Modeling substorm dynamics of the magnetosphere: From self-organization and self-organized criticality to nonequilibrium phase transitions

M. I. Sitnov, A. S. Sharma, and K. Papadopoulos
Department of Astronomy, University of Maryland at College Park, College Park, Maryland 20742

D. Vassiliadis
Universities Space Research Association, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771

(Received 9 April 2001; revised manuscript received 12 July 2001; published 18 December 2001)

Earth’s magnetosphere during substorms exhibits a number of characteristic features such as the signatures of low effective dimension, hysteresis, and power-law spectra of fluctuations on different scales. The largest substorm phenomena are in reasonable agreement with low-dimensional magnetospheric models and in particular those of inverse bifurcation. However, deviations from the low-dimensional picture are also quite considerable, making the nonequilibrium phase transition more appropriate as a dynamical analog of the substorm activity. On the other hand, the multiscale magnetospheric dynamics cannot be limited to the features of self-organized criticality (SOC), which is based on a class of mathematical analogs of sandpiles. Like real sandpiles, during substorms the magnetosphere demonstrates features that are distinct from SOC and are closer to those of conventional phase transitions. While the multiscale substorm activity resembles second-order phase transitions, the largest substorm avalanches are shown to reveal the features of first-order nonequilibrium transitions including hysteresis phenomena and a global structure of the type of a temperature-pressure-density diagram. Moreover, this diagram allows one to find a critical exponent, that reflects the multiscale aspect of the substorm activity, different from the power-law frequency and scale spectra of autonomous systems, although quite consistent with second-order phase transitions. In contrast to SOC exponents, this exponent relates input and output parameters of the magnetosphere. Using an analogy to the dynamical Ising model in the mean-field approximation, we show the connection between the data-derived exponent of nonequilibrium transitions in the magnetosphere and the standard critical exponent $\beta$ of equilibrium second-order phase transitions.

DOI: 10.1103/PhysRevE.65.016116
PACS number(s): 89.75.Fh, 94.30.Lr, 68.18.Jk, 05.45.Tp

I. INTRODUCTION

Earth’s magnetosphere is a huge cavity created by the magnetic field of our planet in the flow of the plasma coming from the Sun (solar wind). Part of the solar wind energy penetrates this cavity due mainly to the reconnection of the magnetic field lines at the magnetopause, accumulates there, and is then suddenly released [1,2]. The most strongly pronounced phenomena associated with these storage and release processes are called magnetospheric substorms. They have their typical time scale (several hours), well-defined separate phases (growth, expansion, and recovery), and distinctive signatures: ground based (marked by a definite level of the so-called auroral indices), near Earth (aurora brightening), and global (formation and tearing of a huge drop of magnetized plasma, or plasmoid, in the tail of the magnetosphere). There are also manifestations of the magnetospheric activity on other temporal scales, both smaller [pseudobreak-ups, magnetohydrodynamic (MHD) turbulence, current disruption phenomena] and larger (convection bays, magnetospheric storms). The magnetosphere is usually far from equilibrium because of the persistent external driving by the turbulent solar wind as well as its own inherently unstable plasma. One of the most distinctive signatures of this out-of-equilibrium state is the variable asymmetric shape of the magnetosphere with the long (around 100 Earth radii) night-side magnetotail region compared to the relatively short (around 10 Earth radii) day-side magnetosphere. Although this configuration changes drastically during storms and substorms, it always remains highly stretched in the direction away from the Sun. Out-of-equilibrium signatures on lesser scales are conventional MHD turbulence [3,4], specific intermittent energy transport phenomena known as “bursty bulk flows” [5], and non-Maxwellian particle distributions [6]. Thus Earth’s magnetosphere represents an open (input-output) spatially extended nonequilibrium system, which, on one hand, is well organized in space and time, and, on the other hand, manifests its activity over many different spatial and temporal scales. This system has been carefully studied for a long time using data from both ground stations and spacecraft missions. It is believed that some of these results may be discussed in a more general context as they reveal the important features of open spatially extended systems far from equilibrium, which are closely related to concepts now extensively studied in many other branches of science.

