IMPULSIVE MAGNETIC RECONNECTION IN THE EARTH’S MAGNETOTAIL AND THE SOLAR CORONA

A. Bhattacharjee
Center for Magnetic Reconnection Studies, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire 03824;
email: amitava.bhattacharjee@unh.edu

Key Words forced reconnection, resistive MHD, Hall MHD, magnetotail substorms, solar flares

Abstract Impulsive reconnection dynamics is characterized not only by fast growth but also by a sudden change in the time derivative of the growth rate. I review recent developments in the theory and simulation of forced impulsive reconnection based on the equations of resistive and Hall magnetohydrodynamics (MHD). Impulsive reconnection can be realized in resistive as well as Hall MHD by the imposition of suitable boundary conditions. However, compared with resistive MHD, Hall MHD reconnection is distinguished by qualitatively different magnetic field and electron and ion signatures in the reconnection layer. Furthermore, nonlinear reconnection rates in Hall MHD are weakly dependent on the Lundquist number. I discuss applications of the physics of impulsive reconnection to substorms in the Earth’s magnetotail and solar flares.

1. INTRODUCTION

Although fast reconnection has been a subject of great interest to laboratory, space, and astrophysical plasma physicists for nearly five decades, the primary emphasis of nonlinear reconnection theory has been on understanding steady, or quasi-steady, reconnection dynamics. In numerical simulations, researchers have often made deliberate attempts to set up initial and boundary conditions that realize a quasi-steady state, thus enabling tests of steady-state models. This is not surprising. The seminal theoretical models of Sweet-Parker (Sweet 1958, Parker 1957) and Petschek (1964), which have provided the compass for our thinking on nonlinear reconnection physics, are steady-state models. Both models, which have been extensively discussed in the literature (see, for example, Biskamp 1993, Priest & Forbes 2000), were motivated by observations of solar flares and are based on resistive magnetohydrodynamics (MHD). To fix ideas, let us consider a sheared equilibrium magnetic field.
\( \mathbf{B}_0 = B_P \tanh \left( \frac{y}{a} \right) \mathbf{\hat{x}} + B_T \mathbf{\hat{z}}, \) \quad (1)

where \( B_P \) and \( B_T \) are positive constants. The poloidal component of the magnetic field changes sign across the so-called neutral line at \( y = 0 \). In the Sweet-Parker model, assuming that the plasma is incompressible, steady-state reconnection occurs in the vicinity of the neutral line on a characteristic timescale \( \tau_{SP} = (\tau_A \tau_R)^{1/2} = S^{1/2} \tau_A \), where \( \tau_A \equiv a/V_A = a(4\pi \rho)^{1/2}/B_P \) is the poloidal Alfvén time, \( \tau_R \equiv 4\pi a^2/\eta c^2 \) is the resistive diffusion time, and \( S \equiv \tau_R/\tau_A \) is the Lundquist number. (Here \( \rho \) is the mass density, \( \eta \) is the resistivity of the plasma, and \( c \) is the speed of light.) In the Sweet-Parker model, the reconnection layer has a \( Y \)-point geometry (Syrovatsky 1971) in the \( x \) – \( y \) plane (Figure 1). For weakly collisional systems such as the solar corona, \( S \) is typically very large (\( \sim 10^{12} – 10^{14} \)), and since \( \tau_A \) is of the order of one second, the timescale \( \tau_{SP} \) is on the order of hours. Because \( \tau_{SP} \) is much too long to account for fast events such as solar flares, Petschek proposed a different steady-state model that maintains an \( X \)-point geometry for all times (Figure 2). In contrast with the Sweet-Parker model, the Petschek model yields a reconnection timescale that has a weak logarithmic dependence on \( S \). For the high-\( S \) solar corona, the Petschek timescale is on the order of minutes, much closer to the observed timescale for flares. For a comparative discussion of relevant timescales and dimensionless numbers for the solar corona and other plasma environments, I refer the reader to the monograph by Priest & Forbes (2000).

Figure 1  A schematic diagram of the \( Y \)-point geometry of the Sweet-Parker reconnection layer. The length of the layer (shaded in black) is typically on the order of the system size and is much larger than the layer width \( \Delta \).

Since the mid-1980s, high-resolution computer simulations have produced some of the most illuminating tests of Petschek’s model (see Biskamp 1993 and references therein). A persistent feature seen in these simulations is that in the

Figure 2  A schematic diagram of the \( X \)-point geometry of the Petschek reconnection layer. The length of the layer (shaded in black) is of the same order as the layer width.
high-$S$ regime, even if one begins with an equilibrium state containing an $X$-point that would appear to favor Petschek, in the nonlinear regime one ends up obtaining an extended reconnection layer with $Y$-point structure typical of Sweet-Parker, whereupon reconnection occurs on the Sweet-Parker timescale. Thus, the classical models of Sweet-Parker and Petschek leave us with a quandary for high-$S$ plasmas. Whereas the Sweet-Parker timescale is realizable dynamically in the high-$S$ regime, it is too slow. On the other hand, the Petschek model, which yields a faster timescale, appears not to be realizable in the high-$S$ regime.

In addition to the issue discussed above, there is a subtler issue that is impossible to resolve within the framework of any steady-state reconnection model. By definition, steady-state models can provide but one timescale—that of steady reconnection (proportional to $S^{1/2}$ for Sweet-Parker and $\ln S$ for Petschek). However, steady reconnection is not a generic condition. It is a strong theoretical assumption, and one that is frequently violated in many dynamical situations of great physical interest. In particular, there are phenomena involving magnetic reconnection in laboratory, space, and astrophysical plasmas where the growth rate is not only fast, but also exhibits a sudden increase in its time derivative. This is often referred to as the “trigger problem”—the magnetic field configuration evolves slowly for a long period, only to undergo a sudden change during a much shorter period. As the classical steady-state reconnection models of Sweet-Parker and Petschek do not include time dependence, they cannot account for the time evolution of the reconnection rate. We characterize these reconnection processes as impulsive or bursty. They are the focus of this review.

