Paradigm transitions in solar–terrestrial physics from 1900: my personal view

S.-I. Akasofu

International Arctic Research Center, University of Alaska Fairbanks, P.O. Box 757340,
Fairbanks, Alaska, USA

Correspondence to: S.-I. Akasofu (sakasofu@iarc.uaf.edu)

Abstract. Solar–terrestrial physics, like any other scientific field, has evolved and developed by replacing older theories with newer theories. Unfortunately, each generation of young researchers tends to learn naturally only the latest, and perhaps the most popular theory and believes that it is the only useful one to pursue. Therefore, they do not necessarily realize that in the past the theory they chose had struggled to reach its presently acceptable state, and that eventually it might be replaced with a new theory. Two generations of scientists or in some subjects even more generations tend to be guided by one particular idea or theory. Thus, among us (namely, one or two generations) a high degree of agreement occurs, both on the theoretical assumptions and on the problem to be solved within the framework provided by the theory. Such an idea or theory was termed paradigm by Kuhn (1970). The purpose of this article is to describe several examples of the transition of paradigms and ideas in the subjects of solar–terrestrial physics. The examples are subjects that experienced a paradigm change after prevailing in the field for a few generations and also some that are perhaps on the verge of the transition. The chosen subjects are (1) Stormer's single particle theory to Chapman's plasma theory (1907–1963), (2) the auroral zone to the auroral oval (1860–1971), (3) the closed to open magnetosphere (1931–1971), (4) the current system controversies (1918–1963) and (1964–present), (5) the fixed pattern concept to the concept of auroral/magnetospheric substorms (1935–1982), (6) the importance of the interplanetary magnetic field (IMF) in the development of geomagnetic storms (1905–1966), (7) the ring current: solar wind protons to oxygen ions from the ionosphere (1933–1977), (8) the storm–substorm controversy (1963–present), (9) substorm onset (1964-present), (10) solar flares (1958–present) and (11) sunspots (1961–present).

1 Introductory remarks

Since the life of major paradigms is about 25–50 years, most of us spend our active research life during the period when a particular paradigm prevails, and thus we do not necessarily realize, in fact, that we ourselves are spending our time in a particular paradigm. We tend to consider together that there is only one main idea or theory to be pursued under the paradigm and do not pay much attention to the other theories; therefore, even when we recognize the need for the transition of a paradigm (because of many unsolved problems associated with it), it tends to take a long time for the transition, except for in cases of unexpected discoveries or crucial observations.

An interesting point in science is that solid theories/facts obtained in old paradigms, which have been thought to be obsolete or no longer adequate, are often useful and can become one of the building blocks of a new paradigm. By learning the history, one may not have to waste time in re-discovering the old solid facts. One good example is that in 1958 James Van Allen identified that motions of trapped particles in the Van Allen belt are identical to those of Stormer's elaborate computation (in the 1910s) of trajectories of single electrons from the sun in a dipole field, which were found to be inadequate for auroral studies. Another example is Chapman's SD (solar disturbance day variation) (2-D) current system (see Sect. 4, Fig. 4e), which was found to be an inadequate description of the storm time current systems, but neverthe-
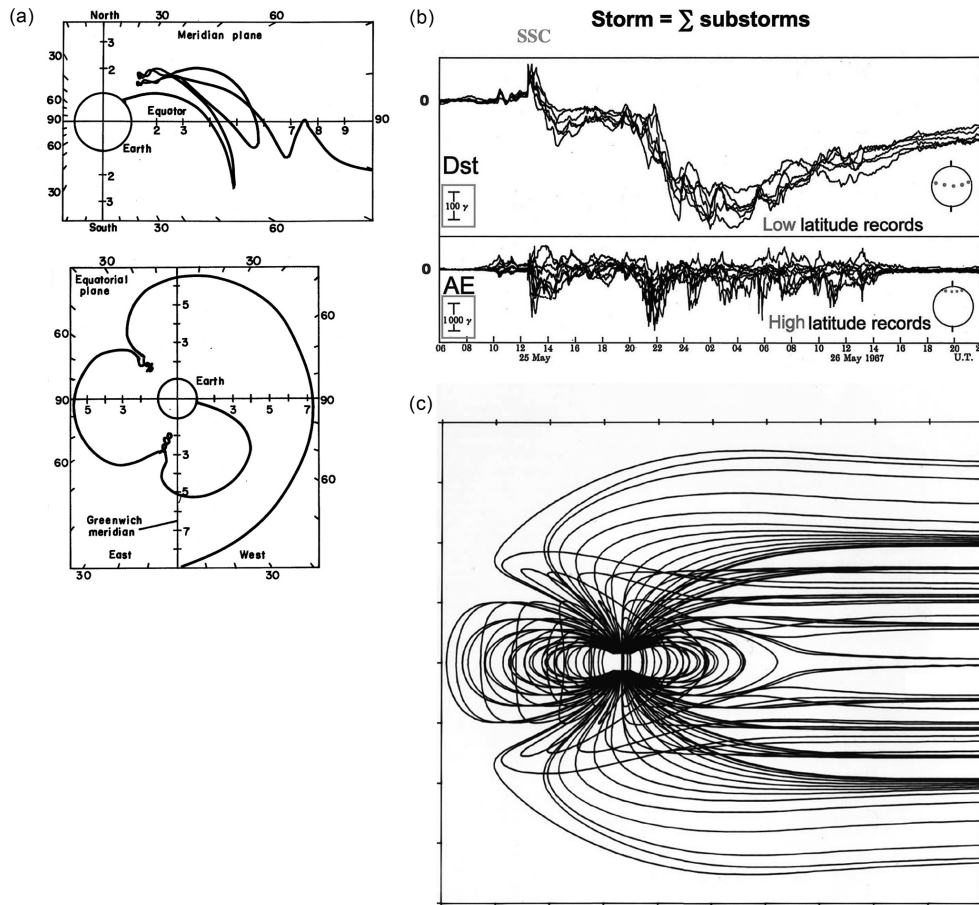
Figure 1. (a) An example of the trajectories of an electron in a dipole field (Stormer, 1955). **(b)** A typical example of a geomagnetic storm. The upper part is a collection of records (the horizontal component) at low-latitude observatories and the lower part at high-latitude observatories; note the difference of the scale in both. A step function-like change indicates the arrival of the shock wave, but Chapman considered that it signaled the arrival of solar plasma flow. **(c)** A non-magnetized plasma flow confines the earth and its magnetic field into a comet-shaped cavity.

less Dungey (1961) and Axford and Hines (1961) took a hint from it to suggest the concept of magnetospheric convection, which is now an important part of our understanding of magnetospheric processes. In a monograph or textbook, the magnetospheric convection is described without mentioning this background.

    I am fortunate to have had an opportunity to experience, participate in and witness a few paradigm transitions. Obviously, however, my point of view of a particular paradigm and of the transition from one paradigm to another can be different from those of other researchers. This is the reason why I added "my personal view" to the title of this paper. Indeed, the transition itself tends to be very controversial, and thus even if one is directly involved in it, it is difficult to describe it completely and without bias.

    The purpose of this paper is to describe my personal view of the transition of several paradigms and ideas which have occurred in solar–terrestrial physics since about 1900, in-

cluding my approach and reflection in overcoming difficulties which I had faced at the time of the paradigm transitions.

2   Stormer's single particle theory to Chapman's plasma theory (1907–1963): the need to account for the morphology of the auroral zone and to invoke charge neutrality of solar wind plasma was such a transition

This particular paradigm began with Stormer's study (1907) of the trajectories of charged particles in a dipole field and ended when the Chapman–Ferraro theory of the magnetosphere formation was confirmed by a satellite observation in 1963.

2.1   Stormer's single particle theory

In the 1800s, the luminescence in an electrical discharge tube suggested to a number of physicists that the aurora was

caused by incoming electrons from the sun. Stimulated by Birkeland's Terrella experiment, Stormer (1907, 1955) began his study of trajectories of electrons in the vicinity of the earth (a dipole field), which had been shot from infinity (the sun; see Fig. 1a). Most researchers in those days considered either single electrons or a swarm of electrons from the sun, until Chapman and Ferraro (1931) successfully theorized the impact of plasma flow on the earth's dipole field; for details of the history up to that time, see Akasofu (2011).

At the time, Stormer was the only person who could claim to explain some details of the aurora. During the Auroral Conference held at the University of Western Ontario in 1951 (Gerson et al., 1954: pp. 386–387), Stormer claimed, during a discussion session after Chapman's presentation, that his theory could explain (a) how auroral particles enter the atmosphere, (b) how the auroral zones around the two geomagnetic poles can be reproduced, (c) how his theory could explain arcs and rays structure of the aurora, etc.

On the other hand, Stormer's theory had two serious difficulties. First, for the energy range of electrons he considered, the computed auroral zones were formed only very close to the geomagnetic poles, not along the auroral zone. Second, a swarm of electrons would be dispersed by its own electrostatic force before reaching the earth (Lindemann, 1919); however, Stormer's theory was considered as the only theory for a long time.

2.2 Chapman–Ferraro theory of plasma flows

Chapman's first theory of magnetic storms assumed either protons or electrons (Chapman, 1918). It was Lindemann (1919) who suggested that solar gas should be what we now call plasma. Based on his suggestion, Chapman and Ferraro (1931) considered a plasma flow around a dipole field and explained successfully the onset of geomagnetic storms, a step-function-like increase of the horizontal component of the field as a result of the impact of the plasma flow (this phenomenon is called the storm sudden commencement, denoted by ssc; see Fig. 1b). The importance of their theory was that it could explain ssc's, and the fact that the earth's magnetic field is confined in a comet-shaped cavity in the plasma flow from the sun (see Fig. 1c). This cavity is what we now call the magnetosphere, so their theory is regarded as the first theory of the formation of the magnetosphere.

It is difficult to determine accurately when the Chapman–Ferraro theory became accepted. However, it is likely that their theory was generally recognized by the late 1950s, since it was the only theory which could quantitatively explain ssc's. However, the actual confirmation of their theory had to wait until 1966 when Cahill and Amazeen (1963) observed the magnetospheric boundary on the dayside from a spacecraft. In fact, the magnetic field intensity just inside the day side boundary was exactly twice the undisturbed earth's dipole field, as predicted by the theory.

2.3 Summary

The transition from Stormer's theory to the Chapman–Ferraro theory was crucial in advancing solar–terrestrial relationship, because it was recognized that the solar wind should be treated as a plasma, instead of a group of solitary particles. However, the Chapman–Ferraro theory could not explain the formation of the auroral zone (more accurately the auroral oval). This problem had to wait until the discovery of the open magnetosphere (Sects. 3 and 4).

3 The auroral zone to the auroral oval (1860–1971): the need for an accurate determination of the distribution of the aurora

This particular paradigm began at the time when Loomis (1860) published his paper on the auroral zone in 1860 and ended when a Canadian satellite imaged the auroral oval for the first time in 1971.

3.1 The auroral zone

Loomis (1860) obtained the sighting locations of the aurora on a geographic map and showed that it was a belt centered around the northwestern tip of Greenland. This belt is called the auroral zone, which had been thought to be the belt where auroral arcs lie; it is an annular ring of radius of about 3000 km in the geomagnetic latitude (geomag. lat.) local time coordinate (LMT); see Fig. 2a.

3.2 The auroral oval

Based on the International Geophysical Year (IGY) all-sky camera data, Feldstein (1963) determined the belt in which auroral arcs actually lie in the geomag. lat. local time coordinate; for details see Feldstein et al. (2014). This belt is called the auroral oval. However, it was difficult for auroral researchers to accept Feldstein's oval for many years.

Since the aurora is relatively a large-scale phenomenon, it had been rather difficult to convince aurora researchers of the validity of the auroral oval without a network of all-sky cameras (cf. Akasofu, 2007: 38–52). Fortunately, in 1971, an imager aboard the Canadian ISIS (International Satellites for Ionospheric Studies) satellite successfully imaged the auroral oval; see Fig. 2b.

Since then, the concept of the auroral oval has provided the natural coordinate for various polar upper atmospheric phenomena; as we see in the next section, the auroral oval delineates approximately the boundary of the open region of the magnetosphere (Sect. 4).

