Storm-Substorm Relationship: Current Understanding and Outlook

A. S. Sharma¹, D. N. Baker², M. Grande³, Y. Kamide⁴, G. S. Lakhina⁵, R. M. McPherron⁶, G. D. Reeves⁷, G. Rostoker⁸, R. Vondrak⁹, L. Zelenyi¹⁰

¹Department of Astronomy, University of Maryland, College Park, Maryland
²Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado
³Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, U. K.
⁴Solar Terrestrial Environment Laboratory, Nagoya University, Toyokawa, Japan
⁵Indian Institute of Geomagnetism, Colaba, Mumbai, India
⁶Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California
⁷Los Alamos National Laboratory, Los Alamos, New Mexico
⁸Department of Physics, University of Alberta, Edmonton, Alberta, Canada
⁹Laboratory for Extraterrestrial Physics, Goddard Space Flight Center, Greenbelt, Maryland
¹⁰Space Research Institute, Moscow, Russia

One of the long-standing issues in solar wind-magnetosphere-ionosphere coupling has been the understanding of the development of the ring current, leading to the main phase of a geomagnetic storm. In the earliest picture the ring current resulted from the accumulation of many substorms, each producing an enhanced westward electric field near the outer ring current boundary. This brought in particles from the plasma sheet and was usually referred to as a substorm injection. These particles were supposed to become trapped on closed azimuthal drift paths, and form the symmetrical ring current. The substorm would promptly enhance the plasma sheet population with ionospheric particles which also then contributed to the ring current. The incompleteness of this picture has been recognized for some time now and recently a new view is emerging in which the ring current develops dominantly from a sustained enhancement of the convection electric field. Most of the ring current magnetic perturbations are now recognized to be due to a partial ring current, which closes in part through the ionosphere and in part through the magnetopause. This view is supported by the recent analysis of ground-based as well as spacecraft observations; in particular the IMAGE spacecraft observations show that the ring current is not symmetrical during storm main phase. The enhanced cross-magnetospheric electric field moves the plasma inward thus energizing it and also moving the ring current closer to the Earth. Only after the enhanced field is reduced do particles find themselves on closed drift paths and the ring current becomes symmetric. Thus, the classical two-stage decay of the ring current during recovery is seen as a prompt initial decay due to plasma on open paths convecting out of the system, followed by a slow decay due to charge exchange within the trapped population. At the same time, substorms
are always accompanied by the injection of energetic particles and how these contribute to the storm-time ring current is not completely clear yet. However considering the electrodynamic nature of the interaction between different regions of the magnetosphere and the dominantly global nature of its dynamics, these two manifestations of geomagnetic activity are not expected to just co-exist, but the detailed way in which they influence each other will need to be understood from new observations and modeling. The Chapman Conference at Lonavala [Sharma et al., 2001] was a remarkable occasion, in that it saw the cementing of a new paradigm for the ring current and the storm-substorm relationship. The accumulating evidence against the substorms being the main constituents of storm main phase and the recognition of the dominant role of partial ring current led to a consensus (Lonavala consensus) marking a turning point in the understanding of storm-substorm relationship.

1. INTRODUCTION

The relationship between storms and substorms is a key problem in solar wind-magnetosphere-ionosphere coupling and one of the long-standing questions has been what causes the development of the ring current, leading to the main phase of a magnetic storm. The relative importance of the substorm occurrence and sustained enhancements of magnetospheric convention in the development of magnetic storms has been an open issue. It is widely believed that the main phase of a magnetic storm is the interval in which many intense substorms must take place successively. Among the questions that have been raised in this context [Gonzalez et al., 1994; Kamide et al., 1998] are: Is this a necessary condition for the occurrence of magnetic storm? How many is “many”? How intense is “intense”? How successive is “successive”? These are the type of practical questions which have not been settled in storm-substorm relationship studies.

In terms of physical processes in the magnetosphere, the two major questions are: (1) What is the physical difference between storm-time substorms and non-storm-time substorms, if any? (2) Is the main phase of magnetic storms a result of (a) the impact of southward interplanetary magnetic field (IMF) which also relates to substorm activity, or (b) the successive occurrence of substorms which also have a direct relationship with southward IMF? Numerous studies in the past have not been conclusive and only recently is a coherent picture emerging.

The Chapman Conference at Lonavala [Sharma et al., 2001] was held at a critical juncture, and it saw the cementing of a new paradigm for the ring current and the storm-substorm relationship. The change has been summarized by Clauer et al. [this volume]. In the past, the conventional idea has been that the ring current results from the accumulation of many elementary disturbances, viz. substorms. Each substorm would produce an enhanced westward electric field near the outer ring current boundary and bring particles in from the plasma sheet, a so-called “substorm injection”. These particles become trapped on closed azimuthal drift paths, and form a symmetrical ring current. The substorm would promptly enhance the plasma sheet population with ionospheric particles which would also then contribute to the ring current. The new view is that the ring current development results directly from a sustained enhancement of the convection electric field. Most of the ring current magnetic perturbations are due to a partial ring current, which closes in part through the ionosphere and in part through the magnetopause. The energetic neutral analyzers aboard IMAGE spacecraft observe that the ring current is not symmetrical during storm onset [C. son Brandt et al., Reeves et al., this volume]. The enhanced cross-magnetospheric electric field moves the Alfvén layers inward thus energizing the plasma and also moving the ring current closer to the Earth. Only after the enhanced field is reduced do particles find themselves on closed drift paths and the ring current becomes symmetric. Thus the classical two-stage decay of the ring current during recovery is seen as a prompt initial decay due to plasma on open paths convecting out of the system, followed by a slow decay due to charge exchange within the trapped population. Substorms can play a role in storm development in different ways, e.g., in initiating ionospheric ion out flow [Grande et al., this volume; McFadden et al., 2001], or injection of energetic ions into the inner magnetosphere [Reeves et al., this volume; Daglis and Kamide, this volume]. The emerging picture of storms consists of a dominant role of convection which brings in the plasma into the inner magnetosphere thus increasing its energy due partly to the conservation of the adiabatic invariants, and substorms which may contribute to the energization of the particles and facilitate their trapping in the ring current region.

