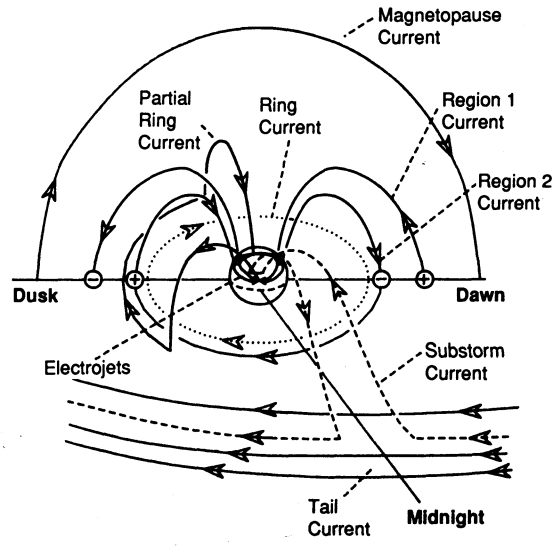

FIG. 13.7. Highly schematic representation of the various current systems linking magnetospheric and ionospheric currents and ultimately responsible for magnetic activity.

term studies of such phenomena as solar-cycle effects. Useful reviews of magnetic indices include those by Davis and Sugiura (1966), Baumjohann and Glassmeier (1986), Lincoln (1967), Mayaud (1980), Rostoker (1972), Troshichev et al. (1988), and Menvielle and Berthelier (1991).

Today, most of the current systems responsible for magnetic activity have been identified and studied. Figure 13.7 summarizes these in a highly schematic manner. They include various currents described either in this chapter or elsewhere in this book, including the magnetopause current, the tail current, and the ring current. Additional currents illustrated include the *partial ring current,* which flows near the equatorial plane principally near dusk, closing through the ionosphere, to which it is linked by field-aligned currents, and the *substorm current wedge,* a diversion of the tail current that also links into the ionosphere through field-aligned currents. The ionospheric portions of these current systems flow in enhanced-conductivity channels at high latitudes and are called the *auroral electrojets*. Sheets of field-aligned currents centered near dawn and dusk are referred to as *region-1* and *region-2 currents*. The higher-latitude region-1 currents flow into the ionosphere from the dawn sector and out at dusk. The lower-latitude region-2 currents have the opposite polarities. Not shown are polar-cusp currents, polar-cap closure of the electrojets, and currents associated with IMF-B_y effects. The primary sources of ground magnetic disturbances during substorms, such as those illustrated in Figure 13.5, are the two electrojets and the substorm current wedge. The sources of the midlatitude storm time variations shown by D_{st} in Figure 13.6 are the magnetopause current, the ring current, and the partial ring current.

13.4 SOLAR-WIND CONTROL OF GEOMAGNETIC ACTIVITY

Before the space age, it was well known that geomagnetic activity was related to events on the sun. How that influence was transmitted was not understood. Early theories due to Chapman and Ferraro (1930, 1931, 1932) suggested that solar flares ejected clouds of particles that traveled through the vacuum of space to interact with the earth's field. Despite obvious successes in correlating magnetic storms with solar flares, the flare concept did not account for perplexing storms that repeated at every 27-day solar rotation. Those storms were attributed to "magnetically effective" regions or M regions on the sun, even though nothing could be seen there to cause them (Chapman and Bartels, 1962, p. 410). It gradually became clear from studies of comet tails and other phenomena that there must be charged particles traveling outward from the sun at all times. With the launch of the first interplanetary probe, that speculation was quickly confirmed (Snyder, Neugebauer, and Rao, 1963), and a theory developed to explain what came to be called the solar wind (Parker, 1958). From that time on, it was obvious that the solar wind was the agent that transmitted the sun's influence to the earth. We begin our study of solar-wind–magnetosphere coupling with a brief review of three different variations apparently originating in changes at the sun: the solar-cycle variation, the annual variation, and the 27-day recurrence tendency.

13.4.1 Variations Related to the Sun

13.4.1.1 SOLAR-CYCLE VARIATIONS For several centuries, telescopic observations of the sun have shown that the number of spots on its surface oscillates, with a period of roughly 11 yr (see Chapters 1 and 3). More recent observations have demonstrated that there is a related oscillation in the pattern of the sun's main magnetic field. At every solar cycle, the main field reverses, so that in two spot cycles or 22 yr it returns to its initial orientation. Geomagnetic activity reflects the numbers of sunspots, as illustrated in Figure 13.8. The top panel shows a time-series plot of yearly means for the magnetic activity index AA (the first-difference time series of daily mean H at midlatitudes). Immediately below is the yearly sunspot number. High geomagnetic activity is obviously correlated with a high sunspot number. The same relation can be seen with monthly means in the bottom panel.

13.4.1.2 ANNUAL VARIATION The rotation axis of the earth is tipped at 23° to the ecliptic plane, causing a strong annual variation in the amount of sunlight that strikes the atmosphere. Because the dayside ionosphere is created primarily by photoionization, it might be expected

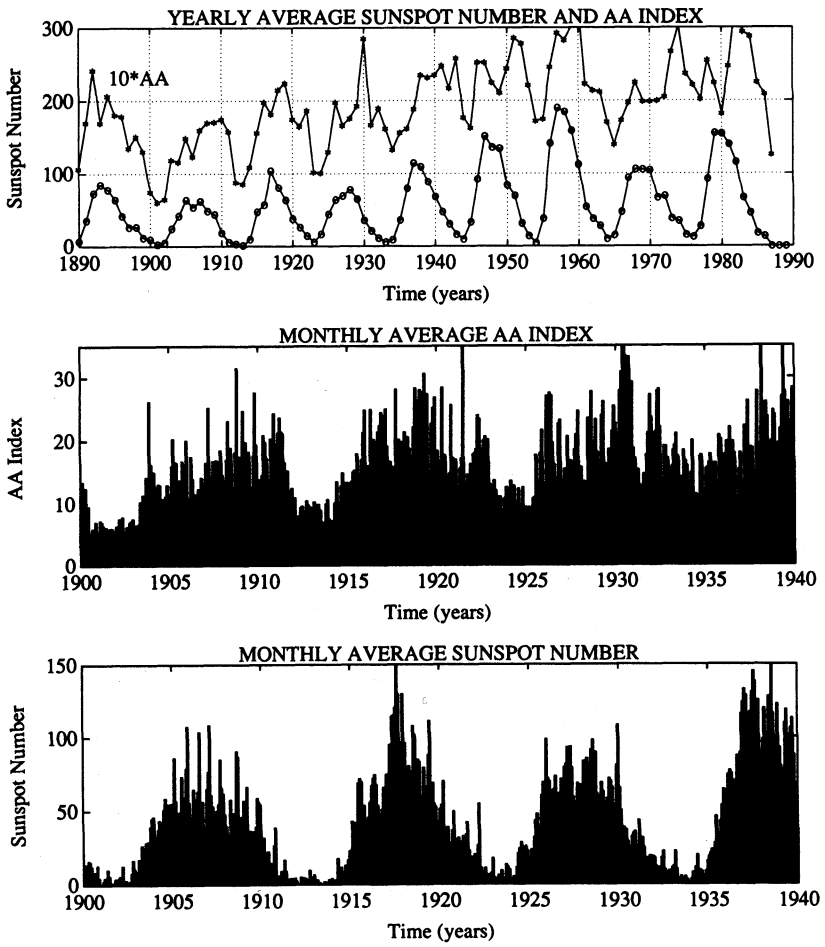

FIG. 13.8. The number of sunspots and the level of magnetic activity change together in both the annual and monthly means of sunspot and activity indices. (From Chapman and Bartels, 1962.)

that the conductivity of the ionosphere in a particular hemisphere would have an annual peak at its summer solstice. Therefore, any geomagnetic-activity index calculated from stations in that hemisphere should also show an annual modulation, peaking at the summer solstice. For indices that equally weight activity in both the summer and winter hemispheres, we might expect a semiannual variation, with peaks at each solstice. For indices independent of ionospheric conductivity, such as the storm time index D_{st}, it is not obvious what to expect. In fact, the ring current and other indices show very strong semiannual variations, as illustrated in Figure 13.9. Rather than one or two peaks per year at the solstices, there are two peaks near the equinoxes. At an equinox (March 22 or September 22), the sun crosses the earth's geographic equator. On March 5 and September 5 the earth crosses the sun's rotational equator. The phase of the activity peaks is not sufficiently well determined to decide whether the modulation in activity is related to the crossing of the sun's equator or to the crossing of the geographic equator.

The explanation of the semiannual variation put forward by Russell

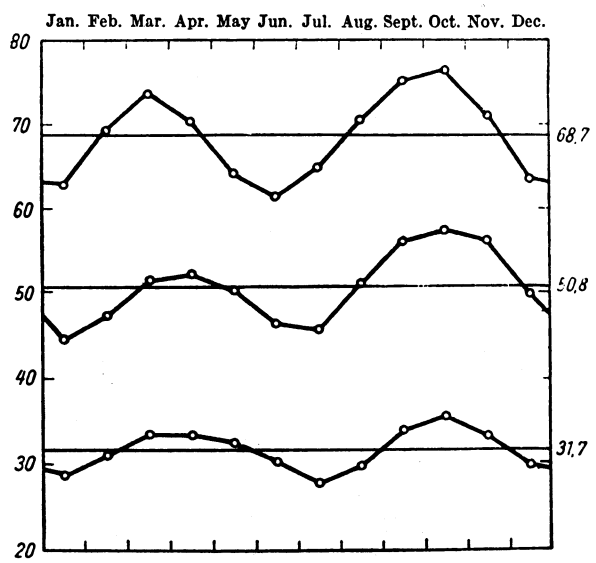

FIG. 13.9. The annual variation of magnetic activity as measured by monthly means of the v_1 index has two peaks near the equinoxes. The v_1 index is defined as the absolute difference between successive daily means in the H component of a station normalized so that the index has the same distribution as sunspot number. (From Chapman and Bartels, 1962.)

and McPherron (1973) is now generally accepted. Their argument is based on the idea that magnetic activity occurs preferentially when the IMF is southward relative to the dipole axis and that it increases with the size of the southward component, a matter that will be discussed later in this chapter. Russell and McPherron proposed that the variation in activity is controlled by the projection of the cross-flow component of the solar-wind magnetic field onto the cross-flow component of the earth's magnetic-dipole axis. At the spring equinox, the earth's dipole axis makes the largest possible angle, 35°, to the ecliptic normal at 2240 UT; at the fall equinox, this happens at 1040 UT. Depending on whether the solar-wind magnetic field **B** points inward toward or outward from the sun, the component of the ideal Archimedean spiral magnetic field, B, perpendicular to the flow will have a component B sin 35° that is either northward or southward relative to the earth's field at the subsolar point. If the projected field is southward there will be significant coupling. At other times of day, the north–south component of a field in the ecliptic plane is smaller, and the level of activity diminishes.

13.4.1.3 SOLAR-ROTATION VARIATIONS A third periodicity in geomagnetic variations is associated with the equatorial rotation rate of the sun as viewed from the moving earth. This is illustrated by the standard plot of the 3-h range index K_p presented in Figure 13.10. In this diagram, K_p values for successive solar rotations (27 days) are plotted one below another. The heights of the individual bars are proportional to the K_p index for a given interval. Intervals of intense activity are wrapped around in this display and are shown as superposed black bars. The sudden commencement of a magnetic storm is denoted by a dark triangle beneath the time line. A tendency for strong activity and weak activity to repeat at roughly the same time in each rotation is evident.

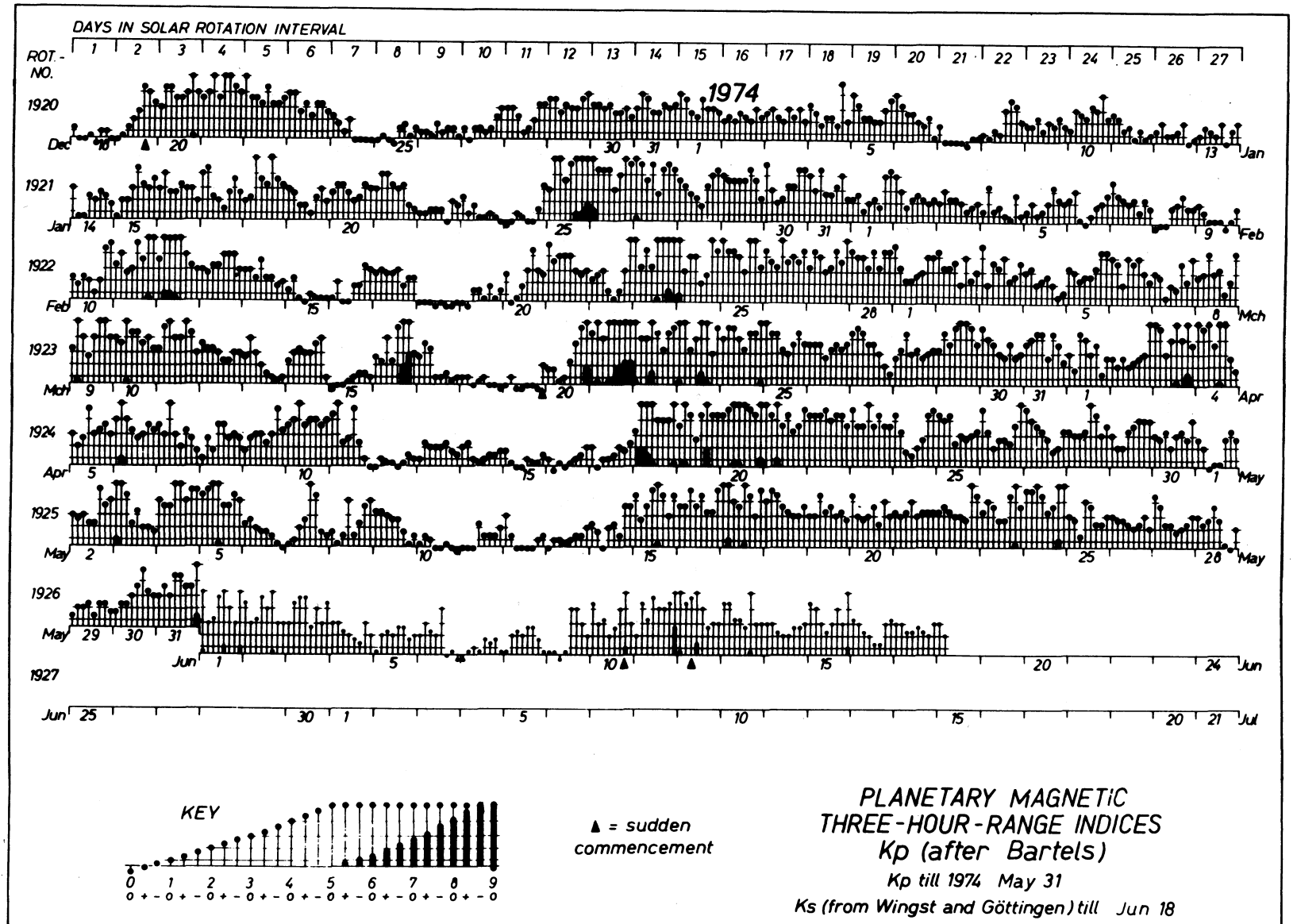

FIG. 13.10. The tendency for magnetic activity to recur with a period of 27 days is illustrated by charts of the planetary range index K_p.

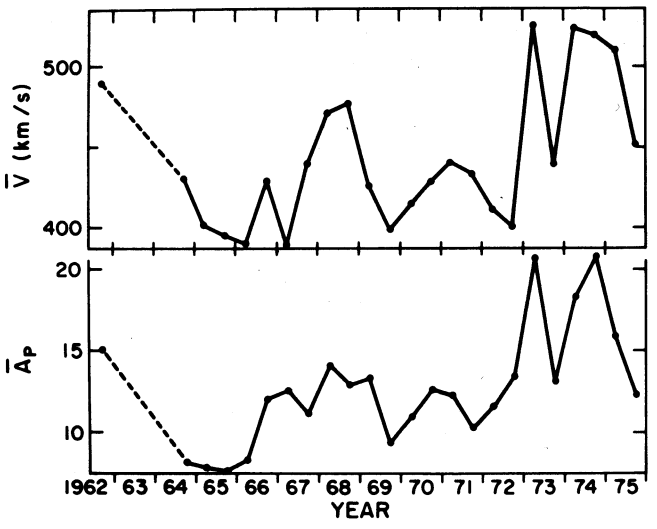

FIG. 13.11. Yearly averages of the A_p index are closely related to the solar-wind velocity, as shown by time-series plots. (From Crooker et al., 1977.)

The primary cause of the solar-rotation periodicity (27 days) in geomagnetic activity is that streams of high-speed solar wind recur at each rotation. As discussed in Chapter 4, large unipolar regions develop on the sun during the declining phase of the solar cycle. These regions produce coronal holes that become the sources of high-speed solar-wind streams. During the declining phase, the current sheet separating the north and south polarities becomes highly inclined to the solar equatorial plane. Thus, as these fast streams propagate toward the earth, they overtake slower plasma emitted from angular sectors of the sun ahead of the unipolar region, including the active streamer belt. An interaction region develops at the interface, and the magnetic field is compressed and tilted out of the plane of the ecliptic. As discussed in the next section, the combination of high solar-wind velocity and a southward IMF is particularly favorable for the development of geomagnetic activity.