A. Self-organization

A considerable group of models of magnetosphere behavior during substorms is based on the assumption of its global coherence. In particular, the near-Earth neutral line model [7,8] explains substorms by the formation of the X line in the magnetic field structure on the night side of the magnetosphere relatively close to the Earth. This process impulsively resolves the imbalance between the rates of reconnection at the day-side magnetopause and the distant neutral line, lead-
ing to the accumulation of magnetic flux in the tail and its stretching. The consequences of the near-Earth neutral line formation are twofold. First, it results in the formation of a plasmoid and its ejection in the direction away from the Sun. On the other hand, it provides a sudden shrinking of the earthward part of the tail and generation of hot earthward plasma flows in that shrinking part, leading to sudden brightenings of the polar aurora. These key features of the global organized behavior of the magnetosphere during substorms have been convincingly confirmed by direct spacecraft measurements including the recent observations by the Interball and Geotail experiments [9–11]. It is tempting therefore to substantiate this organized substorm dynamics on more rigorous mathematical grounds, using in particular modern techniques of data processing and phase space reconstruction [12]. The original idea was to assess the effective dimension of the magnetosphere as a dynamical system in a manner used for many other real systems and nonlinear dynamical models [13]. It was based on the assumption that a considerable part of the complexity of the system behavior is due to the nonlinear dynamics of a few major degrees of freedom (dynamical chaos) and thus the number of these degrees of freedom can be estimated using time delay embedding [12]. Earlier studies [14–19] have actually given clear evidence of the low effective dimension of the magnetosphere. Moreover, further elaboration of this hypothesis has resulted in creating very efficient space weather forecasting tools [20] using local-linear autoregressive moving-average filters [21] and data-derived analogs [22], as well as analog models that explicitly utilize and extend the picture of the magnetosphere as a dripping faucet [23–25].

However, the subsequent analysis [26,27] has cast doubt on this evidence of self-organization in the magnetosphere. It has been found, in particular, that the use of a more appropriate modified correlation integral [28] to assess the effective dimension of the magnetosphere may not reveal any finite value of this dimension. Moreover, the data were shown to share many properties with the colored noise output of a high-dimensional stochastic process. It turned out eventually that the magnetospheric activity might be explained on a basis different from the hypotheses of self-organization and dynamical chaos, namely, as a manifestation of multifractal behavior generated by intermittent processes [29] or turbulence [4,30], or as a colored noise produced by a specific class of cellular automata [31].

B. Self-organized criticality

The idea of using cellular automata to model the magnetospheric activity became popular as observations showed scale-invariant features with a considerable range of scales. It was discovered in particular that the spectra of the magnetic field fluctuations in the tail current sheet [32–35] and those of the auroral indices [27], as well as the probability distributions of auroral blobs [36], obey power laws. The first evidence of scale invariance in the energy releases during substorms was noted in the form of the power-law burst size distribution of the AE index [31]. It has been conjectured [37,31,38–41] that, like other avalanche processes in natural systems such as earthquakes [42] and forest fires [43], the activity of the magnetosphere represents self-organized criticality (SOC) [44,45]. According to the original definition of SOC given in Ref. [44], it differs from the usual criticality, viz., the scale-invariant behavior exhibited by systems near the point of a second-order phase transition [46,47], in that the SOC critical point is an attractor of the dynamics. This kind of criticality arises spontaneously and requires no tuning of the system parameters. Taking explicitly into account the large number of degrees of freedom of the system and their interactions on different scales, the SOC concept seems essentially to complement that of self-organization in modeling open, spatially extended systems. The best-known model of substorms, which emphasizes the multiscale SOC-like aspects of the magnetospheric activity, is the so-called current disruption model [48]. This model is based on the fact that the substorm often starts from a burst of plasma turbulence and the corresponding partial disruption of the cross-tail current in the near-Earth region. Then the global tail reconfiguration including the formation of the X line arises as a macroscopic consequence (inverse cascade) of this relatively small-scale process [49].

However, it is already known that the SOC approach alone cannot describe the whole variety of magnetospheric phenomena. Violations of SOC behavior are detected in observations of particle injections in the near-Earth magnetosphere during substorms [50,51] as well as in the consequences of such injections in the form of VLF whistler mode noise (the so-called substorm-related chorus events) [52]. It has been shown in particular that the intensity and the inter-substorm interval for one-half of the substorms have a probability distribution with a well-defined mean [53] (for more details on the functional form of the distributions, see Refs. [50,51]). Another distinctive non-SOC feature of substorms is their global spatial coherence, exemplified by plasmoids, major substorm current systems, and their rather regular changes recurring in every substorm cycle (see, for instance, Ref. [8]). The simplest mathematical analogs of sandpiles are too simplified to capture this coherent behavior of the magnetosphere as well as its specific features mentioned above.