The debate on the role of substorms in the storm development can now be viewed from a new
perspective of determining the relative roles of substorms and convection. At this stage the role of convection is clearer than that of substorms. The plasma and fields in the different regions of the magnetosphere are coupled strongly and efficiently due to the electrodynamic nature of interaction of the plasma in the anchor dipole field of the Earth [Sharma, 1995]. This coupled with the close proximity of the ring current region, \((4 - 6)R_E\), and the region of substorm initiation, which could be as close or closer to Earth than \(8R_E\), make the development of storms without any influence from the substorms very unlikely. Moreover the substorm injections during storms are ubiquitous during storms [Reeves et al., this volume].

Our understanding of the storm-substorm relationship has been derived mainly from the global characteristics of the magnetosphere. For example, from the observational point of view, the global indices such as Dst and AL have been used extensively to study storms and substorms, respectively. On the theoretical side, energy balance conditions such as the virial theorem have been used to evaluate the partition of stored magnetic energy into the kinetic energy of the ring current and the dissipation in the ionosphere [Siscoe and Petschek, 1997]. This use of the global quantities to describe many essentially local processes is perhaps the main cause of many of the previous misunderstandings. The study of substorm injection during a storm main phase using indices is one such example. Neither the auroral indices such as AL representing substorms, nor the Dst representing the ring current, can yield spatial dependences of a substorm injection, which is a local process. The study of the global as well as the local processes is needed to understand the interrelationship between convection, substorms and storms. Many studies presented in this volume use data from spatial extended locations [e.g., Clauer et al., C:son Brandt et a., Filligim et al., Reeves et al., this volume]

2. THE RING CURRENT: A NEW UNDERSTANDING

Our understanding of the ring current has been derived mainly from the Dst index and its correlation with the solar wind. Another part is obtained from an interpretation of the Dst index in terms of simple models including Chapman-Ferraro theory of the magnetopause, single particle drift in a dipole field, magnetospheric convection, substorm collapse of the tail field and charge exchange loss of ring current protons. Only a small part of our understanding is derived directly from in-situ observation or remote sensing.

A simple model of the ring current, represented by the Dst index, is the Burton et al. [1975] equation, which can reproduce nearly all the variance in hourly values of Dst [McPherron, 1997; McPherron and O’Brien, 2001]. The root-mean-square prediction residual is of order 8 nT, comparable to most estimates of the noise in the solar wind and Dst data. This equation uses only solar wind data to make the prediction, completely ignoring substorms, while the standard paradigm for storm development [Gonzalez et al., 1994] requires a substorm expansion to inject ions and accelerate them to the energies observed in the ring current. This indicates that something is lacking in the standard paradigm since it requires substorms.

In its simplest form the Burton et al. [1975] equation states that the rate of change of dynamic pressure-corrected Dst is a balance of injection controlled by the solar wind electric field, and decay controlled by charge exchange. For data of hourly resolution, the injection filter is a delta function at lag zero, i.e., it is a simple constant of proportionality equal to 4.4 nT/hr per mV/m and the decay time is generally taken to be 8 hours. The interpretation of this empirical relation is that the rectified solar wind electric field drives magnetospheric convection that injects particles into the ring current that then decay away by charge exchange. This alternative model however has a problem as there is no known way for a steady electric field to produce a symmetric ring current. A steady convection electric field drives convection from the tail along drift paths that are open to the magnetopause and thus the particles do not form a symmetric ring current.

It might be thought that this is the role of the substorm expansion. Although the solar wind electric field is steady, the internal convection electric field is not. Each fluctuation in the internal field traps some of the drifting particles onto closed drift paths. Unfortunately the data do not support this conjecture. Iyemori and Rao [1996] have shown that the rate of decrease of Dst becomes smaller after a substorm expansion, not larger. Also Fay et al. [1986] showed that the high time-resolution injection filter is a Gaussian pulse peaked at about 20 minutes delay, a time scale much shorter than the average time for an expansion phase to develop after a southward IMF turning.

The Chapman Conference on Storm-Substorm Relationship [Sharma et al., 2001] coincided with the emergence of a new paradigm for the ring current and the storm-substorm relationship. The change has been summarized by Clauer et al. [this volume] and C:son Brandt et al. [this volume] and presents the ring current and Dst index in a new light. In the standard view the Dst is caused by the symmetric part of the ring current, and asymmetry by a separate partial ring
current. However the new results show that the ring current is not symmetric during the storm main phase. In this view ions simply drift Sunward in the convection electric field until gradient and curvature drifts take over close to the Earth. On the night side the ions drift across equipotentials of the convection electric field gaining energy and increasing their density as they approach the Earth. This produces an effective westward drift current that increases across the night side all the way to dusk. As the ions move into the day sector they drift opposite to the convection electric field, lose energy, move outward, and the current decreases. Of course, the divergence of the westward current must be connected to the ionosphere with an outward current on the nightside and an inward current on the dayside.