13.4.2 Variations Related to the Solar Wind

13.4.2.1 DEPENDENCE ON VELOCITY One of the first major findings in the space era was the discovery of the solar wind. Measurements of its velocity quickly showed that long-term averages of geomagnetic activity were closely related to averages of the solar-wind velocity (Snyder et al., 1963). Figure 13.11 illustrates a result obtained after data on one complete solar cycle became available. The yearly means for the solar-wind speed v and the planetary index A_p closely track one another (Crooker, Feynman, and Gosling, 1977) (see Appendix 13B).

13.4.2.2 DEPENDENCE ON THE IMF Comparison of the waveforms of high-time-resolution magnetic indices and the solar-wind veloc-

ity contradicts the foregoing finding! There is virtually no relation between their joint variations. This perplexing result was not understood until it became clear that the solar-wind magnetic field was also important in controlling geomagnetic activity. That the IMF might be important was first suggested by Dungey (1961). He noted that if the IMF were antiparallel to the earth's field at the subsolar point of the magnetopause, then the two fields might be forced together and merge (see Chapter 9). He viewed this process as one in which the field lines of the earth and the IMF are first cut and then reconnected with a new topology. This view of the process leads to the name "magnetic reconnection." The amount of magnetic flux that reconnects will depend on the rate at which magnetic flux is transported to the subsolar point. One might therefore expect that geomagnetic activity would depend on the quantity uB_z, which is the rate per unit length in the equatorial plane at which the solar wind transports flux to the earth. Here, B_z is the magnitude of the z_{GSM} component. It is also the magnitude of the solar-wind electric field, according to the relation $\mathbf{E}=-\mathbf{u}\times\mathbf{B}$. However, it is expected that not all flux that flows to the magnetopause will merge. Much of it will slide around the magnetosphere. If we suppose that the fraction that merges increases with velocity, then the dependence on u^2 that is observed is not unreasonable.

These considerations led a number of investigators to examine the relationships of magnetic-activity indices to the solar-wind magnetic field. Schatten and Wilcox (1967) were among the first to show that scatter plots of K_p versus B_z revealed a linear dependence of K_p on B_z when B_z was negative (southward). Later, Arnoldy (1971) used correlation analysis between the auroral-electrojet index AE and various solar-wind parameters to obtain the result shown in Figure 13.12a. As shown by the legend, five different parameters, including u, n, B, and hourly integrals of B_n and B_s, were considered. Here n is the density of the solar wind, and B_n and B_s are, respectively, the magnitudes of the z_{GSM} component when it is greater than zero and less than zero. One parameter stood out: the integral of B_s over the hour preceding the hour in which activity was measured. Arnoldy also confirmed the result reported earlier by Hirshberg and Colburn (1969) that the correlation is highest when B_s is calculated in the geocentric solar magnetosphere (GSM) coordinate system. GSM coordinates have an x-axis pointing directly at the sun and an x–z plane that contains the earth's dipole axis (Figure 13.12b). In subsequent work, Murayama and Hakamada (1975) demonstrated that an optimum correlation is obtained with the function $\mathrm{AL}=Cu^2B\ (\cos\theta)\ U(\theta-\pi/2)$, where θ is the clock angle of the IMF about the x-axis (zero northward), U is the unit step function, and C is a constant of proportionality (see Appendix 13B for the substorm index AL). Figure 13.13 shows that there is almost no dependence of AL/u^2 on B_n, whereas there seems to be a linear increase in this quantity as B_s increases.

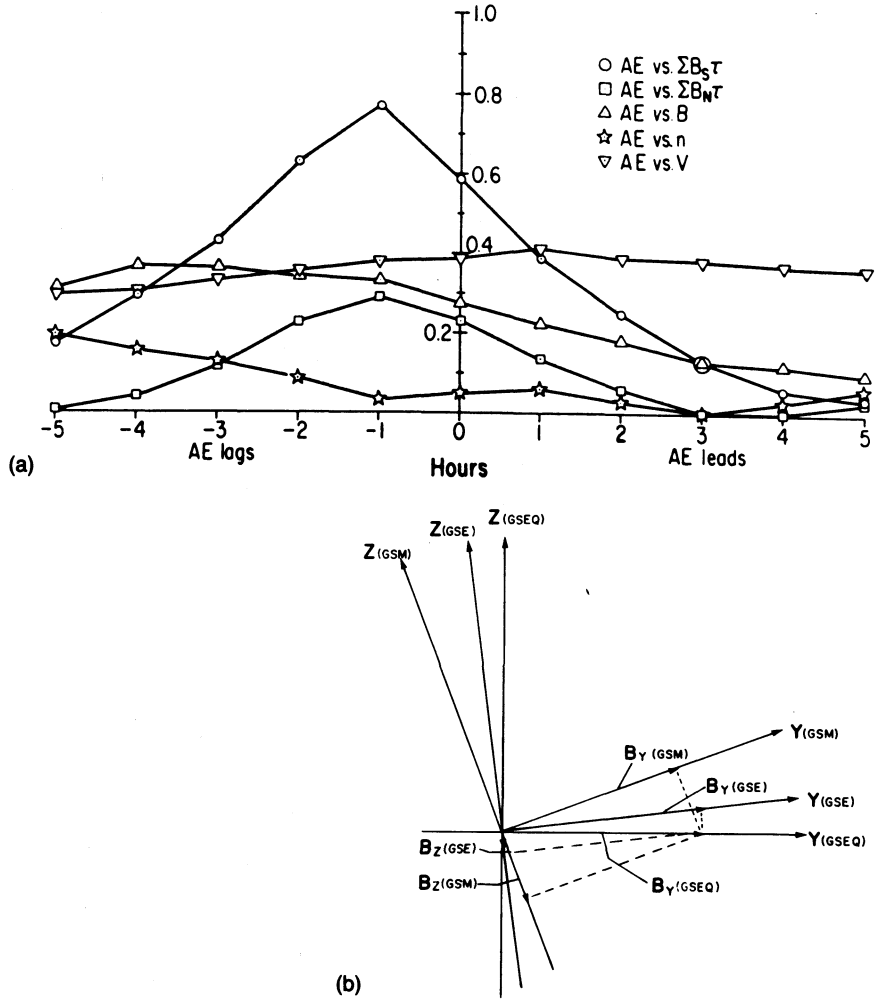

FIG. 13.12. (a) Linear-correlation functions demonstrate that hourly averages of AE are most closely related to the integral of B_s in the preceding hour. (From Arnoldy, 1971.) (b) Relationships between the GSM coordinate system, which maximizes the correlation between IMF B_s and geomagnetic activity, and the solar equatorial and solar ecliptic coordinate systems, which are less effective. The tilts as illustrated would apply on September 21 at 1040 UT, a time of maximum dipole tilt toward negative *y*.

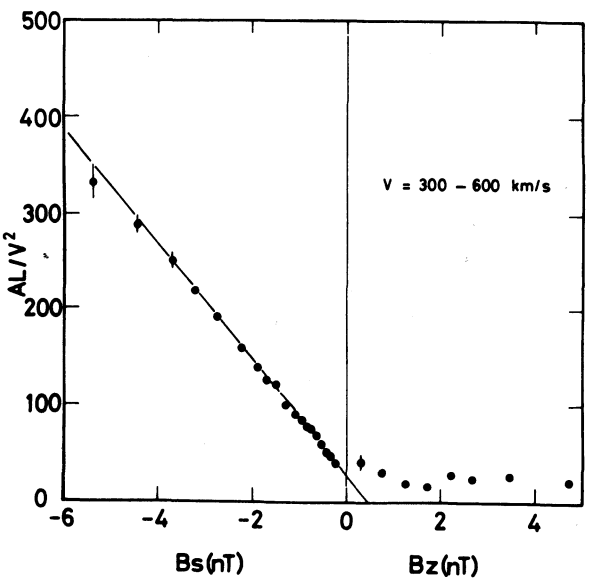

FIG. 13.13. Dependence of substorm magnetic activity, as measured by the AL index, on solar-wind velocity and the north–south component of the IMF. The plot shows the half-wave rectifier response of activity measured by AL and normalized by the square of the solar-wind velocity. The abscissa shows the hourly average B_z component when $B_z > 0$. For negative values, the observations are weighted by their durations to give B_s. (Murayama et al., 1980.)

13.4.3 Solar-Wind Coupling Parameters

13.4.3.1 EMPIRICAL STUDIES The results presented in Figure 13.13 are typical of those obtained in solar-wind–magnetosphere coupling studies. These studies make use of the long history of observations of the solar wind and geomagnetic activity to derive empirical relations. The standard approach uses statistical techniques to define relevant variables, determine the functional dependence on these variables, and then combine the variables into coupling parameters. The analysis assumes that the relation between the solar wind and the magnetic index is linear, or that it can be transformed to a linear relation piecewise by an appropriate choice of the solar-wind coupling function, as shown in Figure 13.13. This assumption allows us to treat the various processes as an electronic "black box" characterized by a transfer function. As discussed later, if this assumption is valid, then it provides a straightforward procedure for predicting geomagnetic activity.

13.4.3.2 DIMENSIONAL ANALYSIS The good correlation between magnetic-activity indices and the properties of the solar wind implies that the solar wind is the source of energy that drives the various magnetospheric processes. That observation led Akasofu (1980) to propose that both the coupling function and the activity indices should be expressed in units of power. Vasyliunas et al. (1982) studied the implications of this constraint and concluded that the principle of dimensional similitude provides a guide for choosing the variables. This principle states that the input function must be the product of (1) an energy flux relevant to the process, (2) the area on which it is incident, and (3) an arbitrary function of dimensionless variables important to the process that transfers energy to the magnetosphere. For magnetic reconnection, the relevant parameters were identified as the solar-wind kinetic-energy flux, the magnetopause cross-sectional area, the clock angle of the interplanetary magnetic field, and the solar-wind Alfvén Mach number M_A. For simplicity, the dimensionless function was taken as a separable function of Mach number and clock angle, so that the power input to the magnetosphere P is given by

$$P = ku(\rho u^2) l_c^2 (M_A^{-2})^\alpha g(\theta)$$

where $u(\rho u^2)$ is the flux of solar-wind kinetic-energy density, l_c^2 is an area equal to the square of the standoff distance to the subsolar magnetopause, $(M_A^{-2})^\alpha$ is the first term in a power-series expansion of a general function of inverse Mach number taken as a small parameter, $g(\theta)$ is the angular gating function for reconnection, and k is a dimensionless constant of proportionality corresponding to the fraction of the total flux incident on l_c^2 that reconnects.

The foregoing assumptions lead to predictions of the functional form of the solar-wind power coupling parameter. Any choice of the coupling

exponent α produces a quantity that is expressed in units of power. For example, $\alpha = 0.5$ corresponds to a coupling parameter of the form

$$P = \rho u^3 l_c^2 M_A^{-1}\, g(\theta) = \text{constant} \cdot p_{SW}^{1/6} E_{SW} g(\theta)$$

where p_{SW} and E_{SW} are, respectively, the solar-wind dynamic pressure and electric field. This expression is close to the quantity uB_s used in many statistical studies. Kan and Akasofu (1982) carried out the analytical procedure recommended by Vasyliunas et al. (1982) and concluded that $\alpha = 1.0$. However, Bargatze et al. (1985) found $\alpha = 0.5$. The discrepancy is explained by the fact that Kan and Akasofu did not allow the size of the magnetosphere to scale with dynamic pressure, as required by the principle of dimensional similitude. Thus, it seems that the best power coupling parameter is more nearly proportional to the solar-wind electric field.

Actually, there seems to be no need to require that the input coupling parameter have energy (or power) units. Most of the available indices of magnetospheric activity, such as AE, do not have these units. These indices will almost certainly not be linearly proportional to energy input. Empirical studies, such as those summarized by Maezawa and Murayama (1986), have demonstrated that the AL index is best predicted by a function proportional to u^2B, which does not have dimensions of power.

13.4.3.3 LINEAR PREDICTION FILTERING Most empirical studies of solar-wind coupling have found input parameters that are approximately linearly related to a given activity index. If the relation is truly linear and the system is time-invariant, then the technique of linear prediction filtering can be used to describe the system. This technique treats the solar-wind–magnetosphere system as a black box. Time histories of the solar-wind input and magnetic-activity-index output can be used to determine a transfer function. The Fourier transform of the transfer function is the impulse response of the system. Fourier-transform theory shows that the impulse response determined from this analysis can be convolved with any input at a later time to predict the output of the system; hence the name "linear prediction filter." As discussed by McPherron et al. (1988), this technique works well for the D_{st} index, but predicts less than half the variance in the AL index. Much of the unpredictable variance is correlated with the onset of the substorm expansion phase, suggesting that this phase of the substorm is not directly driven by the solar wind.

13.4.3.4 SOLAR-WIND TRIGGERING OF SUBSTORMS In most geomagnetic activity it is impossible to determine what causes the sudden increase in dissipation that accompanies the magnetospheric reconfiguration that is known as a substorm expansion. Akasofu (1979) attributed this to an increase in an energy-input parameter above some critical threshold. He speculated that increases in energy input cause

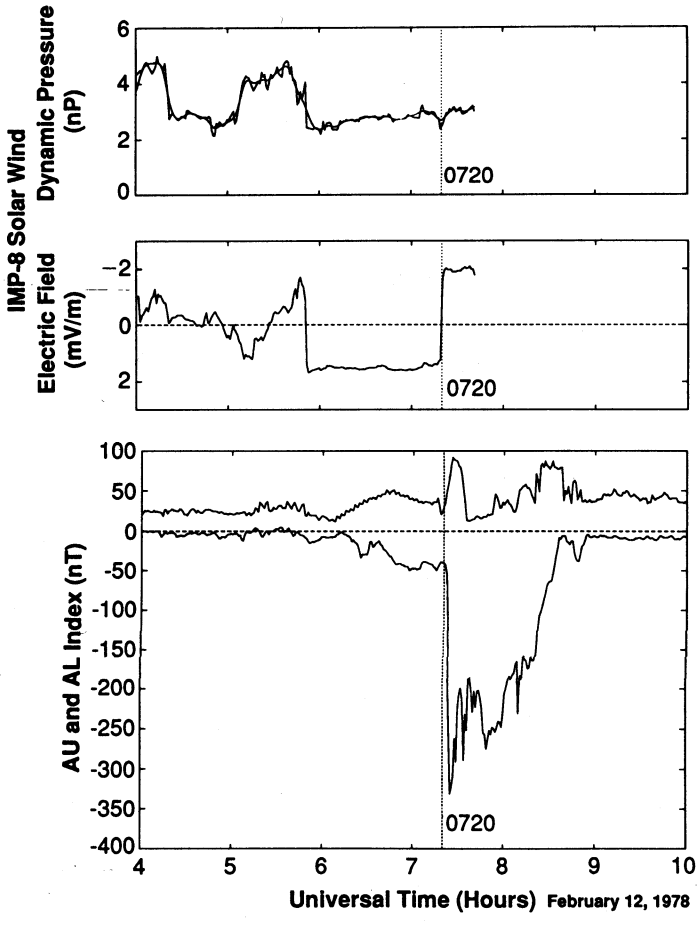

FIG. 13.14. Sudden changes in the solar-wind dynamic pressure and, more frequently, northward turnings of the IMF can trigger the sudden unloading of stored energy, as evidenced by changes in the AE-related indices. A sudden decrease in the AL index is correlated with a northward turning of the IMF, which here appears as an increase of the electric field *E*. Other events show that either phenomenon alone can cause the same effect in AE.

increases in the intensities of field-aligned currents, until they reach a threshold where they become unstable and generate field-aligned potential drops (Akasofu, 1980). These potentials accelerate electrons into the ionosphere, creating discrete arcs and enhanced conductivity. This enhancement draws more current, leading to larger potential drops, greater conductivity, and so forth.

Two processes are known to cause the sudden onset of dissipation. These are sudden solar-wind dynamic-pressure pulses (Burch, 1972; Kokubun, McPherron, and Russell, 1977) and sudden northward turnings of the IMF, which are equivalent to decreases in the dawn–dusk component of the solar-wind electric field (Caan, McPherron, and Russell, 1977; Rostoker, 1983; McPherron et al., 1988). Figure 13.14 shows an example of a sudden change in dissipation apparently triggered by a sudden northward turning. For both types of changes, it has been found that the IMF must have been southward for about an hour prior to the triggering event for an onset to occur. This strongly suggests that there is some threshold of stored energy or strain in the magnetospheric configuration required to produce the explosive phase of a substorm. In

the case of northward turnings, the solar-wind input parameter decreases immediately, and the field-aligned currents driven by the solar wind decrease as well. We would thus expect dissipation to decrease, not explosively increase as is often observed. For such cases there must be some reservoir of stored energy feeding the enhanced dissipation. As we shall show later, this reservoir is the tail magnetic field.