There are numerous examples of impulsive reconnection phenomena in laboratory as well as naturally occurring plasmas, some of which have been studied quite extensively. These include plasmas in high-temperature tokamaks, the Earth’s magnetotail, or the solar and stellar coronae that are characterized by very different plasma parameters, geometry, and boundary conditions but that have one feature in common: Their classical Lundquist number is often very high ($S \geq 10^8$). Studying reconnection phenomena in these plasmas from a common perspective is useful because the diagnostic tools available are different but complementary, and they elucidate different aspects of the underlying physics (see Bhattacharjee et al. 2001 for a recent tutorial that includes tokamak applications). Together, these observations impose powerful constraints on theory.

Magnetic reconnection can be broadly classified into two types: “free” and “forced.” Free reconnection is caused by an instability, driven by the magnetic free energy stored in an equilibrium. Forced reconnection is not an instability. It is driven by perturbations imposed at the boundary that induces topological change in a stable equilibrium. The emphasis of this review is on problems of forced, impulsive reconnection.

We consider forced impulsive reconnection in two classes of fluid models: resistive MHD and Hall MHD. Ohm’s law in resistive MHD is

$$\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} = \eta \mathbf{J},$$

(2)
where $E$ is the electric field, $B$ is the magnetic field, $v$ is the plasma flow velocity, $c$ is the speed of light, $\eta$ is the plasma resistivity, and $J$ is the current density. The classical Lundquist number of most space astrophysical plasmas we consider here is very high. Such plasmas tend to develop thin and intense current sheets of width $\Delta_1$ in the reconnection layer (Figure 1). When the thin current sheet width $\Delta$ is sufficiently narrow, it is not valid to neglect the collisionless terms in the generalized Ohm’s law discussed in textbooks (see Gurnett & Bhattacharjee 2004)

$$E + \frac{v \times B}{c} = \eta J + \frac{4\pi}{\omega_{pe}} \frac{DJ}{Dt} - \frac{\nabla \cdot \vec{p}}{ne} + \frac{J \times B}{ne}.$$  (3)

Here $\vec{p}$ is the electron pressure tensor, $\omega_{pe}(\omega_{pi})$ are the electron (ion) plasma frequencies, $n$ is the electron density, $e$ is the magnitude of the electron charge, and $D/Dt \equiv \partial/\partial t + v \cdot \nabla$ is the total convective derivative. By collisionless terms, we mean the last three terms on the right-hand side of Equation (3). The second term is attributed to finite electron inertia, the third to electron pressure tensor, and the fourth to Hall current. The last two terms in Equation (3) are often referred to collectively as Hall MHD terms. In recent years, it has become apparent that the Hall MHD terms bring about fundamental changes in the physics of the reconnection layer obtained from resistive MHD theory. As discussed later, these changes have significant implications for impulsive reconnection phenomena.

In Section 2, we begin with a theoretical example of impulsive forced reconnection within the framework of two-dimensional (2D), incompressible resistive MHD. In Section 3, we treat the same example using the 2D Hall MHD equations. This comparative discussion demonstrates that although forced impulsive reconnection dynamics is realizable in both resistive and Hall MHD models, impulsive reconnection mediated by Hall MHD effects is distinguished by qualitatively different magnetic field and electron and ion signatures in the reconnection layer. In Sections 4 and 5, we use the Hall MHD model to study the problem of substorm onset in the Earth’s magnetosphere and reconnection in coronal arcades, respectively. We conclude in Section 6 with a summary and a discussion of issues that merit further investigation.

2. IMPULSIVE FORCED RECONNECTION IN RESISTIVE MHD

We begin with an example of impulsive forced reconnection in 2D incompressible resistive MHD. At $t = 0$, we assume that the equilibrium magnetic field, specified by Equation (2), is subject to steady and continuous inward flows of the form $v = \pm V_0(1 + \cos \kappa x)\hat{y}$, imposed at the boundaries $y = \mp a$. Here $V_0$ is a small, sub-Alfvénic velocity, i.e., $V_0 \ll V_A$. This introduces a new timescale in the problem, the so-called forcing timescale $\tau_0 = a/V_0$, which is assumed to obey the inequality $\tau_A \ll \tau_0 \ll \tau_R$. 
The initial-value problem in the linear and nonlinear regimes was treated by Wang et al. (1996) who demonstrated by analysis and simulation that the reconnection dynamics exhibits an impulsive phase. The magnetic and velocity fields are represented as

\[ B(x, y, t) = \hat{z} \times \nabla \psi(x, y, t) + B_z(x, y, t)\hat{z}, \]  
and

\[ v(x, y, t) = \hat{z} \times \nabla \phi(x, y, t), \]
respectively, where \( \psi(x, y, t) \) is a flux function and \( \phi(x, y, t) \) is a stream function.

In an incompressible plasma obeying the 2D resistive MHD equations, the guide field in the \( z \)-direction plays no role in the dynamics. In the results discussed below, we set \( B_z = B_T = 0 \) in the initial equilibrium, and it remains so for all times. The nonlinear resistive MHD equations reduce to the two following coupled equations for \( \psi \) and \( \phi \):

\[ \frac{\partial \psi}{\partial t} + v \cdot \nabla \psi = \frac{\eta c^2}{4\pi} \nabla^2 \psi, \]  
\[ \rho \left( \frac{\partial}{\partial t} + v \cdot \nabla \right) \nabla^2 \phi = \frac{1}{c} B \cdot \nabla J, \]

where \( J \equiv J_Z = (4\pi/c)\nabla^2 \psi \).

The linear phase of the dynamics can be obtained by using the methodology of Hahm & Kulsrud (1985), investigated further by others (Wang & Bhattacharjee 1992, Vekstein & Jain 1998, Fitzpatrick 2003). In analytical treatments of reconnection dynamics, it is standard practice to use boundary-layer theory. One assumes that the plasma is divided into two regions: the “outer region” that comprises most of the plasma and the “inner region” that is the boundary layer centered on the neutral line \( y = 0 \). The inequality \( \tau_A \ll \tau_0 \ll \tau_R \) is invoked to make the simplifying assumption that the outer region adjusts quickly to the boundary perturbation to maintain MHD equilibrium at all times, and that all explicit time dependence can be neglected (i.e., \( \partial/\partial t = 0 \)). In the inner region, one retains time dependency and inertia, and nonideal effects such as resistivity and viscosity, but the governing equations are simplified by exploiting the narrowness of the boundary layer. The inner and outer region solutions are then matched asymptotically. Fitzpatrick et al. (2003) have recently demonstrated that the asymptotic matching procedure outlined above underestimates somewhat the reconnection rate, especially at early times.