In this picture there is no symmetric ring current, only a combination of tail current, partial ring current, and dayside continuation to the magnetopause. However, if the convection electric field begins to decrease slowly, some of the open drift paths will be converted to closed drift paths and a symmetric ring current should begin to develop. A sudden northward turning of the IMF should convert all open drift paths to closed drift paths and eventually the ring current should become completely symmetric. It is easy to demonstrate that this never happens. A plot of the asymmetry index versus the Dst index reveals that the two vary together, and that asymmetry is never zero. In fact Dst can explain over half of the variance in asymmetry. It appears that a substantial fraction of asymmetry is produced by the same current as Dst. The fact that asymmetry is never zero might be explained by the existence of a convection electric field driven by the viscous interaction. Such a field would maintain a background convection electric field and some open drift paths that lead to asymmetry by the mechanism described above.

This model may also explain a perplexing observation that the initial ring current recovery rate after VB, reaches its minimum value is much faster than any possible charge exchange lifetime of observed ring current ions [O'Brien and McPherron, 2000]. For very strong solar wind electric fields (E>10 mV/m) the recovery rate is of order of four hours. This time scale is close to the travel time of particles from dusk to the magnetopause, a loss mechanism that has been named the “flow-out effect” [Takahashi et al., 1990]. Thus the observed dependence of the ring current recovery rate on the solar wind electric field may be a consequence of a gradual transition from the flow-out effect during strong convection to charge exchange further and further from the Earth as the convection field diminishes.

Many papers in this volume emphasize and elaborate different aspects of the point of view that substorms are not essential to the ring current development during the storm main phase. Tsurutani et al. [this volume] have shown that storms are possible without substorms. Korth et al. [this volume] showed that the cycling of O\textsuperscript{+} to the tail was independent of substorm activity. FAST data showed the action of a partial ring current during stormtime [McFadden et al., 2002]. Baker and Li [this volume] showed that the same prediction algorithms relate Dst to solar wind parameters at solar maximum and at solar minimum. It is interesting to consider that if this is so, then ionospheric material, which is largely absent at solar minimum, can not be directly mediating the process. Grande et al. [this volume] showed that on the timescale of order one hour or less there is no major change in the average behavior of the Dst index itself, when ordered by substorm onset, which implies that there is no prompt connection between the injection of energetic particles at substorm onset and changes in the energetic particle population of the ring current.

Overall, it is becoming clear that whereas in the past there was a tendency to think of substorms as the "quanta" of storms, a new view is emerging with a whole range or "family" of coexisting magnetospheric disturbances, including storms, pressure pulse events, enhanced convection interludes, multiple and single onset substorms, pseudo-breakups and enhanced flows. All of these enable the magnetosphere to respond to solar wind drivers, and dissipate energy. The question of to what extent, and under what circumstances, these responses are associated, will form one of the important areas for the field in future. A further major question is the need to understand the build-up of relativistic electrons during storm-time [Baker and Li, this volume]. While the association is clear, the precise correlations are certainly not. It seems clear [Grande et al., this volume] that substorm activity does not directly provide these electrons. However, it may provide the seed population, as in the idea that substorms "precondition" the storm-time plasma.

3. ROLE OF OXYGEN ION AND WAVE-PARTICLE INTERACTIONS

The role of energetic O\textsuperscript{+} ions of ionospheric origin in the development of magnetic storms is among the key unsolved problems [Daglis and Kamide, Grande et al., Korth et al., this volume]. The storm triggers ionospheric upflow, and if it persists long enough, this material finds its way into the plasma sheet, and eventually injected into the ring current. Simulations of ionospheric processes show the formation of patches at altitudes where the oxygen concentration is the highest, thus indicating their potential role in the outflow of oxygen ions [Gondarenko and Guzdar, this volume]. Because of its high mass, the oxygen ions contribute a
large part of the energy density, which can be represented in terms of the Dst index. However, the role of the oxygen ions is not quite clear. The case when it does not play any special role in the process is presented in Grande et al. [this volume]. The other view in which they play a significant role is presented in Daglis and Kamide [this volume]. An important issue is how the thermal oxygen ions are extracted from the ionosphere and then energized to ring current energies of ~ a few keV to hundreds of keV. As the particles convect into the region of stronger magnetic field closer to Earth they gain energy due to conservation of the first adiabatic invariant. Studies of single-particle dynamics in models of magnetic field dipolarizations indicate that low-energy ionospheric-origin O$^+$ ions can be accelerated up to a few hundreds of keV and injected earthward during substorm-related dipolarization events. However this does not yield the required energies and different types of fluctuating fields have been considered as sources of ion energization. Intense plasma waves could provide an efficient mechanism for energy transfer between different ion species and may prove important for selectively heating and accelerating thermal heavy ions [Thorne and Horne, 1994]. In fact, assessing the integrated effect on storm-time ring current losses due to the scattering of ions by waves is one of the unsolved problems concerning magnetic storms [Lakhina and Singh, this volume]. The fact that the occurrence of particular plasma modes is usually limited in time or confined to localized regions casts doubts on the ability of the wave-scattering processes to affect significantly the energy balance of the ring current globally.