3.3 Summary

The transition from the concept of the auroral zone to that of the auroral oval marked the beginning of the modern auroral physics and magnetospheric physics, as demonstrated in
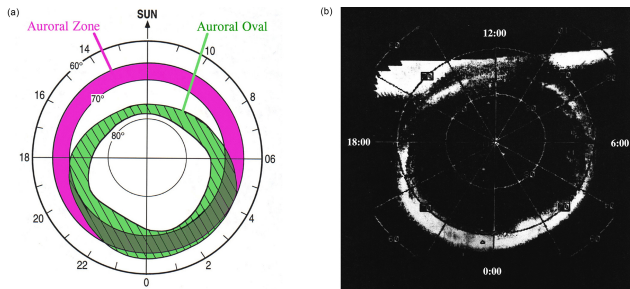
Figure 2. (a) Both the auroral zone and the auroral oval are shown in the geomag. lat. LMT coordinate for the purpose of comparison. **(b)** The first image of the auroral oval obtained by a Canadian satellite (Courtesy of C. Anger, 1972).

Sects. 3, 4 and 5. The IGY all-sky operation played a crucial role.

4 The closed and open magnetospheres (1931–1971): the discovery of the fundamental importance of the relationship between the auroral oval and the structure of the magnetosphere

This particular paradigm began when Chapman and Ferraro published their paper on the formation of the magnetosphere in 1931. Since their plasma was diamagnetic, their magnetosphere was a closed system. However, the discovery of solar energetic electrons within the auroral oval in 1971 made it clear that the magnetosphere is an open system, as suggested earlier by Dungey (1961) and others. Nevertheless, his open model had to wait for confirmation from the spacecraft observations by Vampola (1971).

4.1 The closed magnetosphere

As mentioned earlier, the Chapman–Ferraro theory indicated that the earth's magnetic field is completely confined in a comet-shaped cavity (Fig. 1c). The Chapman–Ferraro theory assumed an interaction between a non-magnetized plasma flow from the sun and the earth's magnetic field. Under this condition, the solar wind plasma cannot enter into the magnetosphere, except perhaps from the two foci of the Chapman–Ferraro current on the front side of the magnetopause. The fact is that the Chapman–Ferraro theory cannot transfer solar wind energy into the magnetosphere.

4.2 The open magnetosphere

Dungey (1961) considered the interaction between the solar magnetic field carried by the solar wind and the earth's magnetic field, as suggested by Hoyle (1949). Dungey (1961) began his paper by stating: "The discovery of a regular interplanetary magnetic field by Pioneer V has reawaken interest in Hoyle's suggestion that the primary auroral particles are

accelerated at neutral points in the combination of an interplanetary magnetic field (IMF) and the geomagnetic field"; see Fig. 3a.

Although his paper has become a guiding principle in the field of magnetospheric physics, his main interests at that time were the acceleration of auroral electrons at the neutral point or line (produced by merging between the IMF and earth's dipole field) and also Chapman's SD current system (Sect. 4); this point will be discussed later in this section. Although his figure showed clearly an open magnetosphere model, this particular point did not get much attention until 1971. Further, although his paper is regarded as the first paper to discuss magnetic re-connection, Dungey (1961) did not mention the term magnetic re-connection or merging in the paper.

Vampola (1971) found that energetic electrons of 376 keV are found uniformly within the entire auroral oval (Fig. 3b). Unless the magnetosphere is open, such electrons cannot penetrate into the earth's dipole field, except perhaps only a few degrees from the geomagnetic pole as Störmer showed or from the two foci of the Chapman–Ferraro current on the magnetopause.

Figure 3c schematically shows the difference of the situation in which the magnetosphere is either closed or open. Figure 3d shows points where the IMF lines and earth's magnetic field lines connect on the magnetopause (a model). The high energy electrons enter into the magnetosphere following the merged field lines. The earth's magnetic field lines connected with the IMF field lines are rooted in the polar cap. Those field lines are called open field lines, and the rooted area is called the open region. It soon became clear that the open region is bounded approximately by the auroral oval and thus that the auroral oval approximately delineates the open field region; this fact added to the credibility of the auroral oval as well (see Fig. 3d).

4.3 Summary

The discovery of the open magnetosphere was the beginning of modern-day magnetospheric physics. The importance of the effect of the interplanetary magnetic field on the structure of the magnetosphere was recognized. The finding of the relationship between the auroral oval and the open magnetosphere became crucial later in the development of auroral physics and magnetospheric physics.

5 The current system controversies: the need to connect the ionosphere and the magnetosphere with electric currents (1935–present)

The determination of the three-dimensional electric current system causing geomagnetic disturbances has been one of the important topics since the beginning of the 20th century. Birkeland (1918) was the first to propose a three-dimensional current system (a streaming electron beam from the sun),
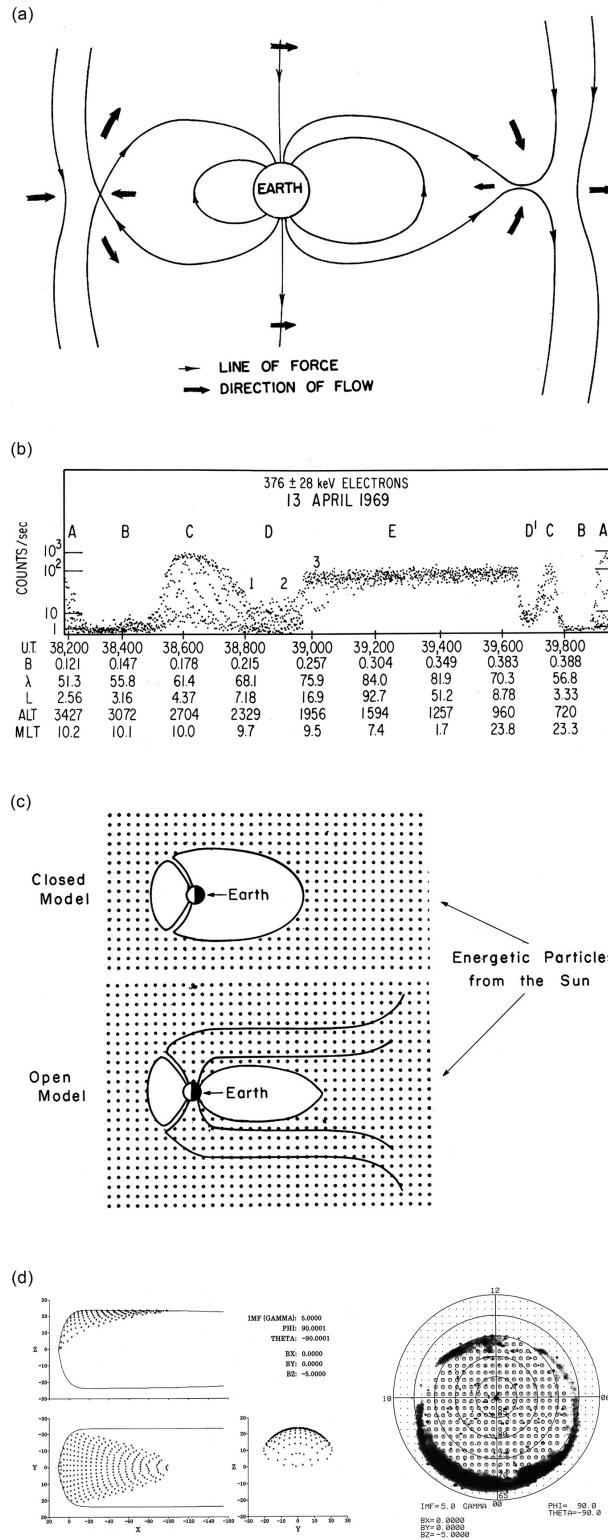
Figure 3. **(a)** Dungey's idea of the interaction between the IMF and the magnetosphere (Dungey, 1961). **(b)** Observation to show that high energy electrons penetrate in the auroral oval uniformly. Note the latitude (λ) where the uniform flux drops in the morning side and the night side (MLT). The lower latitude fluxes are the Van Allen belt electrons (Vampola., 1971). **(c)** Schematic figure to illustrate the difference of the closed and open magnetospheres, when solar energetic electrons reach the earth. **(d)** Left: the entry points of the merged field lines on the magnetopause based on a model. Right: the root of the merged field lines in the polar region is surrounded by the auroral oval.
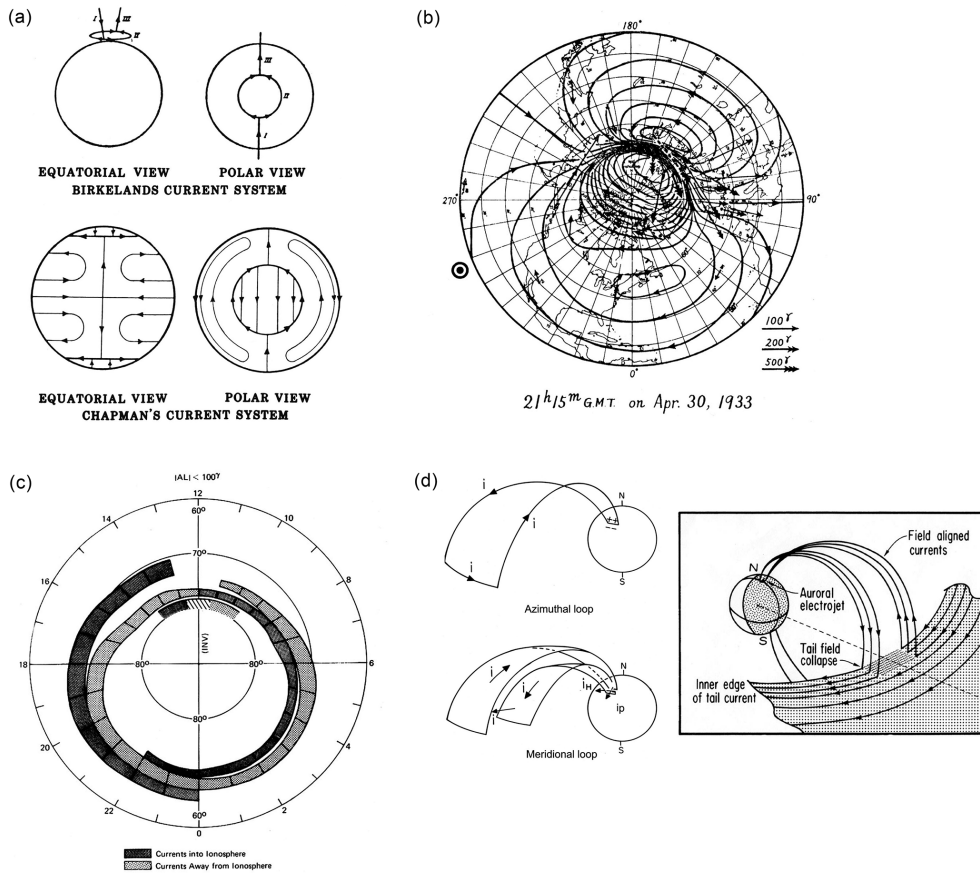
Figure 4. (a) Above: Birkeland's three-dimensional current system. Below: Chapman's two-dimensional current system. Both the equatorial and polar views are shown. (Chapman's polar view is the same as that in Fig. 4e). **(b)** An example of the substorm current system on a spherical shell (Fukushima, 1953). **(c)** The distribution of the field-aligned currents (Iijima and Potmra, 1978). **(d)** Left: Bostrom's current system. Right: the current system by Clauer et al. (1974).

while Chapman (1935) considered a two-dimensional current system, stating that it is not possible to determine the current system on the basis of ground-based observations alone. This "controversy" was in some way a matter of presentation, not physics. Our modern efforts have been trying to find a first approximation three-dimensional current system in understanding the magnetosphere–ionosphere coupling.

5.1 Two-dimensional or three-dimensional?

Analyzing a considerable number of magnetic records, Birkeland (1918) considered that electrons from the sun follow the earth's magnetic field lines and flow into the upper atmosphere; see Fig. 4a. On the other hand, Chapman (1935) was concerned about the uniqueness of determining the current system based on ground-based data alone. Thus, Chapman devised a method of representing the current system on a spherical shell, concentric with respect to the earth. Hence, a controversy about the current system, namely, two-dimensional or three-dimensional, is not a matter of contro-

versy based on physics, but is partly a matter of differences in the presentation.

On the other hand, Chapman (1931) was well aware of the existence of an electrically conductive upper atmosphere, since he published the first theory on the ionization of the upper atmosphere, namely, the formation of the ionosphere. Therefore, it was natural for him and many others to consider that the storm current system was located on a spherical shell. His work was followed by many researchers, including Silsbee and Vestine (1942), Fukushima (1953) and many others; Fig. 4b shows such an example.