Another reason forcing us to go beyond SOC is that, contrary to many other SOC prototypes, Earth’s magnetosphere is essentially nonautonomous. The solar wind input in the magnetosphere is by no means steady, and periods of constant loading, usually connected with the southward orientation of the interplanetary magnetic field (IMF), which might actually resemble the flow of sand onto a pile, are often replaced by periods of practically no loading (northward IMF) or nonsteady input due to the transition of IMF shocks. To better understand the most appropriate analog of the magnetospheric dynamics it is necessary to take into account both the output of the magnetosphere and its input. This has become quite clear as a result of attempts to create practical prediction tools of the substorm activity [21]. Even simple linear input-output filters predict a considerable portion of the activity (much more than the extrapolation of the output alone) [54]. The prediction accuracy as well as the length of prediction can be further increased by using local-linear filters with autoregression [21].
B. Input-output analysis

The reconstruction of the geometry of the dynamics from a limited number of time series is based on the idea of time delay embedding [58]. We will use in particular singular spectrum analysis (SSA) [59]. In this technique a time delay is introduced to construct a multidimensional space from the original time series. Then the resulting extended set of the time series data is sorted to reveal their linear combinations, which are most essential to reproduce the dynamics of the system. SSA can also be generalized to the case of input-output systems as described below.

SSA is based on the singular-value decomposition (SVD) (e.g., [60]) of the so-called trajectory matrix \( Y \), which for the given input and output can be presented in the form of the time series of \( 2m \)-dimensional vectors

\[
Y_i = \{(O(t_1), \ldots, O(t_i - (m - 1)\tau)) ; I(t_1), \ldots, I(t_i - (m - 1)\tau)\},
\]

where \( O(t_j) \) is the \( AL \) index characterizing the state of the magnetosphere at \( t = t_j \) (out parameter), while the input \( I(t_j) = -v(t_j)B_z(t_j) \). The time delay \( \tau \) is 2.5 min corresponding to the given temporal resolution of the data and the value of the embedding dimension \( m = 32 \) is chosen to provide a total delay \( \Delta \tau = 80 \) min comparable to the typical largest substorm scales.

The SVD of the matrix \( Y \),

\[
Y = UVW^T,
\]

provides the expansion of this matrix into a series of projections

\[
P_i = U_iw_i = (YV)_i
\]

corresponding to different eigenvalues \( w_i \) of the appropriate \( 2m \times 2m \) covariance matrix \( Y^TY \). One of the ideas behind the original (autonomous) version of SSA [59] was the hypothesis of the “noise floor,” viz., the threshold magnitude \( w_{fl} \) of SSA eigenvalues, which is much less than \( \max w_i \), \( \gg w_{fl} \). Most of the eigenvalues lie below this threshold, \( w_i \approx w_{fl} \). Then the number of eigenvalues with \( w_i > w_{fl} \) is an estimate of the effective dimension of the system. However, being a linear technique, SSA is indicative only of a dimension assessment. Moreover, in many realistic cases the SSA spectrum has a well-expressed power-law shape assuming no floor at all. This occurs, in particular, for the subset of the first 20 intervals of the data set [54] as shown in Fig. 2. Similar results have been obtained for different subsets of that set including the high-activity region [61].

Nevertheless, a limited number of SSA projections may serve as a good approximation of the system based on the following arguments. Let us consider the task of predicting \( AL \) index based on the given data set and taking into account the effects of autoregression. It is reduced to finding the best-fit \( F \) in the equation

\[
F \cdot Y_k = O(t_{k+1}) = O_k, \quad k = 1, \ldots, N_f.
\]

The number of fitting equations \( N_f \) is either comparable to the number of points in the data set in the case of global
linear fitting or much less than that number in the case of local-linear (nonlinear) fitting. It is known that just SSA applied to $Y$ provides the best linear least-squares fit in this case. In other words, the formula

$$ F_{apr} = V_i [1/w_i] \cdot (U^T \cdot O), $$

(2.5)

where the sum over $i$ is limited by a small number of the largest $w_i$, yields in most cases the best fit for $F$. This feature of SSA/SVD procedures allows one to obtain a reasonable image of the system in the space of a relatively small dimension corresponding to the few largest SSA eigenvalues. Eventually, SSA determines which specific linear combinations of the extended original set of variables are most appropriate for creating a finite-dimensional image of the system’s dynamics.