Interaction between the ring current ions and modes such as electromagnetic ion cyclotron (EMIC) waves is considered as an important ring current loss process. The global impact of ion cyclotron waves on the ring current have been modeled by using the Ring Current-Atmosphere Interaction Model (RAM) which follows the evolution of three major ring current ion species (H$^+$, He$^+$, and O$^+$) considering adiabatic drift motion, Coulomb collisions, charge exchange, and pitch-angle scattering of protons in the field of EMIC waves [Jordanova et al., 1996; Kozyra et al., 1997]. The model produced order-of-magnitude enhancements in the ion precipitation as a result of diffusion in the ion cyclotron waves within the unstable region. However, no significant impact of the wave losses was seen in the global energy balance even though the waves reduced the anisotropy in the proton pitch angle distributions locally. An improved scheme has been used for estimating the global distribution and amplitude of EMIC activity using a warm plasma ray tracing code in the RAM model [Kozyra et al., 1997]. The comparison of the storm development from this model with that from the maps of magnetic field disturbances obtained from ground-based measurements has contributed to a new view of the storm main phase dominated by partial ring current [Clauer et al., this volume].

The O$^+$ ions of ionospheric origin can also excite the electromagnetic helicon mode in the near-Earth plasma sheet region [Lakhina and Tsurutani, 1997, 1998; Singh et al., 2002]. This instability may play an important role by facilitating the excitation of tearing instability, which is important during substorm onset. During storms, plasma sheet oxygen ions can be accelerated by the helicon mode waves and injected earthwards and become part of the ring current. In addition to the EMIC and helicon mode waves, quasi-electrostatic instabilities driven by loss-cone distributions of the ring current ions can also occur during magnetic storms [Lakhina and Singh, this volume]. The role of the wave-particle interaction in storm development can be best assessed when reliable models of the magnetic field in the inner magnetosphere are available.

4. SUBSTORM AND MAGNETOTAIL EFFECTS IN THE INNER MAGNETOSPHERE

The injection of energetic particles from the near-Earth plasmasheet into the inner magnetosphere is one of the key processes during storms and substorms. Injections are so commonly observed in association with the dipolarization of the magnetic field at substorm onset that if a substorm is documented without observing an injection it is often assumed that there was an injection at another location but that there may not have been a suitably-located observing spacecraft. Recent results and understanding of the role of injections during substorms have been reviewed by Reeves [1998].

Among the key issues in storm-substorm relationship is the differences and similarities of storm-time and isolated substorms [Baumjohann et al., Petrukovich, this volume]. “Isolated” substorm injections are produced by localized, inductive electric fields with little or no change in the large-scale, externally-imposed, “convection” electric field. Therefore the substorm injection serves to move particles from open or untrapped drift trajectories to closed, trapped drift orbits. During storms there are changes in both the large-scale convection electric field and superimposed fluctuations of the more localized inductive electric fields (which are often hard to separate observationally). Either process or, more commonly, both together serve to move energetic particles from the magnetotail to the inner magnetosphere. Whether those particles are eventually trapped on closed drift trajectories or lost to the
dayside magnetopause depends on the precise time history of the electric fields experienced by the particles, but, whatever their fate, those particles (particularly the ions) are the particles that carry the storm-time ring current which is the defining feature of a geomagnetic storm.

With the advent of global Energetic Neutral Atom (ENA) imaging the limitations due to the inability to obtain a global picture of the injection and transport of the actual current-carriers through the inner magnetosphere is beginning to be overcome. Energetic neutral atoms are produced by the charge exchange of magnetospheric ions with tenuous, cold, exospheric neutrals. Since the first application of ENA observations to geomagnetic storms [Roelof, 1987] this technique has led to many new advances. Among these is a direct relationship between ENA flux and the Dst index during the recovery phase of storms. This is expected because the charge-exchange of ring current ions which produces the ENA fluxes is also a direct loss process for ring current ions. If charge exchange is the only loss process then the time rate of change of Dst is proportional to Dst. While the ENA emission during the recovery phase is generally proportional to Dst it is not the only loss process [Jorgensen et al., 2001]. In a two-phase recovery, charge exchange accounts for roughly 75% of the decay of the ring current but in the early, rapid recovery phase it accounts for only a small proportion of the loss. The rapid recovery of Dst is produced primarily by loss of ring current particles to the magnetopause – what is often called the “partial ring current” but might better be called the “untrapped ring current”. Liemohn et al. [2001] recently provided compelling observational and model-based evidence for the importance of magnetopause loss in rapid recovery and Reeves et al. [this volume] show ENA evidence that Dst (and SYM-H) can reach values of -100 nT without any substantial flux of ring current ions extending past noon local time.

The injection of energetic particles during storms and substorms need to be studied using simultaneous data from many different observations. Reeves and Henderson [2001] undertook a study which compared 7 isolated substorm injections with 7 storm-time injections using POLAR ENA observations and in situ geosynchronous fluxes in order to gain better understanding of the storm-substorm relationship. For the geomagnetic storms they used the first injection in order to have a clear and unambiguous timing signature. One conclusion was that while main-phase substorms can be difficult to identify, essentially all storms began with a clear substorm and a clear substorm injection. Further they found that the storm-time injections were essentially identical to isolated substorm injections. They were neither larger (in flux or local time) or more intense (e.g., in spectral hardness). What distinguished storm-times from isolated events was (a) continued injection activity for a period of hours following the initial injection, (b) a spreading of the local time extent of ion injection toward dawn - opposite to the direction of ion drift, and (c) an immediate response in Dst for the storm-time injections compared to no measurable response for the isolated events. This study led to the conclusion that it was the presence of large-scale, externally-imposed, “convection” electric fields superimposed on localized, inductive electric fields which differentiated storm-time particle injections from typical substorm injections. More recently Lui et al. [2001] reached a similar conclusion using Geotail ENA observations of the ring current and SuperDARN radar observations of polar cap convection. Two papers in this volume [C:son Brandt et al., and Reeves et al.] present the first IMAGE observations of storm-time substorms and ring current evolution with unprecedented spatial and temporal resolution of the injection and transport of inner magnetospheric ions.