The first detection of field-aligned currents was made by Armstrong and Zumuda (1973). Thus, it became clear that the current system is three-dimensional. The distribution of the field-aligned currents is refined by Iijima and Potemra (1978); Fig. 4c. The field-aligned currents are often called the Birkeland currents. However, Birkekand's field-aligned currents originated at the sun, because he believed that the aurora was caused by a beam of electrons shot from the sun. We understand now that the field-aligned currents
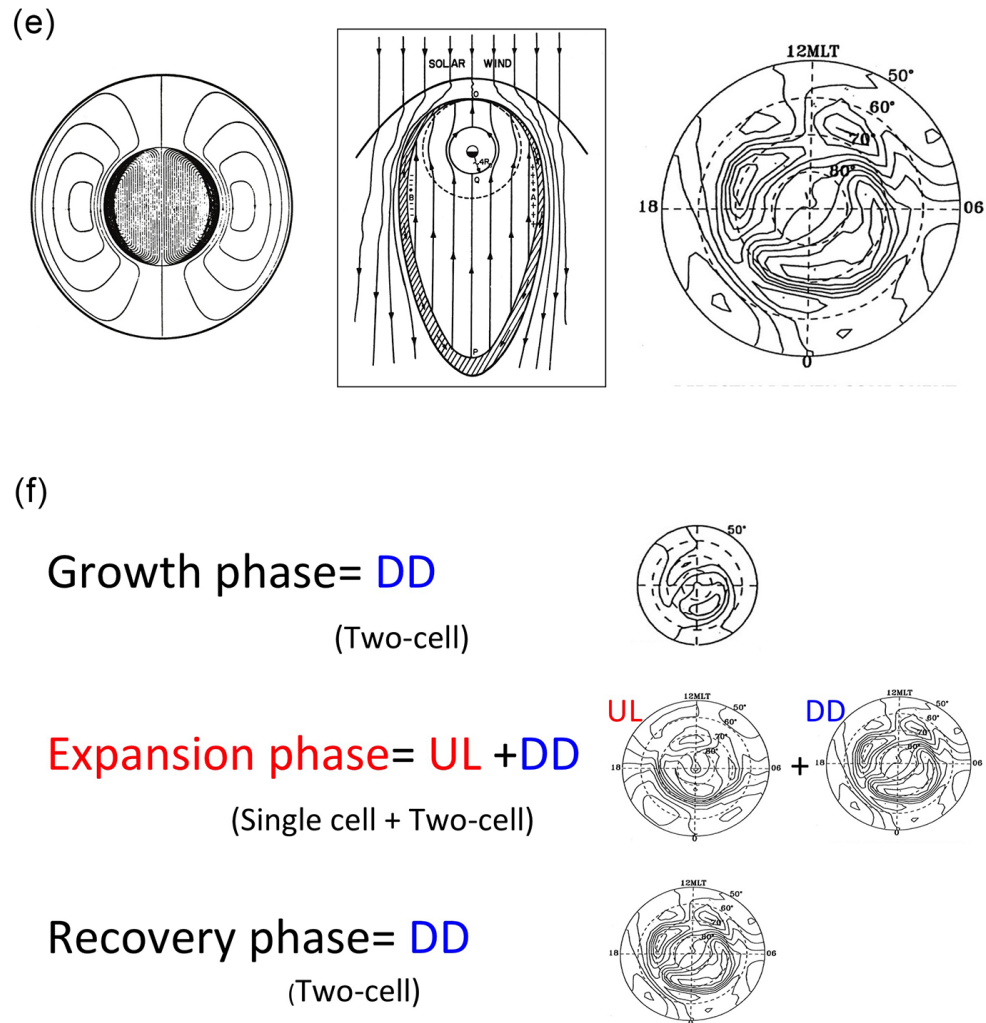
(e)

(f)

Growth phase= DD

(Two-cell)

Expansion phase= UL +DD

(Single cell + Two-cell)

UL   DD

Recovery phase= DD

(Two-cell)

Figure 4. (e) Left: Chapman's SD current (polar view). Middle: the pattern of magnetospheric convection on the equatorial (Axford and Hines, 1961). Right: the actual convection pattern (Sun et al., 1998). **(f)** The current system during substorms consists of two components. The two-cell current (directly driven, DD) exists throughout the life of substorms; however, single-cell current occurs only impulsively during the expansion phase.

result from a coupling between the ionosphere and the magnetosphere.

5.2 Current system for the expansion phase: Bostrom's current system or the diversion of the cross-tail current (the current wedge)

One of the most crucial problems in studying auroral substorms is finding the current system for the expansion phase. Bostrom (1964) proposed a current system, which has two components, the meridional component and the azimuthal component; see Fig. 4d.

Clauer et al. (1974) proposed another model, called the current wedge model. It assumes that the auroral electrojet is the diversion of the cross-tail current to the ionosphere (Fig. 4d). This model has been widely adopted. However, it should be recognized that the westward auroral electrojet during the expansion phase is mainly the Hall current which is driven by an equatorward-directed electric field. If the electrojet results from the diversion of the cross-tail current, it should be the Pedersen current driven by a westward electric field across the magnetotail; the potential drop across the magnetotail is supposed be transfered to the night side of the auroral oval. Incoherent scatter radar observations by Brekke et al. (1974) show that the electric field during auroral substorms is directed equatorward. Therefore, it is unlikely that the current wedge model is adequate, and the actual current system for the expansion has to wait for future studies.

5.3 Magnetospheric convection (1961–present)

As mentioned in the "Introductory remarks", Chapman's SD current system has become the basis for Dungey (1961) and Axford and Hines (1961) to suggest a large-scale convection of magnetospheric plasma, although it is inadequate The convection pattern given by Axford and Hines is shown as Fig. 4e.

In their inference, both Dungey (1961) and Axford and Hines (1961) considered that the ionospheric current is mainly the Hall current and thus plasma above the ionosphere flows along the current lines (but in the opposite direction) shown in the SD current on the basis of the frozen-in field concept, which assumes that plasma and magnetic field lines move together. Therefore, their convection pattern is the projection of the SD current flow onto the equatorial plane along the magnetic field lines. This is a good example that although an earlier result was inadequate, it can be very useful in developing a new idea.

As Fig. 4e shows, the actual current pattern in the ionosphere is greatly distorted from the ideal one (SD) because of anisotropic conductivity and non-uniform conductivity distribution. The actual ionospheric current resulting from the convection (called the directly driven, DD, current; see Fig. 4f) is further discussed in Sect. 9. However, it should be noted that there has so far been no large-scale observation of the convection pattern near the equatorial plane.

5.4 The current system during the expansion phase of substorms

The current system in the ionosphere during substorms has two components: the current driven by the large-scale convection of the magnetospheric plasma (DD) and the current which occurs impulsively only during the expansion phase (UL, unloading); the presence of the single-cell current was discovered by Akasofu et al. (1965). Both are present during the expansion phase (Fig. 4f). However, magnetometers record both together. Sun et al. (1998) successfully separated them.

5.5 Summary

The determination of the true three-dimensional current circuit is the traditional research objective in solar–terrestrial physics and is a very difficult subject, but we are making progress, in spite of many controversies. Bostrom's current system (Fig. 4d) for the expansion phase seems to be a good first approximation. The substorm current system has two components, DD and UL, and the UL current system is associated only with the expansion phase of substorms. The tail-wide earthward plasma flow which is supposed to be generated by magnetic re-connection cannot produce the UL current (Sect. 9).

6 The fixed pattern concept to the concept of auroral/magnetospheric substorms (1935–1982): the need to describe dynamic activities of the aurora

This particular paradigm began when Fuller (1935) published his first paper on the morphology of auroral activities. In his view, the aurora in the evening sector is quiet, active in the midnight sector and patchy in the morning sector, under which the earth and observers rotate once during the course of a night, observing those fixed three phases. His study was followed by Heppner (1954), which confirmed Fuller's view. Based on the IGY all-sky camera project, however, the concept of auroral substorm was developed (Akasofu, 1964), which differs from the fixed pattern concept. The first firm confirmation of the concept of auroral substorms was made by Frank et al. (1982) by a satellite observation from well above the north polar region.

6.1 The fixed pattern

Statistically, the aurora is quiet (mostly in the form of quiet arcs) in the evening hours, active (draperies and coronas) in the midnight sky and patchy (scattered rays) in the morning sky. The earth and observers rotate together under such a fixed pattern of auroral activities, witnessing the sequence of auroral changes from quiet, active and patchy forms during the course of a night; see Fig. 5a. This concept was established by Fuller and Bramhall (1937), Heppner (1954) and Fuller (1935) on the basis of a statistical summary of their observations in Fairbanks, Alaska. The field of view of single observers is very limited, and the earth's rotation carries the observers throughout the night, so that it was not possible to observe simultaneously auroral activities over the whole night hemisphere. The IGY all-sky camera operation made it possible to make a simultaneous observation of auroral activities at a number of locations.

6.2 Auroral substorms and magnetospheric substorms

All-sky camera operation during the IGY allowed us, for the first time, to examine auroral activities simultaneously over the entire dark hemisphere, say Siberia–Alaska–Canada. It was found that a particular type of auroral activity develops and subsides simultaneously all along the auroral oval for a few hours and repeats a few times during the course of active nights. Such an activity is called the auroral substorm (Akasofu, 1964); its development is shown in Fig. 5b. Since this concept differed from the concept of the fixed pattern and was not easily be accepted, although Feldstein and Starkov (1970) supported the new view. The confirmation of the concept of auroral substorms and their details required a satellite that can stay well above the north polar region for several hours. This was accomplished by Frank et al. (1982) with the
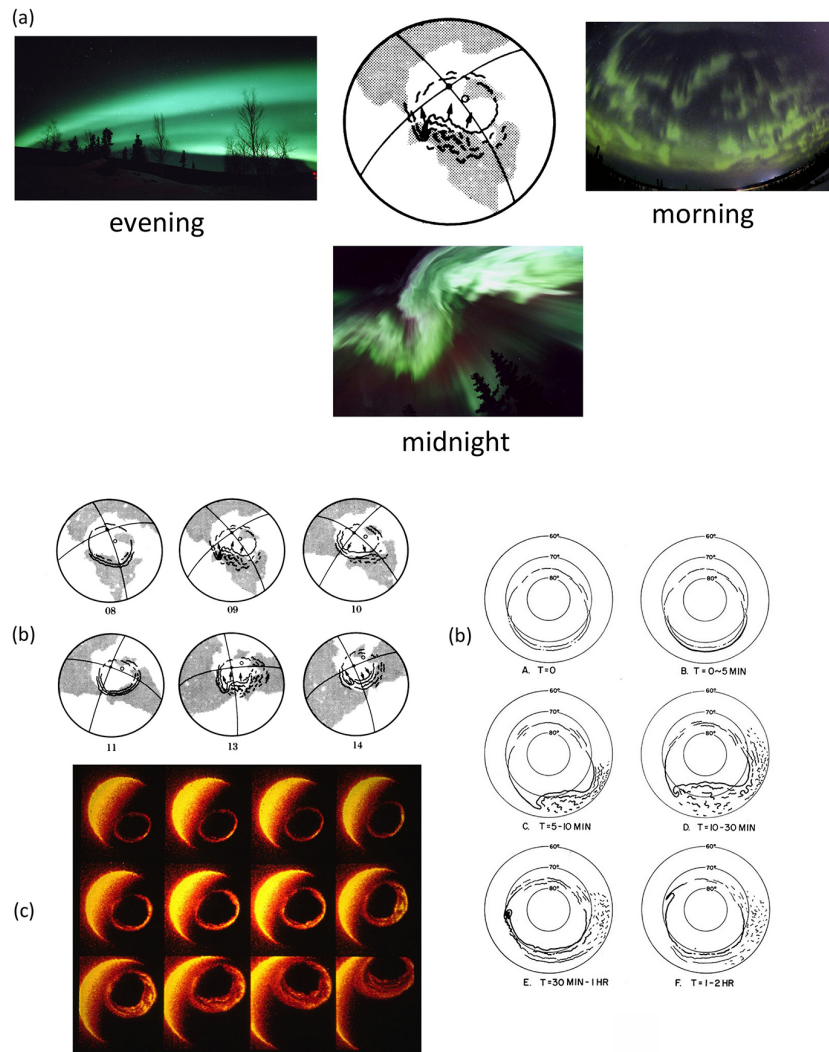
Figure 5. **(a)** The fixed pattern of auroral activities. In the evening sector, the aurora is relatively quiet, very active in the midnight sector and patchy in the morning sector. The earth rotates under such a pattern once a night. **(b)** The concept of auroral substorms is schematically shown. **(c)** The first image of a full development of an auroral substorm (Frank et al., 1982).