A measure of the reconnection rate is given by the perturbed current sheet amplitude at the \( X \)-point, which can be calculated analytically (in certain limits) and numerically. Wang et al. (1996) have demonstrated that the dynamics occurs on a sequence of characteristic timescales. In the linear phase, when \( t \ll \tau_L \equiv \tau_A^{2/3} \tau_A^{1/3} \tau_R^{1/3} \ll \tau_{SP} \), a perturbed current sheet develops at the \( X \)-point with an amplitude growing algebraically as \( J_0(t/\tau_A)^2 \), where \( J_0 \) is a constant independent of \( \eta \). The linear phase is followed by a nonlinear phase in which the reconnection scales
Figure 4 The current density at the X-point as a function of time in units of $\tau_A$ in the resistive MHD simulation (Wang et al. 1996).

as $(t/\tau_N)^{3/2}$ where $\tau_N \equiv (\tau_R \tau_0 \tau_A^3)^{1/5}$ is a new nonlinear timescale, with $\tau_A \ll \tau_0 \equiv a/V_0 \ll \tau_{SP}$. Compared with the Sweet-Parker timescale $\tau_{SP}$ (which is proportional to $\eta^{-1/2}$), the new nonlinear timescale $\tau_N$ has a weaker dependence on $\eta$ (proportional to $\eta^{-1/5}$) as well as a dependence on the forcing timescale $\tau_0$. The growth of the current sheet amplitude exhibits an impulsive enhancement: It grows as $J_0(t/\tau_A)^2$ in the early linear regime and then as $(\tau_R/\tau_0)^{1/2} J_0(t/\tau_A)^{3/2}$ in the nonlinear regime. The multiplicative factor $(\tau_R/\tau_0)^{1/2}$, which can be very large in the high-$S$ regime, gives an indication of the impulsive enhancement in the current sheet amplitude at the X-point and hence the reconnection rate.

These analytical results discussed above have been tested with a 2D, incompressible resistive MHD code. Details are given by Wang et al. (1996), from which we take one example. Figure 3 shows a contour plot at $t = 20\tau_A$ of the flux function $\psi$, the total current density $J = J_z$, the electric field $E_z$ aligned with $J_z$, and the velocity $v_x$. Inward flows, specified above, are imposed at $y = \pm a$, and open boundary conditions at $x = \pm 2a$. The simulations are run with the parameters $\tau_0 = 100\tau_A$ and $S = 10^5$, which yield $\tau_N \approx 25\tau_A$ and $\tau_L \approx 46\tau_A$.

We note the presence of an intense current sheet near the separatrix spanning the Y-points, caused by flux pile-up. The geometry is similar to that envisioned by Sweet (1958), Parker (1957), and Syrovatskii (1971) but very different from that of Petscheck (1964). Figure 4 shows the time evolution of the current sheet density $J_y$ at the X-point. The current density, and hence the reconnection rate, clearly exhibits an impulsive transition from a slow growth phase to a faster phase.

3. IMPULSIVE FORCED RECONNECTION IN HALL MHD

We now consider the same example of impulsive reconnection as in Section 2, except that the governing equations are now given by compressible Hall MHD. We assume that the electron pressure is a scalar, i.e., $\tilde{p} = \rho \tilde{I}$, where $\tilde{I}$ is the
unit dyadic. This assumption is made for reasons of simplicity because it enables us to close the system of equations by providing a simple equation of state for the electron pressure. This assumption is supported in certain circumstances by observations, but it is not valid in general. However, the qualitative features of Hall MHD reconnection emphasized in this review do not depend on the tensor character of the pressure. In dimensionless form, the generalized Ohm’s law can be rewritten

\[ E + v \times B = \frac{1}{S} J + \frac{d_i^2}{n} \frac{dJ}{dt} + \frac{d_i}{n} (J \times B - \beta_p \nabla p), \]

where we have redefined the following variables to make them dimensionless:

- \( cE/(B_P v_A) \rightarrow E \)
- \( B/B_P \rightarrow B \)
- \( v/v_A \rightarrow v \)
- \( p/(n_0 T_e) \rightarrow p \)
- \( n/n_0 \rightarrow n \)
- \( \beta_p \equiv 4\pi n_0 T_e / B_P^2 \)
- \( di/a \rightarrow di \)
- \( de \rightarrow de/a \)
- \( \nabla \rightarrow \nabla \)
- \( 4\pi a J/(cB_P) \rightarrow J \).

In the definitions above, \( n_0 \) is the average ion and electron density (in a hydrogen plasma), and \( T_e \) is the electron temperature. On the right-hand side of Equation (6), there are two terms that can break field lines: the first term proportional to the plasma resistivity, and the second term proportional to electron inertia. In the simulations described below, where \( S = 10^5 \), the effect of resistivity dominates the effect of finite electron inertia that is, therefore, neglected (i.e., we set \( de = 0 \)).

The time-dependent equations for \( B_z \) and \( \psi \) in Hall MHD are obtained by combining Faraday’s induction equation with the generalized Ohm’s law (Equation 6). These equations are

\[ \frac{\partial B_z}{\partial t} = -\nabla \cdot (B_z v) + B \cdot \nabla v_z + \frac{1}{S} \nabla^2 B_z - \frac{d_i}{n} \left[ \nabla \times \left( \frac{J \times B - \nabla p}{\rho} \right) \right], \]

and

\[ \frac{\partial \psi}{\partial t} = -v \cdot \nabla \psi + \frac{1}{S} J_z + \frac{d_i}{\rho} (J \times B)_z. \]

Other relevant equations have been given by Ma & Bhattacharjee (1996).