C. First-order phase transitions

Figures 3–5 represent the set of three eigenvectors $V_i$, $i=1,2,3$, corresponding to the three largest SSA eigenvalues $w_i$, and the approximation of the global dynamics of the magnetosphere by the two-dimensional, (2D) surface in 3D space determined by these eigenvectors. This choice of the embedded 3D space as well as the approximating 2D manifold is not arbitrary. It is based on a direct assessment of the dimension of the system in the space of higher dimensions [59], which while not completely conclusive (as the dimension is not sustained for relatively small scales) is valid on the largest scales.

The eigenvectors in Fig. 3 are not the immediate outcome of the SVD routine. An additional rotation has been made in the chosen 3D subspace of the main eigenvectors to obtain the best image of the substorm dynamics, which is also most suitable for interpretation and comparison to other systems. (The description of the specific rotation algorithm, quantitative parameters, and criteria may be found in [61].) The comparison of the first eigenvector $V_1$ plotted in Fig. 3(a) with Eqs. (2.1) and (2.3), which define the trajectory matrix $Y$ and projections $P_i$, shows that this eigenvector forms $P_1$ by extracting from the data largely the time-integrated parameter $vB_s$, the input parameter of the system. The second projection $P_2$, according to Fig. 3(b), reflects mainly the output $AL$ averaged over time in a similar manner. Figure 3(c) shows that the third dynamical variable $P_3$ is constructed like $P_1$ largely of the input time series $vB_s$. However, in contrast to $P_1$, it is roughly proportional to the time derivative of $vB_s$ or, to be more precise, the appropriate finite difference between a nearly immediate value of $vB_s$ and its value taken approximately one hour earlier. $P_1$ and $P_3$ may be treated therefore as independent dynamical variables like the generalized coordinate and momentum of the system.

The average evolution of the system on the plane of the two main SSA components $P_2$ and $P_1$ corresponding to the eigenvectors $V_2$ and $V_1$, representing the input and output, respectively, is shown in Fig. 4. The circulation flows given by $dP_2/dt$ and $dP_1/dt$ are represented by the arrows. The third reconstructed coordinate $P_3$ is gray scaled. Surface ap-
proximation has been performed using the standard triangulation procedure, which is described in detail in [61]. Figure 5 reflects nearly the same dynamical picture as shown in Fig. 4 (only hysteresis events are deliberately excluded and discussed below separately) on the plane \((P_3, P_1)\), with \(P_2\) being gray scaled. The substorm cycle starts from an increase of the variable \(P_1\) at constant (nearly zero) output \(P_2\) (left part of Fig. 4), which corresponds to the so-called growth phase of substorms [1]. During this phase the third principal component \(P_3\) first increases and then decreases to nearly zero level (right part of Fig. 5). Then the output component \(P_2\) falls from zero down to large negative values at nearly constant input parameters \(P_1\) and \(P_2\) (upper part of Fig. 4; the anomalous phenomenon inside the white frame is discussed below). This corresponds to the so-called substorm expansion phase. The restoration of the system to the original state, corresponding to motion toward the left bottom corner in Fig. 4, involves the decrease of \(P_1\) and \(-P_2\), which is first accompanied by further decrease of \(P_3\) followed by the restoration of this parameter to zero (left part of Fig. 5). This stage is usually called the recovery phase in the substorm phenomenology.

This picture of the magnetosphere’s evolution during substorms has been inferred from the data [54] on the basis of its dimension assessment and small (around 10–20 \% [61]) relative deviation of the actual trajectories of the system in 3D space from the approximating 2D surface. The SSA itself was used in this procedure to reveal three specific dominant directions in the original multidimensional phase space representing the input-output dynamical relationship. Both the
surface itself and the corresponding circulation flows turn out to be surprisingly close to a very simple low-dimensional scheme of the magnetospheric substorm as a cusp catastrophe (inverse bifurcation) first proposed by Lewis [62] and illustrated in Fig. 6. In this model the dynamics of the magnetosphere is described by the following evolution equation for the state parameter $z$:

$$
\frac{dz}{dt} = -\frac{\partial U(z,c_1,c_2)}{\partial z},
$$

(2.6)

where the effective potential

$$
U(z,c_1,c_2) = z^4 + 2c_1z^2 + 4c_2z^2,
$$

(2.7)