13.5 MAGNETOSPHERIC CONTROL OF GEOMAGNETIC ACTIVITY

In the preceding sections we have examined magnetic activity as if the magnetosphere were a passive electric circuit that transforms the solar-wind input into a magnetic-activity index. In fact, the magnetosphere is a highly dynamic system that undergoes a more or less predictable sequence of changes each time the IMF turns southward. This sequence is called a *magnetospheric substorm*. When the IMF remains southward for an extended interval, auroral currents become continually disturbed, and the ring current grows with time. The ring current causes a strong decrease in the surface magnetic field, a signature that is known as a *magnetic storm*. In this section we shall describe the phenomena associated with substorms and storms in greater detail and then present a physical model of what causes them.

13.5.1 The Auroral and Polar Magnetic Substorms

13.5.1.1 THE AURORAL OVAL The disturbances known as substorms are most clearly seen in the auroral ovals. These are two oval-shaped bands roughly centered about the north and south magnetic poles, within which bright, active aurorae and strong magnetic disturbances are observed. The auroral oval as seen from space by the *DE-1* auroral imager is shown in Figure 14.1d. If the location of this band of auroral luminosity is plotted in the invariant magnetic-latitude–magnetic-local-time coordinate system (Wallis et al., 1982), it becomes a circle of about 20° radius centered about a point located 4° tailward of the instantaneous location of the dipole axis (Holzworth and Meng, 1975). In geographic coordinates, however, it is oval-shaped; hence its name. The region poleward of the oval is called the polar cap. Field lines in this region are open to the solar wind and are connected to the lobes of the geomagnetic tail (Paulikas, 1974).

13.5.1.2 DISCRETE AND DIFFUSE AURORAE Two distinct classes of aurorae are observed in the nightside auroral oval. Near the equatorward edge is the diffuse aurora notable for its lack of structure. Farther north and predominantly in the evening sector is the discrete

aurora. Discrete aurorae are made up of long east–west bands of luminosity, with very small north–south extensions. Generally, those extensions are less than the 50–100-km resolution of current imagers, so that they cannot be resolved in images obtained with high-altitude spacecraft. The ionospheric signatures of auroral activity are discussed more fully in Chapter 14.

13.5.1.3 THE AURORAL SUBSTORM For many years it was thought that the type of aurora observed depended only on the latitude and local time of the observer and that changes in aurorae resulted primarily from the rotation of the earth. During the International Geophysical Year (1957), arrays of all-sky cameras were placed around the auroral oval to record what actually occurred. Analysis of the pictures established that the auroral type at a given location depended on universal time (UT), as well as on local time (LT) and latitude. A phenomenological model describing this development was called the auroral substorm (Akasofu, 1964). Figure 13.15 schematically illustrates the main features of this model.

In the model of the auroral substorm, activity begins from a quiet state consisting of multiple arcs drifting equatorward (panel A). The first disturbance is a sudden brightening of a portion of the most equatorward arc somewhere in the premidnight sector (panel B). This event is called the *onset* of the auroral substorm. The brightening expands rapidly westward and poleward (panel C). Within a short time a bright bulge of auroral disturbance forms in a broad region spanning the midnight sector close to where the aurora originally brightened. Within the bulge the aurora is very dynamic. Arcs appear and disappear; patches form and pulsate. Most arcs develop drapery-like folds that rapidly move along the arc. Lower borders of the arcs may become intensely colored. The interval of time during which the disturbed region is growing is called the *expansion phase* of the substorm. Eventually the auroral bulge develops a sharp kink at its westward edge, where it joins with the bright arc extending farther westward. This kink often appears to move westward, becoming more pronounced with time; hence it is called the *westward-traveling surge* (panels C and D). At the eastern edge of the bulge, torchlike auroral forms appear, extending poleward from the diffuse aurora and drifting eastward (panel D). These forms are called *omega bands,* from the shape of the dark regions defining their poleward borders. At the equatorward edge of this eastern region, dim pulsating patches of aurora appear, drifting eastward (panel E). After about 30–50 min, the auroral activity ceases to expand poleward, and the expansive (expansion) phase of the substorm has ended (panel E).

With the end of the expansion phase, auroral activity begins to dim at lower latitudes in the oval, and quiet arcs reappear. To the west, the westward-traveling surge degenerates, and a westward-drifting loop re-

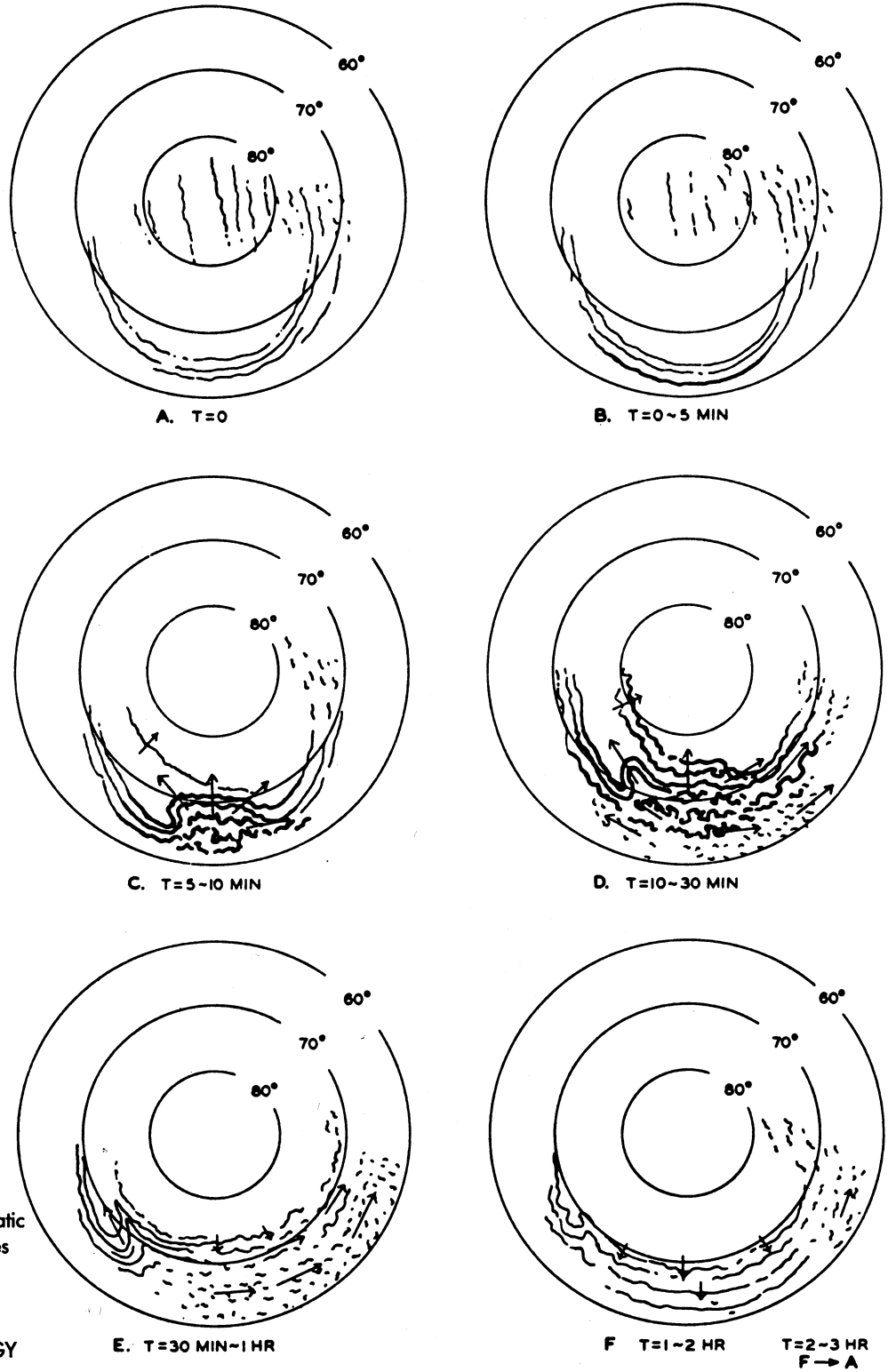

FIG. 13.15. Schematic representation of six stages in the development of an auroral substorm, as determined from all-sky camera data during the IGY (1957). (From Akasofu, 1964.)

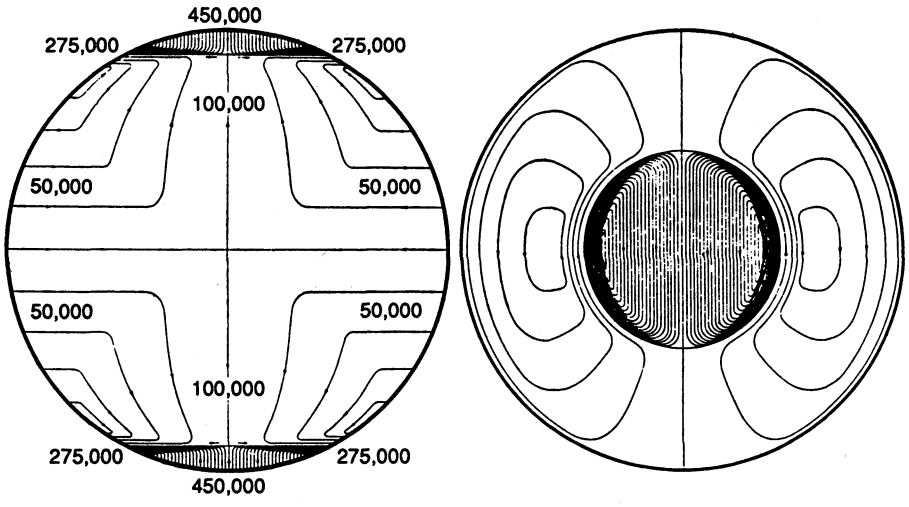

FIG. 13.16. An early view of the form of the ionospheric current system responsible for positive and negative bays in auroral-zone ground magnetograms. (From Chapman and Bartels, 1962.)

places it (panel E). In the morning sector, a pulsating aurora proceeds for some time. This phase of a substorm lasts for about 90 min and is called the *recovery phase*.

More recent observations with spacecraft imagers have modified this picture to some extent. Most significant is the observation that typical substorm expansions comprise multiple intensifications of the aurora, rather than one continuous expansion (Rostoker et al., 1987b). Each of these intensifications produces a new westward-traveling surge. However, the satellite observations demonstrate that frequently the surges do not travel far from where they form, and they do not proceed continuously, but rather in steps (Kidd and Rostoker, 1991). It appears that the original concept was incorrect because of spatial and time aliasing by the limited network of all-sky cameras. A more correct picture is that successive surges tend to form progressively farther west and farther poleward, giving the appearance of continuous expansion. In addition, it now appears that intensifications continue to occur on the poleward edge of the auroral bulge long after recovery has begun at lower latitudes (Hones et al., 1987).

13.5.1.4 THE CONVECTION ELECTROJETS Magnetic activity is associated with the auroral disturbances described by the auroral-substorm model. These disturbances were first extensively studied by Birkeland (1913), who called them elementary polar magnetic storms. Magnetogram recordings of these disturbances became known as geomagnetic bays, because their waveforms in a relative-amplitude-versus-time plot resembled the bays on a coastline. Positive bays in the H component usually are recorded by stations in the afternoon-to-premidnight sector, and negative bays by stations in the midnight-to-morning hours. A typical isolated bay disturbance lasts about 3 h.

Early studies of bay disturbances by a limited network of stations

were statistical. Average disturbances seen by a single station at each hour of local time were calculated. Using stations distributed at different latitudes, a complete pattern of average disturbances occurring in polar and auroral regions was constructed. Figure 13.16 illustrates the result of such averaging, as reported in an early study. Contour lines show the ionospheric current, which could produce the average pattern of magnetic activity. This equivalent current system, called the solar daily variation (SD), is nearly symmetric about the earth–sun line, with two cells centered near the dawn and dusk terminators. The current flows sunward across the polar cap as a sheet of current. But in the auroral oval it is concentrated by high conductivity into the *eastward electrojet* (dusk) and *westward electrojet* (dawn). The right-hand rule is easily used to determine the magnetic perturbations that the two electrojets produce on the ground.

After the International Geophysical Year (1957), enough magnetic observatories existed at high latitudes to determine the instantaneous pattern of magnetic activity. It was found that the two-cell pattern apparent in the averages was also observed instantaneously, at least in the early stages of an auroral substorm (Iijima and Nagata, 1972; Kokubun, 1972). The instantaneous pattern was named DP-2 (disturbance polar of the second type). The axis of symmetry for DP-2 was found to differ from the pattern shown in Figure 13.16. It is tilted along a line from late morning to late evening, rather than along the earth–sun line. An example of this current pattern, as determined by more than 50 stations, is shown in the left panel of Figure 13.17 (Clauer and Kamide, 1985).

The current producing the ground disturbance is a Hall current, meaning that it flows at right angles to the ionospheric electric field. This current is produced by the drift of ionospheric charges in the presence of orthogonal electric and magnetic fields. In the absence of collisions, positive and negative charges would drift at the same rate, and there would be no current. In the weakly collisional ionosphere, ions drift more slowly than electrons, producing a current in a direction opposite to their mutual drift. The existence of the DP-2 current system implies that ionospheric plasma is moving in a circulation system referred to as *ionospheric convection*. The direction of this flow is approximately from noon to midnight across the poles and then back to the dayside through the auroral ovals; thus the two electrojets that establish and maintain the flow frequently are called the *convection electrojets*. The form of this convection pattern provided the initial insight into the mechanism by which the solar wind causes magnetic activity.

13.5.1.5 THE SUBSTORM ELECTROJET AND THE CURRENT WEDGE Auroral-zone magnetic activity is strongest during the expansion phase of an auroral substorm. In fact, once the concept of an auroral substorm (visible excitation of light in the ionosphere) had been

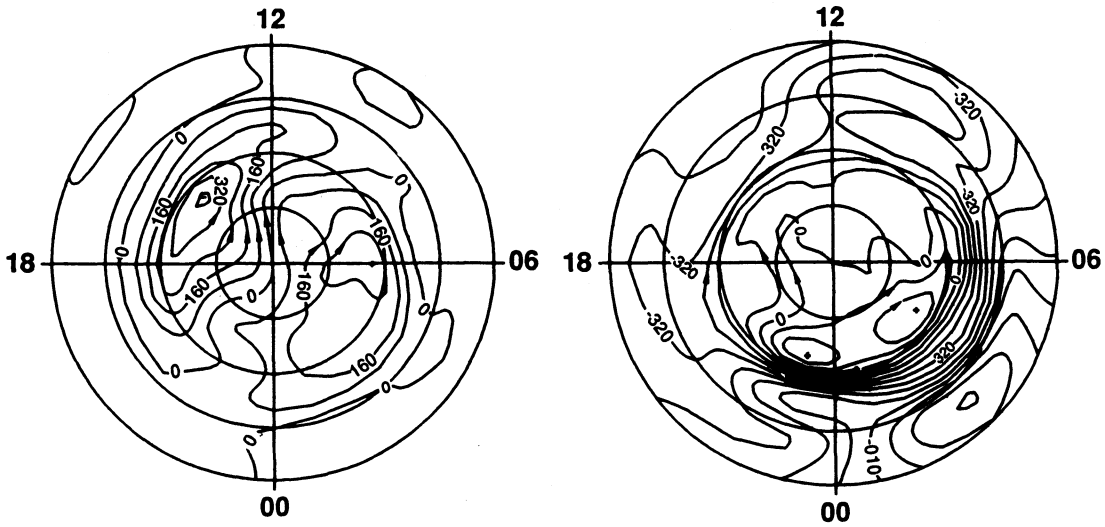

FIG. 13.17. Results from a recent determination of the patterns of ionospheric currents during magnetic disturbance. Closed contours show the flow lines for an equivalent ionospheric current that produces the observed ground magnetic perturbations. Left: The two-cell DP-2 current system present in the substorm growth phase. Right: The single-cell DP-1 current system that dominates during the substorm expansion phase. (From Clauer and Kamide, 1985.)

developed, it became clear that there was an accompanying polar magnetic substorm (current flow and convection). The main characteristic of this phenomenon is the sudden enhancement of the westward electrojet across the midnight sector. According to Akasofu, Chapman, and Meng (1965), as soon as the expansion phase begins, a westward current begins to flow along the newly brightened arc. As the auroral bulge expands poleward, and the surge moves westward, the region occupied by the current expands as well. The poleward edge of the bulge and the westward surge define the limits of the current. As the surge moves toward dusk, the westward end of the current intrudes along the poleward edge of the eastward electrojet. Because this new current is stronger than the eastward current, the associated magnetic perturbations that were initially positive become negative. Eventually, when this current dies away, the eastward current again dominates, leaving the magnetograms with an *H*-component waveform called an indented positive bay. Near midnight, the poleward expansion of the westward current causes the z component to reverse sign as the center of the current passes over a station. Eventually the current begins to decrease. This happens first at the equatorward edge of the auroral oval, and later at higher latitudes. This apparent poleward motion of the equatorward edge of the westward current is the recovery phase as seen in magnetic disturbances.

