18
SOLAR FLARE THEORY
Daniel S. Spicer and John C. Brown

INTRODUCTION

The theoretical interpretation of solar flares has, for many years, been one of the most hotly debated topics in solar physics. In addition, the current theoretical ideas advanced to explain solar flares are being used to help explain stellar flares (Mullan, 1977) and other astrophysical phenomena. However, a theoretical understanding of solar flares is confused by the multitude of models developed over the years to explain flare phenomenology. Until very recently, most of these models have been very qualitative in nature. Indeed, there exists no real model of a solar flare that can be used to compute the observed behavior of a flare, even given the initial conditions, boundary conditions, and pertinent parameters such as $B(x,t)$, where $B$ is the magnetic induction (these conditions and parameters are currently unknown). This situation has its origins in the complexity of the flare phenomenon, both observationally and theoretically, and the lack of the right kinds of data required to model such a situation. Nevertheless, there is a marked tendency by theoreticians, which is fully supported by observers, to develop “models” which contain many free parameters with which observations can be qualitatively compared. Practically all of these models agree qualitatively in some way with the observations. However, when firm predictions are requested, the models are unsatisfactory. They are unsatisfactory because the theoretical emphasis has been to develop rather general pictures of the flare process. Because of this unsatisfactory state of affairs, this review is restricted to a brief description of the models currently in vogue, with emphasis on the physical mechanisms generally invoked in the various models to explain the flare. This approach has many advantages in that it permits a discussion in some detail of the physics of each mechanism and their interrelations. It will also lead to a classification of mechanisms by their drivers, that is, the source of energy on which they feed. In turn, this yields a classification of the multitude of models according to mechanisms and, subsequently, according to drivers. This presentation of the mechanisms invoked to explain solar flares will also allow the interested reader to objectively evaluate these same mechanisms applied to coronal heating (see Wentzel, Chapter 14). This follows from the fact that all of the coronal heating mechanisms which use currents are, in reality, scaled down flare models; for example, the model of Rosner et al. (1978), using anomalous Joule heating, is just a modified Alfvén-Carlqvist model (Alfvén and Carlqvist, 1967; Smith and Priest, 1972; Spicer, 1981a), and the tearing mode model proposed by Galeev et al. (1981) is just the tearing mode flare model proposed by Spicer (1976, 1977a, and 1977b). Thus, all the strengths and weaknesses of a given mechanism used in flare theory can also be examined, when the same mechanisms are applied to coronal heating.

It is the opinion of the authors that this approach, which is radically different from those of previous reviews (e.g., Sweet, 1969), will prove more useful, because it clearly demonstrates, first, that each driver has a given number of possible mechanisms associated with it and, second, that each mechanism is model-independent with respect to the basic physics and predictions as long as there is a driver to excite the mechanism. In this way, the relevant question becomes not “which model is correct to explain

Provided by the NASA Astrophysics Data System
a flare?” but “which mechanism and driver are correct to explain a flare?” Thus, the confusing issue of evaluating models is reduced to the underlying simpler question of mechanisms. However, it must be noted that mechanisms are one stage further removed from making observationally verifiable predictions than are models.

The chapter proceeds as follows. In the second section, there is a discussion of the physical mechanisms important to flares, which, simultaneously, introduces various physical concepts useful for understanding the physics of the flare mechanism. In the third section, there is a critical examination of the question of flare model requirements and flare triggers, which also gives a brief summary of the flare models currently in vogue, which demonstrate how each mechanism can be used in the context of a model. In the fourth section a summary of these issues appears. This review is limited to the first acceleration phase of the flare, and no attempt is made to discuss the second acceleration phase.

**BASIC PHYSICAL PROCESSES**

In this section, the specific mechanisms are examined that are used in the various flare models reviewed in the third section. These mechanisms involve the dissipation of magnetic free energy, that is, of currents. The mechanisms will be divided into those driven by currents drifting parallel and those driven by currents drifting perpendicular to the attendant magnetic field. This will permit the development of a classification scheme for the various mechanisms and, therefore, of models. There are essentially three mechanisms presently proposed for the dissipation of magnetic free energy: (1) magnetic reconnection; (2) magnetic dissipation by (anomalous) Joule heating; and (3) double layers. All three of these mechanisms are considered in what follows.

**The Conservation of Magnetic Flux**

Consider first a number of useful definitions and concepts that will clarify the physics of the mechanisms to be discussed.