has two control parameters $c_1$ and $c_2$. These parameters control the quasistatic changes of $z$, which are possible as long as the condition $\partial^2 U(z,c_1,c_2)/\partial z^2 = 0$ is satisfied. This condition determines the folded surface drawn in Fig. 6. The state parameter in the model [62] is the night-side magnetic field orientation. The first control parameter is $c_1$ = (open) flux + const, which resembles the parameter $-P_1$. The second control parameter $c_2$ = [(night-side) − (day-side)] reconnection rate is very much like the parameter $-P_3$ because, according to Fig. 3(c), $P_3$ represents the difference between the immediate dayside reconnection rate, which is proportional to $\nu B_1$ [positive black bay in Fig. 3(c)] and its delayed value (negative black bay), with the delay being comparable to the time of signal propagation from the subsolar magnetopause to the distant neutral line. This is why the negative black bay in Fig. 3(c) extracts from the $AL-\nu B_1$ time series a parameter that mimics the night-side reconnection rate in spite of the fact that this parameter is not directly measured. The potential (2.7) may have either one or two minima corresponding to different equilibrium states of the system. The onset of the substorm is represented as a local fold catastrophe arising due to the disappearance of the upper potential minimum. Figure 6 also shows a typical substorm cycle including the growth, onset, and recovery phases suggested in the model [62].

This resemblance to the model [62] persists in some more details. In particular, the catastrophe scenario implies that the rate of decrease of the parameter $P_2$ must be largest in the left upper sector of Fig. 4 because the region of the most pronounced catastrophic changes is expected just in this part of the phase space. This expectation is confirmed by mapping the parameter $dP_2/dt$ on the same surface $(P_2, P_1)$ [61]. Another feature of the catastrophe scenario is the hysteresis phenomenon marked by the white frame in Fig. 4. Here, in the vicinity of the local fold catastrophe, the same set of input parameters $P_1$ and $P_3$ results in different values of the output $P_2$ depending on the history of the system. Note that the hysteresis prevents one-to-one surface plotting in Fig. 5. This becomes possible only after removing the appropriate (and rather rare) hysteresis episodes from the original data set.

Figures 4 and 5 clearly show that the evolution of the magnetosphere on the largest scales is quite regular. It is seen particularly well from the circulation flows in Fig. 4 and the surface approximation in Fig. 5, which resembles the temperature-pressure-density (TPD) diagram of equilibrium phase transitions [46], although the latter simplification of the actual dynamical picture requires the elimination of hysteresis episodes. However, the analysis of the data using SSA [54] reveals also significant deviations from the ideal bifurcation/catastrophe picture [63,64], associated with the finite rates of driving and evolution. In particular, the clear TPD structure gradually disappears with increase of the average activity level [61], presumably because of "tunneling" transitions through the effective potential well associated with the local fold catastrophe. But there are also deviations that cannot be described by merely dynamical effects. First, the dimension is not clear for smaller scales (less than ~1/20 of the largest scale). Second, the power-law SSA spectrum itself does not at all imply that the number of essential principal components that are necessary to reveal the main dynamical features is 3 or even some larger but finite number.

We offer, however, another interpretation of the above data, which accounts for these deviations. As was conjectured in [61], the same catastrophelike picture and multiscale features may be created by dynamical phase transitions. The above bifurcation/catastrophe picture then turns out to be just one aspect of a more general phenomenon. This aspect is associated with first-order dynamical phase transitions, while the deviations from the ideal catastrophe picture may be explained by second-order phase transitions near the critical point. Moreover, this first-order phase transition picture suggests the location of the critical point, which can then be used to obtain the appropriate critical exponents [46,65].

D. Criticality

Self-organized criticality models, which represent systems where the input is not essential because of self-tuning properties, provide only one class of critical exponents that relate some parameter of the system such as the energy released to the spatial scale or the characteristic frequency [45]. In con-
FIG. 7. Mean-field picture of the dynamical phase transition according to the Ising model. \( \Delta m_1 \) is the change of the magnetization according to the coexistence (dashed) curve typical for equilibrium phase transitions; \( \Delta m_2 \) is the appropriate width of the spinodal (dash-dotted) curve appearing in the nonequilibrium case; \( \Delta m_3 \) is the change of magnetization in the nonequilibrium system evolving through the marginal stability state under quasistatic evolution of the control parameters \( H \) and \( T \).

Contrast to this rather peculiar critical behavior, general second-order phase transitions have at least one more class of critical exponent that relates the input parameter of the system (such as magnetic field or temperature) to its output (magnetization or density) [46]. The data-derived image of substorms as phase transitions provides an interesting opportunity to check whether we are dealing with second-order phase transitions or with the SOC phenomenon.