Initially there was considerable debate concerning the appropriate equivalent-current pattern for the expansion-phase current. The term

"equivalent current" is used to describe a current system confined to the ionosphere that produces the magnetic signatures observed by a network of ground magnetometers. Ground observations cannot determine whether or not the currents are confined to the ionosphere. If the ionospheric conductivity is uniform, the same ground signatures can be produced by current systems that include field-aligned currents. Akasofu et al. (1965) believed that the pattern was primarily a single cell centered at midnight, with strong currents across the auroral bulge and weaker return currents at both higher and lower latitudes. It was suggested that the twin-cell system was really an artifact of averaging the temporal development of the single-cell system. That was not correct. The two different current systems are characteristic of different phases of the substorm. The expansion-phase current is now referred to as the DP-1 current (disturbance polar of the first type) and also as the *substorm electrojet*. The right panel in Figure 13.17 illustrates the form of this current for the same substorm characterized by DP-2 at an earlier time. Note the strong current in the sector between dawn and 2200 LT (closely spaced contours) that corresponds to the westward electrojet.

The difficulty in interpreting the early observations of the substorm electrojet can be attributed to the assumption that the current was confined to the ionosphere. Birkeland (1913) had believed that field-aligned currents were also present, but Chapman and others argued that they were confined to the ionosphere. Several workers, including Bostrom (1964), Atkinson (1967), and later Akasofu and Meng (1969), suggested that DP-1 was part of a three-dimensional system. Final proof that this was true came from a combination of ground data and synchronous-satellite data. The proof lies in the observation that the east-component magnetic perturbations at midlatitude stations and at the synchronous satellite *ATS1* have the same sign. This is possible only if the causative currents are above the spacecraft. McPherron, Russell, and Aubry (1973) pictured the actual currents, as shown in Figure 13.18. The actual three-dimensional current system that gives the signature observed on the ground is a wedge-shaped sector of magnetic-field lines. Eastward current across the midnight equatorial plane is diverted along field lines into the northern and southern auroral ovals. The current flows westward across midnight and then returns to space. The point at which the outward current leaves the ionosphere is the westward surge. This current system is called the substorm current wedge.

13.5.2 The Magnetospheric Substorm

13.5.2.1 PHASES OF A SUBSTORM The existence of two distinct current patterns during a substorm had no counterpart in the original model of auroral and polar magnetic substorms suggested by Akasofu (1968). Because of that, the model was amended by McPherron (1970), who noted that many phenomena preceded the onset of the

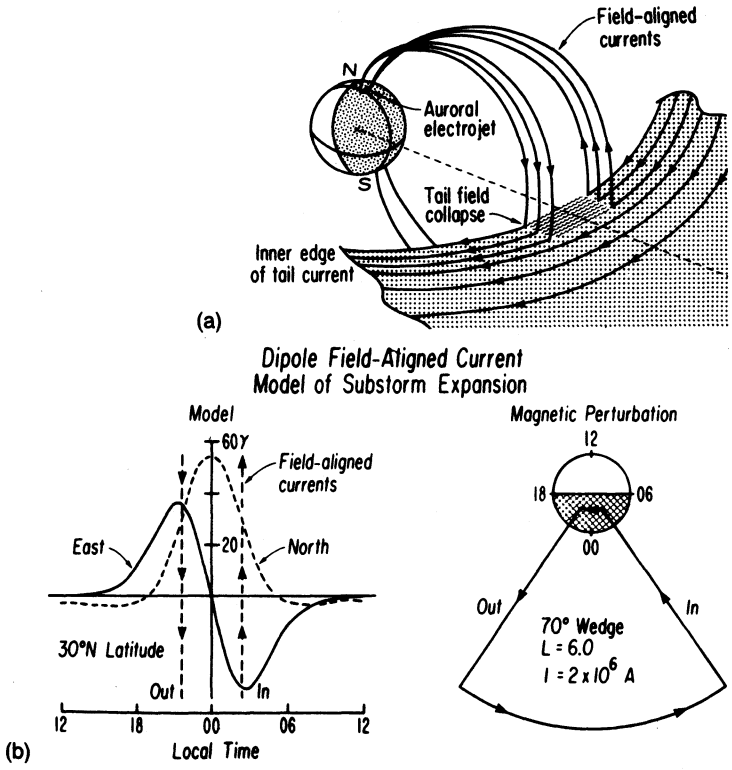

FIG. 13.18. Schematic illustration of the three-dimensional current system that is responsible for the DP-1 current system during the expansion phase of a polar magnetic substorm. (a) Diversion of the cross-tail current through the midnight ionosphere. (b) Magnetic perturbations caused by this current system along a chain of northern midlatitude magnetic observatories. The wedge shape of the projected equivalent current accounts for the name "substorm current wedge." (From Clauer and McPherron, 1974.)

expansion phase in the ionosphere. Foremost among them were weak positive and negative bay signatures similar to those seen later in the expansion phase. There also appeared to be an increasing probability of weak, short-duration intensifications of the aurora and electrojet, accompanied by bursts of ULF waves called Pi-2 bursts. Yet another feature that occurred before expansion was a gradual increase in the size of the polar cap. McPherron interpreted these phenomena as the *growth phase* of a substorm. The growth phase is an interval of time, prior to the onset of expansion, during which energy extracted from the solar wind is stored in the magnetosphere. The expansion phase corresponds to the release or unloading of that stored energy, and the recovery phase is the return of the magnetosphere to its ground state.

Figure 13.19 illustrates the three substorm phases in terms of the AU and AL indices during the substorm shown in Figure 13.17. As discussed in Appendix 13B, AU and AL are the envelopes of the superposed *H*-component traces from a worldwide chain of auroral-zone magnetometers underneath the electrojets. The beginning and end of an isolated substorm are defined by the departure and return of both those indices to background levels defined by quiet-day variations. The growth phase is then the initial interval of slowly growing AU and AL. These changes are created by the DP-2 system. Occasionally during the growth phase, a pseudobreakup will momentarily disturb the development, as if a substorm expansion with its DP-1 system were about to start. Eventually that does happen, and a full expansion phase ensues. Often, several

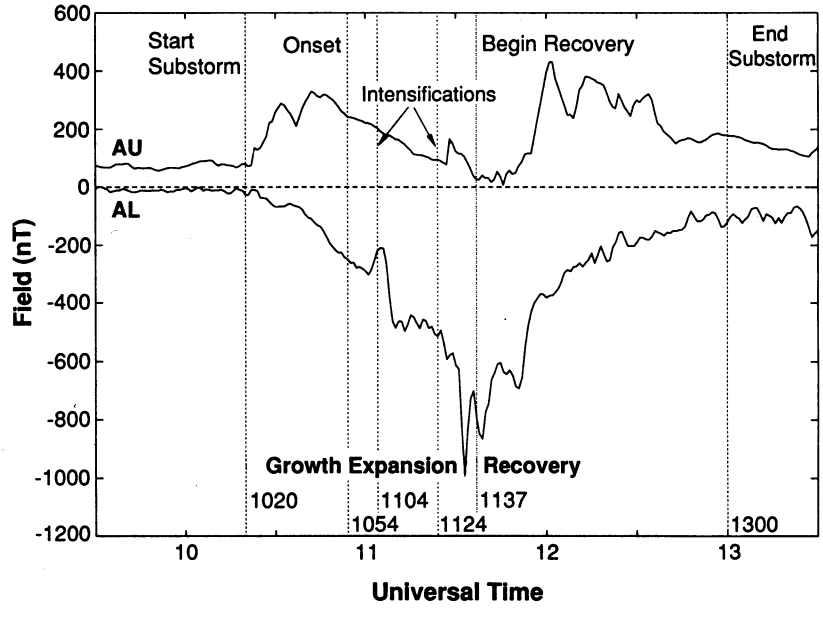

FIG. 13.19. AU and AL indices for a particularly well studied substorm. The three phases of this substorm (growth, expansion, and recovery) can be identified (see labels at bottom) by examination of the slope of the AL index. Times corresponding to the beginning and end of each phase are denoted by labels at the top. The onset of the substorm expansion (1054 UT) is often characterized by a sudden increase in the rate of decrease of AL. Subsequent increases are called intensifications. Note the indented positive bay (reduction in AU) during the middle of the substorm. Vertical lines are drawn at times when the B_z component of the tail field increased sharply. (From McPherron and Manka, 1985.)

distinct intensifications of the AL index occur during the expansion phase. These are caused by new substorm current wedges forming, each with a surge at its western end, and each closing westward through the ionosphere as a westward current filament. The initial onset and some of the intensifications usually are apparent in the AL trace as a change in slope. Eventually the AL index reaches a minimum and begins to recover. The interval of increasing AL (decreasing magnitude) is usually called the recovery phase, although some intensifications or surges continue to occur at high latitudes. As the DP-1 current system dies away, the DP-2 current system reappears, but later it too disappears. The entire sequence of events consisting of the three phases is the substorm. The rise and fall of the substorm electrojet constitute only the expansion and early recovery phase, although many people mistakenly refer to that rise-and-fall process as a substorm.

13.5.2.2 MAGNETOSPHERIC EFFECTS During a typical isolated substorm, the magnetosphere undergoes a distinct sequence of changes in its magnetic field and plasma, associated with the changes in the auroral current systems. We shall identify the key features here and expand the description in the next section. The IMF turns southward. The dayside magnetopause is eroded, and the associated magnetic flux is transported to the tail lobes. The plasma sheet thins, and the tail current moves earthward. A connected pair of X- and O-type neutral lines form in the near-earth plasma sheet. Magnetic reconnection at this X-line forms a bubble of plasma in the plasma in the plasma sheet. This bubble is disconnected and pulled out of the center of the tail. The extra flux in the lobes reconnects earthward of the bubble and convects back

to the dayside. Particles energized at the X-line are injected into the inner magnetosphere and drift in the radiation belts. Eventually the near-earth portion of the X-line moves tailward, establishing a distant X-line. These events constitute the three phases of a substorm as seen in the magnetosphere, and we shall refer to the complete sequence as the *magnetospheric substorm.*

13.5.3 Magnetic Storms

13.5.3.1 PHASES OF A MAGNETIC STORM An isolated substorm is created by a brief (30–60-min) pulse of southward IMF. When the IMF remains southward for longer times, activity becomes more complex. There is a series of overlapping auroral-zone activations, each injecting particles into the inner magnetosphere. The injected particles drift in a ring around the earth. Protons drift westward, and electrons eastward, creating a westward current called the *ring current.* Some particles from each activation are accelerated by drift across the enhanced magnetospheric electric field. The stronger the electric field, the greater their energy, and the closer the ring current is to the earth. In addition, particles are accelerated out of the ionosphere into the equatorial plane, so that heavy ions such as oxygen become important in the ring current, as discussed in Chapter 10. The ring current causes large decreases in the H component over most of the earth's surface. This effect is known as a *magnetic storm.* As long as injection of particles continues, the ring current will grow toward some asymptotic value in which the rate of injection equals the rate of loss. The time during which the ring current is growing is called the *main phase* of the magnetic storm. However, as soon as the IMF weakens, or turns northward, the ring current stops growing, and the ground perturbations begin to decrease. The ground perturbations decrease principally because particles are lost from the ring current. The loss process occurs in several steps. First, the rate of dayside reconnection decreases, and the convection boundaries move to larger radial distances. The ionosphere begins to refill flux tubes within the new boundary. As the cold ionospheric plasma encounters the ring-current plasma, ion-cyclotron waves begin to grow, and these waves scatter the ring-current protons into the loss cone. Other ring-current ions charge-exchange with the cold neutral hydrogen. Ring-current ions become energetic neutral atoms and are lost to the atmosphere or outer space. The low-energy ions that replace them contribute little current, and so the strength of the ring current decreases with time. This is the *recovery phase* of the storm. Many storm recoveries occur in at least two stages. The first stage results from the rapid loss of oxygen ions, and the second from the slower loss of protons.

Some magnetic storms are preceded by an initial phase of enhanced H component in ground magnetometer records. This effect is unrelated

to the ring current and is caused by an enhancement of the magnetopause current. Many magnetic storms follow solar flares or coronal mass ejections. In either case, a high-speed parcel of the sun's atmosphere sweeps through slower solar wind, compressing and distorting the magnetic field ahead of it. That parcel of gas then encounters the earth's field and compresses it, enhancing the magnetopause currents and thereby producing positive perturbations in H at the earth's surface. This compression of the field is called a *sudden impulse*. In many storms, this phase will last for 4–16 h, as long as the IMF is northward. Eventually the IMF will turn southward, and there will be a sequence of substorms producing a magnetic storm. The H component that results from superposition of the enhanced magnetopause current (positive ΔH) and the ring-current perturbation (negative ΔH) is negative, even though the solar-wind dynamic pressure may remain elevated for some time.

13.6 PHENOMENOLOGICAL MODELS OF SUBSTORMS

13.6.1 The Ground State of the Magnetosphere

The occurrence of large-scale, systematic changes throughout the ionosphere during a substorm suggests that events are simultaneously occurring in space on the field lines connected to the aurora and electrojets. A phenomenological model has been developed to describe the associated magnetospheric phenomena that compose the magnetospheric substorm (Coroniti, McPherron, and Parks, 1968; Akasofu, 1968; Rostoker et al., 1980, 1987a). Detailed descriptions of this model have been given by McPherron et al. (1973), Akasofu (1977), McPherron (1979, 1991), Hones (1979), and Baker et al. (1984). In this section we present an abbreviated description of this model using schematic illustrations to represent the findings from many experimental studies. We begin with a description of the ground state of the magnetosphere as it is observed during relatively quiet conditions.

Figure 10.4 represents the topology of the magnetosphere in its ground state between substorms. Three types of field lines are present. There are interplanetary magnetic-field lines connected to the sun at both ends and generally excluded from the magnetosphere by the currents on the magnetopause. There are dipolelike field lines connected to the earth at both ends. These pass through the earth's equatorial plane, although they may be highly distorted by the tail and magnetopause currents. Finally, there are open field lines connected at one end to the earth and at the other to the sun. The open field lines have their footprints at latitudes poleward of the auroral oval and form the *polar caps*.

The topology of the earth's magnetic field defines many of the important regions of the magnetosphere. The polar cusp is the region of open field lines just poleward of the boundary between open and closed

field lines on the dayside (see Figure 9.19). Near the magnetopause, these field lines link to a region of weak magnetic field where solar-wind pressure creates an indentation in the magnetopause. These open field lines provide a path for solar-wind plasma to reach the ionosphere. On the nightside, the boundary between open and closed field lines is connected to the distant X-line. The X-line is the separatrix between open and closed field lines. Above and below the X-line the field lines are open, connected to the polar cap and to the solar wind. Earthward in the equatorial plane they are closed and connected to the earth. Tailward they disconnect from the earth and hence are solar-wind field lines, despite their shape. The closed field lines of the tail confine the plasma sheet. This is a region of weak magnetic field filled with charged particles. Drift of these charges in the tail magnetic field produces the tail current that in turn creates the magnetic field in which they drift. The open field lines from the polar caps pass through the lobes of the tail. These regions contain few particles, because particles are easily lost either to the polar cap or to the solar wind. Close to the earth there is a rapid transition from dipole-like to tail-like field lines on the nightside. This occurs at the inner boundary of the plasma sheet. This boundary exists because charged particles of most energies in the tail are unable to drift arbitrarily close to the earth. The region earthward of the boundary is a "forbidden region" for these particles. For cold particles of either sign, the earthward boundary is the separatrix of the magnetospheric electric-potential distribution (see Chapter 10). This separatrix is symmetric around the dawn–dusk meridian, bulging outward somewhere near dusk. During steady conditions, the separatrix corresponds to the plasmapause, the boundary separating a region of high-density cold plasma originating in the ionosphere from lower-density plasma drifting earthward from the magnetotail. For higher-energy particles, the boundaries are more complex. Very roughly, the region accessible to tail particles of all energies and charge states has a horseshoe shape in the equatorial plane, with the open ends nearly meeting at local noon. Inside of this boundary lies the plasmasphere discussed in Chapter 10.

The locations of the inner edge of the plasma sheet and the convection separatrix depend on the electric field applied to the magnetosphere by the solar wind. The stronger the electric field, the closer to the earth the boundary will be. In the ground state of the magnetosphere, the boundary is typically at $10R_E$ near midnight. Fluctuations in the convection electric field enable particles from the tail to penetrate into the forbidden region, and in fact these fluctuations are the main process that populates the forbidden region with energetic particles. The energetic particles that penetrate the separatrices constitute the outer radiation belt, and their drift creates the ring current.

In the ionosphere, the nightside portion of the auroral oval contains particles lost from the plasma sheet. These particles hit the ionosphere

and cause emission of light. The dayside portion of the auroral oval contains particles lost from open field lines in the polar cusp and from the closed field lines of the low-latitude boundary layer, as discussed later. This region is also horseshoe-shaped, but points tailward. Thus the oval has two main parts: one produced by particles lost from field lines with their feet on the dayside, and one produced by particles lost from closed field lines with their feet on the nightside.