The ideal MHD approximation requires that magnetic flux be conserved; that is, the magnetic flux

$$\Phi_s = \frac{1}{c} \int_B \cdot dS$$  (18-1)

must satisfy the condition

$$\frac{d\Phi}{dt} = 0$$  (18-2)

if the flux is to be conserved, where the integral goes over the open surface $S$ (Figure 18-1) and $B$ denotes the magnetic field. This condition can only be satisfied if there exists no dissipative mechanisms present such as finite resistivity. To prove this, one needs only to expand Equation (18-2) using Faraday’s equation and the simplest form of Ohm’s law.
\[
\begin{align*}
J &= \frac{\mathbf{V} \times \mathbf{B}}{\eta} + \frac{\mathbf{E}}{c} \\
\eta &= \frac{\partial \mathbf{B}}{\partial t} - \frac{\partial \mathbf{E}}{\partial \mathbf{r}} = \frac{\eta c^2}{4\pi} \nabla^2 \mathbf{B},
\end{align*}
\]

where \( \eta \) has been taken to be constant. Taking

\[\nabla^2 \mathbf{B} \approx \frac{\mathbf{B}}{(\delta \xi)^2},\]

where \( \delta \xi \) is the scale length of the spatial variation of \( \mathbf{B} \), gives

\[\frac{\partial \mathbf{B}}{\partial t} = \frac{\eta c^2}{4\pi} \frac{\mathbf{B}}{(\delta \xi)^2},\]

which integrates to

\[\mathbf{B} = \mathbf{B}_0 e^{\xi^2/\tau},\]

where

\[\tau = \frac{4\pi(\delta \xi)^2}{\eta c^2},\]

which is the characteristic time for free (i.e., not externally driven) diffusion of the magnetic field into the plasma or vice versa, and is sometimes referred to as the resistive skin time.

The time, \( \tau \), can also be interpreted as the characteristic time for magnetic field dissipation. The field lines induce currents as they move through the plasma, and these induced currents result in Joule heating. This energy is transferred
from the field to the plasma, and the field weakens unless new flux replenishes that dissipated, that is, unless it is driven. The energy lost per cubic centimeter in time \( \tau \) is \( \eta J^2 \). Now

\[
|J| = c \frac{\nabla \times B}{4\pi} \approx \frac{cB}{4\pi \delta \ell},
\]

so that

\[
\eta J^2 \tau = \eta \frac{\left( \frac{B}{4\pi \delta \ell} \right)^2}{\eta c^2} = \frac{B^2}{4\pi}, \quad (18-9)
\]

or

\[
\tau = \frac{1}{2} \frac{B^2}{8\pi \eta J^2} . \quad (18-10)
\]

Thus, \( \tau \) is the time it takes for the magnetic field energy to be dissipated into Joule heat. Further, note that in terms of "lumped" circuit parameters, \( \tau = L/R \), where \( L \approx \ell_o/c^2 \) is the total inductance of the system, assumed constant in the present argument, \( R \approx \eta \ell_o/[\pi (\delta \ell)^2] \) is the total resistance, and \( \ell_o \) the length of the system.

Returning to Equation (18-8), it is apparent that there are basically two ways to decrease the dissipation time: either decrease the scale length of the spatial variation of \( B \) (steepen the field gradients), or increase the resistivity. In fact, both can occur together. As will be shown, reconnection mechanisms decrease \( \tau \) by driving the scale length for the magnetic field variation, \( \delta \ell \), to smaller values, while anomalous Joule heating mechanisms further decrease \( \tau \) by increasing the effective resistivity. The reduction of \( \tau \) is the key to understanding reconnection and anomalous Joule heating flare mechanisms.

**The Concept of \( \beta \)**

The concept of \( \beta \) and its relationship to plasma containment and to force-free magnetic fields are essential to understanding the numerous possible effects of solar magnetic fields. The parameter, \( \beta \), plays an important role in distinguishing between flare mechanisms that are driven by currents perpendicular to \( B(J) \) and those driven by currents parallel to \( B(U) \). Consider the static MHD equilibrium equations in the absence of gravity:

\[
\nabla \times B = \frac{J \times B}{c} \quad , \quad (18-11)
\]

\[
\nabla \cdot B = 0 \quad , \quad (18-13)
\]

It is trivial to show that

\[
\nabla \cdot B = 0 \quad , \quad (18-14)
\]

\[
\nabla \times B = \frac{J \times B}{c} \quad , \quad (18-12)
\]

and

\[
\nabla \times B = \frac{B \cdot \nabla B}{8\pi} \quad , \quad (18-15)
\]

\[
\nabla \cdot B = 0 \quad , \quad (18-13)
\]

It is trivial to show that

\[
\nabla \cdot B = 0 \quad , \quad (18-14)
\]

\[
\nabla \times B = \frac{B \cdot \nabla B}{8\pi} \quad , \quad (18-15)
\]

and

\[
J = \alpha(r) B \quad , \quad (18-16)
\]
where $J = J_{\parallel} + J_{\perp}$,

$$\beta = \frac{8\pi P}{B^2}, \quad (18-17)$$

and $\alpha(r)$ is an arbitrary scalar function. Equations (18-14) to (18-17) illustrate the importance of the quantity $\beta$. If $\beta \ll 1$, Equations (18-14) and (18-16) yield

$$\nabla \cdot B = \frac{B^2}{2}, \quad (18-18)$$

and

$$J \approx \alpha(r)B = J_{\parallel}, \quad (18-19)$$

or, ideally, a force-free magnetic field. It should now be evident that determining $\beta$ yields information about whether $J_{\parallel}, J_{\perp}$, or both are important in a particular problem. For example, coronal magnetic loops are generally thought to be, in part, force free in the corona, because mechanical body forces in the corona are not capable of sustaining appreciable Lorentz forces ($\beta \ll 1$) and thus, only $J_{\parallel}$ is of importance. However, the contrary is true in a neutral sheet where $B \to 0$ and thus, $\beta \to \infty$; here $J_{\perp}$ is the dominant component of $J$. Determining $\beta$ also allows us to say something about the local stability of a given magnetic field configuration, since in low $\beta$ systems those displacements that do not change the vacuum magnetic field (the so-called pressure driven interchange modes) are unimportant. In high $\beta$ systems, however, pressure gradients can drive local instabilities (see Ideal MHD Instabilities).