Dynamic Explorer satellite. Figure 5c shows the first satellite images which confirmed the concept of auroral substorm.

The concept of auroral substorm provided the time frame of reference on a study of other polar upper atmospheric phenomena, such as magnetic and ionospheric changes, VLF emissions and X-ray precipitations. Thus, the concept of polar substorms was established. Then, it became soon apparent that these phenomena are manifestations of various changes in the magnetosphere or the interaction between the ionosphere and the magnetosphere (the magnetosphere–ionosphere coupling), and thus the concept of magnetospheric substorm was developed (cf. Akasofu, 1968). It has been found that there are a number of phenomena observed by satellites, which can also be interpreted in terms of the substorm concept.

6.3 Summary

The transition from the concept of the fixed pattern to the concept of the auroral substorm was based on data from the IGY all-sky camera network. The concept of auroral substorms has become a guiding principle in studying the dynamics of the auroral and magnetospheric substorms.

7 The importance of the interplanetary magnetic field (IMF) in the development of geomagnetic storms: the need for some unknown parameter other than the solar wind pressure (1905–1966)

It was Maunder (1905) who found that the sun is responsible for geomagnetic disturbances. It was naturally thought that the intensity of the solar plasma flow determined the inten-

sity of geomagnetic disturbances. Theoretical studies of geomagnetic storms until about 1963 were based mostly on the concept of the average geomagnetic storm or a typical geomagnetic storm with the three phases, the storm sudden commencement (ssc), initial phase and the main phase. However, Akasofu and Chapman (1963a) found that the intensity of the main phase of geomagnetic storms does not much depend on the magnitude of ssc's, (which is proportional to the solar wind pressure). Thus, they proposed that some unknown parameter determines the intensity of the main phase and the ring current. The unknown parameter was found to be the interplanetary magnetic field (IMF) which was discovered by Fairfield and Cahill (1966).

7.1   The unknown parameter of the solar wind

Hoyle (1949), Alfvén (1950) and later Dungey (1961) suggested the importance of the interplanetary magnetic field (IMF) in understanding geomagnetic storms. However, they were considering mainly an electrical discharge process (acceleration of auroral particles) at the neutral point or line which is formed by the interaction between the IMF and the geomagnetic field, not specifically in terms of geomagnetic storms.

Facing the theoretical difficulty that there is no simple way for the solar wind energy to enter the magnetosphere, Akasofu and Chapman (1963) examined individual magnetic storms, rather than the average storm or a typical storm (Fig. 1b), which consists of ssc, the initial phase and the main phase. Figure 6a shows the variety of the ways by which geomagnetic storms develop. In some cases, intense plasma flows do not produce storms, and they are called sudden impulse (si) (see (1) in Fig. 6a); further, intense storms can develop even without ssc's and they are called gradually commencing storms (sg), exceptional geomagnetic storms (see (7) in Fig. 6a); these storms were not considered as storms on the basis of the average storm concept (because they are not typical) and thus were excluded in studying geomagnetic storms. Between those two storms, there are a great variety of ways by which geomagnetic storms develop.

This case is a good lesson to be learned; the concept of intense plasma flows causing geomagnetic storms (a paradigm) can exclude any anomalous or unusual cases, such as si and sg storms, which can provide crucial clues to the solution of the problem as described below.

Akasofu and Chapman (1963) suggested that there must be some unknown parameter in the solar wind that determines the intensity of geomagnetic storms. In fact, they realized, after all, that the Chapman–Ferraro theory indicates that solar plasma simply flows around the magnetosphere and thus solar particles cannot get into the magnetosphere, except perhaps from the foci of the Chapman–Ferraro current on the magnetopause. However, many theorists were puzzled by the unknown parameter, because they could not consider anything other than an intense plasma flow. Fortunately, Fairfield

and Cahill (1966) found such a parameter, stating: "It is suggested here that this parameter is the southward component field."

7.2   The solar-wind–magnetosphere dynamo

Perrault and Akasofu (1978) and Akasofu (1981) made a quantitative determination of the control of the southward component of the IMF on geomagnetic storms in terms of the solar-wind–magnetosphere dynamo, because the kinetic energy of the solar wind is converted into an electrical power for auroral substorms. The power of the dynamo is given by

$$\varepsilon = VB^2 \sin^4(\theta/2) l^2,$$

where V, B and θ denote the solar wind speed, the IMF magnitude and the polar angle, respectively, and l a constant. Figure 6b shows the relationship between the power ε and the total dissipation rate UT of the magnetosphere.

7.3   Summary

The identification of the southward component of the interplanetary magnetic field and its importance was the most crucial element in the development of studies of geomagnetic storms and auroral substorms. This relationship was quantified as the power ε of the solar-wind–magnetosphere dynamo.

8   Ring current particles, solar wind protons to singly ionized oxygen ions from the ionosphere (1933–1997): the need to find the particles of the observed ring current

The present understanding of the ring current has had a very interesting course of development, which began with a paper by Chapman and Ferraro (1933). Assuming that solar plasma can get into the magnetosphere from its boundary, they assumed that those particles would have a circular motion around the earth. The discovery of the Van Allen radiation belt made it clear that ring current particles are trapped particles as computed by Stormer. However, until 1972, it was thought that the ring current particles were of solar wind origin. The discovery of singly ionized oxygen ions in the ring current by Shelley et al. (1972) made it clear that the major part of the ring current particles during intense geomagnetic storms are of ionospheric origin. This is an example of paradigm transition which resulted from an unexpected discovery.

8.1   Chapman's ring current (1933–1958)

After the success of theorizing ssc's, Chapman and many theorists were trying to explain the main phase of geomagnetic storms. The Chapman–Ferraro theory showed that the ring current must be located within the comet-shaped cavity
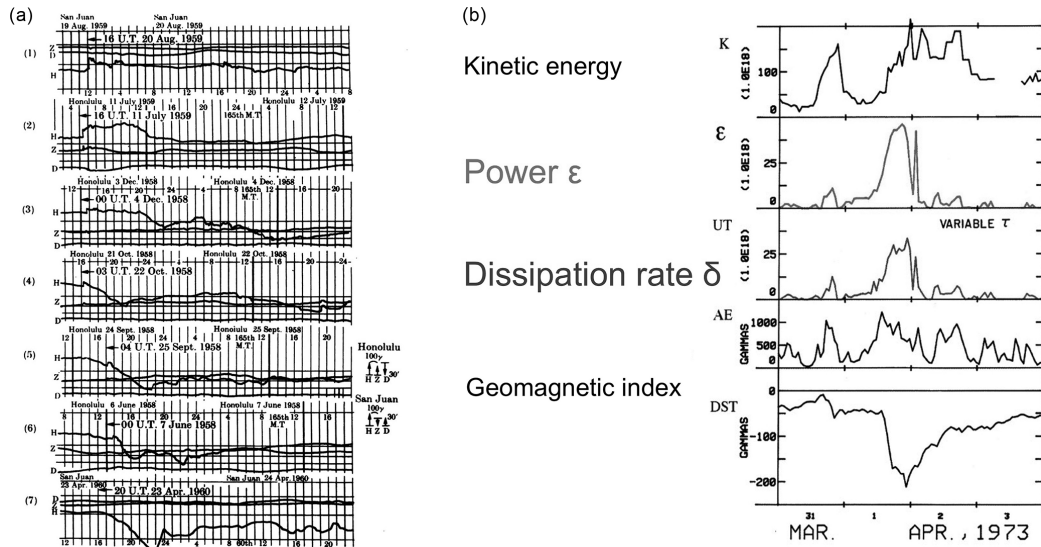
Figure 6. (a) Variety of the development of geomagnetic storms. **(b)** The relationship between ε and the total dissipation rate of the magnetosphere is shown. It can be seen that there is no clear relationship between the kinetic energy of the solar wind and UT, because the IMF polar angle θ plays a crucial role. UT is determined on the basis of the two geomagnetic indices AE and D_{st}.
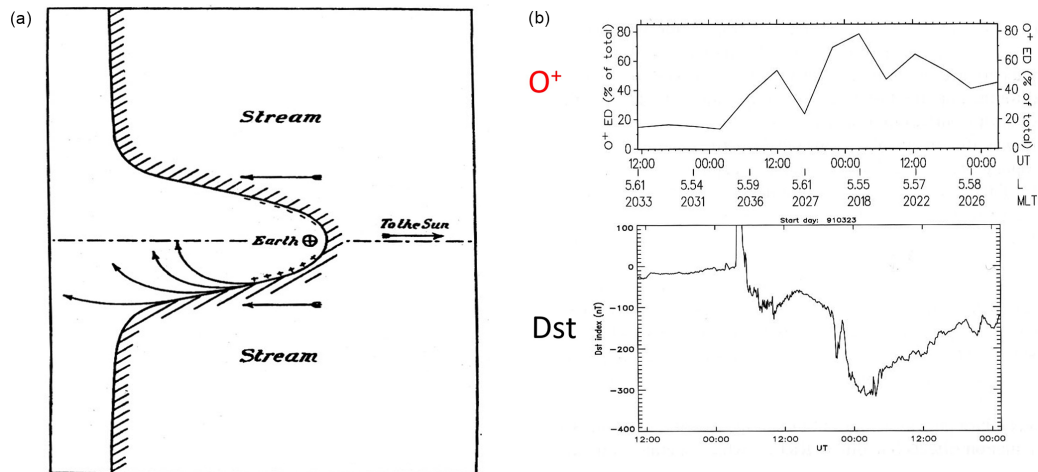

Figure 7. (a) Chapman considered that there occurs a charge separation on the magnetopause and the released particles circulate around the earth, forming the ring current (Chapman and Bartels, 1940: p. 868, Fig. 5). **(b)** Time variations of the flux of singly ionized oxygen ions and the D_{st} storm index (Daglis, 1997).

(Fig. 7a). Chapman and Ferraro (1941) considered the stability of a group of the particles circulating around the earth after released from the magnetopause, so that their ring current became the standard model.

After the discovery of the Van Allen radiation belts in 1958, it became clear that the ring current particles are particles that are trapped in the magnetosphere and execute the guiding center motions originally computed by Stormer. For trapped particles of an isotropic pitch-angle distribution, the current is basically the diamagnetic current that flows westward in the outer part of the belt and eastward in the inner

part (Akasofu et al., 1961). It was also found that protons, instead of electrons, are mainly responsible for the current.

In spite of such an increased understanding of the nature of the ring current, the serious problem of how the solar wind protons can get into the magnetopause remained until 1997. Thus, until then, the problem was how solar wind protons can get into the magnetosphere or how the kinetic energy of the solar wind can be transferred into the magnetosphere. Dessler et al. (1960) thought that hydromagnetic waves are generated at the boundary of the magnetosphere and heat plasma inside the magnetosphere. The struggle experienced by researchers in those days may be represented by a state-

ment by Parker (1962): the thermal energy for a $10^2\gamma$ [$=nT$] main phase is thus equivalent to the accumulation of this enhanced flux for less than 10 s. The enhanced solar wind, as indicated by the active phase of the storm, has a duration of several hours. Thus, only a tiny fraction of the energy incident on the outer boundary is accumulated as thermal energy deep in the geomagnetic field.

A decade later, McIlwain (1972, 1974) showed, based on observations of particles by geosynchronous satellites, that protons and ions in the plasma sheet are convected earthward from the night side into the ring current region from the magnetotail during substorms. The convection was suggested earlier by Axford and Hines (1961). Protons drift westward after arriving there, becoming ring current particles. This was a major change in considering the source of ring current protons, and this concept is considered to be valid today. The convection accelerates protons before being injected into the ring current belt.

8.2   The discovery of oxygen ions from the ionosphere

An unexpected discovery was made by Shelley et al. (1972) who found outward flows of singly charged ions (such as He^+ and O^+) from the ionosphere to the inner magnetosphere during a geomagnetic storm. They noted: they may also contribute significantly to the storm-time magnetic field depression (D_{st}), since at the same time they represent an energy density greater by a factor 4 than that of the protons. Their finding was further studied by Shelley (1979) and Geiss et al. (1978).

Subsequently, Daglis (1997) showed that singly ionized oxygen ions are dominant ring current particles during an intense main phase. Figure 7b shows that the flux of O^+ varies with the D_{st}; see also Daglis (2000).

8.3   Summary

It became clear that singly ionized oxygen ions from the ionosphere play the major component of the ring current particles during intense geomagnetic storms. See the next section for further discussions.

