RE-EXAMINATION OF DRIVEN AND UNLOADING ASPECTS OF MAGNETOSPHERIC SUBSTORMS

D. N. Baker,* A. J. Klimas,* T. I. Pulkkinen** and R. L. McPherron***

* NASA/Goddard Space Flight Center, Greenbelt, MD 20771, U.S.A.
** Finnish Meteorological Institute, Helsinki, Finland
*** IGPP/UCLA, Los Angeles, CA 90024, U.S.A.

ABSTRACT

Magnetospheric substorms represent a global interaction between the solar wind, the magnetosphere, and the ionosphere. Energy extracted from the solar wind is episodically stored in the magnetosphere, with a large fraction of the energy often being in the form of excess magnetic flux in the magnetotail lobes. This stored energy is then explosively dissipated in the near-Earth, nightside region at substorm onset. It is generally accepted, therefore, that substorms consist of both directly driven and loading-unloading processes. However, a recent study has presented results in which nearly 90% of the auroral electrojet (AE) variation was directly predictable from the solar wind variations alone. This would suggest that only a small residual in the AE variability is due to internal magnetospheric dynamics. We consider nonlinear dynamical models of the global solar wind-magnetosphere interaction. In the present work we use the observed, highly variable solar wind electric field (VB5) to drive the Faraday loop analogue model. We find that it is critically important to include magnetotail unloading in the model in order to replicate the main features of geomagnetic activity: With just the driven response in the model, we do not obtain realistic time behavior of the model AL index.

INTRODUCTION

In a 1979 paper, Akasofu /1/ drew the distinction between magnetospheric activity that was "directly driven" by the solar wind and activity that was due to an "unloading" process. For the driven activity, it was suggested that a measure of magnetospheric response such as the total magnetospheric energy dissipation rate /2,3/ would smoothly and faithfully track the solar wind energy input to the magnetosphere. In such a simply driven system, the response would probably show some time delay, but would otherwise reflect all of the major variability of the solar wind velocity and interplanetary magnetic field (IMF) fluctuations that determine the solar wind-magnetosphere coupling strength /3/.

For an unloading process, one would in general expect quite different time dependences for solar wind energy input and the resultant magnetospheric response. In principle, for this kind of coupling the magnetosphere could store energy for long periods of time (~hours) and then could dump or dissipate the energy on time scales very different from those characteristic of the loading time. The distinction between driven and unloading processes is shown in a schematic summary in Fig. 1.
Studies in the last several years /3,4/ have tended to favor a compromise between purely driven or purely unloading processes. It is recognized, for example, that the very process of loading energy into the Earth's magnetotail must be accompanied by some ionospheric dissipation as magnetic flux tubes are convected from the dayside to the nightside of the Earth and are, therefore, dragged through the conducting polar ionosphere. Whether the dissipation associated with the loading process is large or small compared to unloading dissipation levels seems to vary from case to case and also depends on general activity levels /5/.

Researchers repeatedly return to the practical question of just how "predictable" magnetospheric response to solar wind changes can be. A recent study by Goertz et al. /6/ suggested that by appropriate modeling of dayside and nightside magnetospheric processes during substorms, one could obtain linear cross-correlation coefficients between a predicted auroral electrojet (AE) index and the actual measured index in excess of 0.9. Such correlations for any geophysical system are remarkable, implying that over 80% of the variance in the system can be predicted from knowledge of the solar wind properties alone. This would leave little variance due to internal magnetospheric dynamics /3,4/ and would tend strongly to favor the driven activity scenario. We wish to revisit the issue of the relative roles of driven and unloading aspects of substorms using recently developed nonlinear analogue models of substorm dynamics /7,8/.