Energetically, a low $\beta$ configuration is preferable for solar flare mechanisms that dissipate magnetic free energy, with the exception of neutral sheet mechanisms. This follows because only a small fraction of the magnetic field needs to be dissipated in order to heat a plasma to high temperatures by dissipation of magnetic free energy. For example, suppose there exists a $\beta \approx 10^{-3}$ and it is possible to dissipate by some specified process, only 10 percent of the magnetic free energy, all of which goes into heating the plasma within some fixed volume. The temperature of the plasma can be increased by a factor of $\sim 100$ and the $\beta$ increased to 0.1 if the density remains fixed and the plasma will still be contained, as is observed in small flare loops (Widing and Spicer, 1981). Note also that the characteristic time for Joule heating $\tau_{JH}$ to raise $T$, where $\tau_{JH} \approx NkT/r_{\perp}2 \approx 2\beta r$, is shorter the smaller the $\beta$ for a fixed $J$, where Equation (18-8) and Ampere’s law have been used.

The Origins of Resistivity and Electron Runaways

Consider the equation of motion of a test particle of charge $q = -e$ and mass $m_e$ drifting with a velocity $V_D$ with respect to a stationary ion background. This is

$$m_e \frac{dV_D}{dt} = -eE - m_e V_D \nu(V), \quad (18-20)$$

where $E$ is the macroscopic electric field, taken to be constant, and $\nu(V)$ is the collision frequency given by (Spitzer, 1962),

$$\nu(V) = \frac{4\pi N_e e^4 \ln \Lambda}{m_e^2 r^3}, \quad (18-21)$$

where $N_e$ is the background electron density, $V$ the particle’s net rms velocity, and $\ln \Lambda = \ln(4\pi N_e \lambda^3_D)$, with $\lambda_D$ the electron Debye
length. In the absence of collisions, the electrons are freely accelerated relative to the ions so that

\[ V_D = -\frac{eEt}{m_e} \]  \hspace{1cm} (18-22)

When the electrons suffer collisions with the more massive ions or with slow moving field electrons, a steady state can be achieved, such that,

\[ V_D = -\frac{eE}{m_e\nu(V)} \]  \hspace{1cm} (18-23)

The current density is related to the electron drift by \( J = -N_e eV_D \). Using this relation and Equation (18-23) yields

\[ J = \frac{N_e e^2 E}{m_e \nu(V)} = -\eta E \]  \hspace{1cm} (18-24)

where \( \eta \) is defined as the electrical resistivity given by

\[ \eta = \frac{4\pi\nu(V)}{\omega_{pe}^2} \]  \hspace{1cm} (18-25)

where the plasma frequency \( \omega_{pe}^2 = 4\pi N_e e^2/m_e \).

If the drift velocity of the electrons is less than the electron thermal velocity, \( V = (kT_e/m_e)^{1/2} \), then \( V = (V_D^2 + V^2)^{1/2} \approx V \). So that Equation (18-21) reduces to

\[ \nu = \frac{4\pi N_e e^4 \ln \Lambda}{m_e^{5/4} (kT_e)^{3/2}} \]  \hspace{1cm} (18-26)

which is the well-known Spitzer result (Spitzer, 1962). Using \( \lambda_{De} = V_{Te}/\omega_{pe} \), Equation (18-26) can be cast into the form

\[ \nu = \frac{\ln \Lambda}{\omega_{pe}} \approx \frac{\omega_{pe}}{\Lambda} \]  \hspace{1cm} (18-27)

to be used in the discussion below.

Notice that the drag force on the electrons \( (F_{De} = m_e V_D \nu \approx V_D^3 / V_{Te}) \) in the region \( V_D < V_{Te} \) increases with \( V_D \), implying that a steady state can be achieved according to Equation (18-23), while in the regime \( V_D > V_{Te} \), the drag force decreases as \( V^{-2} \), implying that a steady state cannot be achieved. The following physical picture then emerges. In the low velocity regime \( V_D < V_{Te} \), where the drag force is greater than the electric force, the electron motion is essentially random, and a steady state can be achieved. However, as the dynamic friction becomes weaker at higher velocities, there is a critical velocity beyond which electrons will be accelerated faster than collisions can drag them. As a result, they gain more energy, while, at the same time, their friction is reduced still further as they reach these higher velocities, and so on. Eventually, the friction becomes negligible so that the electrons are freely accelerated by the dominant electric force until some other energy loss and momentum loss mechanisms become dominant, such as radiation, or a finite spatial extension of the applied electric field, both of which will appear as a cutoff in the distribution function at higher energies. These freely accelerated electrons are the so-called runaway electrons. The critical velocity, \( V_c \), at which they start to run away, may be obtained by balancing the two opposing forces to yield

\[ V_c = \frac{4\pi N_e e^3 \ln \Lambda}{m_e E} \]  \hspace{1cm} (18-28)