The problem of finding the critical exponents other than simple frequency or size distributions is complicated, however, because of the dynamical and nonequilibrium character of the corresponding transitions. In particular, the hypothetical coexistence curve (Fig. 7), which is often used in the equilibrium case to assess the exponent \( \beta \) [46], cannot be used in our case because of a number of hysteresis phenomena associated with “overheated” and “overcooled” states of the magnetosphere [61]. Moreover, even the spinodal curve, which explicitly takes into account these metastable overheated and overcooled states (see, for instance, [66]), cannot be identified in our case because of the finite and varying rate of the external driving. Nevertheless, an exponent similar to the classical \( \beta \) exponent can be proposed and inferred from the data. This is the exponent \( \beta_H \), which relates the envelope of the output variation rate during the transition to the conventional input parameter. To reveal the relationship between these two exponents let us consider a simple model of dynamical phase transitions, namely, the Ising model in the mean-field approximation [67]:

\[
\frac{dm(t)}{dt} = -m(t) + \tanh([m(t) + H(t)]/T) \tag{2.8}
\]

This equation describes the time evolution of the dimensionless magnetization \( m \), which is controlled by the imposed magnetic field \( H \) and the temperature \( T \) of the system. The critical point in the space \((T,H,m)\) is \((1,0,0)\), and the spinodal and coexistence curves are located in the region \( T < T_\ast = 1 \).

Equation (2.8) describes in particular the slow changes of the state parameter \( m \) under quasistatic changes of \( H \) and \( T \) on the surface:

\[
F(m,T,H) = -m(t) + \tanh([m(t) + H(t)]/T) = 0, \tag{2.9}
\]

which is similar to the cusp manifold in Fig. 6 [in fact, Eq. (2.8) and Eqs. (2.6),(2.7) belong to the same class of evolution equations representing cusp catastrophes]. The system evolves on the manifold as the trajectory moves from the lower to the upper level following a path away from the fold. Once in the upper part, the system approaches the fold region, at the level of the spinodal curve it leaves the surface (2.9). The change of the parameter \( dm/dt \) during this dynamical transition is determined by Eq. (2.8). Assuming that the change of \( m \) during the transition is fast compared to the changes in the parameters \( H \) and \( T \), one can show that the rate \( dm/dt \) actually has an envelope in this interval given by the equation \( dF(m,T,H)/dm = 0 \). Let this envelope be the maximum. Then one considers only the transitions with an increase of magnetization. This maximum \( v_{\text{max}} \) of the rate \( dm/dt \) can be found using the expansion of \( F \) in a power series with respect to \( (m + H)/T \) close to the critical point \( T \to 1 \) with \( T < 0 \),

\[
v_{\text{max}} = \max\{dm/dt\} \approx 2(T/\sqrt{3})(1 - T)^{3/2}, \tag{2.10}
\]

while the “height” of the transition along the direction \( m \) (the parameter \( \Delta m_3 \) in Fig. 7) may be assessed as

\[
\Delta m_3 \approx T\sqrt{3}(1 - T). \tag{2.11}
\]

Similarly, one can verify that close to the critical point \( \Delta m_1 \approx \Delta m_3 \). As a result, the appropriate maximum value of velocity \( v_{\text{max}} \) is connected to the magnetization as

\[
v_{\text{max}} \approx (\Delta m_3)^3 \approx (\Delta m_1)^3. \tag{2.12}
\]

[Note that for the other, non-phase-transition-like class of potential models governed by the equation, with \( d^2m/dt^2 \) instead of \( dm/dt \) in the left hand side of Eq. (2.8), the exponent in Eq. (2.12) would be equal to 2.] Then the exponent that relates the velocity envelope to the input (the magnetic field),

\[
v_{\text{max}} \approx (1 - T)^{\beta_d}, \tag{2.13}
\]

turns out to be connected [in the mean-field approximation (2.8)] to the standard critical exponent \( \beta \) [46] in the relation \( \Delta m_1 \approx (1 - T)^{\beta} \) as

\[
\beta = \beta_d/3. \tag{2.14}
\]