One of the unique features of ENA observations is that they remotely sense the dominant current carriers in the inner magnetosphere and yet are completely insensitive to other magnetic perturbations such as magnetopause currents, substorm current wedges, or ionospheric electrojets all of which have added ambiguity to understanding the storm-substorm relationship. With the pace of recent developments it is not difficult to imagine the day when the most popular proxy for ring current intensity, Dst, is superseded by actual, global, time-dependent measurements of the ring current ions themselves.

5. RELATIVISTIC ELECTRONS DURING STORMS

As noted by Baker and Li [this volume], most major geomagnetic storms give rise to relativistic electron enhancements in the Earth's outer radiation belt. However, some large storms do not show such electron enhancements [Reeves, 1998]. Thus, it is an area of active research to try to understand in detail how high-energy electron acceleration occurs in the magnetosphere during the course of strong geomagnetic activity. Long-term studies of relativistic electrons in the magnetosphere have shown many of the occurrence characteristics. A very obvious role is played by solar wind speed in producing subsequent relativistic electron enhancements. In fact, the solar wind speed is the single biggest determinant of electron enhancement. However, there is also a key role played by the north-south component of the IMF. There typically must be a significant interval of southward IMF along with a period of high (V\textsubscript{SW}=845
km/s) solar wind speed. Thus, it is generally thought that enhancement in geomagnetic activity (e.g., magnetospheric substorms) is a key first step in the acceleration of magnetospheric electrons to high energies. A second step is then thought to be a period of powerful low-frequency waves that is closely related to high values of $V_{SW}$. In this picture, substorms provide a "seed" population, while high-speed solar wind drives the acceleration to relativistic energies in a two-step geomagnetic storm scenario. This picture seems to apply to most storms examined whether associated with high-speed streams or with CME-related events.

The terrestrial magnetosphere clearly is an efficient accelerator and effective trapping device for energetic particles. The acceleration and transport processes for energetic electrons remain primary issues in magnetospheric physics even four decades after the discovery of the radiation belts. High energy electrons hold special interest because of their continuous presence in the magnetosphere and their effect on human technology. Present-day spacecraft missions have given a remarkable view of energetic particle phenomena. Long-term measurements have unveiled many interesting features of relativistic electrons and have also presented a great variety of new challenges in understanding the dynamics of these particles in the Earth's magnetosphere. Examination of the 10-year record of SAMPEX data [Baker and Li, this volume] shows that the highest electron fluxes were seen in late 1993 and in 1994. The 1993-94 period was a time of very prominent high-speed solar wind streams and was also the period of most extreme relativistic electron radiation in the past solar cycle. There was a clear and prominent 27-day periodicity in the electron flux enhancements. This was well associated with solar wind velocity enhancements. Thus, during the approach to sunspot minimum, high-energy electrons are at their highest levels throughout the outer radiation belt and this population is well associated with recurrent geomagnetic storms. Many authors have studied mechanisms that might account for acceleration of electrons to relativistic energies during geomagnetic storms. An important correlation has been found between electron flux enhancements and ULF waved power in the magnetosphere [Baker et al., 1998a]. Data show increases from quiet day wave power by as much as a factor of 1000 in the frequency range ~1.0 to 20 mHz. Based on correlation studies, it is argued that these ULF waves can play an active role in electron acceleration [e.g., Rostoker et al., 1998; Hudson et al., 2000]. Another important point, however, is that there needs to be a "seed population" of electrons available on which the ULF waves (or other agents) act [e.g., Baker et al., 1998a]. Using plasma wave and particle data from the CRRES satellite, Meredith et al. [2002] suggested that the gradual acceleration of electrons to relativistic energies during geomagnetic storms can be effective only when there are periods of prolonged substorm activity following the main phase of the geomagnetic storm. Thus, magnetospheric substorms are essential to providing the seed population [Baker et al., 1998b]. Baker and Li [this volume] suggest that magnetospheric substorms and geomagnetic storms are closely related to one another when it comes to energetic electron phenomena. They note that it would be remarkable if a southward turning of the IMF that opens the magnetosphere to energy input would lead to two totally separate and unrelated phenomena. The original view that storms are merely a superposition of substorms was clearly too limited. But, on the other hand, it seems unlikely that storms could occur without substorms.

Baker and Li [this volume] espouse the belief that substorms are an important step along the way to geomagnetic storms. The magnetosphere crosses many thresholds in its progression of development and it begins to admit many new forms of energy dissipation as it is driven harder and harder by the solar wind. Substorms are an elementary (and essential) component in this progression. Substorms have many important properties like nonlinearity, complexity, self-organization, and even criticality [Sharma et al., this volume]. As the magnetosphere progresses toward major storms, however, the external driver (the strong flow of the solar wind energy) overwhelms and drives the magnetosphere into a mode of powerful direct response. This strong driving of magnetospheric convection, in turn, produces the conditions that, very frequently at least, produces highly relativistic electrons during geomagnetic storms.