Defining a quantity \( E_D \), the Dreicer electric field
(Dreicer, 1959), at which runaway occurs for a thermal electron, for which \( V_e = V_{Te} \), yields

\[
E_D = \frac{4\pi N_e e^3 \ln \Lambda}{m_e V_{Te}^2} = \frac{e \ln \Lambda}{\lambda^2_{De}}.
\] (18-29)

Physically, this is the electric field at which the energy picked up in one collision time is equal to the thermal energy. Further, using Equations (18-28) and (18-29), it is evident that runaway occurs for any electron with a velocity that satisfies

\[
V \geq V_r = \left( \frac{E_D}{E} \right)^{\frac{1}{6}} V_{Te}.
\] (18-30)

So far, the motion of single particles has been treated. However, the behavior of the bulk electron distribution changes remarkably in the presence of an external electric field, as a result of the reduction of \( n(V) \) with higher velocities. In the collision-dominated portion of the distribution, one would expect, intuitively, that the electrons would have a slightly skewed Maxwellian distribution and that they would be drifting relative to the ions with a drift velocity given by Equation (18-23). However, at velocities much greater than \( V_r \), electrons pick up more momentum from the electric field than they lose by momentum exchange with the ions (or field electrons) through collisions. This results in the electrons moving in the antiparallel direction of the electric field; the distribution therefore develops a long tail concentrated on an axis parallel to the electric field. The whole electron distribution then appears as a skewed drifting Maxwellian bulk with a very long and highly anisotropic tail antiparallel to \( E \). This anisotropic drifting distribution represents an excess of free energy, which can excite various collective micro-instabilities, which may, in turn, inhibit the extent of the tail.

If the drift velocity of the skewed Maxwellian has a magnitude such that

\[
(V_{Ti}, C_s) \leq V_D \leq V_{Te}
\] (18-31)

(see Table 18-1), it can excite various current driven collective microinstabilities, where \( C_s = (kT_e/m_i)^{\frac{1}{6}} \) is the ion sound velocity and \( V_{Ti} \) is the ion thermal velocity. That is, instabilities are driven by the bulk of the current carrying electrons, as opposed to just the tail of the distribution, which may possess a bump. When this occurs the phenomenon called anomalous resistivity occurs as a result of charge clumping caused by a microinstability. The plasma can be unstable to the generation of various types of waves (the source of clumping), notably electrostatic waves, which are normal modes of the plasma and grow at the expense of the free energy associated with the drift energy of the electrons. The scattering of the drifting electrons by these turbulent wave electric fields will show up as an additional momentum and energy loss, hence, the term “anomalous resistivity.”

To illustrate physically the relationship between classical collisions and enhanced scattering caused by various collective effects, it is useful to return to Equation (18-27) which relates the classical collision frequency with the plasma frequency and the quantity \( \Lambda = 4\pi N_e \lambda^3_{De} \), that is, number of electrons in a Debye sphere. The quantity, \( \Lambda \), is also a measure of the ratio of the electric field energy density in thermal fluctuations \( <\delta E^2>/8\pi \), to the thermal energy density \( N_e kT_e \), that is,

\[
\frac{<\delta E^2>}{8\pi N_e kT_e} \approx \frac{1}{\Lambda}.
\] (18-32)

(Krall and Trivelpiece, 1973), hence, Equation (18-27) becomes
<table>
<thead>
<tr>
<th>Name</th>
<th>Drift Velocity Turn on Conditions</th>
<th>Field Gradient Scale Length Turn on Conditions</th>
<th>Effective Collision Frequency</th>
<th>Growth Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buneman</td>
<td>$V_C \geq V_{Te}$</td>
<td>$\delta \xi \leq $</td>
<td>$\nu_{eff} \approx \left(\frac{m_e}{m_i}\right)^{1/3} \frac{1}{\omega_{pe}}$</td>
<td>$\gamma \approx \frac{\sqrt{3}}{2^{4/3}} \left(\frac{m_e}{m_i}\right)^{1/3} \omega_{pe}$</td>
</tr>
<tr>
<td>Ion Acoustic</td>
<td>$V_D \geq c_s \left[1 + \left(\frac{m_i}{m_e}\right)^{\delta} \exp(-1.5-0.5\xi)\right]$</td>
<td>$\delta \xi \leq \frac{c}{\omega_{pe} \left(\frac{1}{\beta_{pe}}\right)^{1/3}} \times \frac{1}{\omega_{pe}}$</td>
<td>$\nu_{eff} \approx \left(\frac{m_e}{m_i}\right)^{1/3} \omega_{pe}$</td>
<td>$\gamma \approx \frac{m_e^{1/3}}{m_i} \omega_{pl}$</td>
</tr>
<tr>
<td>$\xi \equiv \frac{T_e}{T_i} \gg 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electric Ion Cyclotron</td>
<td>$V_D \geq c_s$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \beta_{pe} \equiv \frac{8\pi n_e kT_e}{\delta B^2} \]
Table 18-1 (continued)