Two other important regions of the magnetosphere are depicted in Figure 9.19. One is the *polar mantle*. The polar mantle exists because the solar-wind electric field is present on the open field lines of the polar cusp as well as on closed field lines. This electric field causes the particles to drift across field lines, initially poleward while they are entering and leaving the cusp. Fast particles do not drift far across the field before they leave the cusp and stream tailward along open field lines close to the magnetotail boundary. Lower-energy particles drift farther poleward and emerge from the cusp on field lines deeper in the tail lobes. As the mantle particles stream tailward, they continue drifting across the tail lobes toward the center of the tail. Eventually these particles reach the midplane of the tail. It is generally believed that the distant X-line is located at the point where particles with the characteristic energy of the cusp reach the midplane. This probably occurs at about 100–200 R_E downstream.

The plasma-sheet boundary layer is also depicted in Figure 10.4. This region forms on field lines near the upper and lower boundaries of the plasma sheet. The plasma-sheet boundary layer contains structured beams of particles streaming along closed field lines. Isotropization of the pitch angles of these particles produces the central plasma sheet, which fills the center of the tail.

A word of caution is needed here. It is not clear that either the polar mantle or the plasma-sheet boundary layer exists during quiet conditions associated with a northward IMF. The reason for this is that they are associated with open field lines in the polar cusp and with an electric field in the plasma sheet. Both of these are produced primarily by magnetic reconnection at the subsolar point and at the distant X-line, as discussed later. The mantle, as depicted in the drawing, may exist only during substorms while the IMF is southward. Similarly, the plasma-sheet boundary layer may be sharply defined only while reconnection is occurring at the distant X-line and a strong electric field exists inside the plasma sheet. We believe this is primarily during the recovery phase of a substorm.

It is important to view the magnetosphere in equatorial projection as well. Figure 13.20 portrays this view. The inner edge of the plasma sheet wraps around the earth from the nightside. The separatrix between corotating and convecting plasma (the plasmapause) is shown as a boundary closer to the earth. This drawing is not quite as described earlier. The reason is as follows: Most of the particles in the plasma

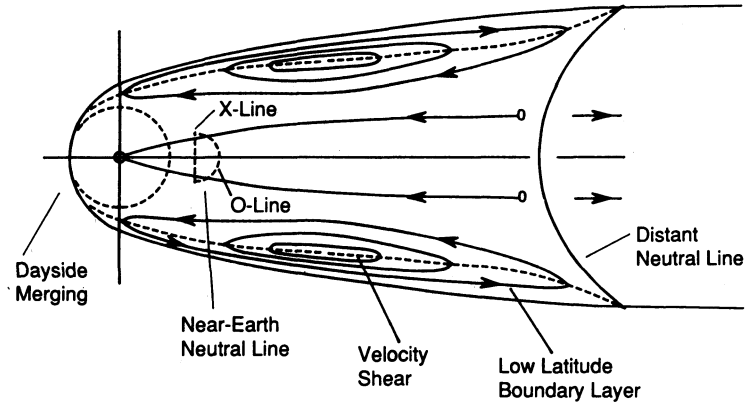

FIG. 13.20. Equatorial projection of the quiet magnetosphere. A dashed line defines the inner edge of the plasma sheet. The distant X-line should be farther from the earth at the center of the tail, and it should reach the flanks at nearly twice the distance of its center point.

sheet have finite energy and thus are subject to gradient drift and curvature drift. Because of these drifts, the forbidden boundary for electrons is located progressively farther outside of the plasmapause as energy increases. Roughly speaking, the boundary for protons is farther inside as energy increases. Thus, there is a gap between the inner edge of plasma-sheet electrons and the plasmapause, whereas the plasma-sheet ions reach and penetrate the plasmapause. These features are illustrated in Figures 10.25 and 10.26.

Data from the *ISEE 3* spacecraft suggest that the X-line in the distant tail is parabolic in shape, opening away from the earth; it may be as close as $100R_E$ at the center of the tail, but probably is $200R_E$ distant at the flanks.

On the flanks of the magnetosphere is the last major region of the magnetosphere to be defined, the low-latitude boundary layer. This region is generally believed to be on closed field lines, with plasma moving tailward at velocities that decrease away from the magnetopause. These two layers are thought to be produced by some type of viscous interaction or by reconnection with a northward IMF. The viscous interaction may be supplied by processes such as particle scattering, wave penetration, pressure-induced bumps on the boundary, or surface waves that transfer momentum from the flowing magnetosheath plasma to the outer portions of the magnetosphere. The inner edge of the boundary layer is a stagnation line at which the direction of convection reverses. Between the dawn and dusk convection reversals, plasma, and field normally flow sunward. Streamlines of flow in the equatorial plane form closed loops about the region of convection reversal. These streamlines project onto the two-cell ionospheric convection pattern and DP-2 current system mentioned earlier. In the drawing, we depict the ends of the boundary layers as merging with the ends of the distant X-line, although it is not certain that this is correct. If reconnection is occurring at the X-line, then field lines that close at the X-line flow sunward in the center of the plasma sheet, separating the return flows from the two boundary layers.

13.6.2 The Driven Model

In the ground state of the magnetosphere described in the preceding section, open field lines connect the polar caps to the solar wind. This configuration can exist only after magnetic reconnection interconnects the magnetic fields of the solar wind and the earth. Once interconnected, the interplanetary electric field is directly applied to the earth's ionosphere, driving electric currents. Wherever there are divergences in the electric field, or discontinuities in the ionospheric conductivity, there will be field-aligned currents. Changes in the ionospheric field-aligned currents and magnetospheric currents cause geomagnetic activity. The driven model of substorms was introduced by Perreault and Akasofu (1978) to explain the high correlation between the magnetic indices that are sensitive to these currents and the solar wind. To quantify the dependence on u, B, and field direction around the earth–sun line θ, Perreault and Akasofu developed the energy-coupling parameter ϵ (a power), defined as $\epsilon = l_0^2 u B^2 \sin^4(\theta/2)$, where $l_0^2 = (6R_E)^2$ is the area on the magnetopause through which magnetic energy (Poynting flux) enters the magnetosphere. In the driven model, a southward turning of the IMF (increasing θ) enhances the coupling of the solar-wind electric field to the ionosphere and hence the strength of the various currents. To explain the sudden enhancement of the aurora characteristic of the expansion phase, Akasofu (1979, 1980, 1981) postulated that an outward field-aligned current near midnight (located in the Harang discontinuity) develops a field-aligned potential drop when its density exceeds a threshold value. This potential accelerates electrons downward, increasing the current, as required by the increased solar-wind coupling, and simultaneously increasing the ionospheric conductivity. The higher conductivity demands more current for a fixed solar-wind electric field, further increasing the field-aligned current and its potential drop. This instability and its subsequent saturation correspond to the substorm expansion phase.

The driven model is implicitly a reconnection model, although the details of magnetic reconnection are never discussed. In particular, the existence of two or more X-lines on the dayside and nightside is not mentioned. No significance is attributed to delays between the onset of dayside and nightside reconnection, nor to the changes in magnetospheric configuration brought about by these delays.

13.6.3 The Boundary-Layer-Dynamics Model

Another reconnection model for substorms is the boundary-layer-dynamics model (BLD) developed by Rostoker and Eastman (1987). This model was developed to explain why fast flows in the magnetotail during substorm expansions are most frequently directed earthward (Eastman, Frank, and Huang, 1985). The authors of the model thought

this meant that the source of the substorm expansion was distant from the earth. The model used a magnetic-field mapping in which the high-latitude ionospheric convection reversal at dusk and the Harang discontinuity near midnight were mapped to the same topological feature of the magnetosphere. Because the convection reversal at dusk clearly mapped to the inner edge of the low-latitude boundary layer on the duskside of the plasma sheet, continuity of mapping forced the authors to conclude that the Harang discontinuity at midnight mapped to a location close to the distant X-line. They therefore explained the expansion phase as a consequence of the sudden onset of reconnection at this distant X-line. Plasma energized by reconnection will jet earthward, enhancing the velocity shears at the inner edge of the low-latitude boundary layer. These shears become unstable to the growth of the Kelvin-Helmholtz instability, wrapping portions of the interface into vortices, preferentially on the duskside. Field-aligned currents flow out of the velocity shear on the dawnside of the plasma sheet, and into the center of vortices on the duskside. The three-dimensional current system produced by this process is the substorm current wedge. Inside the wedge, the currents cause an increase in the vertical component of the magnetic field (dipolarization), and outside the wedge they cause a decrease (more taillike field). As in the driven model, the authors postulate that the outward field-aligned current requires a field-aligned potential drop that accelerates electrons downward, creating the westward-traveling surge. It also accelerates heavy ionospheric ions outward. The authors attribute the occasional observations of tailward plasma flows, threaded by a southward magnetic field in the plasma sheet, to the effects of these ions in a region just west (duskward) of the outward field-aligned current.

The BLD model is based on a magnetic-field mapping that is known to be incorrect (e.g., Fairfield and Mead, 1975). The Harang discontinuity near midnight maps close to synchronous orbit, not the distant X-line. Furthermore, observations of tailward flow accompanied by a southward field in the plasma sheet occur predominantly at midnight, which is the center of typical substorm current wedges, not at earlier local times west of the westward-traveling surge.

13.6.4 The Thermal-Catastrophe Model

Another model that invokes reconnection to drive convection, but otherwise attributes to it no essential role, is the *thermal-catastrophe model* of Smith, Goertz, and Grossman (1986) and Goertz and Smith (1989). In this model, energy enters the magnetotail from surface-wave perturbations on the tail magnetopause. The wave energy propagates into the plasma-sheet boundary layer, where it is resonantly absorbed at field lines for which the natural frequencies match the incident frequencies. The absorbed energy heats the plasma, which convects to the

central plasma sheet and eventually earthward. In this model the substorm growth phase begins when the intensity of compressional waves crossing the lobes increases. Power input to the boundary layer increases, and for a given level of convection, more heat remains in the local plasma, causing the temperature to rise. However, it is a characteristic of the resonant-wave absorption process that the opacity of the boundary layer maximizes at a fixed temperature dependent on the properties of the plasma. At this temperature the boundary layer becomes totally opaque, and all incident wave energy is absorbed. The plasma temperature rises explosively, as convection is unable to maintain an equilibrium. The plasma sheet then dynamically adjusts to a geometry appropriate to the elevated temperature. At the higher temperature, the opacity decreases, and convection is again able to carry off the energy deposited by waves. The expansion phase is attributed to the adjustment of the state of the boundary-layer plasma from one temperature to another at constant energy. This is the so-called thermal catastrophe. There have been attempts to correlate wave power in the lobes with substorm activity, but the evidence of a connection remains questionable.

13.6.5 Magnetosphere–Ionosphere Coupling Models

The models reviewed earlier emphasize the magnetospheric aspects of substorms. The members of another class of models focus on the ionosphere and its coupling to the magnetosphere. Kan (1990) and Rothwell et al. (1988) are proponents of such models, which are referred to as magnetosphere–ionosphere coupling (MIC) models. These models emphasize the positive feedback that changes in ionospheric conductivity can have on the sources of field-aligned current in the magnetotail. In particular, these models appear to provide possible explanations for the dynamic development of various auroral features, such as the surge, Pi-2 pulsation burst, and poleward bulge. The Kan (1990) model considers the consequences of instantaneously imposing an enhanced two-cell convection pattern on the magnetosphere. Alfvén waves reverberate between the ionosphere and the magnetotail, creating the distribution of field-aligned currents. This model is basically the driven model mentioned earlier. It makes no attempt to consider the phenomenology of the magnetotail and does not ascribe any particular importance to reconnection in the tail, other than as a source of convection electric field.

The Rothwell et al. (1984, 1989) model postulates an existing substorm current wedge and considers its ionospheric consequences. This model is very similar to one developed earlier for auroral arcs by Sato (1978) and Miura and Sato (1980). It invokes a mechanism called the feedback instability, which is a consequence of the finite delays imposed on the system by the inductance of field-aligned currents. It allows the

ionospheric and magnetospheric developments of currents and fields to become uncoupled to some extent and results in poleward and westward expansion of the ionospheric portion of the substorm current wedge. A unique feature of this model is the development of a sheet of outward field-aligned current on the poleward edge of the auroral bulge that goes unstable in the same way as the outward current at the western edge of the current wedge. This potential drop becomes so strong that it alters plasma flow in the plasma sheet.

13.6.6 The Near-Earth Neutral-Line Model

The best-developed model of substorms is the near-earth neutral-line (NENL) model (McPherron et al., 1973; Russell and McPherron, 1973; McPherron, 1991). This model attempts to provide an internally consistent explanation for most magnetospheric phenomena, but does not attempt to explain many ionospheric observations. The unique feature of this model is the formation of a *plasmoid,* or bubble of closed field lines, that is ejected from the plasma sheet during the expansion and recovery phase of a substorm. The formation of the plasmoid is a direct consequence of the substorm growth phase that initiates the substorm.

In the NENL model, the substorm begins when a southward turning of the IMF activates dayside reconnection. This occurs in a ground-state magnetosphere (see Figure 10.4) that has a distant X-line separating the closed field lines of the plasma sheet (auroral oval) from the open field lines of the tail lobe (polar cap). This X-line is assumed to be inactive, with little or no reconnection occurring. Dayside magnetic flux from the earth connects to the IMF and is transported over the polar caps by the solar wind, where it is added to the outer portions of the tail lobes. The removal of this flux from the dayside initiates a convective flow of plasma toward the reconnection region, but the flow is retarded by the finite conductivity of the ionosphere at the foot of the field lines. Because of this, the magnetospheric return flow is unable to balance the rate at which flux reconnects, and the dayside magnetopause erodes earthward. The onset of convective flow propagates around the earth into the plasma sheet as a rarefaction wave, initiating earthward flow. Dayside erosion increases magnetopause flaring, thus increasing the dynamic pressure on the boundary. The dynamic pressure then reduces the flaring angle, compressing the tail until a corresponding increase in tail-lobe magnetic pressure balances the external pressure. This enhanced pressure is applied to the plasma sheet, and in combination with rarefaction brought about by flow toward the dayside, it causes the nightside plasma sheet and the tail current to thin. Simultaneously, the increased drag on the magnetotail caused by the newly opened field lines passing through the magnetotail boundary is balanced by the tail current moving earthward, closer to the earth's dipole moment. These changes increase the lobe field intensity, cause the plasma sheet to thin, and its

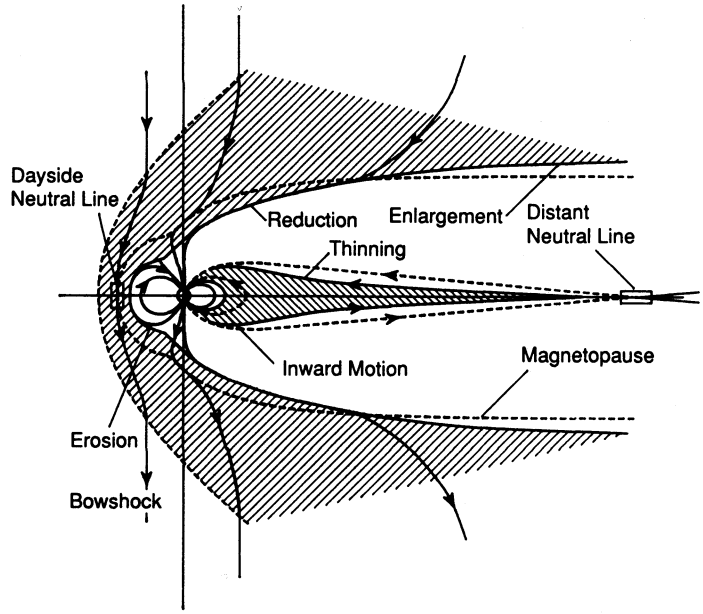

FIG. 13.21. Schematic illustration of the changes in magnetic field and plasma sheet expected in the situation where the reconnection rate on the dayside exceeds that on the nightside. Increased magnetopause flaring, plasma-sheet thinning, and earthward motion of the tail current are the main effects.

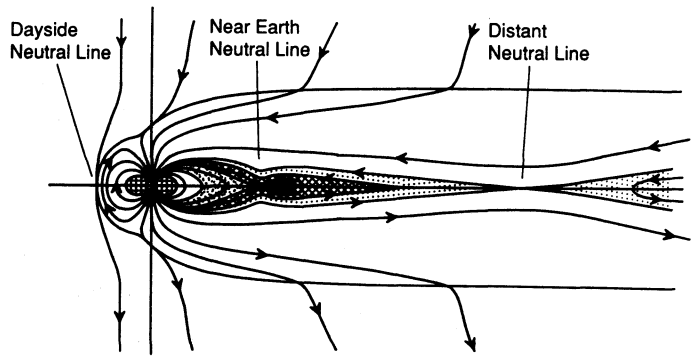

FIG. 13.22. Initial stage of the substorm expansion phase illustrating the first step in the formation of a substorm plasmoid.

inner edge to move earthward. These changes are summarized in Figure 13.21 and have been described quantitatively by the model of Coroniti and Kennel (1972).