We now present Hall MHD simulation results, using the same initial condition as in the resistive MHD study, discussed in Section 2. In the early stages of the simulation, the geometry resembles that from a resistive MHD simulation (Figure 3) transiently. The Hall MHD terms come into play when the thickness \( \Delta \) of the Sweet-Parker current sheet becomes smaller than the ion skin depth, i.e., when \( \Delta < d_i \). Then, the geometry as well as the dynamical features of the reconnection layer changes qualitatively. Figure 5 shows contours of \( \psi \), \( J_z \), collisionless \( E_z \) and \( v_x \), respectively, for a typical run with \( d_i = 0.1, s = 10^5 \) at \( t = 20\tau_A \) in the Hall MHD simulation. The dotted lines in Figure 6 show, respectively, the time dependence of the current sheet amplitude \( J_z \) and the growth rate of the magnetic flux \( (=d\ln\psi/dt) \) at the X-point. For comparison, we show in Figure 6 the results from a resistive MHD simulation with the same initial conditions. We note the
Figure 6  Time development of $J_z$ (top frame) and $d \ln \psi / dt$ (bottom frame) in resistive MHD (solid line) and collisionless Hall MHD (dotted line) runs with the same initial conditions. Time is measured in units of $\tau_A$. We note the impulsive enhancement in $J_z$ and $\psi$ at about $12 \tau_A$ in the Hall MHD simulation, which exceeds the enhancement in the resistive MHD simulation by nearly an order of magnitude (Ma & Bhattacharjee 1996).

impulsive enhancement in $J_z$ and $\psi$ at approximately $12 \tau_A$ in the Hall MHD simulation, which exceeds the enhancement in the resistive MHD simulation by nearly one order of magnitude.

By contrasting the Hall MHD and resistive MHD simulations, a few distinguishing features emerge. First, although both simulations show an impulsive signature, the enhancement in the Hall MHD simulation exceeds that in the resistive MHD simulation by nearly an order of magnitude. Second, the geometry of the reconnection region in the Hall MHD simulation changes qualitatively as the $Y$-points of the transient Sweet-Parker current sheet shrink rapidly to produce a current sheet of significantly shorter length. Consequently, there is a rapid enhancement in the reconnection rate. Third, there is a decoupling of the spatial scale of $E_z$ from the spatial scale of $J_z$ that does not occur in resistive MHD. Whereas the $E_z$-layer has a spatial scale of the ion skin depth $d_i$, the $J_z$-layer has a narrower spatial scale, determined by the high value of $S$. This decoupling of the spatial scales occurs
because Hall MHD effects decouple electron and ion motions in the reconnection layer (Biskamp et al. 1995, Ma & Bhattacharjee 1996, Shay & Drake 1998, Wang et al. 2000, Birn et al. 2001). Whereas the electrons remain frozen to the magnetic field down to the spatial scale of $J_z$, the ions decouple from the electrons on the spatial scale of the ion skin depth.

The results mentioned above prompt comparisons with the Petschek model, although the nonlinear dynamics in our simulation are nonsteady, the reconnection is essentially dominated by collisionless effects, and both of these features are qualitatively beyond the scope of Petschek’s original model. Whereas Petschek obtained reconnection rates that depend logarithmically on the resistivity, the present model obtains reconnection rates that are near-Alfvénic and weakly dependent on the resistivity. Whereas Petschek obtained such high reconnection rates by assuming that the reconnection dynamics sustains an $X$-point geometry for all times, the present model evolves in the high-$S$ regime from a $Y$-point geometry in the early nonlinear stage, characteristic of the models of Sweet (1958), Parker (1957), and Syrovatskii (1971), to a geometry in which the length of the current sheet is significantly shorter.

Impulsive reconnection of the type discussed above has been seen not only in resistive and Hall MHD simulations but also in particle-in-cell (PIC) simulations in which the electron pressure is not a scalar but a tensor (Horiuchi & Sato 1997, Bessho et al. 2003). When the electron pressure is a tensor, it provides dominant support for the electric field $E_z$ at the $X$-point. In contrast, a scalar pressure contributes nothing to $E_z$ at the $X$-point. Despite this distinction, the fact that impulsive reconnection occurs in resistive MHD, Hall MHD as well as fully kinetic PIC simulations suggest that it depends fundamentally on the nature of the forcing process rather than the specific details of fluid or kinetic physics that resolve the inner region. Although the magnitude of the impulsive enhancement does depend on the underlying fluid or kinetic physics of the inner region, its basic occurrence is a robust qualitative feature present in fluid as well as fully kinetic models.

How large is the maximum collisionless reconnection rate after the onset of the impulsive phase, and how does it scale with plasma parameters? These are interesting questions about which there is significant controversy. Some make the strong claim that the nonlinear reconnection eventually asymptotes to a universal fast phase that is on the order of the tenth of an Alfvén speed (Shay et al. 1999). Others have questioned the validity of this strong claim with analytical and numerical results that demonstrate that although reconnection rates as large as a tenth of the Alfvén speed may be possible with special initial conditions, the maximum reconnection rate can often be significantly lower, depending on plasma parameters (such as $d_e$) and the properties of the boundary perturbation (Wang et al. 2000, 2001; Dorelli 2003; Craig et al. 2003; Fitzpatrick 2004). There are also more specific questions regarding the precise scaling of the nonlinear reconnection rate with resistivity (equivalently, $S$) or electron inertia (equivalently, $d_e$), the mechanisms that break field lines, in the limit $S \to \infty$ or $d_e \to 0$. This is an important point of principle even if the dependence is weak. (We recall for
comparison that the Petschek model, based on resistive MHD, depends on \( S \) logarithmically.) In Hall MHD, although some initial conditions appear to produce asymptotic reconnection rates that are independent of \( S \) (Wang et al. 2001, Dorelli 2003, Fitzpatrick 2004) or \( d_e \) (Shay & Drake 1998, Hesse et al. 1999), in other initial conditions the reconnection rate has been demonstrated to depend on \( d_e \) algebraically (Grasso et al. 1999, Porcelli et al. 2002). It is evident that the structural details of the thin current sheet embedded in the broader reconnection layer do depend on \( S \) or \( d_e \). Therefore, the question of the scaling dependence of the nonlinear reconnection rate on \( S \) or \( d_e \) is fundamentally connected to the question of how sensitive the reconnection electric field is to the current sheet structure that is dominated by electrons and has a characteristic spatial scale much smaller than that of the electric field. This question continues to be a subject of current research.