6. GROUND BASED DATA: THE NEED FOR A PROPER INTERPRETATION

The community of researchers working on the storm-substorm relationship has, for years, tried to define the global behavior of particles and fields in the magnetosphere that lead up to substorm expansive phase and to explore the global response of the system as it goes through the development of the expansive phase and the ensuing recovery. The primary reason that these issues have not yet been successfully resolved has been the lack of observation points in the vast volume of space in which the expansive phase develops. The research community is almost always in the position of being unable to separate spatial and temporal effects when trying to analyse satellite data, simply because the magnetosphere is not a homogenous medium in which all perturbations detected can be identified as temporal changes. In an
attempt to alleviate this problem, researchers use ground-based data obtained from instruments such as magnetometers and auroral imagers. These can provide continuous two-dimensional imaging of the auroral ionosphere and the input required by models to infer the three-dimensional structure of the electric current distribution and particle populations. Ground-based magnetometer data, in particular, have been used to identify substorm onsets and to track the evolution of the phases of substorm activity. However, these data are rarely used to the full extent possible with the consequence that some conclusions reached are not warranted. Two such situations are identified below to elucidate how treatment of only a limited part of the ground-based magnetometer data available may lead to incorrect conclusions.

The first case is on the use of the auroral electrojet indices for individual event studies. It is very common to see the state of activity in the magnetosphere quantified by a plot of the AE or AL index, and sometimes onset times will be determined from sudden increases in the value of either of the aforementioned indices. In terms of the storm-substorm relationship, the problem arises because the indices contain contributions of two different types of activity - directly driven and storage-release (cf. Rostoker et al., 1987 for a review of these processes). It is well to remember that the AL index does not reflect the disturbance at any particular point in space. It is derived by superposing all available records of the north-south (H) component of the magnetic perturbations from a specific set of stations (normally 12) at average auroral zone latitudes distributed as uniformly as possible around the world. The maximum negative value from all these records is chosen and AL is assigned that value. The problem that arises is best seen referring to Figure 1. In that figure, the slowly changing perturbation represents a single H-component magnetogram featuring a rise and fall of the directly driven system, whose peak

**Figure 1.** Schematic diagram showing midnight sector and dawn sector H-component magnetograms during a substorm disturbance, together with the AL index that would be formed if these two magnetograms provided the largest disturbance at any time during the three hour interval portrayed. Similar scenario was also discussed in Kamide and Kokubun [1996].

**Figure 2.** North-south component magnetograms from selected CANOPUS stations for a four hour interval on January 10, 1997 (after Rostoker, 2001). The four bottom traces are from the Churchill meridian line from PINA in the south to RANK in the north, while CONT and FSMI are auroral oval stations approximately two hours in local time west of the Churchill line. The attention of the reader is drawn to the ESKI and RANK disturbances at ~0430 UT and ~0630 UT, which look quite similar to one another and might be identified as substorm expansive phase onsets. The ~0430 UT event is, in fact, an expansive phase disturbance.
negative magnetic perturbation at any time is normally found in the dawn sector. A second trace represents a single H-component magnetogram showing a substorm expansive phase onset, which normally occurs in the midnight sector. The third trace is what would appear as the time series of AL values for this event. It is immediately evident that the expansive phase onset would only be a small short-lived increase in AL occurring some time after the start of the increase in the index value. Even more important to note is that, if the peak magnetic perturbation of the midnight sector magnetogram had been less than or equal to the perturbation due to the directly driven activity, the expansive phase onset could not have been identified from the time series of the index. It is clear that for substorm expansive phase identification, AL is unsuitable for either statistical or individual event studies.

The second case deals with the source region of substorm-like disturbances. One point is that there are actually two regions in the auroral oval in which substorm-like perturbations can be found, namely the near the equatorward edge of the oval (expansive phase onsets and pseudo-breakups) and at the poleward edge of the oval (poleward border intensifications or PBIs). In establishing the storm-substorm relationship, it is important to first understand the origin of the belief that the development of the storm main phase is always accompanied by substorm activity.

Figure 2 gives some indication of the nature of this problem. For the event shown, there are two clear disturbances that look very similar in terms of their signatures in the north-south component of the magnetic field. It is only when one examines the evolution of this episode of substorm activity, that one recognizes that the disturbance around 0430 UT is initiated at the equatorward edge of the oval (i.e. is a normal expansive phase onset) while the disturbance at ~0630 UT is at the poleward edge of the oval (i.e. is a PBI). The reader is referred to Rostoker [2002] for a full description of this event. Figure 3 shows latitude profiles for the 0630 UT event which illustrate this point unambiguously. Panel a shows a latitude profile near local midnight just before the onset of the disturbance in question. Panel b shows a differential profile using 0630 UT as the baseline from which the perturbations are measured. Prior to the intensification there is a strong broad westward electrojet centered at ~66.5° N (PACE) with a poleward border at ~67.5° N. The intensification is centered north of the poleward border of the pre-existing westward electrojet with the poleward edge of the latter coinciding with the equatorward edge of the new electrojet (marked by a vertical line in the figure). The weak perturbations (<100 nT) in the latitude regime of that pre-existing electrojet seen in Panel b clearly show the new system to be an independent entity, with the old system not responding significantly at the time of the ~0633 UT intensification. Clearly this event was a poleward border intensification, with little or no response in the equatorward portion of the electrojet (which maps close to the earth near the inner edge of the plasma sheet).

We now can see why the claim that the main phase of a magnetic storm is accompanied by substorm activity must be treated with care. PBIs, as seen in Panel b, can occur at the poleward edge of the auroral oval.
normal magnetograms, look very much like expansive phase onsets. Therefore, it is quite possible that the strong magnetic perturbations seen at average auroral zone locations during storm main phase development are really PBIs. In fact, since the polar cap expands during the development of a magnetic storm, one would expect the poleward edge of the oval to move to latitudes normally occupied by the equatorward edge of the oval. Thus it is entirely possible that the large magnetic perturbations detected at average auroral zone stations do not reflect substorm expansive phase onsets.