At some time during the late growth phase, the vertical component of magnetic field across the plasma sheet becomes sufficiently small that ions in the cross-tail current no longer behave adiabatically. Coroniti (1985) and Baker and McPherron (1990) postulate that at that point, magnetic reconnection begins in the central plasma sheet. Reconnection proceeds slowly at first, cutting closed field lines of the plasma sheet at a new X-line. As successive field lines are cut, they form closed loops within the plasma sheet, centered about an O-line located tailward of the X-line, as illustrated in Figure 13.22. Azimuthal localization of the X-line requires that the X-0 and O-type lines connect at their end points (Russell and McPherron, 1973). As time progresses, the current sheet becomes thinner, and the rate of reconnection increases until it becomes explosive. If reconnection severs the last closed field lines at the edge of

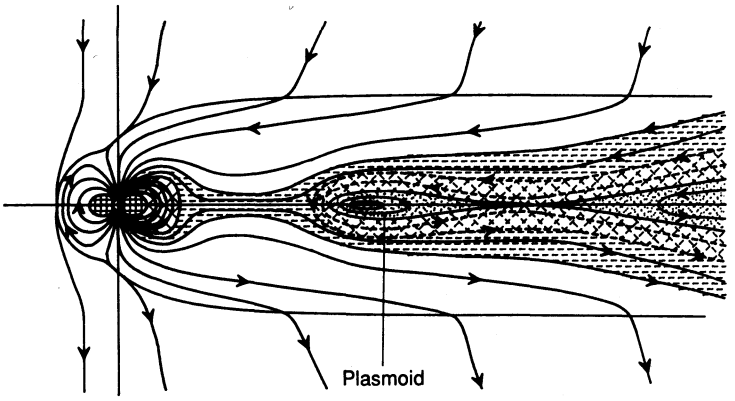

FIG. 13.23. Intermediate stage in the substorm expansion phase, showing a plasmoid moving away from the earth as a consequence of severance of the last closed field line originally connected to the distant neutral line.

the plasma sheet, a full-fledged substorm expansion takes place. Otherwise the disturbance is quenched, and the disturbance is called a pseudobreakup.

Once the open field lines of the tail lobe reconnect, they wrap around the plasmoid formed in the earlier stage and begin to pull it down the tail. The combination of magnetic tension and pressure from the plasma flowing tailward from the X-line accelerates the plasmoid away from the earth. As it leaves, the plasma sheet behind the X-line collapses to a thin sheet, as illustrated in Figure 13.23. Subsequently, open flux of the tail lobe reconnects at the X-line, forming closed field lines earthward of the X-line and open interplanetary magnetic-field lines tailward of the X-line. This process continues for some minutes until the balance of forces in the plasma sheet changes and the X-line begins to move down the tail. Earthward of the X-line, the plasma sheet thickens, and strong earthward flows are observed. As the X-line moves toward its distant location, the currents and aurora begin to die at the lower edge of the auroral bulge, although sporadic intensifications are observed on the poleward edge. This is the beginning of the recovery phase. With time, all disturbances die away, the substorm is over, and the magnetosphere returns to its ground state.

13.6.7 The Current-Sheet-Disruption Model

In 1984, the AMPTE *CCE* spacecraft was launched into an orbit with apogee of $8.8R_E$ and an inclination of 4.8°. That mission began to explore a region of the magnetosphere just beyond synchronous orbit that had been little studied by previous spacecraft. Substorm onsets observed in the regions near apogee in the midnight sector revealed signatures that had not been anticipated. Takahashi et al. (1987) reported on an event in which the magnetic field rotated slowly from nearly northward to quite taillike. Then, in a 3-min interval, oscillations with periods of about 13 s appeared suddenly in all components of the field. The amplitudes of individual fluctuations exceeded 40 nT, with some cycles of the wave

producing field orientations near 80° southward. Ion fluxes in the range 100 keV to 1 MeV increased at *CCE* by an order of magnitude, and a nearby spacecraft reported injection of energetic electrons. Following that burst, the field orientation was nearly vertical, and within another 15 min its magnitude reached 60 nT, close to the dipole value at the spacecraft location. In another event, when *CCE* was very close to midnight and almost exactly in the magnetic equator, Lui et al. (1988) reported that large fluctuations produced southward fields, and in 4 min the field magnitude increased from 10 nT to the local dipole value of 40 nT. Energetic ions appeared, and their distribution showed strong asymmetries, with enhancements alternately earthward and tailward of the spacecraft. For the first event, Takahashi et al. (1987) offered an interpretation based on the assumption that the spacecraft observed radial oscillations of the position of a nearby neutral line, an interpretation that accounted for the reversal of the sign of the magnetic field. Lui et al. (1988) argued that their observations were not consistent with an X-type neutral-line geometry because of inconsistencies of particle and field signatures. They suggested, instead, that instabilities that occur in a thin current sheet can produce the signatures observed. This suggestion has been developed into a new model of the substorm, the *current-sheet-disruption model.*

In the model as described by Lui (1991a, b), a thin current sheet develops in the inner magnetosphere during the substorm growth phase, for the same reasons as in the near-earth neutral-line model. As the current sheet thins, the ions become nonadiabatic and begin to stream across the current sheet in serpentine orbits. The streaming ions interact with adiabatic electrons drifting in the opposite direction (the kinetic cross-field-streaming instability), producing lower hybrid waves (Lui et al., 1990). At the same time, the density gradient on the boundary of the plasma sheet drives the lower-hybrid-drift instability. The combination of the two types of waves produces an anomalous resistance in the plasma sheet that disrupts the cross-tail current. However, because of the high inductance of the tail circuit, the current must continue to flow. It accomplishes this by diversion along field lines, particularly those of the substorm current wedge.

According to observations reported by Lopez and co-workers (Lopez et al., 1988a–c; Lopez and Lui, 1990), the current-sheet disruption begins close to synchronous orbit and expands radially outward into the tail. This view is supported by two spacecraft observations of delays in substorm-related changes in the tail-lobe field. Jacquey, Sauvaud, and Dandouras (1991) and Ohtani, Kokubun, and Russell (1992) have both modeled these variations as the results of tailward propagation of current-sheet disruption. Recent Viking spacecraft auroral observations are also consistent with an origin very near the earth for the substorm expansion (Elphinstone et al., 1991). Using average magnetic-field mod-

els, the auroral oval and expansion onsets are found to project to distances just outside synchronous orbit, presumably the region in which the current-sheet disruptions take place. Furthermore, Murphree et al. (1991) and Murphree and Cogger (1992) have presented evidence that the expansion onset, or at least its initial stage, takes place on closed field lines. Such observations have led Lui, Lopez, Murphree, Ohtani, and others to conclude that the substorm expansion does not involve reconnection on the open field lines of the tail lobe. It has been suggested, instead, that the current disruption launches a rarefaction wave down the tail that induces plasma-sheet thinning and reduction in B_z. At some point down the tail, late in the expansion phase, or perhaps at the beginning of the recovery phase, these effects initiate reconnection and the subsequent generation of a plasmoid.

13.7 CONCLUSIONS

In this chapter we have described the important role of the solar wind in the generation of geomagnetic activity. We have continued the discussion showing that the energy coupled to the earth's magnetic field is transported and controlled by internal processes that produce the variety of magnetic variations known as magnetic activity. Some of this energy eventually reaches the earth's atmosphere, while the rest returns to the solar wind. The primary phenomenon organizing this energy transport is the magnetospheric substorm. The substorm is created by the superposition of two types of processes. One type is directly driven by the solar wind, and the other type corresponds to the unloading of energy stored in the magnetosphere by the driven processes. Isolated substorms are caused by short intervals of a southward IMF (30–60 min). Isolated substorms have identifiable and repeatable structures consisting of three phases called growth, expansion, and recovery. It is generally accepted that the growth phase is produced by changes in the configuration of the earth's field caused by unbalanced reconnection. Until nightside reconnection begins and returns magnetic flux to the dayside, the configuration must evolve as the amounts of open and closed flux in various regions change. Eventually this leads to a flux catastrophe that we believe is the expansion phase.

At the present time, different models posit quite different mechanisms for the expansion and recovery phases of substorms. The driven model discusses neither nightside reconnection nor its effects. The boundary-layer-dynamics model starts reconnection at the distant X-line at expansion onset. The thermal-catastrophe model starts it at the distant X-line soon after it begins on the dayside. The substorm expansion is then an explosion of the plasma-sheet boundary layer due to absorption of ULF waves originally generated on the high-latitude boundaries of the magne-

totail. The near-earth neutral-line model starts reconnection close to the earth on closed field lines in the late growth phase. In this model, expansion onset is the time at which severance of the last closed field line occurs. The current-sheet-disruption model argues that the expansion phase is caused by an instability of the near-earth current sheet that diverts current through the ionosphere. This disruption apparently does not alter open field lines until the recovery phase begins. The magnetosphere–ionosphere coupling models suggest that the ionosphere is the cause of cross-tail current disruption and, like the disruption model, suggest that the reduction in cross-tail current then leads to a formation of an X-line and reconnection. Regardless of which picture of the onset is ultimately found to be appropriate, the onset of the expansion appears to be an internal process seldom directly driven by the solar wind.

From the wide variety of models postulated for the expansion phase, it becomes apparent that we do not yet understand substorms. Kan (1990) has suggested that greater effort be made to develop a global model of substorms that will incorporate all of the well-established observations. In this regard, we have noted a distressing tendency for many researchers, both young and old, to ignore observations that are inconsistent with their particular models. For example, the boundary-layer-dynamics model and current-disruption model tend to ignore observations in the central plasma sheet of tailward flow threaded by a southward field within a minute or two of expansion onset. These models also ignore the fact that the magnitude of the lobe field decreases significantly during the expansion phase, indicating that a reduction in the amount of open flux is occurring in this phase, not just during the recovery phase. Furthermore, observations have established that plasmoid release must occur before the beginning of substorm recovery, not at the beginning of recovery as in the disruption model. On the other hand, the near-earth neutral-line model does not provide an explanation for the auroral-arc breakup at low latitudes that maps to the current sheet just outside synchronous orbit. Neither can it explain how an entire expansion phase can occur on closed field lines, as suggested by some of the recent Viking auroral images. But if this interpretation of the images is correct, how is open flux of the lobe converted to closed flux on the dayside without obvious manifestations?

The conclusion of our discussion is that there are still problems in our models of substorms. Several models are useful in systematizing a selected subset of observations, but no model has yet proved capable of accounting for all of the observations. In the next decade we can expect to obtain additional data from spatial regions that have not been fully investigated. New instruments will improve our knowledge of particle responses to substorm dynamics. Missions involving several spacecraft will give us insight into the three-dimensional structure of the current and field perturbations. We can be optimistic that these new measurements will enable us to reject some proposed models and improve oth-

ers. We can be certain that research on the complex phenomenon that we call a substorm will continue to be challenging and rewarding.

Appendix 13A INSTRUMENTS FOR MEASURING MAGNETIC FIELDS

13A.1 INTRODUCTION

MAGNETIC VARIATIONS MEASURED by instruments on the ground and in space provide information about the solar wind and the earth's magnetic field. These fields are produced by the superposition of the effects of a variety of different current systems. The processes that produce these currents are the subjects of interest to space physicists. A number of different instruments have been developed to measure magnetic fields. Originally, those instruments were basically variations of the compass. More recently, such instruments have begun to be replaced with more sophisticated devices based on other principles, including magnetic hysteresis, proton precession, and the Zeeman effect. In this appendix we describe some of the most common magnetic instruments in use today.

13A.2 MAGNETIC VARIOMETERS AND STANDARD OBSERVATORIES

Magnetic fields can be measured in a variety of ways. The simplest measurement device still in use today is the compass. A compass consists of a permanently magnetized needle, balanced to pivot in the horizontal plane. In the presence of a magnetic field and the absence of gravity, a magnetized needle would align itself exactly along the magnetic-field vector. When balanced on a pivot in the presence of gravity, the needle aligns with a component of the field. In the familiar compass, this is the horizontal component. A magnetized needle may also be pivoted and balanced about a horizontal axis. If this device (called a dip meter) is first aligned in the direction of the magnetic meridian, as defined by a compass, then the needle will line up with the total-field vector and measure the inclination angle I. Finally, it is possible to measure the magnitude of the horizontal field by use of the oscillations of the compass needle. It can be shown that the period of oscillation of a compass needle depends on properties of the needle and the strength of the field.

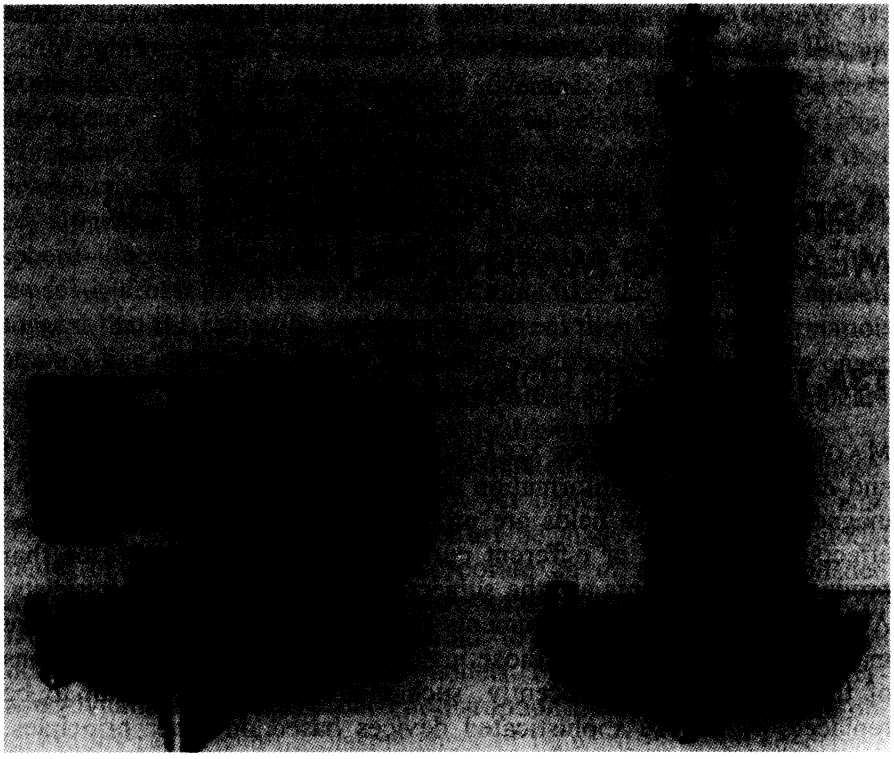

FIG. 13A.1. Examples of older instruments used to measure variations in the earth's magnetic field. Magnets suspended by quartz fibers reflect light beams onto photographic paper.

Magnetic observatories continuously measure and record the earth's magnetic field at a number of locations. In the older observatory instruments, magnetized needles with reflecting mirrors are suspended by quartz fibers inside of brass-and-glass housings like those shown in Figure 13A.1. Light beams reflected from the mirrors are imaged on a photographic negative mounted on a rotating drum. Variations in the field cause corresponding deflections on the negative. Typical scale factors used for such instruments correspond to 2–10 $nT \cdot mm^{-1}$ vertically and 20 $mm \cdot h^{-1}$ horizontally. A print of the developed negative is called a magnetogram. Typical magnetograms are illustrated in Figure 13A.2. The three traces in each panel are respectively the *D*, *Z*, and *H* components, as defined in Section 13.2.1. The bottom panel shows a storm's sudden commencement at 0826 UT on 22 March 1979 that is followed by a long chain of Pc-5 magnetic pulsations (see Chapter 11). A strong negative bay disturbance produced by a substorm began at 1054 UT. Definitions of these phenomena are given in Section 13.2.