9   The storm–substorm controversy (1963–present): the need to relate the two obviously concurrent phenomena

Until about the early 1960s, the main phase decrease (the ring current formation) and polar magnetic substorms were thought to be independent (Chapman and Bartels, 1940: chapter X), although they occur at the same time. Akasofu and Chapman (1963) found that an intense main phase occurs when intense polar magnetic substorms occur frequently, suggesting that substorms are a cause of the main phase decrease. However, there has been some doubt about such a
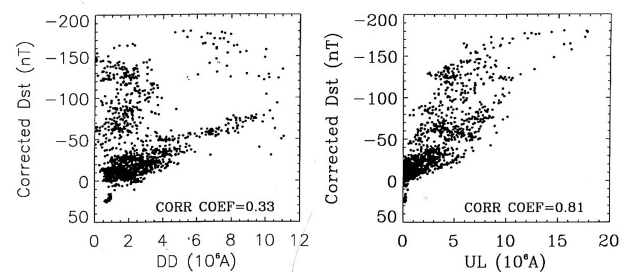

(a)
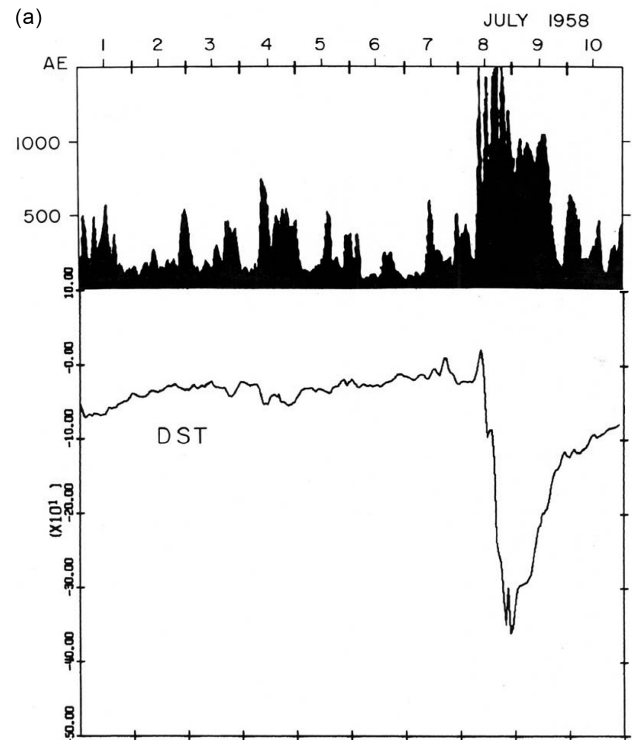
(b)   Storm-substorm controversy

Figure 8. (a) The figure shows the simultaneous occurrence of intense substorms and an intense ring current. **(b)** The relationship between the DD current and the D_{st} index and also between the UL current and the D_{st} index.

close cause–effect relationship (cf. Sharma et al, 2003) and this issue has not been solved yet.

There is little doubt about the fact that an intense geomagnetic storm coincides with the period when intense substorms occur frequently; see Fig. 8a. A careful examination of Fig. 8a shows that before the major storm weak substorms cause a weak D_{st} decrease, namely, a weaker main phase decrease, therefore, substorms contribute to the formation of the ring current or basically *a substorm is a mini-storm*, perhaps associated with the protons convected from the night side, as discussed in the previous section (McIlwain, 1972, 1974). As discussed in Sect. 7, it is likely that O^+ ions are

injected from the ionosphere into the ring current belt during intense substorms. There are no major geomagnetic storms without intense substorms and vice versa.

Actually, it is quite obvious that both substorms and the magnetospheric convection contribute to the formation of the ring current. In fact, the convection is a part of substorm processes too. This can be seen in Fig. 8b, in which the D_{st} index depends on both the two components of substorms, DD and UL; note that the dependence of D_{st} is more clearly shown by UL (the expansion phase) than by DD (the convection), indicating that the expansion phase indeed plays an important role in building the ring current belt.

It may be that a weak ring current is caused by protons which can be convected into the ring current belt from the magnetotail during substorms, even before expansion onset of substorms (the growth phase) after the IMF turns southward and the convection begins (but before expansion onset), while a stronger ring current is caused by O^+ ions, which become the dominant component, after being injected from the ionosphere during intense substorms. The ratio of O^+ / H^+ increases for greater geomagnetic storms (D_{st}).

9.1 Summary

It is quite obvious that both the convection and substorm processes contribute to the formation of the ring current. Substorms are actually mini-storms.

10 Theories of substorm onset (1964–present): the need for a new process other than magnetic re-connection

The substorm onset has been the most controversial subject in magnetospheric physics. The dominant paradigm is that magnetic re-connection powers the expansion phase.

The concept of magnetic re-connection has been the greatest paradigm in solar physics and magnetospheric physics, lasting for almost a half century as the only conversion process of magnetic energy for solar flares and auroral substorms. It presumes that an anti-parallel magnetic configuration annihilates itself, producing explosively a large amount of energy for solar flares and auroral substorms. Thus, practically all observations associated with the conversion of magnetic energy have been interpreted in terms of magnetic re-connection.

10.1 Magnetic re-connection

The concept of magnetic re-connection as a conversion process of the magnetic energy was originally conceived by Sweet (1958) who considered a collision of two sunspot pairs, producing an anti-parallel magnetic field configuration. Since then, most solar physicists have been trying to explain solar flares in terms of magnetic re-connection. However, after carefully examining Sweet's theory, Parker (1963)

found that Sweet's mechanism requires that the merging field should be exactly anti-parallel and that none of the known mechanisms (including resistivity and diffusion) are sufficiently rapid enough to account for solar flares as a result of the annihilation of magnetic fields. Thus, he concluded: "The observational and theoretical difficulties with the hypothesis of magnetic field annihilation suggest that other alternatives for the flare must be explored."

Soon afterward, Petschek (1964) proposed that Parker's difficulty can be removed by having the X-line in the anti-parallel configuration, where the merging could take place. It is assumed that around the X-line, the electrical resistivity is different from what is given by the standard formula, namely, equivalent resistivity. Since then, solar physicists consider that the only way to convert magnetic energy is to have a magnetic configuration that includes anti-parallel magnetic field lines with the X-line (or lines). For this reason, it appears that many researchers – theorists and observers alike – are bound to prove only this particular premise of magnetic re-connection. Most simulation studies assume equivalent resistivity in a collisionless plasma, but the physics involved has not been clarified. It appears that, as the term annihilation implies, anti-parallel magnetic field lines in a thin current sheet are considered as if they are almost the case of collision of the matter and anti-matter. Alfvén (1977, 1986) was very concerned about validity of the paradigm of magnetic re-connection.

The year 1964 was a very interesting year. In the same year that Petschek (1964) published his paper, Ness et al. (1964) discovered the magnetotail where the magnetic field lines are nearly anti-parallel. As mentioned in Sect. 6, the concept of auroral substorms was published also in the same year (Akasofu, 1964). Since there are some phenomenological similarities between solar flares and auroral substorms, together with a nearly anti-parallel magnetic field configuration in the magnetotail, many magnetospheric physicists have considered that both are caused by magnetic re-connection by which magnetic energy is supposed to be released rapidly; see Fig. 9a.

For the above reasons, it has been claimed that magnetic re-connection is one of the most important processes that may occur in many parts of the universe. Thus, the theory of magnetic re-connection has become the single guiding principle in the study of solar flares and auroral substorms for about a half century. Vasyliunas (1975) stated: "The process variously known as magnetic merging, magnetic field annihilation or magnetic field re-connection (or re-connection) plays a crucial role in determining the most plausible, if not the only, way of tapping the energy stored in the magnetic field in order to produce large dissipative events, such as solar flares and magnetospheric substorms."

10.2   Auroral substorms

In considering auroral substorms, it is important to recognize that auroral substorms are various manifestations of an electrical discharge process. Therefore, first of all, an electric power supply is needed (not just southward turing of the IMF Bz), together with currents/circuits to transmit the power to the ionosphere to dissipate the power. The primary dynamo results from the interaction between the solar wind and the magnetosphere (the S–M dynamo). The power of this dynamo is given by ε (Sect. 7.2). A typical power during substorms is about 5×10^{18} erg s^{-1}. This dynamo produces a voltage V of about 100 kV across the magnetotail, resulting in a large-scale $\boldsymbol{E} \times \boldsymbol{B}$ convection of magnetospheric plasma (Sect. 4), which drives a two-cell (the directly driven, denoted by [DD]) current system in the ionosphere during the whole period of auroral substorms (when ε is greater than 10^{18} erg s^{-1}); see Sect. 7.2 and Fig. 6b.

However, during the first hour or so after the power ε rises above 10^{18} erg s^{-1} (namely, before the expansion phase onset, when the ionization of the ionosphere is low), the ionosphere is not conductive enough to dissipate the incoming power; as a result, the DD current is very weak (little dissipation). Thus, it is during this period when the energy produced by increasing power ε is accumulated as magnetic energy in the inductive circuit of the magnetosphere. This period is called the growth phase; the total magnetic energy W accumulated is about 2.0×10^{22} erg s (about 5×10^{18} erg s$^{-1} \times 1$ h [note $W = (1/2)J^2 L$; $J = 5 \times 10^6$ A, $L = 1.6 \times 10^2$ H; $\varepsilon = VJ = 100$ kV $\times 5 \times 10^6$ A]). This is the amount of energy needed for the expansion phase.

During the expansion phase, a single-cell current system develops impulsively; this current is called the UL (unloading) current; see Fig. 4f. The UL current is the ionospheric part of the substorm current system (azimuthal component; see Fig. 4d) that is responsible for the expansion phase. Figure 9b shows the relationship between the power ε, the DD current and the UL current. It is important to note that the DD and UL currents are independent, indicating that substorms are not due to an enhancement of the DD current and the plasma flow produced by magnetic re-connection.

Further, it is not difficult to show that the magnetotail does not have enough magnetic energy for substorms. Even the total magnetic energy between 10 and 20 earth radii is less than what is needed for the expansion phase (Akasofu, 2013). Although satellite observations show that magnetic re-connection occurs within a distance of 20 earth radii (cf. Angelopoulos et al., 2008), recent satellite observations indicated that the maximum flow energy from the magnetotail is about 2 orders of magnitude less than what is needed.

Thus, it is likely that the energy of the expansion phase must be stored but primarily in the main body of the magnetosphere, not in the magnetotail. Since the magnetic field configuration is not anti-parallel there, and magnetic re-connection does not occur (Ge and Russell, 2006), a new process of converting the stored magnetic energy, other than magnetic re-connection, is needed.

Therefore, when the magnetic energy accumulated during the growth phase is released, it must generate a dynamo in the main body of the magnetosphere in order to supply electric power for the auroral discharge. There are several requirements for the expansion phase dynamo. During a brief period of the expansion phase (about 1 h or less), the UL current system develops (Bostrom, 1964); see Fig. 4d. Unlike the DD current produced by the $\boldsymbol{E} \times \boldsymbol{B}$ convection (Fig. 4f), this current system in the ionosphere is a single-cell current, so that it is not an enhancement of the DD current (Fig. 9b). The UL current system is driven by an earthward electric field.

For these several reasons, it is uncertain whether or not the paradigm of magnetic re-connection is the main process in converting magnetic energy for auroral substorms. Since the UL current system is responsible for the expansion phase, regardless of onset theories, our past efforts of finding the onset process come down to the question stated in the following Sect. title.

10.2.1   How can the accumulated magnetic energy be converted into the dynamo power for the earthward electric field?

One possibility is that a charge separation takes place in the main body of the magnetosphere, when a deflation process occurs at substorm onset (the accumulated magnetic energy is spent) after the magnetosphere is inflated during the growth phase (Lui and Kamide, 2003; Akasofu, 2007: p. 117). This requires breakdown of the frozen-in field condition that was observed (Lui, 2011). Figure 9c shows schematically this situation.

10.3   Summary

The paradigm of magnetic re-connection which is supposed to generate the auroral substorm expansion phase has a number of serious difficulties that are described in detail here in order for the readers to examine and judge its adequateness as the theory of the auroral substorm expansion phase.