4. THIN CURRENT SHEET DYNAMICS IN THE EARTH’S MAGNETOTAIL AND SUBSTORM ONSET

The examples discussed in Sections 2 and 3 prepare us to study the problem of substorm onset in the Earth’s magnetotail. The key to substorm onset lies in the growth phase when the magnetotail is prepared for the relaxation dynamics that follow. Researchers now widely accept that during the growth phase the cross-tail magnetotail current intensifies, resulting in substantial stretching of the nightside magnetic field lines at near-Earth distances (<10 \( R_E \) where \( R_E \) denotes the radius of the Earth) (Kaufman 1987). Observations show that the thickness of the thin current sheet can be reduced to less than 1 \( R_E \) in the near-Earth region before the onset of the expansion phase. The onset of the expansion phase involves a sudden reduction (usually referred to as disruption) of the cross-tail current density in the near-Earth region (7–11 \( R_E \)). The current sheet disruption region is localized (<1 \( R_E \)) and the disruption of the current is also partial, typically involving about 20% of the cross-tail current (Lui 1978, Ohtani et al. 1992). At the onset of the expansion phase, the magnetic field configurations at low latitudes become dipolar, and the plasma sheet expands.

Observations of current disruption at near-Earth distances (Ohtani et al. 1992) provide significant constraints on theoretical models of substorms. From local magnetic field and particle measurements, it is inferred in these observations that, after a period of sluggish growth (~0.5–1.5 h), the cross-tail current density exhibits rapid, impulsive growth during a short interval (<1 min) just before the onset of the expansion phase. Following the impulsive growth phase, the current disrupts on a very short timescale (~10 s). These results, which have been summarized qualitatively by Ohtani et al. (1992) in a schematic diagram (shown here as Figure 7), are of great interest because they show the presence of two timescales in the growth phase, followed by a third timescale during which rapid disruption of the tail current occurs.
In this section, we present the results of a Hall MHD simulation of 2D substorm dynamics in the Earth’s magnetotail using the compressible Hall MHD equations beginning from a realistic magnetotail equilibrium (Ma & Bhattacharjee 1998). Following magnetospheric physics conventions, we assume that $y$ is an ignorable coordinate. (In other words, the role of $y$ and $z$ are exchanged.) To maintain adequate spatial resolution in the high-$S$ regime, we consider a 2D domain on the nightside, extending from $-6R_E$ to distances just above $-46R_E$ down the tail. All variables are cast in dimensionless form: In particular, distances are scaled by $1R_E$, and time is scaled by the characteristic Alfvén time $\tau_A = 1R_E/V_A (\sim 6$ s) where the Alfvén speed is computed at the corner $x = -6R_E, z = 12R_E$ of the computational domain. The magnetic field $B$ is represented as $B = \hat{y} \times \nabla \psi + B_y \hat{y}$. The initial equilibrium is obtained by solving the Grad-Shafranov equation (see Gurnett & Bhattacharjee 2004)

$$\nabla^2 \psi = J_y(\psi) = -\frac{d\psi}{d\psi}, \quad (9)$$

numerically with $B_y = 0$. (The inclusion of a $B_y$ field in the initial equilibrium does not change our results qualitatively.) We start with a flux function $\psi_0$ that reproduces the magnetic field in the Earth-noon meridian plane of the Tsyganenko (1989) model and specify the cross-tail current $J_y(\psi)$ as a flux function,

$$J_y(\psi) = -J_0 - J_d \psi_d^3/(\psi^2 + \psi_d^2)^{3/2}, \quad (10)$$

where $J_0, J_d$ and $\psi_d$ are constants. We then iterate numerically the equilibrium equation $\nabla^2 \psi_{n+1} = J_y(\psi_n)$ beginning from the initial source term $J_y(\psi_0)$ until the
convergence condition \(|\nabla^2 \psi_n - J_y(\psi_n)| < 10^{-9}\) is satisfied. Such an equilibrium is generally stable with respect to resistive tearing instabilities at near-Earth distances owing to the presence of a significant \(B_z\) component. Because free reconnection is thus ruled out, we are motivated to investigate the role of forced reconnection. An electric field, given by

\[
E = E_0[0.6 + 0.4 \tanh\left((3(x_1 + x) + 2x_2)/x_2\right)],
\]

is imposed in the \(x\)-direction at the upper and lower boundaries at \(z = \pm 12R_E\), where \(E_0 = 0.006\), \(x_1 = 6R_E\), and \(x_2 = 40R_E\). The functional form (Equation 11) is a simple representation of an electric field that peaks at near-Earth distances and decays monotonically with increasing \(|x|\) (Birn & Hesse 1991). The maximum equatorward inflow \(v_z\) owing to the electric field (10) is on the order of 60 km/s, which is about 6% of the characteristic Alfvén speed (~1000 km/s), roughly consistent with observations. Free boundary conditions are used at the Earthward as well as tailward boundaries of the simulation box for all dependent variables except the flux function \(\psi\) which obeys \(d\psi/dt = 0\).

In Figure 8, we show a color-coded plot of the level surfaces of the cross-tail current density. In early stages of the simulation, an \(X\)-type neutral line with an extremely small separatrix angle is formed in the region \(x = -30R_E\) where the \(B_z\) field is initially weak. Evidence for the formation of such an \(X\)-line during the growth phase is strong in recent GEOTAIL observations (Nagai et al. 1998) and consistent with one of the predictions of the so-called near-Earth neutral line model of substorms (Hones 1979). After a period of slow algebraic growth in time, the cross-tail current density at near-Earth distances exhibits a sudden impulsive enhancement at a sub-Alfvénic growth rate that is weakly dependent on the value of the Lundquist number. This is demonstrated in Figure 9 where we plot the time evolution of the cross-tail current density \(J_y\) at a typical near-Earth distance \(x = -12R_E\) near the center of the plasma sheet for \(S = 10^5\) (solid line), \(2 \times 10^5\) (dotted

![Figure 9](image_url)

**Figure 9** Slow growth, near-explosive preonset, and local disruption of the cross-tail current density at a near-Earth distance of \(12R_E\). The three traces correspond to three different values of the Lundquist number (Ma & Bhattacharjee 1998).
line), and $10^4$ (dashed line). We note in all cases an impulsive enhancement of the amplitude of the current density at $t \sim 300 \tau_A$, with the rate of enhancement weakly dependent on the magnitude of the Lundquist number. Because of the inclusion of the Hall terms in the generalized Ohm’s law, the time of impulsive enhancement is on the order of 1 min, which is comparable with the timescale reported by Ohtani et al. (1992). This feature of impulsive enhancement is qualitatively similar to that found in our resistive and Hall MHD simulations of a Harris sheet, discussed in Sections 2 and 3.