As one can see from Fig. 8, the envelopes of the velocity time series \( v(t) = dP_2/dt \) do actually exist. However, they turn out to be qualitatively different for positive and negative values of the parameter \( dP_2/dt \). While the envelope of positive \( dP_2/dt \), corresponding to the recovery phase of substorms, essentially coincides with a straight line, the array of negative \( dP_2/dt \), corresponding to the expansion phase of substorms (decrease of the \( AL \) index), has a curved envelope, which resembles the nonanalytical dependences typical
of second-order phase transitions. In general the straight envelope may be slightly inclined with respect to the ordinate axis. However, the three eigenvectors in Fig. 3 and the corresponding principal components are determined to within the arbitrary rotation in the 3D subspace created by these eigenvectors, and the specific choice of them reflected by Figs. 3–5 was made after trial and error to get the best 3D images of the regular low-dimensional aspect of substorm dynamics (Figs. 4 and 5). These rotational degrees of freedom can be used once again to match the straight envelope with the ordinate axis as shown in Fig. 8. In fact, this is a slight adjustment of eigenvectors \( V_1 \) and \( V_2 \) compared to their original set used in Figs. 3–5. To find the lower (curved) envelope in this figure we binned the data points on the axis \( P_1 \) into 136 cells on a logarithmic scale and left only the maximum values of \(-dP_2/dt\) in each bin. The result is shown in Fig. 9. This plot may be approximated by a formula of the type of (2.13):

\[
\max(-dP_2/dt) = C(P_1)^{\beta_\ast}
\]  

(2.15)

with the exponent \( \beta_\ast = 0.6374 \pm 0.035 \), which suggests the existence of the appropriate critical exponent \( \beta \approx 0.21 \). For comparison, fluids have \( \beta = 0.34 \) [46], while the critical exponent obtained from the 3D Ising model using the renormalization technique is \( \beta = 0.312 \) [68]. A more fluidlike result (\( \beta \approx 0.275 \)) can be obtained if the linear fitting of the plot in Fig. 8 is restricted to only the medium-scale input parameter values \( 0.01 < \log_{10}(P_1) < 1 \). It should be noted that the relationships (2.10)–(2.13) are obtained under the mean-field approximation and exponents such as those given by Eqs. (2.14) would be obtained more appropriately using the renormalization-group approach [65].

The relationship (2.15) drastically differs from the earlier power-law findings [33,34,31,35,40] because it relates the state parameter of the system not to the frequency but to the actual control parameter of the system. This is analogous to an equilibrium second-order phase transition, with \( P_2 \) being the analog of the density and \( P_1 \) of the temperature difference \( T_c - T \). Also, in contrast to those earlier results but in agreement with the conventional phase transition picture, this scale-invariant behavior has a well-defined location in the space of control parameters of the system. These considerations lead to the conclusion that the substorm dynamics of the magnetosphere resembles more the conventional set of first- and second-order phase transitions rather than any standard SOC or catastrophe model.

This data-derived picture is essentially consistent with the concept of forced and/or self-organized criticality (FSOC), which has been put forward recently in Refs. [69–71] to generalize the original SOC formulation [44], as well as the underlying advanced SOC models [39,41]. The effect of solar wind driving on the scale-free properties of geomagnetic activity has also been emphasized in Ref. [72] based on a comparison of the burst lifetime distributions of \( AU \) and \( |AL| \) indices with similar distributions of the solar wind parameters \( vB_s \) and \( e \) [73]. Moreover, it has been conjectured, based on the similarity of the input and output distributions, that the scale-free properties of geomagnetic indices arise from the solar wind and may not be an intrinsic property of the magnetosphere. It appears, however, that in the latter case the critical exponents \( \beta \) and/or \( \beta_\ast \) would be either integer or relatively simple rational numbers.

### III. Conclusion

In spite of some favorable evidence such as power-law spectra, effective dimensions and TPD diagrams, neither SOC nor self-organization models taken separately can explain the whole variety of the magnetospheric activity on substorm scales. On the other hand, our analysis appears to reveal a more general class of models, which combine these seemingly incompatible concepts in a consistent manner analogous to the equilibrium phase transition theory [46]. As we have shown in this paper, the behavior of the Earth’s magnetosphere resembles very closely that of real sandpiles (e.g., Refs. [74,45]). Both systems reveal scale-invariant be-
behavior for relatively small avalanches and first-order phase-transition-like behavior for the largest avalanches. Thus the power-law spectra of magnetospheric activity including that inferred from our SSA are indicative of second-order phase transitions (in general dynamical and nonequilibrium) rather than being SOC manifestations. This conjecture is confirmed in this paper by the power-law input-output relationship and the appropriate critical exponent $\beta_\ast$. Using an analogy to the dynamical Ising model in the mean-field approximation, we have shown the connection between this data-derived exponent and the standard critical exponent $\beta$ of equilibrium second-order phase transitions, which has no analog in SOC.