1. The magnetotail does not have enough magnetic energy for auroral substorms.

2. The predicted earthward flow cannot generate the UL current system for the expansion phase.

3. The physics of the so-called equivalent resistivity is not clear.

4. The concept of the frozen-in field lines condition breaks down, so that the MHD (Magnetohydrodynamics) theory is not applicable in this process.
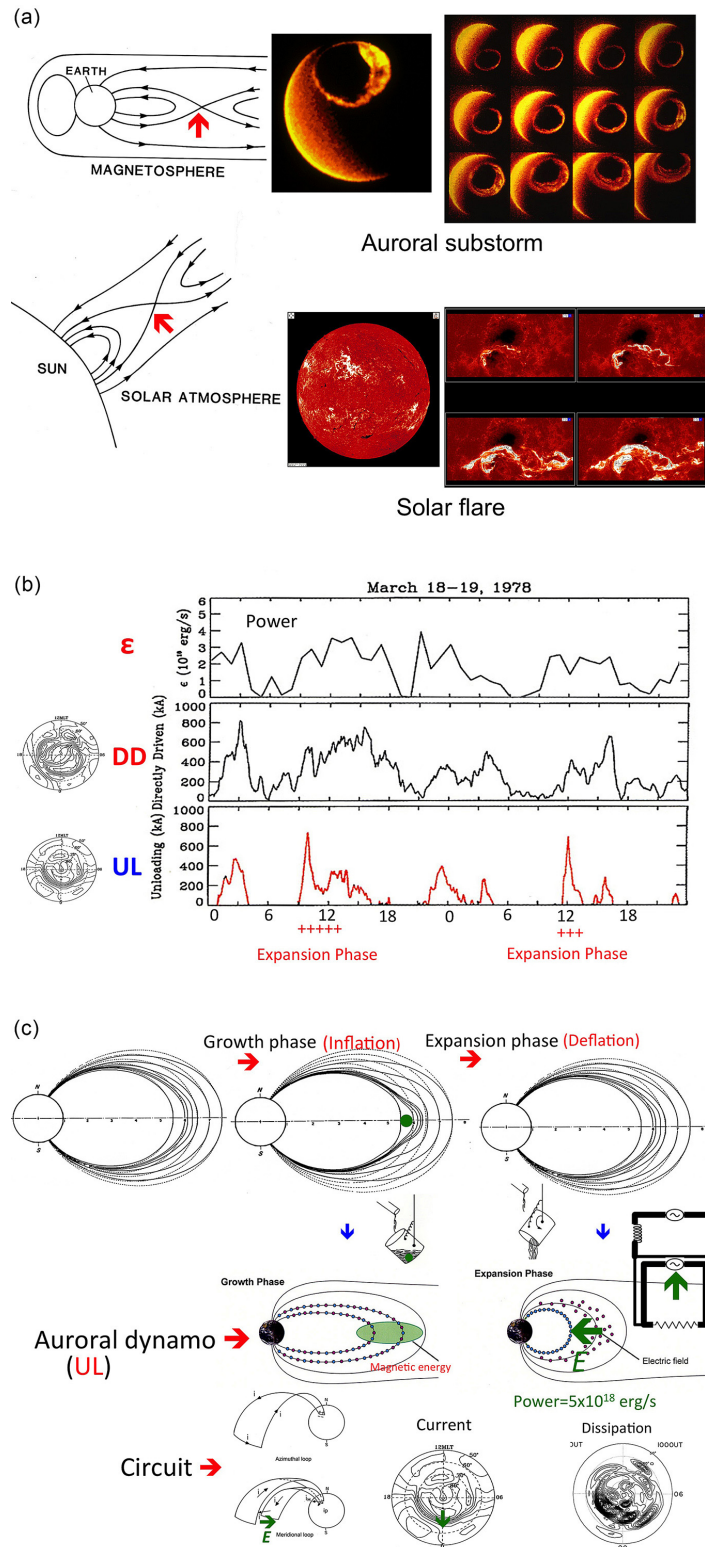
Figure 9. (a) Images of auroral substorms and solar flares, as well as their time sequences. Both the magnetotail and the cusp configuration are also schematically shown. **(b)** shows from the top, the solar-wind–magnetosphere dynamo power $\varepsilon(t)$, the directly driven (DD) component of the ionospheric (equivalent) current and the unloading component (UL). The figure shows that there occurs no prominent current other than DD and UL (or dissipation), indicating that magnetic re-connection does not produce any prominent extra energy (or dissipation). **(c)** The figure illustrates how the charge separation might occur when the inflated magnetosphere is deflated; electrons are tied to contracting field lines, while protons remain, because of the breakdown of the frozen-in field condition.

11   Solar flares: the need for examining the magnetic re-connection theory

The early history of the concept of magnetic re-connection was given in the previous section.

In considering solar flares in terms of magnetic re-connection, it is important to recognize that there are solar flares without sunspots. They are called spotless flares and occur more often than solar flares associated with sunspot pairs (Dodson and Hedman, 1970). However, they have received little attention perhaps because they are not associated with the spectacular phenomena accompanying flares from sunspot pairs. However, they may be the most fundamental flare process, because flares are associated with the two-ribbon emission in the way solar flares are defined (because other phenomena do not always occur simultaneously and because the process of two-ribbon flares should be the same for spotless flares and sunspot associated flares).

Among many flare models, one of the most widely adopted models was originally proposed by Hirayama (1974), in which magnetic re-connection is the energy supply process that is supposed to occur above a magnetic arcade–cusp system that is located over the neutral line, which is the boundary of two unipolar magnetic regions of opposite polarity. Figure 10a shows his model at the maximum phase of flares. The X-line location at an altitude of 44 000 km or above has been identified as "Energy Release (X-type Reconnection)" by Aschwanden (2005; see Fig. 13.35).

Coronal observations are crucial in examining this magnetic re-connection model of solar flares, because the needed magnetic energy is assumed to be in the corona. There have been a few observations which have not received much attention in this regard.

1. Sheeley et al. (1976: p. 178) observed: "– these fields usually interact by changing their flux linkage, much as they do in a vacuum".

2. Hudson and Khan (1996) stated: "The observational situation is clear: there is almost no evidence for inward flows" which are supposed to occur around the expected X-line.

3. There has been no measurement of the magnetic field intensity in the corona around the X-line.

4. Raymond et al. (2012) showed that there is not enough electron density around the X-line.

Since the intensity of the magnetic field B above the cusp has not been measured and since Sheeley's observation suggests that the coronal magnetic distribution is approximated by a potential field in the corona at an altitude of 50 000 km (a force-free magnetic field tends to have a helical structure and sheet currents must close by themselves or connected to other currents). Let us consider the magnetic arcade–cusp

system, in which a 2-D dipole-like field ($B = 12$ G) is lying horizontally on the photosphere as a first approximation, and the cusp height is twice the distance between the feet of the magnetic arcade. Therefore, the magnetic field above the cusp will be $B = (6 \, \text{G}/2^3) \sim 1$ G. Since we do not expected any strong field (current) other than that shown in Fig. 10a, it is likely that the field around the X-line is expected to be about 1 G; note that the magnetic field there is supposed to be approximately a force-free field which tends to have a helical structure. Thus, the volume in the solar corona needed for solar flares with energy of 10^{30}–10^{32} erg s corresponds to a sphere of about 0.4–1.2 of the solar radius, so that it is quite likely that the magnetic energy above the top of the cusp is too small to be useful for solar flares.

There are also important observations which have not received much attention:

1. Sheeley et al. (1975: p. L13) also observed: "reconnection occurs much more often than flares, thus usually occurs without them".

2. Similarly, summarizing the Yokho observations, Tuneta (1996) noted: "There are ubiquitous neutral sheet structures with scale size – 10^3 to 10^5 km almost everywhere on the sun. It appears that most of these neutral structures are static or dormant, do not show explosive reconnection. Only the dynamically formed neutral structure – appears to have fast reconnection." Thus, their identification of the simultaneous flares could be accidental. In fact, only dynamically formed flares may be caused by photospheric motion, as discussed below in terms of a photospheric dynamo.

3. Obviously, an anti-parallel magnetic configuration does not release automatically magnetic energy by magnetic re-connection, particularly because the observation by Hudson and Kahn (1996) showed no sign of the inflow toward the X-line. Thus, their identification of the simultaneous occurrence of flares might be accidental. In fact, dynamically formed feature may be caused by motions in the photosphere, as discussed below.

4. In studying complex natural phenomena, I believe that it is important to try to understand them first in the simplest form which have often receive little attention. Spotless fares and single spots are both good examples.

Since the solar corona is unlikely to provide energy for solar flares, a photospheric dynamo should be considered as a possible alternative. Figure 10b shows the geometry associated with a dynamo developed by Choe and Lee (1996a, b). Their model assumes an anti-parallel plasma flow ($V = 2 \, \text{km s}^{-1}$) along the center line of the arcade, which is the boundary of two unipolar regions of opposite polarity of the magnetic field intensity $B = 12$ G ($= 6 \, \text{G} + 6 \, \text{G}$).

An additional calculation, based on their model, shows here that the field-aligned currents flow along the arcade
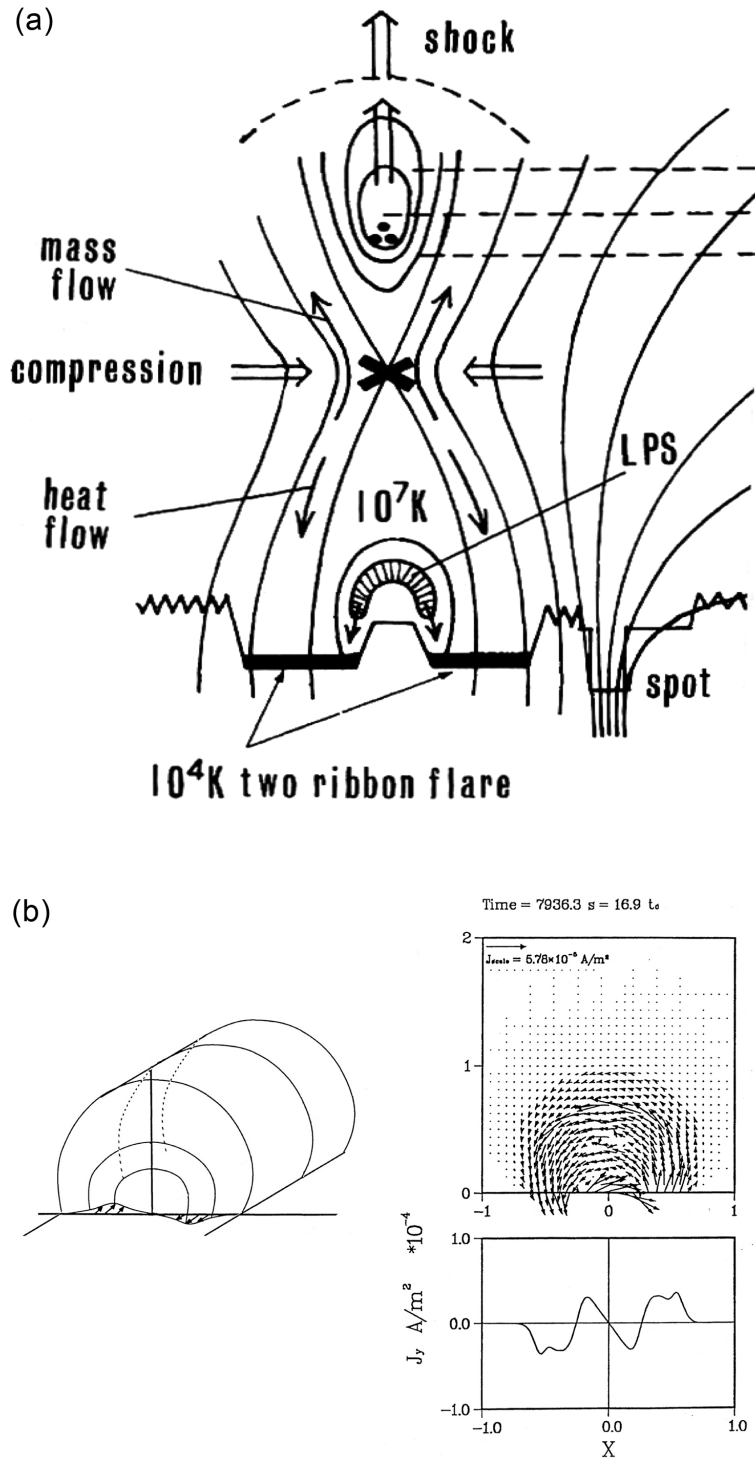
Figure 10. (a) A magnetic arcade–cusp model of solar flares (Hirayama, 1974). **(b)** The photospheric dynamo model developed by Choe and Lee (1996a, b).