Electrostatic Instabilities Driven Perpendicular to B

<table>
<thead>
<tr>
<th>Name</th>
<th>Drift Velocity Turn on Conditions</th>
<th>Effective Collision Frequency</th>
<th>Growth Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified</td>
<td>$V_D &gt; \left( \frac{m_e}{m_i} \right)^{\frac{3}{4}} V_T$</td>
<td>$\nu_{\text{eff}} \approx \frac{0.14 \omega_{pi}}{\Omega_{LH}} \left( \frac{k_{\parallel}}{k_\perp} \right)^{2/3}$, if $\frac{k_{\parallel}}{k_\perp} \leq 1$</td>
<td>$\gamma_{\text{Max}} \approx \Omega_{LH}$</td>
</tr>
<tr>
<td>Two Stream</td>
<td></td>
<td>$\nu_{\text{eff}} \approx \frac{0.14 \omega_{pi}^2}{\Omega_{LH} k_{\perp}} \left( \frac{k_{\parallel}}{k_\perp} \right)^{\frac{3}{4}}$, if $\frac{k_{\parallel}}{k_\perp} &gt; 1$</td>
<td>$k_{\perp \text{Max}} \approx \Omega_{LH}/V_D$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\Omega_{LH} = \frac{\omega_{pi}}{\left( \frac{\omega_{pe}}{1 + \Omega_{LH}} \right)^{\frac{3}{4}}}$</td>
<td>$k \cdot B \approx 0$</td>
</tr>
<tr>
<td>Lower Hybrid Drift</td>
<td>$\frac{m_e}{m_i} V_T \leq V_D \leq \left( \frac{m_e}{m_i} \right)^{3/4} V_T$</td>
<td>$\nu_{\text{eff}} \approx \frac{\pi}{8} \frac{m_1}{m_e} \frac{1}{\Omega_{LH}} \frac{\langle \delta E^2 \rangle}{8\pi NkT_i}$</td>
<td>$\gamma_{\text{Max}} \approx \Omega_{LH} \cdot k \cdot B = 0$</td>
</tr>
<tr>
<td>Drift</td>
<td></td>
<td>$\langle \delta E^2 \rangle \leq \frac{1}{8\pi} \frac{m_e N e V_D^2}{8\pi \rho_e}$</td>
<td>$k_{\perp \text{Max}} \approx \sqrt{\frac{T_e}{\rho_e}} \left( \frac{T_i}{T_e} \right)^{\frac{3}{4}}$</td>
</tr>
</tbody>
</table>

| ion acoustic  | see $J_{\parallel}$               |                                |                                |
| Buneman       | see $J_{\parallel}$               |                                |                                |
\[ \nu \approx \omega_{pe} \frac{\langle \delta E^2 \rangle}{8\pi n_e kT_e} . \] (18-33)

The form of Equation (18-33) correctly suggests that, if it were possible to increase the magnitude of electric field fluctuations through which the drifting electrons are scattered, it would be possible to increase the collision frequency over that of a thermal plasma and, therefore, modify the transport coefficients, for example, the resistivity. In fact, there are a number of microinstabilities that can result in enhanced electric field fluctuations which are orders of magnitude greater than their thermal level. However, just any high frequency microinstability is not sufficient to cause an increased resistivity. This follows because, for the bulk of the drift electrons to experience friction, the phase velocity of the waves produced by some microinstability must be small; that is,

\[ V_p = \frac{\omega}{|k|} \ll (V_{Te}, V_D), \]

which implies, in turn, that the frequency of the waves must satisfy \( \omega \ll \omega_{pe} \). Physically, this has two causes: (1) resistivity is not a resonance phenomenon between wave and particle, and (2) if \( V_p \gg (V_{Te}, V_D) \) the electrons and clumps of charge caused by collective effects would move together with little or no momentum exchange. Hence, \( V_p \ll (V_{Te}, V_D) \) implies that the electrons see a fixed scattering center and that the momentum exchange will be large. Also note that the wave frequency cannot get too low, since the electrons with their small inertia will not even experience the waves. For example, Alfvén waves cannot, of themselves, cause anomalous resistivity.

In Table 18-1 are listed the various known instabilities that are capable of producing anomalous resistivity. They are separated into those that are driven by currents perpendicular to and by currents parallel to the magnetic field. This split in the source or driving current is useful for a number of reasons. For example, it facilitates a recognition of which instabilities might be important in a force-free loop (those driven by \( J_\parallel \)) and which instabilities that might be important in a neutral sheet (those driven by \( J_\perp \)). The reader should notice that the threshold conditions on the drift velocities required to excite these instabilities imply very steep magnetic field gradients. These gradient scale lengths are also included in Table 18-1.