Fig. 1. An illustration of the driven and unloading processes in substorms. The parameter $\varepsilon$ is the solar wind input rate and $D$ is the energy dissipation rate of the magnetosphere /1/.
SOLAR WIND COUPLING PROCESSES

We adopt a view of the solar wind-magnetosphere interaction that may essentially be termed the "standard" model. As shown in Fig. 2, we assume that the energy coupling to the magnetosphere is effectively determined by the interplanetary dawn-dusk electric field, i.e., the product of the solar wind speed, V, and the southward component of the IMF, B_s. By a process of dayside merging at or near the subsolar magnetopause, this drives large-scale field-aligned currents over a large part of the polar regions and leads to transport of magnetic flux from the dayside to the magnetotail /9/.

![Fig. 2. A schematic diagram of the magnetospheric interaction with the solar wind /9/.

For weak dayside merging, the loading of flux into the magnetotail is quite modest and reconnection at the distant neutral line can balance the dayside merging rate. However, for strong B_s and/or high solar wind speed the dayside merging completely overwhelms the reconnection rate and return of flux from the distant X-line. This imbalance forces near-Earth neutral line development and substorm expansion phase onset. Reconnection in the near-tail rapidly returns magnetic flux to the dayside and rids the tail of large amounts of energy by release of the substorm plasmoid /10/.

A major form of energy dissipation during substorms is the currents which flow from the magnetotail and close through the nightside ionosphere. These "unloading currents", as sketched in Fig. 2, produce high levels of Joule heating in the auroral zone near local midnight and can be monitored quite effectively by magnetic indices such as AE or AL.

In an effort to study the relative roles of driven and unloading behavior during substorms, Bargatze et al. /5/ assembled an extensive solar wind and geomagnetic activity data set. They examined several dozen different intervals when the IMP-8 spacecraft was in the interplanetary medium during 1973-74. These intervals of complete, continuous solar wind coverage were typically several days long. The concurrent geomagnetic activity as measured by the AL index was then examined for these intervals. The 30 time periods selected for final inclusion in the data set were chosen so that geomagnetic activity for each of them went from quiet behavior (AL > -50 nT) through a more disturbed level and then back to quiet behavior. In effect, this complete quiet-to-quiet progression allowed the merging of the 30 separate intervals together as a single time series. The solar wind input parameter VB_s was then produced at 2.5-min resolution and was associated with the corresponding AL index (also 2.5-min resolution) for the same times. The composite AL index time series and the VB_s time series arranged from weakest to strongest disturbance levels are shown in Fig. 3a and 3b, respectively.
The data in Fig. 3 are very compressed, but one can readily see the correspondence between increased amplitudes of VBs in the lower panel and the larger values of AL in the upper panel. This relationship clearly supports the idea that solar wind electric fields ultimately drive geomagnetic activity as measured by the electrojet indices. However, Bargatze et al. /5/ used much more sophisticated time series analysis methods to assess the nature of the solar wind driving.

\[ AL(t) = \int g(\tau) I(t - \tau) \, d\tau \]

The linear prediction filter, g(\(\tau\)), gives the most general linear relationship between I(t) and O(t). As noted above, Bargatze et al. used I = VBs and assumed O = AL. The solid curve in Fig. 4 corresponds to a period of about 50 hours duration near hour 500 in Fig. 3. The dashed curve in Fig. 4 corresponds to a period of similar length near hour 1500 in Fig. 3. These are appropriately considered moderate and strong activity intervals, respectively.

We see in Fig. 4 that the moderate activity filter has two distinct peaks as a function of time lag. The first, lower amplitude peak occurs at about 20-25 min lag while the second, stronger peak occurs at 60-70 min lag time. In contrast, the dashed curve in Fig. 4 shows one dominant peak at about 20 min lag with a broad shoulder extending to higher lag times. The interpretation of these results /3,4,5/ is that the 20-min peak in the filters corresponds to directly driven geomagnetic activity. It is due to dayside reconnection and global magnetospheric convection. It is manifested in all of the auroral electrojet indices (e.g., AL and AE), but is particularly evident in...
Fig. 4. Linear prediction filters for moderate and strong activity levels as labeled /5/.