magnetic field lines, which are generated by this dynamo process. It is about 10^{-4} A m^{-2}, corresponding to the electron flux of 10^{14} m^{-2} s^{-1} (a typical auroral electron flux is 10^{11} m^{-2} s^{-1}); the total current for a typical geometry of two-ribbon flares is estimated to be about 10^{11} A. Thus, the energy carried by the electrons can be about 10^{25} erg s^{-1}, as they must be accelerated in order to penetrate into the chromosphere. Large Hα flares are associated with the energy of
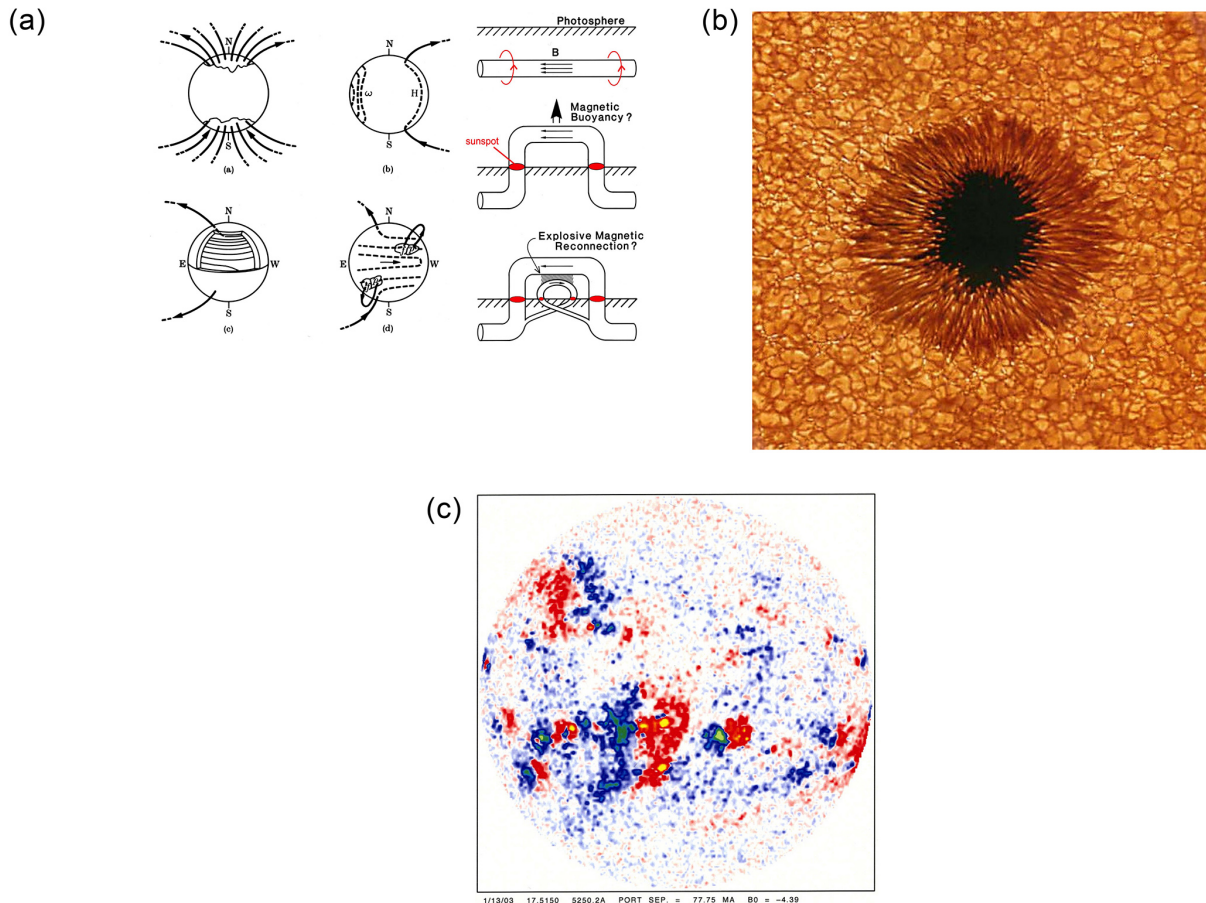
Figure 11. (a) Babcock's model of the sunspot pair formation. **(b)** An example of single spot (courtesy of the Big Bear Observatory). **(c)** The figure shows that single spots (of positive polarity) are formed in the unipolar magnetic regions (of positive polarity), while sunspot pairs are formed at the boundary of two unipolar magnetic region of opposite polarity (courtesy of Mt. Wilson Observatory).

about 10^{26} erg s^{-1} (Svestka, 1976: p. 13). Thus, the photospheric dynamo proposed by Choe and Lee can explain the two-ribbon emission of spotless flares, as well as complex motions of prominence as they demonstrated.

The presence of spotless flares indicates that two-ribbon flares can occur without sunspots (including emerging sunspots or sunspot groups). Since the process of causing the two-ribbon emission of spotless flares is not expected to be different from those that occur in a complex sunspot group, the physics involved in spotless flares seems to have a fundamental importance in all flares. In view of this, spotless flares are basic, and a large range of the energy of flares (10^{30}–10^{32} erg s) depends mainly on V and B through the dynamo process (VB^2) along the neutral lines, instead of B and its volume (in the case of magnetic re-connection).

11.1 Summary

The paradigm of magnetic re-connection has been prevailing in solar flare studies. However, the solar corona does not seem to have enough energy for solar flares, since the coro-

nal field could be approximated by a potential field. Thus, it is suggested that the simplest solar flares are spotless flares and that a moderate photospheric dynamo can provide the necessary power and the field-aligned current. This section is provided for the readers to examine and judge the presented materials for future studies of solar flares.

12 Sunspots (1961–present): the need for a new idea of the formation of sunspots

This particular paradigm began when Babcock (1961) succeeded in scanning the solar disk by a magnetometer, and his theory of the formation of sunspot pairs is still the guiding principle for explaining many active solar phenomena. However, there are a number of problems on the basic morphological phenomena of sunspots, which are difficult to explain with Babcock's theory.

Many generations of students have been taught that the non-uniform rotation of the sun's winds up the dipolar field lines and produces a tube of azimuthal magnetic field flux

beneath the photosphere. As magnetic flux in the tube increases above some critical level by squeezing plasma inside ($B^2 / 8\pi > p$), the tube breaks through the photosphere by magnetic buoyancy. Two cross-sections of the tube on the photosphere are said to be identified as a sunspot pair (Babcock, 1961); see Fig. 10a. It appears that this concept is well accepted even today in describing sunspots.

However, McIntosh (1981: p. 43) found: "– sunspots form preferentially near the borders of large-scale structures on, or near, neutral lines with polarity arrangement appropriate for the Hale law for their hemisphere". This particular aspect is difficult to explain if the rise of magnetic flux tubes occurs randomly beneath the photosphere. There are other well-known morphological features which are hard to explain in terms of Babcock's theory. For example, p spots tend to appear 5 days earlier than f spots (Kiepenhauer, 1953: pp. 375 and 434).

Further, there are isolated spots or single spots. This is the simplest type of sunspots and is hardest to explain in terms of Babcock's model. Figure 11b shows such an example. There is so far no definitive explanation of single spots.

It is well established that there are unipolar magnetic regions which are weak, but widely distributed. If one carefully examined single spots, positive single spots would appear in a positive unipolar magnetic region and negative in a negative region (Akasofu, 2014a). Thus, McIntosh's finding seems to suggest that a sunspot pair is a special case in which positive and negative single sunspots form at the boundary of positive and negative uni-polar magnetic regions, respectively. Figure 11b shows those points.

Therefore, it seems that the formation of a single sunspot is more basic than a pair of spots. Further, the unipolar magnetic regions are also fundamental in understanding solar magnetism, rather than considering them as remains of old sunspot pairs (Akasofu, 2014b). In fact, McIntosh (1981: p. 45) noted: "the [unipolar magnetic field] patterns could not have formed from the scattered remains of sunspot magnetic fields".

12.1 Summary

There are a number of features of single spots and spot pairs that have not been addressed in recent years in association with unipolar regions. Since single spots are basic in understanding solar activities, their origin should be urgently studied.

13 Concluding remarks

A chronological description of the history of scientific achievement is obviously useful and important, but it becomes more so if it describes why and how old paradigms had to be changed. This article may be considered as a historical review of the development of several subjects in solar–terrestrial physics, specially paying attention to the transition of paradigms. It is interesting to find that many transitions take a long time, at least two generations or more. During such periods, researchers believe that a particular theory they choose is the right one (and nothing else) and elaborate further on it, either observationally or theoretically. There is nothing wrong in this, but progress could be accelerated if everyone was aware of the possibility of a paradigm change. A scientific field makes progresses in two ways, by elaborating the presently dominant paradigms and by occasional transitions of paradigms. Needless to say, both are important.

Scientists are eager to contribute to an epoch-making advance in their own field. However, in attempting to do so, their efforts will encounter many issues to be learned from the history of their field. This is partly because an advanced or older paradigm tends to see new ideas heretical, and any observations that do not seem to support or agree with the theory are rejected as trivial exceptions or simply ignored. Thus, a great effort is needed to overcome such difficulties. For these reasons, it is useful to learn how to deal with such difficulties from the history. It is also useful to learn how the particular theory they chose had been established, as well as recognizing difficulties and anomalies that the present theory faces. One possible way in making an important advance, I experienced, is to pursue trivial facts which do not agree with a prevailing paradigm and thus received little attention; in fact, the simplest trivial facts could be of fundamental importance. It may be quite likely that in the rapidly advancing field of solar–terrestrial physics, some of the presently accepted theories and facts may change in the future.

There are many other fascinating subjects in solar–terrestrial physics that may be described in terms of the transition of paradigms, such as the acceleration of auroral particles, the magnetosphere–ionosphere coupling, solar wind theories and coronal mass ejections (CMEs), but they are beyond my experience. Thus, the collection of work in this article is limited to those subjects within my own experience and ability.

References

Akasofu, S.-I.: The development of auroral substorm, Planet. Space Sci., 12, 273–282, 1964.

Akasofu, S.-I.: Polar and Magnetospheric Substorms, D. Reidel Pub. Co., Dordrecht, The Netherlands, 288 pp., 1968.

Akasofu, S.-I.: Energy coupling between the solar wind and the magnetosphere, Space Sci. Rev., 28, 121–190, 1981.

Akasofu, S.-I.: Exploring the Secrets of the Aurora, Springer, 288 pp., 2007.

Akasofu, S.-I.: A historical review of the geomagnetic storm-producing plasma flows from the sun, Space Sci. Rev., 164, 85–132, 2011.

Akasofu, S.-I.: Where is the magnetic energy for the expansion phase of auroral substorms accumulated?, J. Geophys. Res., 118, 7219–7225, doi:10.1029/2013JA019042, 2013.

Akasofu, S.-I.: Single spots, unipolar magnetic regions, and pairs of spots, Geophys. Res. Lett., 41, 3698–3700, doi:10.1002/2014GL060319, 2014a.

Akasofu, S.-I.: Single spots, unipolar magnetic regions, and pairs of spots: II. The development of sunspot pairs and the Hale boundary, Geophys. Res. Lett., 42, doi:10.1002/2014GL062887, 2014b.

Akasofu, S.-I. and Chapman, S.: The simultaneous development of the main phase (DR) and of polar magnetic substorms (DP), J. Geophys. Res., 68, 3155–3158, 1963.

Akasofu, S.-I., Cain, J. C., and Chapman, S.: The magnetic field of a model radiation belt, numerically computed, J. Geophys. Res., 66, 4013–4026, 1961.

Akasofu, S.-I., Chapman, S., and Meng, C.-I.: The polar electrojet, J. Atmosp. Terr. Phys., 27, 1301–1305, 1965.

Alfven, H.: Cosmical Electrodynamics, Oxford Univers. Press, 237 pp., 1950.

Alfven, H.: Electric currents in cosmic plasmas, Rev. Geophys. Space Phys., 15, 271–284, 1977.

Alfven, H.: Double layers and circuits in astrophysics, IEEE Trans. Plasma Phys., PS-14, No. 6, 1986.

Angelopoulos, V., McFadden, J. P., Larson, D., Carlson, C. W., Mende, S. B., Frey, H., Phan, T., Sibeck, D. G., Glassmeier, K.-H., Auster, U., Donovan, E., Mann, I. R., Rae, I. J., Russell, C. T., Runov, A., Zhou, X.-Z., and Kepko, L.: Tail reconnection triggering substorm onset, Science, 321, 931–933, 2008.

Armstrong, J. C. and Zumuda, A. J.: Triaxial magnetic measurements of field-aligned currents at 800 kilometers in the auroral region: Intial results, J. Geophys. Res., 78, 6802–6807, 1973.