As is apparent from Table 18-1, the growth rates for these instabilities are huge compared with the characteristic growth time of overall flare energy release. Furthermore, detailed analytic and numerical studies of these various instabilities show that they saturate in a few growth times, that is, in a time many orders of magnitude shorter than the characteristic time scales of a flare. How then can one really believe that such instabilities play a role in the flare process? The answer is that microinstabilities can be important in flares for long periods in a steady state only if some transient driving mechanism, external to the instability, continuously drives the system toward instability for a time characteristic of a flare. This marginal stability hypothesis, as it is sometimes called, has led to several accurate predictions in laboratory phenomena (Manheimer and Boris, 1977). Since this hypothesis will be invoked through the remainder of this review, it is appropriate to illustrate it here, using the ion acoustic instability listed in Table 18-1. Following that, the question of approach to marginal stability is considered.

Suppose an electric field is applied to a plasma so that a current forms, with a drift velocity on the order of the ion sound velocity, \( C_s \). Suppose further that \( T_e/T_i \gg 1 \), so that (Table 18-1) the conditions are satisfied for excitation of the ion acoustic instability. The marginal stability of the ion acoustic instability implies that the turbulent field fluctuation level adjusts to give an electron wave scattering frequency \( \nu_{eff} \) for which

\[ C_s = \frac{eE}{m_e \nu_{eff}} . \] (18-34)
Hence, if $\nu_{\text{eff}}$ and $E$ are known, then $C_1$ and, thus, the electron temperature, $T_e$, of the plasma can be calculated.

Envision now how marginal stability is approached, following the arguments of Manheimer and Boris (1977). Suppose the magnitude of the external driving mechanism forcing the plasma toward instability is gradually raised. For a very weak electric field, $\nu_{\text{eff}}$ will be determined classically. As $E$ is increased toward the marginal stability point, the level of fluctuations starts to increase, increasing $\nu_{\text{eff}}$. As $E$ increases still further, $\nu_{\text{eff}}$ will increase until the instability saturates, giving a maximum $\nu_{\text{eff}}$. What these arguments show is that if the system is driven hard, until the instability saturates, the instability approaches marginal stability from above, as opposed to the case in which it is driven weakly, so that the instability approaches marginal stability from below. This shows that one will obtain different values of $\nu_{\text{eff}}$ depending on how the system approaches marginal stability.

**Particle Acceleration by Electric Fields in the Presence of Anomalous Resistivity Microturbulence**

On first examination, it might be expected that the microturbulence which results in anomalous resistivity would inhibit efficient particle acceleration by electric fields. However, as was noted in the previous section, the turbulence which causes anomalous resistivity must necessarily have low phase velocities. Thus, an electron accelerated by an electric field sees essentially a fixed scattering center. Electrons with velocities greater than the phase velocity of the turbulent waves will experience dynamic friction that decreases as their velocity becomes greater. It should therefore be expected that the scattering frequency of the waves would decrease with higher electron velocities just as in the Coulomb case Equation (18-21). Indeed, this is exactly what occurs, and the scattering frequency for ion acoustic waves scales as

$$\nu^*_{\text{eff}} = \frac{\nu_{\text{eff}}}{V} \left( \frac{V_T}{V} \right)^3,$$  \hspace{1cm} (18-35)

for particles in the electron distribution with velocities $V \gtrsim V_T$ (Kaplan and Tsytovich, 1973). Hence, electron runaway can still occur in the presence of the turbulence; the only difference is that the classical Dreicer field for runaway is replaced by

$$E_{\text{D}}^* = \frac{m_e V T_e \nu_{\text{eff}}}{e},$$ \hspace{1cm} (18-36)

and electrons with velocities

$$V_R \geq \left( \frac{E_{\text{D}}^*}{E} \right)^{\frac{1}{2}} \left( \frac{k T_e}{m_e} \right)^{\frac{1}{2}},$$ \hspace{1cm} (18-37)

will runaway, where $T_e$ is the electron temperature of the plasma at marginal stability (Spicer, 1981a). An estimate of the production rate of runaway electrons is given by Kruskal and Bernstein (1964):

$$\gamma_0 = 0.35 \nu_{\text{eff}} \left( \frac{E_{\text{D}}^*}{E} \right)^{3/8},$$ \hspace{1cm} (18-38)

$$\exp \left\{ - \left( \frac{2E_{\text{D}}^*}{E} \right)^{\frac{1}{2}} - \frac{E_{\text{D}}^*}{4E} \right\},$$

where the electron-ion classical frequency, $\nu_{\text{ei}}$, has been replaced with $\nu_{\text{eff}}$ and the classical Dreicer field, $E_{\text{D}}$, has been replaced with $E_{\text{D}}^*$. The total number of electrons to runaway in a time $\Delta t$, during which the driving electric field is applied, is just

$$N_T = \int \int_{\Delta V \Delta t} N_e \gamma_0 x^2 dxdt \approx N_e \gamma_0 \Delta V \Delta t,$$ \hspace{1cm} (18-39)
where $\Delta V$ and $\Delta t$ represent the incremental volume and time, respectively, in which and during which the electric field is applied. In other words, $\gamma_0 \Delta t$ undergoes runaway in any volume $\Delta V$ and time $\Delta t$.