The eastward electrojet (as measured prominently by AU). This 20-min timescale represents the period it takes for the magnetosphere to readjust to new interplanetary conditions. Fundamentally this represents the DP-2 current system associated with global transport of flux during the substorm growth phase.

The 1-hour peak in Fig. 4 is taken as the time scale of tail unloading. It represents the delay time between an enhancement of VB₅ in the solar wind and the formation of the near-Earth neutral line. It corresponds, therefore, to the substorm current wedge formation and to the DP-1 current system on the nightside of the Earth.

If the above interpretation of the linear filter elements in Fig. 4 is correct, then these results suggest that the magnetospheric response to solar wind input changes quite fundamentally as the system goes from weak to relatively strong driving conditions. For weak and moderate activity the magnetosphere exhibits both a driven and an unloading character whereas for strong activity the unloading response tends to be washed out. Baker et al. /7/ have suggested that this change in the nature of the magnetospheric response represents an evolution toward essentially chaotic internal magnetospheric dynamics during strong solar wind driving conditions.

Based upon this interpretation and upon "phase-space reconstruction" methods, nonlinear dynamical models of the magnetosphere have been developed /7,8/. These models attempt to replicate the internal magnetospheric dynamical evolution including loading of energy from the solar wind and dumping of energy from the system by plasmoid formation. The Faraday loop model of Klimas et al. /8/, in particular, provides a simple, but dynamically complete representation of the substorm cycle including energy storage in the magnetotail (growth phase), sudden unloading of tail energy (expansion phase), and a return toward the equilibrium or ground state (recovery phase).

ANALOGUE MODELS: TIME-DEPENDENT LOADING

Klimas et al. /8/ considered a closed loop in the magnetotail passing in the dawn-dusk direction
through the cross-tail current sheet and then closing over the top of the tail lobe. They determined the total magnetic flux through this loop and calculated the changing flux through it according to Faraday's law: \( \frac{1}{c} \frac{d\Phi}{dt} = -\int \mathbf{J} \cdot d\mathbf{E}. \) With a current sheet width of \( d \), a cross-tail electric field \( (E_y) \) in the current sheet, and an externally imposed electric field \( (E_0) \) at the magnetopause, one has \( \frac{1}{c} \frac{d\Phi}{dt} = (E_0-E_y) \cdot d. \) By estimating the lobe flux from the integral of \( \mathbf{B} \cdot d\mathbf{A} \) and relating this to the cross-tail current density \( j_y \), one can relate the time rate of change of flux to the area \( (A) \) of the lobe and the half-thickness \( (h) \) of the current sheet:

\[
\frac{d}{dt}(hA j_y) = \frac{c^2}{4\pi} \left( E_0 - E_y \right) d
\]

Klimas et al. then considered that the cross-tail current had three components, viz., a pressure balance term \( (j_0) \), a resistive term \( (j_R) \), and a polarization current \( (j_p) \).

Substituting plausible forms for each of the current components, the total current \( j_y = j_0 + j_R + j_p \) is expressed in terms of solar wind parameters, geometric quantities in the tail, and time variations of the cross-tail electric field. In dimensionless form, the nonlinear dynamical equation for the Faraday loop model is

\[
\frac{d}{d\tau} \left[ a(\alpha \beta + \nu E + dE/d\tau) \right] = (E_0 - E_y) \cdot \sqrt{a}
\]

Here \( \tau = \omega t \) is scaled according to the natural frequency of the system, \( a = (d/D) \) is the dimensionless area of the tail lobe, and \( \alpha = (\sqrt{D/h}) \) is taken as a geometrical constant of the tail configuration. Note that \( D \) is the average diameter of the tail. Klimas et al. used the solar wind \( \mathbf{V} \times \mathbf{B} \) electric field to scale \( E_0 \) and \( E_y \); thus, \( e_0 = E_0/E_{sw} \) and \( E = E_y/E_{sw} \). Finally, the damping coefficient \( \nu \) above is dependent on the plasma properties of the plasma sheet and the natural tail oscillation frequency, \( \omega \). The parameter \( \beta \) is the dimensionless tail lobe field strength necessary for pressure equilibrium. To close the equations, flux is loaded into the tail at a rate \( \beta = \beta_L \) for loading conditions and \( \beta = \beta_D \) for unloading. Thus, \( \beta_D \) represents a much more rapid dissipation of the lobe flux which was added to the tail during the growth phase.