Aschwanden, M. J.: Physics of the Solar Corona, Springer, 892 pp., 2005.

Axford, W. I. and Hines, C. O.: A unified theory of high latitude geophysical phenomna and geomagnetic storms, Canadan J. Phys., 39, 1433–1464, 1961.

Babcock, H. W.: The topology of the sun's magnetic field and the 22-year cycle, Astrophys. J., 133, 572–587, 1961.

Birkeland, K.: The Norwegian Auroral Polaris Expedition, 1902–1903, 801 pp., edited by: Aschehoug, H., Christiania, 1918.

Bostrom, R.: A model of the auroral electrojets, J. Geophys. Res., 64, 4983–4999, 1964.

Brekke, A., Doupnik, J. R., and Banks, P. M.: Incoherent scatter measurements of E region conductivities and currents in the auroral zone, J. Geophys. Res., 79, 3773–3790, 1974.

Cahill, L. J. and Amazeen, P. G.: The boundary of the geomagnetic field, J. Geophys. Res., 68, 1835–1843, 1963.

Chapman, S.: An outline of a theory of magnetic storms, Proc. R. Soc., 97, 61–83, 1918.

Chapman, S.: The absorption and dissciation or ionizing effect of monochromatic radiation in an atmosphere on a rorating earth, Proc. Phys. Soc., 43, 426–445, 1931.

Chapman, S.: The electric current systems of magnetic storms, Terr. Magn., 40, 349–370, 1935.

Chapman, S. and Bartels, J.: Geomagnetism, Oxford Univ. Press, 1049 pp., 1940.

Chapman, S. and Ferraro, V. C. A.: A new theory of magnetic storms, Terr. Magn., 36, 77–97, 1931.

Chapman, S. and Ferraro, V. C. A.: A new theory of magnetic storms, Part-II – The main phase, Terr. Magn., 38, 79–96, 1933.

Champan, S. and Ferraro, V. C. A.: The geomagnetic ring current: I—its radial stability, Terr. Magn., 46, 1–6, 1941.

Choe, G. S. and Lee, L. C.: Evolutionof solar magnetic arcades. I. Ideal MHD evolution under footpoint shearing, Asyrophys. J., 472, 360–371, 1996a.

Choe, G. S. and Lee, L. C.: Evolution of solar magnetic arcades. II. Effect of resistivity and solar eruptive processes, Astrophys. J., 472, 372–388, 1996b.

Clauer, C. R. and McPherron, P. L.: Mapping the local time development of magnetospheric substorms using the mid-latitude magnetic variations, J. Geophys. Res., 79, 2811–2820, 1974.

Daglis, I. A.: Missions study oxygen ions, EOS, 78, 245, 250–251, 1997.

Daglis, I. A., Kamide, Y., Mouiki, C., Reeves, G. D., Sarris, E. T., Shiokawa, K., and Wilken, B.: "Fine structure" of the storm-substorm relationship: Ion injections during Dst decrease, Adv. Space Res. 25, 2369–2372, 2000.

Dessler, A. J., Francis, W. E., and Parker, E. N.: Geomagnetic storms sudden commencement rise time, J. Geophys. Res., 65, 2715–2719, 1960.

Dodson, H. W. and Hedman, E. R.: Major Hα flares in centers of activity with very small or no spot, Solar Phys., 13, 401–419, 1970.

Dungey, J. W.: Interplanetary magnetic field and the auroral zone, Phys. Rev. Lett., 6, 47–48, 1961.

Fairfield, D. H. and Cahill Jr., L. J.: Transitio region magnetic field and polar magnetic disturbances, J. Geophys. Res., 71, 155–169, 1966.

Feldstein, Y. I.: Some problems concerning the morphology of auroras and magnetic disturbances at high latitudes, Geomag. Aeronom., 3, 183–192, 1963.

Feldstein, Y. I. and Starkov, G. V.: The auroral oval and the boundary of closed field lines of geomagnetic field, Planet. Space Sci., 18, 501–508, 1970.

Feldstein, Y. I., Vorobjev, V. G., Zverev, V. L., and Förster, M.: Investigations of the auroral luminosity distribution and the dynamics of discrete auroral forms in a historical retrospective, Hist. Geo Space. Sci., 5, 81–134, doi:10.5194/hgss-5-81-2014, 2014.

Frank, L. A., Craven, J. J., Burch, J. L., and Winningham, D. J.: Polar view of the earth's aurora wit Dynamic Explorer, Geophys. Res. Lett., 9, 1001–1004, 1982.

Fukushima, N.: Polar magnetic storms and geomagnetic bays, J. Fac. Sci. Tokyo Univ., 8, 293–421, 1953.

Fullerf, V. R.: A report of work on the aurora borealis for the year 1932–1935, Terr. Magn., 40, 269–275, 1935.

Fuller, V. R. and Bramhall, E. H.: Auroral research at the University of Alaska, 1930–1934, Univ. Alaska Press, 1937.

Ge, Y. S. and Russell, C. T.: Polar survey of magnetic field in near tail: Reconnection are inside 9Re, Geophys. Res. Lett., 13, L02101, doi:10.1029/2005GL024574, 2006.

Geiss, J., Balsiger, H., Eberhardt, P., Walker, H. P., Weber, L.
    Young, D. T., and Rosenbauer, H.: Dynamics of magnetospheric
    ion compositions as observed by the GEOS mass spectrometer,
    Space Sci. Rev., 22, 537–566, doi:10.1007/BF00223940, 1978.

Gerson, N. C., Keneshea, T. J., and Donaldson Jr., R. J. (Eds):
    Proceedings of the Conference on Auroral Physics, Geophys.
    Res. Pap. 30, AFCRC-TR-54-203, Dept. of Phy. Univ. of West.
    Ont., London and Geophys. Res. Dir., Air Force Cambridge Res.
    Cent., Cambridge, Mass., 1954.

Hirayama, T.: Theoretical model of flares and promi-
    nences,1:Evaporationg flare model, Solar Phy., 34, 323–338,
    1974.

Hoyle, F.: Some Recent Research in Solar Physics, Cambridge
    Univ. Press, 134 pp., 1949.

Heppner, J. P.: Time sequences and spatial relations in auroral ac-
    tivity during magnetic bays at College, Alaska, J. Geophys. Res.,
    59, 329–338, 1954.

Hudson, H. S. and Khan, J. I.: Observational problems for flare
    models based on large-scale magnetix reconnection, 135–144, in:
    Magnetic reconnection in the solar atmosphere, edited by: Bent-
    ley, R. D. and Mariska, J. T., ASP Conference series, Astronom
    Soc. Pacific, Vol. 111, 409–419, 1996.

Iijima, T. and Potemra, T. A.: Large-scale characteristics of field-
    aligned currents associated with substorms, J. Geophys. Res., 83,
    599–615, 1978.

Kiepenhauer, K. O.: Solar activity, in: The Sun, edited by: Kuiper,
    G. P., Univ. Chicago Press, 322–463, 1953.

Kuhn, T. S.: The Structure of Scientific Revolutions, 2nd edition,
    Univ. Chicago Press, 210 pp., 1970.

Lindemann, F. A., Note on the theory of magnetic storms, Phil.
    Mag., 38, 669–684, 1919.

Loomis, E.: On the geographic distribution of auroras in the north-
    ern hemisphere, Amer. J. Sci. Arts, 30, 89–94, 1860.

Lui, A. T. Y.: Reduction of the cross-tail current during near-earth
    dipolaruzation with multisatellie observations, J. Geophys. Res.,
    116, A12239, doi:10.1029/2011JA017107, 2011.

Lui, A. T. Y. and Kamide Y.: A fresh perspective of the substorm
    current system and its dynamo, Geophys. Res. Lett., 30, 1958,
    doi:10.1029/2003GL017835, 2003.

Maunder, E. W.: Magnetic disturbances, 1882–1903, as recorded at
    the Royal Observatory, Greenwitch, and their association with
    sunspots, Mon. Not. R. Astron. Soc., 65, 2–34, 1905

McIntosh, P. S.: The earth and evolution of sunspots: observations,
    in: The Physics of Sunspots, edited by: Crom, L. E. and Thomas,
    J. H., Sacramento Peak Observatory, Sunspot, 7, 7–54, New
    Mexico, 1981.

McIlwain, C. E.: Plasma convection in the vicinity of of the geosyn-
    chronous orbit, Earth's Magnetospheric Processes, edited by:
    McCormac, B. M., p. 268, D. Reidel Pub. Co., Dordrecht, The
    Netherlands, 1972,

McIlwain, C. E.: Substorm injection boundaries, p. 413, Magneto-
    spheric Physics, edited by: B. M. McCormac, D. Reidel Pub. Co.,
    Dordrecht, The Netherlands, 1974.

Ness, N. F., Scearce, C. S., Seek, J. B.: Initial results of the Imp
    1 magnetic field experiment, J. Geophys. Res. 69, 3571–3569,
    1964.

Parker, E. N.: Dynamics of the geomagnetic storm, Space Sci. Rev.,
    1, 62–99, 1962.

Parker, E. N.: The solar-flare phenomenon and the theory of recon-
    nection and annihilation of magnetic field, Astrophys. J. Supple-
    ment 8, 177–184, 1963.

Perrault, P. and Akasofu, S.-I.: A study of geomagnetic storms, Geo-
    phys. J. Roy. Astron. Soc., 54, 547–573, 1978.

Petschek, E. H.: Magnetic field annihilation.425-439, in The
    Physics of Solar Flares, Proceedings of the AAS-NASA sym-
    posium held 28-30 October, 1963 at the Goddard Space Flight
    Center, Greenbelt, MD, edited by: Hess, W. H., Washington, DC,
    NASA, Science and Technical Information Division, 1964.

Raymond, J. C., Krucker, S., Lin, R. P., and Petrosian, V.: Ob-
    servational aspects of particle acceration in large solar flares,
    Space Sci. Rev., 13, 197–221, doi:10.1007/s11214-0214-012-
    9897, 2012.

Sharma, A. S., Kamide, Y., and Lakhina, G. S. (Ed): Disturbances in
    Geospace: The storm-substorm Relationship, AGU Monograph
    142, AGU, Washington, DC, 2003.

Shelley, E. G.: Heavy ions in the magnetosphere, Space Sci. Rev.,
    23, 465–497, 1979.

Shelley, E. G., Johnson, R. G., and Sharp, R. D.: Satelite obser-
    vations of energetic heavy ions during a geomagnetic storm, J.
    Geophys. Res., 77, 6104–6110, 1972.

Sheeley Jr., N. R.: Energy released by the interaction of coronal
    magnetic fields, Solar Phys., 47, 173–180, 1976.

Sheeley Jr., N. R., Bohlin, J. D., Bruecker, G. E., Purcell, J. D.,
    Scherrer, V., and Tosey, R.: XUV observations of coronal mag-
    netic fields, Solar Phy., 40, 103–121, 1975.

Silsbee, H. C. and Vestine, E. H.: Geomagnetic bays, thgeir fre-
    quency and current-system, Terr. Magn., 47, 195–208, 1942.

Stormer, C.: Sur les trajectories des corpusecules electrises dans
    l'esace sous l'action du magnetime terrestre avec application aux
    auroras boreales, Arch. Sci. Phys. Nat. Geneve, 4, period, 24, 5–
    18, 1907.

Stormer, C.: The Polar Aurora, 403 pp., Oxford Univ. Press, 1955.

Sun, W., Xu, S.-Y., and Akasofu, S.-I.: Mathematical separation
    of directly-drven and unloading components in the ionospheric
    equivalent currents during substorms, J. Geophys. Res., 103,
    11695–11700, 1998.

Svestka, Z.: Solar Flares, D. Reidel Pub. Co., Dordrecht, The
    Netherlands, 399 pp., 1976.

Sweet, P. A.: The neutral point theory of solar flares, in: Electro-
    magnetic Phenomena Cosmical Physics, edited by: Lehnert, B.,
    Cambridge Univ. Press, 123–134, 1958.

Tsuneta, S.: Magnetic reconnection: Open issues, in: Magnetic re-
    connection in the solar atmosphere, edited by: Bentley, R. D. and
    Mariska, J. T., ASP Conference series, Astronom. Soc. Pacific,
    San Francisco, Vol. 111, 409–418, 1996.

Vampola, A. L.: Access of solar electrons to closed field lines, J.
    Geophys. Res., 76, 36–43, 1971.

Vasyliunas, V. M.: Theoretical models of magnetic field line merg-
    ing. Rev. Geophys. Space Phys., 13, 303–336, 1975.