In principle, the critical behavior of discrete SOC models has already been shown to require external tuning [75]. However, the situation is much less clear in the case of continuous models [76,77], and our results provide independent arguments in favor of a second-order (tuned) critical behavior of the magnetosphere. Note that even the specific location of the critical point in our data-derived phase transition picture ($P_1 = P_3 = 0$) is consistent with the results [75], and in particular with the requirement of a small enough rate at which a conserved quantity is added to the system.

Recently, a number of attempts have been made to consistently combine global and multiscale aspects of dynamics in modeling dissipative spatially extended nonlinear systems. Some of them (Refs. [39,78,79]) are advanced SOC models that describe internal scale-invariant avalanches and system-wide events distributed around a characteristic mean. These models are quite consistent with the data-derived phase transition picture of substorms presented above, and yet they seem to somewhat oversimplify the specific features of the global substorm dynamics, including the formation of a plasmoid or first-order phase-transition-like behavior. They are also essentially autonomous, in contrast to the explicitly nonautonomous behavior of the magnetosphere. A more general approach, involving the SOC concept and containing an explicit description of the global dynamics in a manner close to the mean-field approximation of phase transitions, was proposed in [80] on the basis of Ginzburg-Landau theory. This describes the global behavior in terms of a subcritical bifurcation, while the multiscale dynamics is introduced through a diffusive relaxation process. A similar and even more realistic model, showing both scale-invariant behavior and global avalanches with a well-defined mean power, was proposed in [81] on the basis of a one-dimensional MHD model of magnetic field reversal corresponding to the tail region of the magnetosphere.

The general problem of the conditions under which we may use concepts dealing with equilibrium statistical mechanics phenomena, like equilibrium phase transitions, to describe the behavior of an out-of-equilibrium system such as the magnetosphere remains open. One way to resolve it is to extend the static equilibrium description in terms of specific dynamic and nonequilibrium effects such as dynamic critical exponents and hysteresis [82,66,83]. Another direction is the inverse dynamical chaos problem [12]. The latter has provided the techniques of time delay embedding and singular spectrum analysis to reconstruct a sufficient number of system variables, analogs of the temperature, pressure, and density of the equilibrium water-steam system, for the cases where these variables are either not evident or not available because of lack of the appropriate number of original time series [14–21,61]. In the case of magnetospheric studies this has led to the data-derived TPD diagram of the nonequilibrium transitions during magnetospheric substorms, which includes hysteresis effects [61]. The third interesting direction of modeling transitions in nonequilibrium systems is connected with SOC and FSOC models with the main emphasis on the nonequilibrium modifications of classical critical phenomena, their possible self-organization, and the appropriate autonomous scale-free properties of the system. The application of these concepts to the magnetosphere was recently reviewed in Refs. [70,79]. In this paper we have tried to combine the approaches of the two first directions mentioned above to distinguish between the SOC regime and those more resembling FSOC [70], driven SOC [72], and second-order phase transitions. We have shown in particular how to assess the analog of the equilibrium critical exponent $\beta$ in the nonequilibrium case. This exponent cannot be obtained directly in nonequilibrium systems like the Earth’s magnetosphere because of their unsteady loading and hysteresis effects. Moreover, in contrast to many other nonequilibrium systems explored (see, for instance, Refs. [83–86]), the magnetosphere cannot be precisely tuned to infer critical exponents from the hysteresis properties themselves. This is a reason why the rather simple and straightforward approach proposed in this paper to assess an analog of the critical exponent $\beta$, which allows one in particular to distinguish between SOC and genuine phase-transition behavior, may be of interest beyond magnetospheric research. It should be especially advantageous when the system allows only passive observations without any tuning.

ACKNOWLEDGMENTS

The authors acknowledge useful discussions with D. N. Baker, T. Chang, S. C. Chapman, C. Goodrich, A. J. Klimas, A. T. Y. Lui, G. Milikh, V. A. Sergeev, V. Uritsky, and J. A. Valdivia. They also thank Y. Avitzour for discussions of the dynamical Ising model. This work was supported by NSF Grants No. ATM 9901733 and No. ATM 0001676.


As shown in Ref. [72], nondiverging means may be a direct consequence of the exponential cutoffs in solar wind burst lifetime distributions rather than an intrinsic property of the magnetosphere. They appear also in some special classes of the SOC models [39]. We describe, in fact, the third situation when the nonsteady solar wind driving with a finite driving rate transforms an otherwise pure SOC system into a nonequilibrium phase transition system according to mean-field theory [75].