If the Buneman instability is excited by an electric field which satisfies $E \geq m_e \omega_{pe} C_s/e$, a driven two-stream instability can occur when the electric field is applied for durations greater than a few growth periods (see e.g., Spicer, 1977b, and references therein). The physics is straightforward. The electric field accelerates the electrons up to $V_D = V_{Te}^\gamma$, the system goes unstable to the Buneman instability, and the instability takes energy from the current and heats the electrons, making $V_{Te} > V_D$ again, thereby quenching the instability. The applied electric field, however, continues to accelerate the electrons and the process repeats. This model has been confirmed by detailed particle simulations (Boris et al., 1970; Morse and Nielson, 1971), from quasilinear theory (Hamasaki et al., 1977), and also by laboratory experiments (Hirose and Skarsgard, 1976). The result of this whole process is bulk electron heating combined with bulk electron runaway, that is, acceleration at some fraction of the free streaming velocity. The total number of runaways can be approximated from Equations (18-38) and (18-39), with $v_{eff}$ for the Buneman instability obtained from Table 18-1, replacing that of the ion acoustic instability.

It should be noted that Equation (18-38) is not valid for $E \geq E_D$, and therefore some other means of estimating the rate of runaway production by $E$ is needed. When $E > E_D$, the acceleration is essentially collisionless so that an electron will be freely accelerated through a distance, $D$, in which the voltage is dropped, in a time given by

$$\gamma_R^1 = \left( \frac{Dm_e}{eE} \right)^{1/2} \quad \text{(18-40)}$$

Hence,

$$N_T = \gamma_R N_e \Delta V \Delta t \quad \text{(18-41)}$$

The above results will be used in the section, Flare Models.

One very important point associated with particle acceleration and anomalous resistivity should be emphasized: if the current density is high enough to cause the current to become unstable to the various instabilities that can cause anomalous resistivity it must have a driving electric field which is also large enough to cause particles to be accelerated. However, the total number in the tail of the distribution will not be much greater than that found in a skewed Maxwellian. Conversely, if the field can accelerate large numbers of particles, the drift current must cause anomalous Joule heating.

Ideal MHD Instabilities

This section elucidates ideal MHD instabilities relevant to solar flare theory, with the aim of introducing various concepts associated with these instabilities that will allow one to differentiate them from resistive MHD instabilities and microscopic instabilities.

There are basically two sources of driving energy for MHD instabilities—ideal and dissipative. Currents perpendicular to $B$ caused by pressure gradients are responsible for driving the so-called interchange instabilities. These instabilities cause one portion of the plasma to exchange places with another portion. Since they depend only on the local conditions near the line of force, they are therefore called local instabilities (or modes) as opposed to global instabilities (or modes). As a general rule, interchange instabilities, being local, do not necessarily imply global instability, and so some level of local instability is tolerable and usually appears as turbulence, for example, in solar features such as prominences (Spicer, 1979a). The second source of driving energy is the current parallel to $B$. Macroscopic instabilities driven by $J_\parallel$ are spread out over the plasma volume. They are generically called kinks, or helical instabilities. Kinks can be further subdivided into free-boundary kinks (external kinks) and internal kinks. The external kink involves motions of the entire plasma magnetic field configuration. The erupting prominence
appears to be a classical example of a free-boundary kink (Sakurai, 1976; Spicer, 1976, 1979a). On the other hand, internal kinks involve distortions and motions within the plasma magnetic field configuration (PMFC) and are not necessarily visible to an external observer. As will be noted later, internal kinks have lower thresholds than external kinks, but external kinks are more violent.

As a general rule, both ideal and dissipative MHD instabilities occur when a perturbation to a PMFC does not bend or stretch the magnetic field lines. Such perturbations, thus, do not result in magnetic restoring forces, which are necessary to restore the perturbed equilibrium, and the perturbation continues to grow. It is necessary to identify more precisely those effects which are capable of causing instability and those which are capable of stabilizing.* This is achieved by utilizing the energy principle (Bernstein et al., 1958) and examining the potential energy $\delta W$, dividing $\delta W$ into positive and negative parts so that the stabilizing and destabilizing terms can be identified. $\delta W$ can be written in the form (Furth et al., 1966)

$$\delta W = \delta W_F + \delta W_V + \delta W_S , \quad (18-42)$$

where $\delta W_F$ is the change in potential energy resulting from the perturbation of the plasma given by

$$\delta W_F = \frac{1}{2} \int_{\text{plasma}} d^3x \left\{ \frac{\delta B^2}{4\pi} - \frac{B_0 \xi \cdot \nabla P_0}{4\pi} + \frac{4\pi}{B_0^2} \right\}^2$$

(Alfvén)