The cross-tail current density is enhanced by an order of magnitude compared with its value at the end of the slow growth phase and the thickness of the current sheet is reduced approximately to $0.2 R_E$. The geometry of the thin current sheet and reconnection region undergoes rapid temporal changes, evolving from an extended Y-point structure to a structure shorter in length and thinner than that seen in a resistive MHD simulation run with the same initial condition (Ma et al. 1995).

In the present simulation, the impulsive growth phase ends with a partial current disruption brought about by near-Earth reconnection. However, two significant discrepancies with observations remain. First, the timescale of current disruption and dipolarization in our simulation is on the order of several minutes, whereas observations appear to suggest that the timescale is on the order of 10 s. In other words, the timescale observed in our simulation at near-Earth distances is too slow by nearly an order of magnitude when compared with observations, and this discrepancy cannot be cured by simply appealing to a larger Lundquist number. Second, the magnetic field in the simulations does not truly dipolarize. Because our simulation is 2D (that is, $y$ is an ignorable coordinate), three-dimensional (3D) instabilities that break the $y$-symmetry are excluded, and such instabilities can alter qualitatively the dynamics seen in the late stages of our simulation. For example, we have demonstrated elsewhere (Bhattacharjee et al. 1998, Zhu et al. 2003) that because of the buildup of strong pressure gradients in the impulsive growth phase at near-Earth distances, the thin current sheet is unstable to a rapidly growing linear ballooning instability (with very large wave number along $y$). On the basis of an asymptotic reduction of the full MHD equations using the ballooning-mode representation, Hurricane et al. (1997) suggest that such a linear instability will grow explosively in the nonlinear regime. Although these results are promising, more analytical and numerical work needs to be done with the full nonlinear and Hall MHD equations in realistic magnetotail geometry before one can claim that ballooning is indeed a mechanism for dipolarization and current diversion at substorm onset.

5. IMPULSIVE RECONNECTION IN CORONAL ARCADES: IMPLICATIONS FOR SOLAR FLARES

Solar flares often lead to the liberation of energies on the order of $10^{25}$ Joules or more over a timescale varying between 100 and 1000 s. Most present-day models of flares are based on the principle that they are powered by a sudden release
of magnetic energy stored in the corona. The continual emergence of new flux from the convection zone and the shuffling of the footpoints of closed field lines causes stresses to build up in the coronal field. Eventually, these stresses exceed a threshold beyond which a stable equilibrium cannot be maintained, and the field erupts. If the eruption is sufficiently powerful, then part of the coronal plasma will be ejected into interplanetary space as a coronal mass ejection (CME). If the eruption occurs in a region where there are strong magnetic fields (i.e., an active region containing sunspots), then bright flare emissions on the solar surface will be produced. The most powerful eruptions produce both a flare and a CME.

The motions of flare ribbons and loops provide clear evidence of magnetic reconnection in the solar atmosphere. Doppler-shift measurements show conclusively that these motions are not due to mass motions of the plasma within the ribbons and loops, but rather to the continual propagation of an energy source onto new field lines (e.g., Schmieder et al. 1987). Currently, reconnection is thought to be the most important mechanism to account for such a propagating source.

Although one can argue that the loop and ribbon motions are strong evidence of reconnection, they are, nevertheless, indirect evidence. X-ray observations of sufficient spatial resolution and sensitivity can, in principle, provide direct evidence. During the last few years, high-resolution images (such as Figure 10) obtained from the Hard X-ray Telescope and the Soft X-ray Telescope on Yohkoh show several features that are consistent with a reconnection site in the corona. These features include the following:

1. A hard X-ray source located above the soft X-ray loops (Sakao et al. 1992, Bentley et al. 1994)
2. Cusp features suggestive of either an X-type or a Y-type neutral line (Acton et al. 1992, Doschek et al. 1992, Tsuneta 1993)
3. Bright features at the top of the soft X-ray loops (Tsuneta 1993, McTiernan et al. 1993)
4. High temperature plasma along field line mapping to the tip of the cusp (Tsuneta 1993, 1996)

These observations have inspired a great deal of theoretical research. At present, there are about four different classes of flare models (Van Ballegooijen & Martens 1989, Low 1990, Forbes & Isenberg 1991, Mikic & Linker 1994, Antiochos et al. 1999, Lin & Forbes 2000). In nearly all these models, magnetic reconnection is invoked either directly or indirectly as a mechanism that mediates topological change and can trigger the conversion of an enormous reservoir of magnetic free energy into thermal and bulk kinetic energy.

A frequently observed feature of an impulsive solar flare is its sudden growth from a relatively quiescent background (see Hudson et al. 1994). Numerical simulations of flares are generally based on resistive MHD equations, employing either constant or anomalously enhanced resistivity. Although these simulations attempt to incorporate many of the physical effects relevant to a flare, they do not
appear to address the issue of the trigger, which is the focus of the present review. Observations of the impulsive phase impose a significant constraint on magnetic reconnection models: Not only is the timescale fast, but the time derivative of the growth rate increases suddenly.

The classical Lundquist number, $S$, of the solar corona is very high. To make estimates, we consider a preflare coronal loop with length $a \sim 10^9$ cm, magnetic field strength $B \sim 300$ G, density $n \sim 10^{10}$ cm$^{-3}$, and temperature $T \sim 2 \times 10^6$ K (Tandberg & Emslie 1988). (These numbers are also approximately consistent with numbers given in table 1.2 of the more recent monograph by Priest & Forbes 2000.) The classical Lundquist number is then given by $S = \tau_R/\tau_A = (4\pi a^2/\eta c^2)/(a/V_A) \sim 10^{14}$. Such plasmas can essentially be classified as “collisionless” and, as seen in the simulations discussed in Section 2, tend to develop thin and intense current sheets in the reconnection layer. When the thin current sheet width $\Delta$ falls in the collisionless range $d_e \equiv c/\omega_{pe} < \Delta \leq d_i \equiv c/\omega_{pi}$, it is not valid to neglect the collisionless terms in the generalized Ohm’s law. On the basis of the parameters given above, we estimate that the typical width for a Sweet-Parker current sheet is $\Delta \sim S^{-1/2} a \sim 10^6$ cm, and the electron and ion skin depths are given, respectively, by $d_e \sim 5$ cm and $d_i \sim 225$ cm. Because the inequality $d_e < \Delta < d_i$ is indeed satisfied in the corona, we conclude that the collisionless terms in the generalized Ohm’s law should not be neglected. [See table 1.2 of Priest & Forbes (2000), who use slightly different numbers and somewhat different arguments but arrive at exactly the same conclusion regarding the importance of the Hall current term.]