Magnetic observatories have recorded data in this manner for well over 100 yr. Their magnetograms are photographed on microfilm and submitted to World Data Centers, where they are available for scientific or practical use. Among these uses are the creation of world magnetic maps for navigation and surveying, correction of data obtained in air,

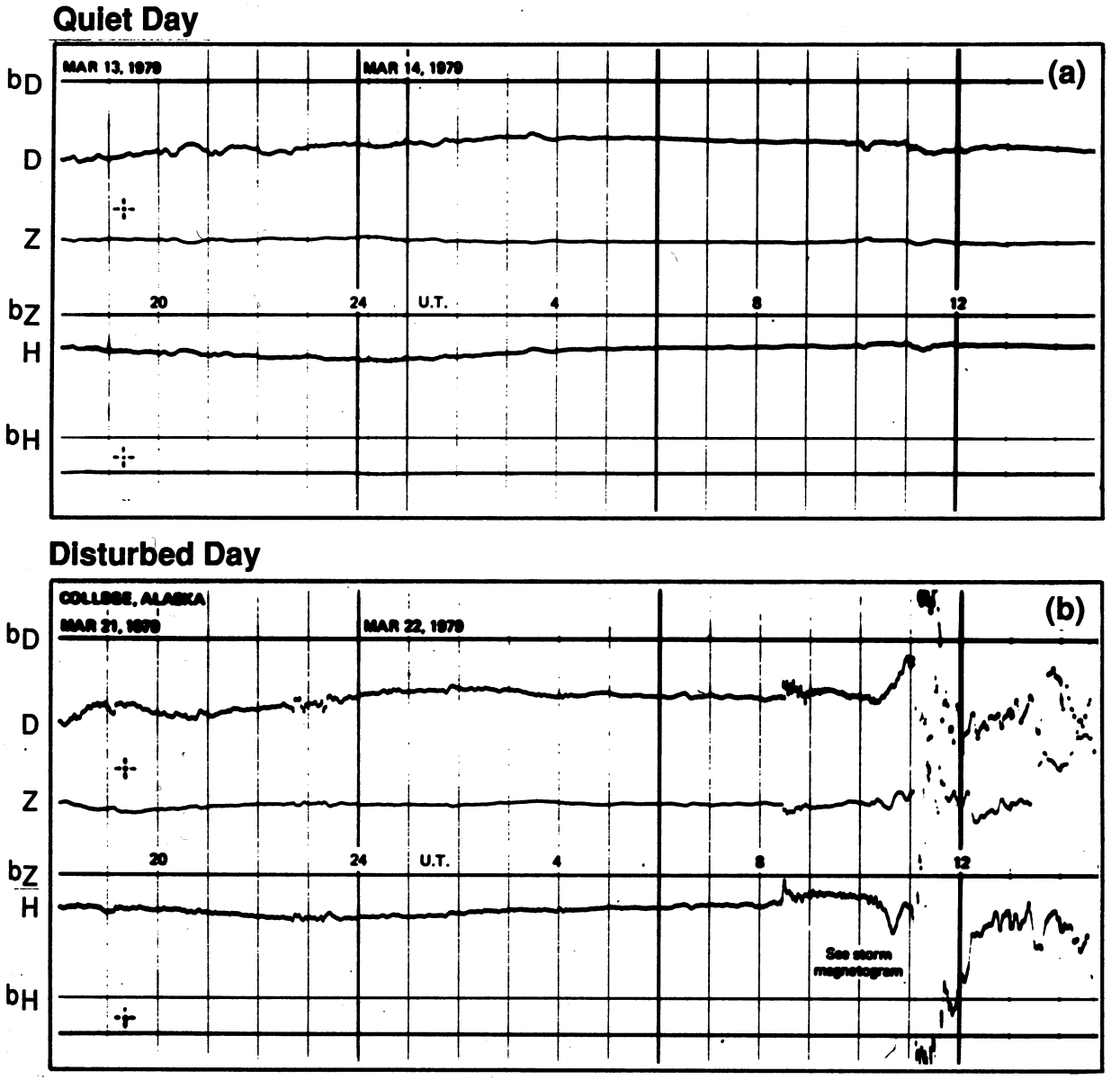

FIG. 13A.2. Example of magnetograms from a standard magnetic observatory at College, Alaska. The top panel displays a quiet day (14 March 1979), and the bottom panel a disturbed day (22 March 1979). Horizontal lines are baselines for the traces produced by the three variometers. The magnetogram spans two days of universal time, because the record at this observatory is changed at 8 A.M. local time.

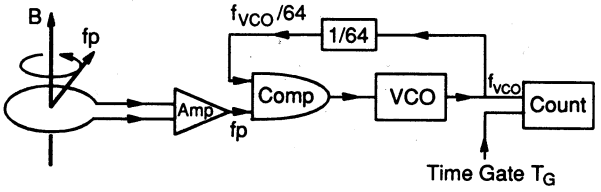

FIG. 13A.3. Schematic illustration showing how absolute measurements of the total magnetic field are made using the principle of proton precession.

land, and sea surveys for minerals and oil, and scientific studies of the interaction of the sun with the earth.

13A.3 VAPOR AND LIQUID MAGNETOMETERS

Today, other methods of measuring magnetic fields are more convenient, and the older instruments are being gradually replaced. One new type of instrument is the proton-precession magnetometer. The proton magnetometer takes advantage of the fact that the magnetic moment of a proton makes it act like a small bar magnet. Using this property, it is possible to temporarily align a proton in an external magnetic field created by passing a strong current through a coil wrapped around a container of liquid containing protons (water or kerosene). When the polarizing field is turned off, the protons try to align themselves with the earth's field. However, because they are spinning as well as magnetized, they initially behave like gyroscopes and precess around the earth's field. The precession frequency f_p is proportional to the magnitude of the earth's field B through a constant called the gyromagnetic ratio ($f_p = B/g_p$). This frequency can be measured by detecting magnetic effects of the gyrating protons with the polarizing coil. When the polarization field is suddenly released, all the protons are aligned in the same direction and begin to gyrate in phase, producing a time-varying magnetic field at the precession frequency. Very quickly the protons interact with other particles and are scattered in direction and phase. These interactions cause the signal strength to decay with time. Figure 13A.3 summarizes this technique of measurement. The exponentially damped sinusoid detected by the polarizing coil is amplified (Amp) and presented to a phase comparator (Comp). The voltage output of the comparator is input to a voltage-controlled oscillator (VCO) running at approximately 64 times the input frequency (f_{VCO}). The output of the VCO is fed back through a circuit that reduces the frequency by 64 and inputs it to the phase comparator. As long as the two inputs to the comparator have different frequencies, the output voltage of the comparator will continue to change the VCO frequency until it matches the precession frequency. The output of the VCO is also the input to a digital counter that counts the number of cycles of the VCO over a specified time interval (time gate). The various constants can be adjusted to force the output count to exactly equal the field in nanotesla, or any desired fraction thereof.

A proton-precession magnetometer is an absolute instrument that uses a fundamental property of matter to obtain an unbiased measurement of the strength of a magnetic field. To obtain the direction of the field, some other method must be devised. A method that uses the proton-precession magnetometer as a sensor consists of a coil system with three orthogonal axes surrounding a proton sensor. Accurately known currents are successively passed through each of the three coils, and the total field, consisting of the earth's field plus that of the coil, is measured. A total of four measurements, one for each coil and one without any field in the coils, is sufficient to calculate the three components of the vector field.

Instruments similar to the proton-precession magnetometer, but operating at much higher speed and with greater resolution, have been developed. These instruments are based on the Zeeman effect. In the presence of a magnetic field, electromagnetic radiation given off by an atom is split into two or more slightly different frequencies. This happens because the energy levels of the outermost atomic electron are split into a number of sublevels, depending on the orientation of the electron spin axis relative to the magnetic field. The separation depends on the strength of the magnetic field and is much less than the separation of the major levels. When an electron drops from an excited state to a lower level, the frequency of light given off depends on which sublevel the electron occupies at the beginning and the end of the transition.

A magnetic sensor that uses the Zeeman effect is constructed in the following way: Two cells containing atomic vapor (e.g., rubidium) and a photocell are aligned. The first cell is heated, causing light to be generated at all frequencies corresponding to the possible transitions between the two lowest atomic levels. Light from the first cell passes through a circularly polarizing filter and into the second cell. The polarized light is absorbed by the vapor in the second cell, raising its electrons to specific sublevels of the higher energy state. The electrons then fall to the lower level, giving off light. However, because the light is circularly polarized, transitions are allowed only between certain sublevels. After a brief interval, all electrons occupy the highest sublevel of the lowest energy state, and no more absorption is possible, because transitions from the high level are forbidden. While absorption is in progress, the amount of light falling on the photocell decreases, but when all electrons are "pumped" to the highest sublevel, it returns to normal. The pumped electrons can be redistributed to lower sublevels by applying a radio-frequency signal corresponding to the frequency difference between the sublevels of the higher state. This is done with a coil surrounding the second cell. If the frequency corresponds exactly to the Zeeman splitting, the electrons change levels, and the pumping process can be repeated. The required frequency is generated by using the output of the photocell to control the frequency of an oscillator. If proper phase shifts are introduced between the photocell and coil, the system oscillates

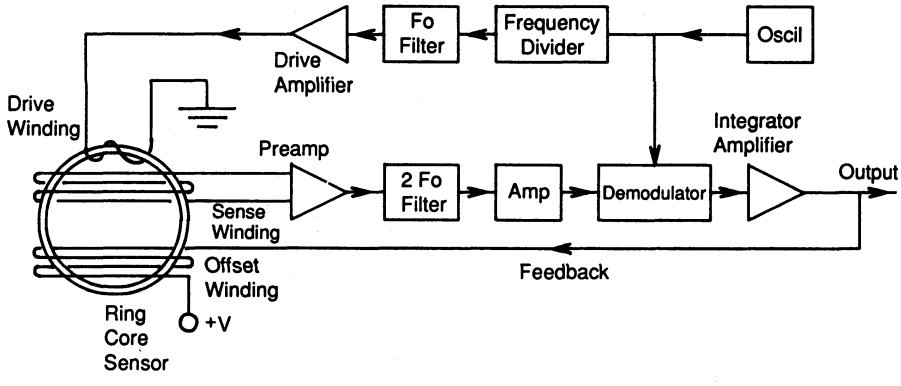

FIG. 13A.4. Schematic illustration showing the basic components of a fluxgate magnetometer.

between absorption and radiation. The field strength is measured by determining this oscillation frequency with a counter and converting it to an equivalent magnetic-field strength.

13A.4 FLUXGATE MAGNETOMETERS

Today the simplest and most common method of measuring vector magnetic fields is with a fluxgate magnetometer like that displayed as a block diagram in Figure 13A.4. The sensor in this instrument is a transformer wound around a high-permeability core (ring-core sensor). The primary winding of the transformer is excited by high-frequency current (5 kHz). The permeability of the core and the strength of the current are chosen so that the core is driven into saturation on each half cycle of excitation. The secondary winding detects a time-varying voltage that is related to the input through the hysteresis curve of the core material. For high-permeability materials, this curve is very nonlinear, and the output signal is highly distorted, containing all harmonics of the input signal. In the special case of no external magnetic field along the axis of the transformer, the hysteresis loop is executed in a symmetric manner. For this case it can be shown that only odd harmonics of the drive frequency are present in the output. If, however, an external field is present, the output reaches saturation in one-half cycle sooner than it does in the other half. This lack of symmetry introduces even harmonics into the output signal. The amplitude and phase of all even harmonics are proportional to the magnitude and direction of the field along the transformer axis.

In practice, the strength of even harmonics is very weak relative to that of odd harmonics. To amplify and detect these weak signals, the odd harmonics must first be eliminated. This is accomplished with the race-track or ring-core transformer, which is designed with two identical, parallel cores. Each core is excited by separate coils carrying equal currents in opposite directions. One secondary coil wound around both primaries is used to detect the output. For zero external field, the

two coils induce exactly equal but opposite effects in the secondary, producing zero output. For nonzero field, the odd harmonics still cancel, and only even harmonics appear in the output. The second harmonic is amplified and detected, giving a voltage proportional to the field along the transformer axis. Three components of a vector field are measured by three separate sensors with their transformer axes in mutually orthogonal directions.

The electronics needed to measure one component of the magnetic field is constructed as shown in Figure 13A.4. A precision oscillator (Oscil) generates a string of pulses at a frequency $2f_0$, where f_0 is the final drive frequency. This signal is passed to the demodulator circuit as a reference signal, and to a frequency divider ($\frac{1}{2}$). The output of the divider (f_0) is passed through a narrowband filter to the drive amplifier. The string of pulses from the drive amplifier is applied to the primary (Drive) winding of the transformer. A secondary winding around the transformer (Sense) detects the total induced signal and passes it to a preamplifier (Preamp). The output of the preamp passes through a narrowband filter of frequency $2f_0$ and is further amplified. A strong signal of frequency $2f_0$ is then presented to the synchronous demodulator. This is simply an electronic double-pole, double-throw switch. Each time the input waveform starts to change sign, the switch is activated by the reference signal, making the output signal positive or negative depending on whether the input lags or leads the reference by 180°. The output waveform thus has a frequency of $2f_0$. The output of the demodulator is input to an integrator/amplifier that smooths over many cycles of the rectified waveform, producing a near-dc voltage, with amplitude proportional to the amplitude of the second harmonic component output by the sensor, and with sign depending on the phase of the second harmonic relative to the reference signal. These two quantities are, respectively, proportional to the magnitude and direction of the component of the external magnetic field along the axis of the transformer. This low-frequency signal is the magnetometer output. The output voltage is also used to supply a current to an offset winding wrapped around the transformer. The coil constant of this winding and a feedback resistor are chosen so that the current that flows in the offset winding will exactly cancel the magnetic field along the coil axis. Thus, the fluxgate magnetometer serves only as a null detector, making the entire instrument very linear over a large dynamic range.

A fluxgate magnetometer is not an absolute instrument like the proton-precession magnetometer, and so it must be calibrated against standards. For measurements of the accuracy required by magnetic observatories, this calibration is difficult and includes sensor offset, sensitivity, temperature coefficients, and alignment angles. Such calibrations require large nonmagnetic test facilities with three-axis calibration coils, proton magnetometers, and optical theodolites.

13A.5 DIGITAL MAGNETIC OBSERVATORIES

Modern magnetic observatories usually include both proton-precession and fluxgate magnetometers mounted on granite pillars in nonmagnetic, temperature-controlled rooms. The outputs from the instruments are electrical signals that are digitized and recorded on magnetic media by systems with components as summarized in Figure 13A.5. Many observatories also transmit their information almost immediately to central facilities, where it is collated with data from other locations in a large computer database.

A typical digital data-acquisition system has the components shown in Figure 13A.5. A voltage output from some detector circuit is presented to a multiplexer (MUX). This device is simply an electronic selector switch with the added feature that it captures and holds the voltage for the time it takes to convert it to a digital value. The output of the MUX is applied to an analog-to-digital converter (ADC) that produces a digital representation of the input voltage. The rate at which the ADC converts a voltage is controlled by a precision clock. This clock is also an input to a microprocessor controlling the rate at which it performs calculations. The microprocessor produces control signals that switch the MUX from one input channel to another and initiates data conversion in the ADC. The output of the ADC is sent to the microprocessor, which stores it in memory while formulating an output record. This record is written to disk memory for temporary storage. Periodically, disk memory is dumped to some other device such as a modem (modulator/demodulator) that outputs to a radio or telephone, a tape device, or a computer network connection.

Magnetic measurements are often made at locations remote from the fixed observatories. Such measurements are part of a survey to better define the earth's main field or to detect anomalies in it. They are routinely carried out by people on foot or in ships, aircraft, and spacecraft. For surveys near the earth's surface, the proton-precession magnetometer is also always used, because it need not be precisely aligned.

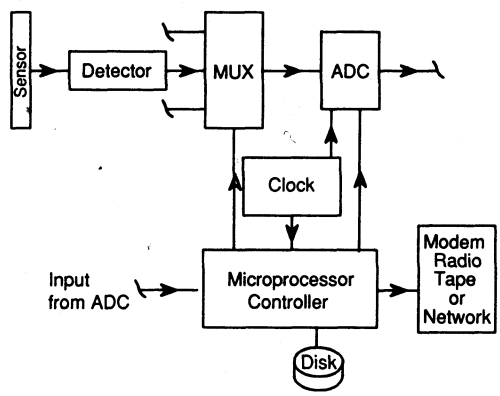

FIG. 13A.5. Block diagram showing the major components of a system to obtain and record digital data from the analog voltage output of an electronic magnetometer.

Well above the earth's surface, the main field decreases rapidly, and the requirement of precise alignment is less stringent. Thus, fluxgate magnetometers are generally used on spacecraft. Calculation of components of the vector field in a coordinate system fixed with respect to the earth requires knowledge of the location and orientation of the spacecraft.

Appendix 13B STANDARD INDICES OF GEOMAGNETIC ACTIVITY

13B.1 INTRODUCTION

MAGNETIC INDICES are widely used in studies of solar-terrestrial physics. Ideally, an index can be simply derived and is a monotonic function of some important physical parameter related to the phenomenon causing the disturbance. The original indices were simply averages of subjective impressions of the level of disturbance. With time, various components of magnetic activity were identified, and new indices were defined to isolate these specific phenomena. In this appendix we describe some of the most common indices in use today.