We have taken the analogue (Faraday loop) model and have considered highly time-dependent solar wind driving of the model. Thus, in contrast to earlier treatments of the model in which continuous or highly idealized step-function input was used, we have now gone back to the real solar wind to calculate the model inputs. Hence, in Fig. 3 we have used the observed \( \mathbf{V}_{B_S} \) from the Bargatze et al. data set as a driver for the Faraday loop model. In Fig. 5, for example, we show the result of this work for a portion of the overall time series shown in Fig. 3. Specifically, the lower panel of Fig. 5 shows the value of \( \mathbf{V}_{B_S} \) for \( t = 325 \text{ h} \) to \( t = 535 \text{ h} \). In the upper panel, the trace shows the measured value of \( AL \) (see Fig. 3a) for the same time interval. The trace in the second panel of Fig. 5, on the other hand, shows the value of \( AL \) computed from the Faraday loop model using \( \mathbf{V}_{B_S} \) as the driver but with unloading turned off: This corresponds to a strictly driven response. The third panel shows the model response with both driven and unloading included.

We see from Fig. 5 that just including the driven part of the analogue model response fails to replicate the large, abrupt negative spikes in the measured \( AL \) time profile. These big, sudden changes in \( AL \) represent the clear occurrence of substorm expansion phase onsets. Thus, without unloading in the model, one can fit reasonably well the low frequency response of the magnetosphere to the solar wind, but one cannot very well fit the sharp substorm onsets or the large amplitude changes in \( AL \). These discrete features are the essence of substorms. Thus the analogue model with realistic solar wind input driving shows results very similar to prior studies of the real magnetosphere, viz., that the directly driven aspect of substorms cannot account for a substantial portion of the variance in auroral activity time series /3,4,11/.

On the other hand, Fig. 5 shows that with both driven and unloading aspects included, the analogue model provides a reasonably good replication of the observed \( AL \) time series.
Fig. 5. The measured value of AL (top panel) for a portion of the data set shown in Fig. 3 and the response of the Faraday Loop analogue model (middle panels) due to the VBs shown in the lower panel. The second panel from the top shows a model response with only the driven response included while the third panel shows both driven and unloading responses.
CONCLUSIONS

Magnetospheric substorms represent a global mode of interaction between the solar wind, the magnetosphere, and the ionosphere. Energy extracted from the solar wind is episodically imparted to the magnetosphere, with a large fraction of the energy often being stored (or loaded) as excess magnetic flux in the magnetotail lobes. This energy is then explosively dissipated (or unloaded) in the near-Earth, nightside region at substorm onset. The physical mechanism that is thought to play an essential role in the release of magnetotail energy is reconnection and plasmoid formation. The ionosphere plays a key role by supporting global scale current systems that are fundamental to the substorm development.

Work presented in this paper utilized a simple nonlinear dynamical model of the global substorm processes. The results are represented by computational simulations using realistic solar wind input time series. The present work shows that realistic replications of the AL index result from the nonlinear model simulations if unloading is permitted in the model. If unloading is not included, i.e., if only a driven response is permitted, then the resulting simulation of AL is unrealistic in its amplitude and temporal evolution. This is quite consistent with other types of studies of the natural magnetospheric system /4,11/. The model by Goertz et al. /6/ which is in some sense a directly driven system, can be made to reproduce the magnetic indices very well. The Faraday loop model, however, appears to require a type of unloading in its formulation in order to reproduce the data. Future work should focus on how to reconcile the different models and their interpretations.
REFERENCES