Because the Ohmic energy dissipation at such high values of the Lundquist number is too low, Spicer (1982) has considered a number of ways that anomalous resistivity owing to microinstabilities of the thin current sheets at multiple sites of reconnection can provide a mechanism for enhancing the dissipation rate. In such a scenario, the microinstabilities are triggered when the current sheets become thin enough, and the anomalous resistivity thus produced broadens the thin current sheet to evolve it into a state of marginal stability. In the process, energy dissipation owing to anomalous resistivity can be enhanced one to two orders of magnitude over the classical value. If the anomalous resistivity is turned on suddenly at the onset of the relevant microinstability and turns off when the instability becomes marginally stable, it can provide an explanation of the trigger if the resistivity is large enough to match the required reconnection rate. We do not question the plausibility of such a scenario, but we explore the alternative possibility that collisionless effects outside the scope of the simple Ohm’s law can play an important role and provide a viable mechanism for the trigger without requiring that the resistivity be anomalous.

Observations of chromospheric fibrils (Martin et al. 1985) and horizontal fields in the photosphere suggest that sheared fields play an important role in the occurrence of large flares. Consequently, considerable work has been done recently to build models that produce flares by shearing the footpoints of an arcade of magnetic loops (Steinolfson 1991, Mikic & Linker 1994). Footpoints on one side of
the arcade are moved parallel to the length of the arcade, while footpoints on the other side are moved in the antiparallel direction. Such shearing causes magnetic energy to be stored in the form of a force-free current flowing up one side of the arcade and down the other. As the shearing increases, the arcade expands upward. Theoretically, it is not yet established whether shearing of a simply connected arcade leads to instability or a loss of equilibrium. However, numerical simulations have demonstrated that a sheared arcade can erupt if rapid reconnection is allowed (Inhester et al. 1992, Mikic & Linker 1994). If a large amount of shear is applied and if the electrical resistivity of the plasma is sufficiently high, then reconnection leads to the formation and ejection of a magnetic flux rope. Because the resistivity used in the simulations is much larger than the value thought to occur in the corona, the question of whether an eruptive flare can be produced simply by shearing an arcade remains open.

To determine the role of photospheric footpoint shear in the context of Hall MHD, we will begin with the configuration (shown in Figure 11) that has been investigated extensively within the resistive MHD framework (Low & Wolfson 1988, Finn & Lau 1991, Vekstein & Priest 1992, Ma et al. 1995). The coronal loop lies in the $x$–$y$ plane, with the footpoints of the magnetic field lines intersecting the photosphere. The $z$-axis lies in the plane of the photosphere which is perpendicular to the plane of the paper. The contour lines represent the flux surfaces for a numerically generated vacuum-field solution of a coronal loop.

![Flux Function](image)

**Figure 11** Flux surfaces for a numerically generated force-free solution of a 2D coronal arcade (Ma et al. 1995).
In this 2D configuration, reconnection occurs near the separatrix (encompassing the $X$-point) when the motion of photospheric footpoints twists the vacuum field. At the bottom boundary that represents the photosphere, we impose shearing motion of the footpoints according to the relation,

$$v_z(x, y = 0) = v_0 \exp \left[ -\left( \frac{x - x_0}{\Delta x} \right)^2 \right],$$  

(12)

where $v_0$ is a small fraction of the Alfvén speed, $x = x_0$ is the point at which the outer separatrix intersects the $x$-axis in the simulation box (which is half of the physical domain), and $\Delta x$ characterizes the spatial extent of the sheared profile.

Using nonlinear resistive MHD equations, Wang & Bhattacharjee (1994) have shown that the current density $J_z$ at the $X$-point grows algebraically in time according to the relation

$$J_z \sim J_0 \eta^{-1/2} (t/\tau_{SP})^{3},$$

(13)

where $J_0$ is a constant independent of $\eta$ and $\tau_{SP} \equiv (\tau_A \tau_R)^{1/2}$ is the characteristic Sweet-Parker timescale. These analytical results agree well with our recent, high-resolution numerical simulations (Ma et al. 1995) based on the fully compressible resistive MHD equations. However, these results do not agree with an important generic feature of impulsive coronal dynamics: the sudden increase in the time derivative of the growth rate.

With the same initial and boundary conditions, we now present simulation results using the Hall MHD equations. As mentioned above, the Hall terms come into play when the thickness of the current sheet becomes of the order of or smaller than the ion skin depth. Before this time, magnetic reconnection is controlled by the resistivity. Figure 12 shows the time dependence of the maximum current sheet amplitude for the Hall MHD run with $d_i = 0.1$. Although the peak amplitude of

![Figure 12](image_url)
the Hall MHD current sheet is somewhat smaller than the resistive MHD current sheet, the former shows a much greater degree of impulsiveness. We remark that the nonmonotonic time evolution seen in the Hall MHD run can be explained as the outcome of two competing tendencies. On the one hand, the photospheric footpoint motion tends to build up the flux near the separatrix which reconnects impulsively under the influence of Hall MHD effects. On the other hand, owing to the buildup of the flux, the arcade tends to expand outward, and this outward motion tends to reduce the flux and the reconnection rate near the separatrix.

Figure 13 shows image plots of $J_y$ and $E_y$ for the resistive and Hall MHD runs at $t = 30\tau_A$. Two features stand out. First, the Hall MHD simulation shows a significant change in the geometry of the reconnection region by reducing the length of the Y-type current sheet seen in the resistive MHD simulation. Second, there is a separation of spatial scales between $J_y$ and $E_y$ in the Hall MHD simulation, absent in the resistive MHD simulation. This decoupling of spatial scales in Hall MHD is due to the decoupling of ion and electron motions.

The initial geometry used in this simulation is simple, and we do not claim that we have simulated a flare. Any model of magnetic reconnection that proposes to account for flare dynamics must show a convincing way to account for the energy liberated as well as the large number of energetic particles produced, and this has not been done. We have focused instead on the issues of fast timescale and trigger and demonstrated that reconnection models based on Hall MHD have the potential to account for these signatures.