The magnetic field variations measured on the ground are driven by the solar wind and in specific cases the relevant solar wind variable can be identified. In the case of January 1997 magnetic cloud event the ground magnetic variations are found to be strongly correlated with the solar wind plasma density [Kamide et al., this volume].

7. FORECASTING STORMS AND SUBSTORMS: ROLE IN SPACE WEATHER

Storms and substorms are the major geomagnetic events of interest for space weather and arise from an efficient coupling of the solar wind energy and momentum to the magnetosphere and ionosphere. They can cause severe damage to our technological systems in space as well as on Earth, and forecasting storms and substorms is essential for protecting these systems. Nonlinear dynamical models of the magnetosphere derived from observational time series data using phase space reconstruction techniques have yielded new advances in the understanding of its dynamics [Sharma, 1995]. In particular, it is now recognized that the dynamics has a strong global component which leads to overall coherence in the magnetosphere. This forms a basis for the predictability of the magnetospheric behavior and space weather. These techniques have been successful in developing nonlinear models for predicting the global dynamics in terms of the geomagnetic indices, such as the auroral electrojet index AL, or the storm time disturbance index Dst, from the spacecraft data of the incoming solar wind.

The importance of nonlinear dynamical studies to space weather arises from its ability to reconstruct the dynamics from the observational data of a limited number of variables. The reconstructed phase space of the system yields a dynamical description independent of particular modeling assumptions, and embodies the information in the past observational data [Sharma et al., this volume].

Predictability is a natural consequence of the low dimensionality, which represents the global aspects of the magnetospheric dynamics. The input-output nature of the solar wind-magnetosphere interaction is incorporated into the local-linear techniques by including the solar wind input along with the geomagnetic response. In input-output studies, the local linear technique has been successful in yielding simple predictive models of the global magnetospheric dynamics by using the main features of the system. In the study of storm-substorm relationship, such a model was developed from the high resolution data set for 1979 using AL as the input and Dst as the input. This model [Kamide et al., 1998] showed that AL can predict Dst well. Another issue that was examined using the input-output models is the relative timing between storms and substorms from AL and Dst indices data. The linear prediction filters relating AL to Dst for data averaged over 10-50 min showed no delay between AL and Dst indices [Sharma et al., 1998]. This indicates that there is no relative delay between substorms and storms, consistent with the conclusion that solar wind, rather than the substorms, is the main driver for storms. The nonlinear dynamical techniques thus yield predictive models and in the same time provide insights on the relationships between different processes. Many techniques are now used to study the solar wind – magnetosphere coupling and among these is the wavelet analysis [Cade et al., this volume].

The effects of geomagnetic storms on the equatorial ionosphere are enhanced during active periods and this has important implications for space weather forecasting [Sastri et al., this volume].

8. SUN-EARTH CONNECTION

Although the role of storms and substorms in the basic structure and dynamics of the magnetosphere has been identified, our knowledge and understanding is still incomplete. Because storms and substorms are temporal reconfigurations, not objects, a successful integration of knowledge requires a paradigm that identifies the sequence of significant events. Some specific unresolved questions are as follows.

The primary question is the initiation of storms and substorms. Geomagnetic storms result from changes in the solar wind that modify the dynamics of the magnetosphere. However, the relative effectiveness of solar wind pressure gradients and interplanetary magnetic field variations is not well established. Similarly, substorms may result from solar wind drivers or they may be initiated within the magnetosphere. So, important progress will be made when we understand the entire sequence of events that initiates a substorm or a storm. Furthermore, there is insufficient understanding of the detailed processes that trigger a substorm, or even how a substorm might be connected to a storm [Reeves et al., this volume].
The energy budget of the magnetosphere during the course of a storm or substorm is another important issue. Although the solar wind is the primary energy source of both storms and substorms, we do not understand how that solar wind energy is coupled to the magnetosphere and how it is dissipated, e.g., by means of the ring current, high-latitude Joule Heating and particle energy input, acceleration of energetic trapped particles, magnetopause losses, etc. [Ostgaard and Tanskanen, this volume]. We need to establish the efficiency of energy transfer from the solar wind into the magnetosphere for a variety of conditions. Understanding this transfer will enable predictive models of the response of the magnetosphere to extreme solar wind conditions.

The understanding of the entire Sun-Earth coupled system is the main issue. Ultimately we need a predictive capability for space weather that allows us to go from observations of the solar surface to ultimate consequences at Earth. Current predictive ability is mainly in solar wind-magnetosphere coupling [Baker and Li, Sharma et al., this volume].

The solutions to these problems require many ingredients. However two key elements can be identified. The first is the development of a comprehensive paradigm of magnetospheric dynamics in which the sequence of events that occur during a geomagnetic storm and substorm are represented. This paradigm should be a synthesized picture that contains the essential features and processes. It would be a modern, contemporary version of the auroral substorm sequence developed by Syun Akasofu in the early 1960's [Akasofu et al., this volume].

His paradigm was based only on ground-based observations, yet it revolutionized our understanding of the auroral current systems and of magnetospheric dynamics. Such an integrated picture need not include all the details, but should be sufficiently comprehensive so as to describe the significant events, the controlling processes, and the most important consequences.

This difficulty in reconstructing an accurate, integrated picture of the essential characteristics of storms and substorms is reminiscent of the classic story of the several blind men who encounter an elephant. Each provides an accurate description of a single part of the elephant, but their combined descriptions are misleading and confusing. In a similar way we have developed very precise characterizations of the various pieces of storms and substorms, but we have yet to develop a satisfactory representation of the entire storm/substorm elephant that lives within the magnetosphere. In fact, there are almost certainly several types of magnetospheric elephants, which only compounds our difficulty in reconstructing the true characteristics of each kind.