13B.2 THE CHARACTER FIGURES CI AND C9

One of the simplest magnetic indices, and one with nearly the longest record, is the character figure C. This index is generated daily at a specified set of collaborating observatories. It consists of a subjective determination of the level of disturbance observed in each Greenwich day at each observatory. A number (0, 1, 2) is assigned depending on whether the record appears to be quiet, moderately disturbed, or severely disturbed. The average over all observatories is termed the international character figure Ci. A nonuniform tabular transformation of this average to a scale from 0 to 9 is called the C9 index. Figure 13B.1 shows the standard tabulation of this index for a 3-yr interval approaching solar maximum in 1980. Daily values are plotted horizontally, with data from successive solar rotations placed below those for the preceding rotation. Notice that the sizes of figures increase with their numerical size, thus allowing a ready appreciation of the intervals of disturbance. Often the effects of an active region on the sun, or a region of high dynamic pressure in the solar wind corotating with the sun, will stand out in such plots as a pronounced 27-day recurrence tendency (e.g., see the last half of 1973 near the end of a rotation interval). The daily sum K_p index defined in the next section has been calibrated against the historical

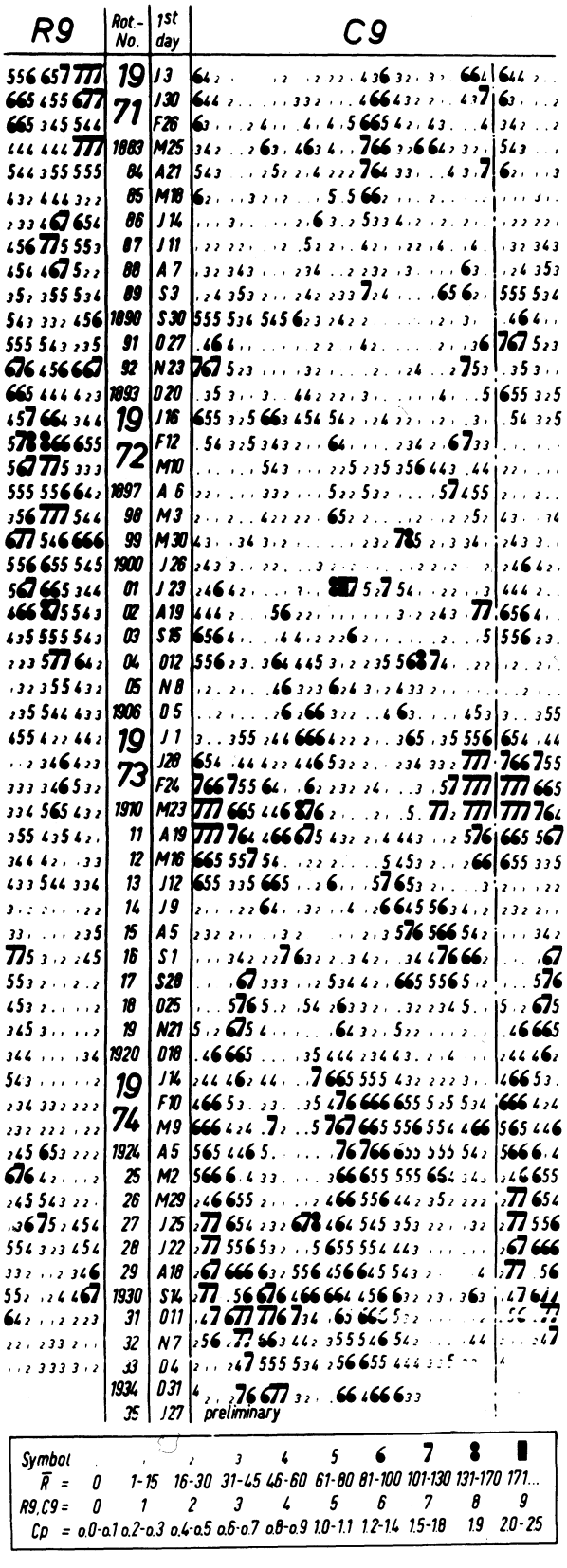

FIG. 13B.1. International character figure C9 displayed in standard form. The first major column displays the 3-day mean sunspot number, R9. The second column is the solar-rotation number or the calendar year. The third column shows the month and day of the first day of each solar rotation. The main column presents 27 consecutive days of the C9 index. The final column shows C9 for the first six days of the next rotation. The sizes of the figures are adjusted to correspond to the numerical value of the index, to emphasize disturbance.

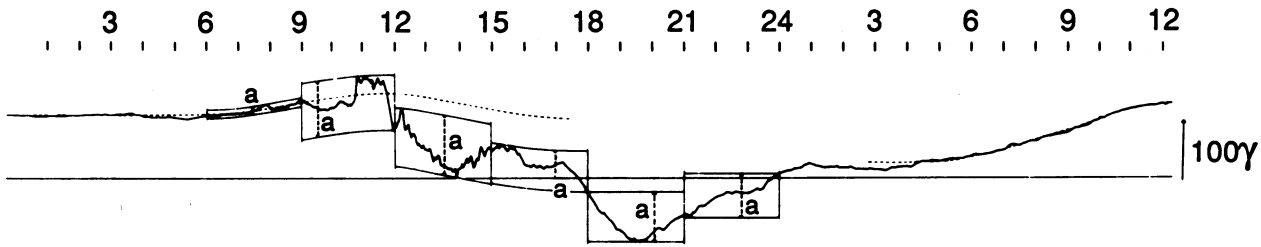

FIG. 13B.2. Illustration of the procedure by which the 3-h range index K is determined at a single station.

record of C9 and provides almost identical data. As a basis of comparison, it should be noted that the two magnetograms shown in Figure 13A.2 correspond to C9 values of 2 (quiet) and 7 (severely disturbed), respectively.

13B.3 THE RANGE INDICES K, K_p, A_k, and A_p

The K index is a range index for a field component at a given station that nominally measures the magnitude of disturbances caused by phenomena other than the diurnal variation and the long-term components of the storm time variation. It is calculated in the manner illustrated in Figure 13B.2. The Greenwich day is divided into 3-h intervals. For each interval, the diurnal variation appropriate to the season, phase of the moon, station, and component is visualized. This diurnal variation is translated vertically to the level for which it touches the minimum value of the trace in the interval and thence to the level for which it touches the maximum value of the trace in the interval. The vertical separation between the translated diurnal variations is the range of the component in that interval. The range is then converted to a quasi-logarithmic K index by a table specific to the observatory. The first transformation table was created for a European observatory and was used to generate K values for several years. Subsequently, the range values for the year 1938 at that reference observatory were tabulated and compared to range values at other observatories in the same year. Separate tables were produced for each observatory, such that their distribution of K values was the same as the distribution at the reference observatory.

It is very useful to have a single index like C9 derived from the K values of observatories distributed over the entire earth. Because in any interval these are located at different distances from the source of disturbance, they measure quite different K values. To create a planetary K index, K_p, the station indices are first standardized to K_s indices. This is done through tables that create equal distributions of K_s values of each station in every 3-h interval of every season. These tables, specific to each observatory, translate the integral (0–9) K values into 28 fractional K_s values quantized to units of $\frac{1}{3}$ ($0, \frac{1}{3}, \frac{2}{3}, \ldots, 9$). The K_p

index is then defined as the arithmetic average of the K_s values at 13 standard observatories. Originally, ranges were calculated for all three components, and the most disturbed component was used to define K. From 1964 onward, the vertical component was no longer used because it was shown to be sensitive to underground conductivity anomalies.

For long-term studies, it is useful to define daily average indices. Because of the logarithmic nature of the K indices, they are difficult to average. Consequently, linear range indices have also been defined. The A_k index at a station is obtained by multiplying the central value of the range used to define the integral K by a factor obtained by dividing the lower limit of the range corresponding to $K=9$ by 250. This number is typically of order 2; hence the A_k range index may be thought of as being twice the average amplitude of disturbance in the interval. A similar procedure is used with the K_p index to define a planetary A_p index with finer graduations.

13B.4 THE POLAR-CAP INDEX PC

The polar-cap index, PC, has only recently been proposed (Troshichev et al., 1988). This index is designed to be a measure of the strength of the sheet current flowing sunward across the polar cap, closing the two auroral electrojets. Thus, it provides a measure of the penetration of the solar-wind electric field into the magnetosphere. It is calculated using the following formula for ΔF:

$$\Delta F = \Delta H \sin\beta + \Delta D \cos\beta, \qquad (\beta = \lambda + \delta + \mathrm{UT} + \varphi)$$

In this relation, ΔH and ΔD are deviations of the H and D components from quiet levels, δ is the average declination of the station, λ is the geographic station longitude, UT is converted to degrees by multiplying by 360/24, and φ is the angle the average current vector makes with respect to local noon. ΔF is essentially the magnetic perturbation perpendicular to the average current direction. Figure 13B.3 illustrates this pattern when the IMF is southward, with B_y nearly zero.

The polar-cap conductivity and hence the return current are strong functions of solar illumination. Troshichev et al. (1988) have tried to determine this dependence using linear regression against various solar-wind coupling parameters. They have found the highest correlation (~0.7) for the linear relation

$$\Delta F = \alpha E_M + K = \alpha u B_T \sin^2\theta/2 + K$$

where α and K are constants, E_M represents the merging electric field, θ is the clock angle of the IMF about the x-axis, B_T is the magnitude of the projection of the IMF into the GSM y–z plane, and u is the solar-wind speed. Those authors propose that the PC index be defined by $\mathrm{PC} = \Delta F/\alpha$.

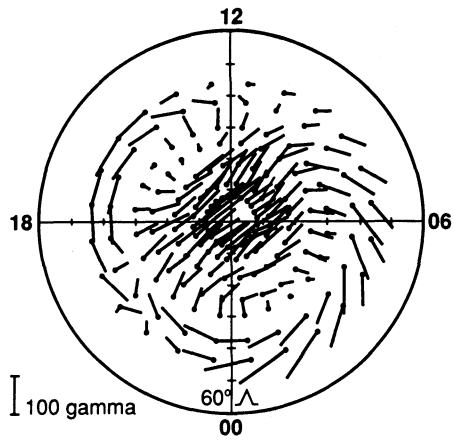

FIG. 13B.3. Polar-cap current pattern driven by a southward IMF. The strength of the ground perturbation normalized for ionospheric conductivity variations is the Pc index. (From Vennerstrom and Friis-Christensen, 1987.)

The advantages of the proposed PC index are several. First, it contains almost the same information about the solar wind as the AL index, with which it is highly correlated. Second, it can be calculated with data from only one station. Third, it is not contaminated by effects of the substorm current wedge, and therefore it appears to measure the driven response of the magnetosphere (the difference between driven and unloading processes was discussed earlier in this chapter). Its disadvantage is its sensitivity to changes in conductivity and its dependence on IMF B_y effects. It remains to be determined whether or not this index will be adopted as a worldwide standard and will produce new and interesting scientific conclusions.

13B.5 THE SUBSTORM INDICES AU, AL, AE, AND AO

The auroral-electrojet indices were defined by Davis and Sugiura (1966) to obtain a measure of the strength of the auroral electrojets relatively uncontaminated by effects of the ring current. The technique for calculating them can be understood by reference to Figure 13B.4. The third set of traces in this diagram displays the H-component traces from a worldwide chain of auroral-zone magnetic observatories. Monthly mean values are first subtracted from each station's trace to give a base value of zero. The traces are then plotted with respect to a common baseline, and upper and lower envelopes are calculated. The AU (auroral upper) index is defined at any instant of time as the maximum positive disturbance recorded by any station in the chain. Similarly, AL is defined as the minimum disturbance defined by the lower envelope. If the disturbances were caused by an infinite sheet current, then AU and AL would be proportional to the maximum overhead current density in the two electrojets. A single measure that approximates the total effect of both

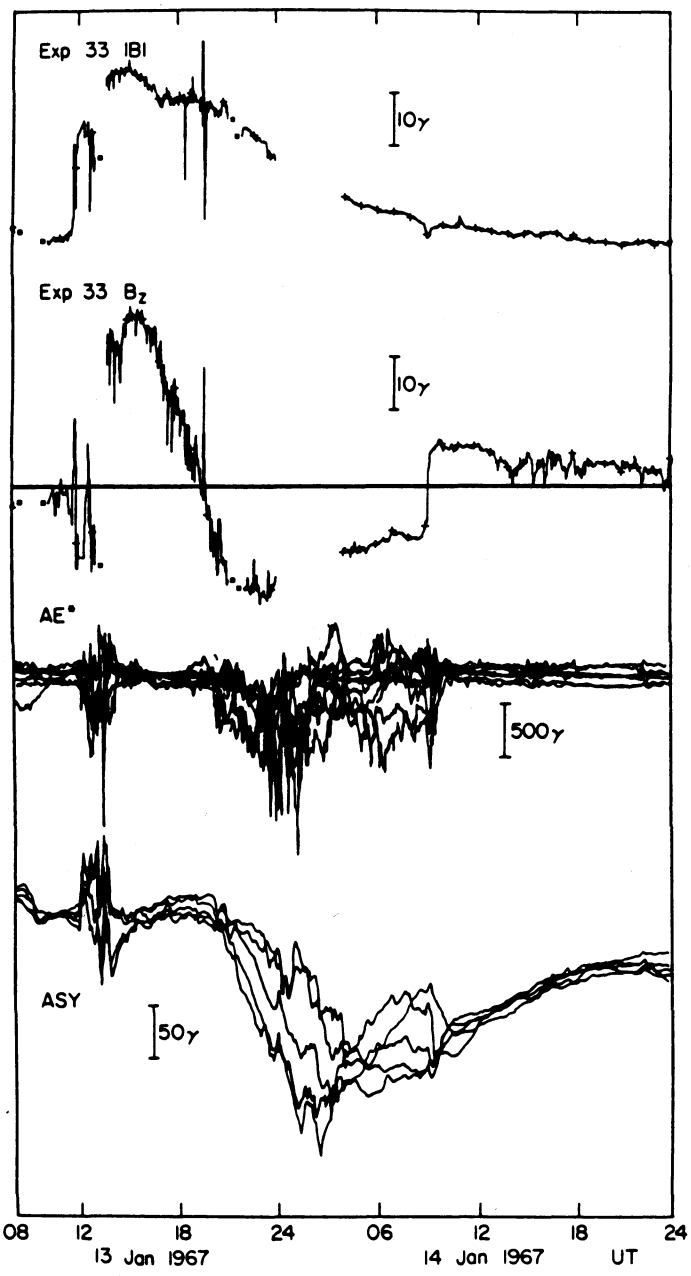

FIG. 13B.4. Illustration showing the type of data used to create the auroral-electrojet indices (AU, AL, AO, AE) and the disturbance storm time and asymmetry indices (D_{st}, A_{sym}). AU is the upper envelope of auroral-zone deviation of H (third panel, labeled AE) from a reference value; AL is the lower envelope, AO is the average, and AE is the separation of envelopes. At midlatitudes, D_{st} and A_{sym} are, respectively, the average deviation of H from a quiet day and the separation of the upper and lower envelopes (bottom panel, labeled ASY). The top two traces show that magnetic activity is produced by a strong interplanetary magnetic field pointing southward ($B_z<0$) and parallel to the earth's dipole axis. (From Wolf et al., 1986.)

electrojets is defined as AE = AU − AL. For completeness, AO is defined as the average of AU and AL: AO = (AU + AL)/2.

13B.6 THE STORM INDICES D_{st} and A_{sym}

A measure of the strength of the ring current is the disturbance storm time index D_{st}. This index is calculated by a technique similar to that used in the auroral-electrojet indices, with several refinements. These

are required because secular variations (long-term changes in the main field) and diurnal variations at each station can be as large as the storm time disturbance. The basic problem is to define a sequence of quiet values that can be used to define the trend and the seasonally dependent diurnal variation. Unfortunately, quiet days often occur in the recovery phase of magnetic storms. At such times the H trace is depressed, but is increasing exponentially with time. If these days are included in the determination of the secular trend, the result will be biased to low values. A similar problem arises in obtaining an average quiet day appropriate to a season. Storm recovery tends to make the values of H at the ends of days higher than at the beginnings. It is also difficult to define a quiet day. The quietest day in a month may contain some intervals of disturbance that can bias the calculation of an average variation. For the secular trend, the best that can be done is to take a sequence of midnight values (minimum diurnal variation) that occur during rare instances of no activity, well separated from magnetic storm recoveries. A polynomial fit to these values can then be subtracted from all data acquired by a station in a given year. From these data, one then selects quiet intervals identified by AE or some other index. These intervals are corrected for recovery trends and offset to zero at local midnight to remove storm bias. The data are then arranged in a two-dimensional matrix with rows for each day of the season and columns for each hour of the day. A two-dimensional Fourier analysis is performed, and only the low harmonics are retained. The trend coefficients and Fourier harmonics can then be used to predict a quiet H at any time of day and year.

After the quiet day is removed, the amplitude of the residual is adjusted by dividing by the cosine of the station's magnetic latitude. This is equivalent to the assumption that the measured H perturbation is the projection of an axial disturbance onto the tangential plane at the observatory (see Figure 13.1). The modified residuals can then be plotted against a common baseline, as illustrated in the bottom traces of Figure 13B.4. The D_{st} index is then defined as the instantaneous average around the world of the adjusted residuals (see the bottom trace). The A_{sym} index is a measure of the departure of the H perturbations from the axial symmetry expected for a ring current. It is defined as the instantaneous separation of the upper and lower envelope of H traces in the bottom set of traces. It is entirely equivalent to the AE index in the auroral zone. Similarly, D_{st} is similar to the AO index. As shown in Chapter 10, D_{st} is proportional to the total energy in the drifting particles that create the current (Dessler and Parker, 1959).

ADDITIONAL READING

Akasofu, S.-I. 1968. *Polar and Magnetospheric Substorms*. Dordrecht: Reidel.

Chapman, S., and J. Bartels. 1962. *Geomagnetism*, vol. 1. Oxford: Clarendon Press.

Coroniti, F. V., and C. F. Kennel. 1972. Changes in magnetospheric configuration during the substorm growth phase. *J. Geophys. Res.* 77:3361–70.

Kamide, Y., and J. Slavin (eds.). 1986. *Solar Wind Magnetosphere Coupling*. Tokyo: Terra.

Kan, J., T. A. Potemra, S. Kokubun, and T. Iijima. 1991. *Magnetospheric Substorms,* Geophysical Monograph 64. Washington, DC: American Geophysical Union.

McPherron, R. L. 1991. Physical processes producing magnetospheric substorms and magnetic storms. In *Geomagnetism,* vol. 4, ed. J. Jacobs (pp. 593–739). London: Academic Press.

Russell, C. T., and R. L. McPherron. 1973. The magnetotail and substorms. *Space Sci. Rev.* 11:111–22.