6. SUMMARY

This review’s main emphasis has been on impulsive reconnection dynamics, which involve not only a fast timescale but also a trigger. Through a simple model of forced reconnection, we have demonstrated that fast and impulsive reconnection can be realized in resistive as well as Hall MHD. Although it is not necessary to invoke Hall MHD effects to realize impulsive reconnection dynamics, Hall MHD is the preferred system because in the high-Lundquist-number magnetosphere or solar corona, it is not justified to ignore the collisionless terms in the generalized Ohm’s law. In this review, we have discussed two examples where Hall MHD effects are important: substorms in the Earth’s magnetotail and flares in the solar corona. We have shown through analysis and simulation that the incorporation of collisionless, Hall MHD effects brings theory closer to observations in the sense that it provides a fast (i.e., sub-Alfvénic) reconnection timescale as well as a trigger. Hall MHD does not explain everything about these complex phenomena; in each case, we have pointed out some of the important questions that remain. However, the fact that two rather distinct phenomena in two different space environments can be viewed from a common perspective underscores the importance of Hall MHD effects at a fundamental level in reconnection dynamics in weakly collisional plasmas.

One of the principal limitations of the examples discussed in this review is that they are essentially 2D. As discussed in Section 3, even at this level of simplicity,
there are open questions regarding the maximum realizable reconnection rate and how it scales with plasma parameters. However, phenomena such as magnetotail substorms or solar flares are very likely 3D (i.e., without any ignorable spatial coordinates). Our understanding of 3D reconnection dynamics is in its infancy, and there are questions regarding geometry and dynamics that remain wide open. Resistive, Hall MHD, and PIC simulations are presently being developed and are beginning to produce results that need to be tested against a broad spectrum of observations from laboratory (including fusion), space, and astrophysical plasmas. Thanks to the confluence of several important laboratory experiments, space missions, and new developments in the theory and simulation of reconnection, we are likely in the next decade or so to arrive at definitive resolution of some of the important unresolved questions mentioned in this review.

ACKNOWLEDGMENTS

The author thanks his colleagues and collaborators at the Center for Magnetic Reconnection Studies: N. Bessho, B. Chandran, J. Dorelli, R. Fitzpatrick, T. Forbes, K. Germaschewski, T. Linde, Z. Ma, C. Ng, R. Rosner, X. Wang, and P. Zhu. This research is supported by the Department of Energy and the National Science Foundation.

LITERATURE CITED

Figure 3  Contour plots of the flux function $\psi$, the current density $J_z$, the electric field $E_z$, and the velocity $V_x$ in a resistive MHD simulation of forced reconnection driven by inward boundary flows.
Figure 5  Contour plots of the flux function $\psi$, the current density $J_z$, the electric field $E_z$, and the velocity $V_x$ in a Hall MHD simulation of forced reconnection driven by inward boundary flows (Ma & Bhattacharjee 1996).
Figure 8  Time evolution of the cross-tail current density during substorm evolution from a Hall MHD code (Ma & Bhattacharjee 1998).
Figure 10  An arcade of flare loops observed end-on by Soft X-ray Telescope on Yohkoh. The cusp-shaped form of the loop apexes is one of the key indicators that reconnection is occurring (Courtesy of ISAS).
Figure 13 Image plots of $J_y$ and $E_y$ for resistive and Hall MHD simulations with the same initial conditions at $30\,\tau_A$. 
# Annual Review of Astronomy and Astrophysics

## Volume 42, 2004

### CONTENTS

- **FRONTISPIECE, Adriaan Blaauw**  
  xii
- **MY CRUISE THROUGH THE WORLD OF ASTRONOMY,**  
  Adriaan Blaauw  
  1
- **ASTROPHYSICS WITH PRESOLAR STARDUST, Donald D. Clayton and Larry R. Nittler**  
  39
- **THE FIRST STARS, Volker Bromm and Richard B. Larson**  
  79
- **ISO SPECTROSCOPY OF GAS AND DUST: FROM MOLECULAR CLOUDS TO PROTOPLANETARY DISKS, Ewine F. van Dishoeck**  
  119
- **NEUTRON STAR COOLING, D.G. Yakovlev and C.J. Pethick**  
  169
- **INTERSTELLAR TURBULENCE I: OBSERVATIONS AND PROCESSES, Bruce G. Elmegreen and John Scalo**  
  211
- **INTERSTELLAR TURBULENCE II: IMPLICATIONS AND EFFECTS, John Scalo and Bruce G. Elmegreen**  
  275
- **GRS 1915+105 AND THE DISC-JET COUPLING IN ACCRETING BLACK HOLE SYSTEMS, Rob Fender and Tomaso Belloni**  
  317
- **IMPULSIVE MAGNETIC RECONNECTION IN THE EARTH’S MAGNETOTAIL AND THE SOLAR CORONA, A. Bhattacharjee**  
  365
- **ABUNDANCE VARIATIONS WITHIN GLOBULAR CLUSTERS, Raffaele Gratton, Christopher Sneden, and Eugenio Carretta**  
  385
- **DYNAMICS OF LUNAR FORMATION, Robin M. Canup**  
  441
- **EROS AND FAINT RED GALAXIES, Patrick J. McCarthy**  
  477
- **FINE STRUCTURE IN SUNSPOTS, John H. Thomas and Nigel O. Weiss**  
  517
- **PLANET FORMATION BY COAGULATION: A FOCUS ON URANUS AND NEPTUNE, Peter Goldreich, Yoram Lithwick, and Re’em Sari**  
  549
- **SECULAR EVOLUTION AND THE FORMATION OF PSEUDOBULGES IN DISK GALAXIES, John Kormendy and Robert C. Kennicutt, Jr.**  
  603
- **YOUNG STARS NEAR THE SUN, B. Zuckerman and Inseok Song**  
  685
CONTENTS

INDEXES

Subject Index 723
Cumulative Index of Contributing Authors, Volumes 31–42 749
Cumulative Index of Chapter Titles, Volumes 31–42 752

ERRATA

An online log of corrections to Annual Review of Astronomy and Astrophysics chapters may be found at http://astro.annualreviews.org/errata.shtml