<table>
<thead>
<tr>
<th>Combination of Factors</th>
<th>Storms</th>
<th>Substorms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $B_{IMF}$, $B_g$, I</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Magnetosphere of Earth &amp; Giant planets</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. $B_{IMF}$, $B_g$, I = 0</td>
<td>Yes?</td>
<td>No</td>
</tr>
<tr>
<td>Hermean magnetosphere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. $B_{IMF}$, $B_g$ = 0, I</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Venus magnetosphere</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. $B_{IMF}$, $B_g$ = 0, I = 0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Moon like interactions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. $B_{IMF}$ = 0, $B_g$, I</td>
<td>No (weak driving)</td>
<td>Yes (viscous interaction)</td>
</tr>
<tr>
<td>Table 1. Combinations of factors that can lead to substorms and storms in magnetospheres with different characteristics</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Perhaps the first thing would be to provide a better parameterization of the stormtime current as observed on the ground. This could be done by combining simple models for the magnetopause current, the tail current, the symmetric ring current, the partial ring current, and the substorm current wedge. Roughly 15-20 parameters might be determined as a function of time, and used for further analysis. A more sophisticated inversion of these currents could be done using the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) technique [Richmond and Kamide, 1988]. These models should be motivated by more detailed statistical surveys of satellite magnetometer and plasma data during various phases of a storm. The newly emerging technique of Energetic Neutral Atom imaging should be used to develop a picture of where ions are charge exchanging with atmospheric neutrals as a function of storm phase. In conjunction with these experimental measurements there should be continued improvement of the numerical simulations. In particular these simulations need to be made self consistent so that the effects of the storm time currents are fed back to alter the magnetic field in which particles drift. The inner
magnetospheric disturbances. Similarly, they should provide comprehensive measurements of the energy budget and its partition, as well as other significant physical parameters.

Eventually there must be more measurements in the ring current region by multiple spacecraft. Several spacecraft in each of several orbits separated in local time are needed to track the growth and decay of the ring current. Measurements spanning the inner magnetosphere in the equatorial plane, as well as stereoscopic measurements in energetic neutrals, are needed. However, overall, it is apparent that a new consensus has emerged in storm-substorm relationship, which will form the basis for future considerations of the field.

It is of general interest to examine the cause-effect relationships in a broader context of the interaction of solar wind with planetary environments. In the absence of adequate data in this area, we can use a Gedanken experiment to reach reasonable conclusions. There are three main components which control the solar-planetary relations: solar wind supersonic flow with frozen in interplanetary magnetic field $B_{\text{IMF}}$, internal magnetic field of the planet, $B_p$, and planetary atmosphere and the ionosphere, I. We can then examine the effect of switching on and off one or two of these factors. The more familiar cases are those of Venus (developed ionosphere and almost absent internal magnetic field) and Mercury (significant internal $B_p$ and almost absent ionosphere). Also, one could imagine that the interaction of solar wind with these planets occurs during the (exceptional) intervals when the intensity of IMF field is very low ($B_{\text{IMF}} \rightarrow 0$).

The question is what types of dynamical phenomena in planetary magnetospheres one might expect during such hypothetical (although partially realistic) interactions? Would the resulting response be storms, substorms, both or something else? Do we expect, for example, that storm–type and or substorm-type phenomena might occur in the Hermean or Venusian magnetospheres? A list of the possible combinations is shown in Table 1 in terms of possible “Yes” or “No” answers. However, it is not so easy to reach conclusive answers and this Table is meant to stimulate further discussion.

10. SUMMARY

The prevalent view in storm-substorm relationship has been that substorms are the main building blocks of storms [Chapman, 1962]. This view was questioned, based mainly on the important role of the solar wind as the driver of geomagnetic activity [Kamide, 1992]. Advances in modeling and measurements have intensified the pace of development in this arena [Siscoe, 1997]. A new understanding of the ring current, responsible for the magnetic disturbances during storms, was reached at the Chapman Conference on Storm-Substorm Relationship [Sharma et al., 2001]. During the storm main phase the magnetic disturbances are caused by a partial ring current driven mainly by convection, and the fast decay of this current is due to the losses to the ionosphere and the magnetopause. The ring current becomes symmetric as the recovery phase develops and its decay is governed mainly by charge exchange processes. The accumulated results from recent modeling studies and the agreement with the ENA imaging observations have led to this consensus.

However there many issues that need to be settled and some of these are: What role, then, do substorms
play? What does substorm injections do or accomplish? In the broader context, the more fundamental than the storm-substorm relationship is: How do two different processes in the magnetosphere with the same origin, the solar wind, and occurring in the same spatial volume (in a highly coupled system) evolve and how they affect each other?

Acknowledgements. The authors gratefully acknowledge many constructive discussions with many colleagues and grant support by many funding agencies. The supporting grants at the University of Maryland are NSF grants ATM-0001676 and ATM-0119796.

REFERENCES


Baker, D.N., and X. Li, Relativistic electron flux enhancements during strong geomagnetic activity, this volume, 2002.


Lui A. T. Y., R. W. McEntire, and K. B. Baker, A New Insight on the Cause of Magnetic Storms A New Insight on the


Ostgaard, N., and E. Tanskanen, Dynamics of Isolated and Storm-time Substorms, this volume, 2002.