$\delta W_V$ is the change in potential energy resulting from the perturbation of any vacuum magnetic field, given by

$$\delta W_V = \frac{1}{2} \int d^3x \frac{|\delta B^2|}{4\pi} , \quad (18-44)$$

and $\delta W_S$ is the change in potential energy associated with any surface currents present, given by

$$\delta W_S = \frac{1}{2} \int dS \cdot \left[ \frac{P_0 + \frac{B^2}{8\pi}}{8\pi} \right] (\kappa \cdot \xi)^2 , \quad (18-45)$$

where $\xi$ is the fluid displacement, $\gamma$ is the ratio of the specific heats, $n$ is a unit vector normal to the equilibrium magnetic surface, $\delta B$ is the perturbed magnetic field, and

$$\kappa = \frac{1}{2B^2} \left[ B_0 \times \nabla \left( \delta P_0 + \frac{B_0^2}{8\pi} \right) \right] \times B_0 \quad (18-46)$$

is the curvature of the magnetic field. The vacuum potential energy change $\delta W_V$ is always

*The reader should note that the arguments to be given are geometry-independent, that is, kinks can occur in any geometry, although with differing growth rates and physical manifestations.
stabilizing, and $\delta W_f$ will vanish if no surface currents exist. The first three terms in $\delta W_f$ are stabilizing, while the last two are destabilizing. The first of the two destabilizing terms results from currents flowing parallel to $B$, and the second destabilizing term results from the interaction of the pressure gradient with the field curvature. These latter two terms arise from $J_{\|}$ and $J_{\perp}$, respectively. In solar PMFC $\beta$ is generally believed to be much less than unity (except in neutral sheets where $B \to 0$); hence, $J_{\perp} \approx 0$, and one should therefore expect only $J_{\|}$ to play an important role in the energetics of such configurations, although local weak turbulence caused by $J_{\perp}$ may be present (Spicer, 1979a). However, just the opposite is true inside neutral sheets where $\beta \to \infty$, since $B \to 0$, so that $J_{\perp}$ becomes very important; hence, interchanges should be expected in neutral sheets if finite curvature effects exist. Indeed, Uchida and Sakurai (1977) have recently argued that such interchanges might play an important role in the energetics of neutral sheet reconnection during a flare.

The first and second stabilizing terms arise because energy is required to stretch and shift lines of forces if the direction of the magnetic field is changed by the perturbation. Contained within this term is the global magnetic shear, the average shear over the entire magnetic surface, and the local shear, the amount that a field line must be stretched if it is to exactly replace a neighboring field line in the course of the perturbation (Ware, 1965). Magnetic shear stabilizes, because in a sheared magnetic field the direction of the magnetic field changes its direction with position; hence, it becomes very difficult to replace a field line with another neighboring line which is at an angle with respect to another, unless the line is bent and/or stretched. Since bending and stretching requires work be done on the field, $\delta W_f$ increases rather than decreases; hence, shear is stabilizing. It should be emphasized, at this point, that shear is stabilizing in the ideal MHD limit because flux is conserved; that is, field lines cannot be broken. However, dissipation effects such as resistivity allow the lines of force to break and reconnect. Thus, shear in the presence of dissipation is not as effective a stabilizing influence as it is in the ideal MHD approximation.

Equation (18.43) contains two compression terms, $\gamma P_0 (\nabla \cdot \xi)^2$ and $(B_0 \xi \cdot \nabla P_0) / B_0^2$, both of which are stabilizing. This follows because they are a measure of the net energy absorbed by the PMFC in compressing the magnetic field and the plasma. As with stretching, a finite amount of compression is necessary if one field line is to exactly replace another. Notice also that both terms become ineffective in the limit of $\beta \to 0$.

The destabilizing term,

$$J \cdot B_0 \left( \frac{B_0 \times \xi}{B_0^2} \right) \delta B,$$

is the term responsible for driving kink instabilities (Vosla and Callebaut, 1962; Green and Johnson, 1962; Raadu, 1972) in force-free fields by means of forces resulting from the interaction of the current parallel to $B_0$ with $\delta B$. Energy is released by lowering the net current along the magnetic field. The constraint that the magnetic flux within a given flux surface be conserved is satisfied by bending and stretching the field lines into a helical or screw shape (Kruskal and Kulsrud, 1958). The decrease in the magnitude of $B$ inside the flux surface is balanced by an increase in the cross section of the bounding surface, which is why a prominence expands (Sakurai, 1976; Spicer, 1979a).

The term $2 \xi \cdot \nabla P_0 \cdot \kappa$ in Equation (18.13) is related to the curvature and, thus, to the tension of the lines of forces and is responsible for driving the interchange instability.* This tension results in a force which is proportional to $B^2$, so that work must be done to move lines of force against this tension.

One additional stabilizing effect not obvious from the discussion so far is line tying. In the ideal MHD limit, line tying results when one

*The Rayleigh-Taylor instability is essentially an interchange instability, with gravitational acceleration $g$ as the driving term instead of $\kappa$. It arises as a result of a coupling between the pressure gradient and $g$. 

426
insists that the finite length field lines are rigidly tied to the end boundaries. Hence, the only way field lines can be interchanged is by twisting the field lines by rotating one or both boundaries. However, this increases rather than decreases the total energy of the field, so line tying represents a stabilizing effect. Line tying is included in $\delta W$ when minimizing $\delta W$ with respect to $\xi$ by insisting that the component of $\xi$ perpendicular to $B$ vanish at the end boundaries. This gives rise to a positive definite stabilizing term which varies inversely as $L_0^{-1}$, where $L_0$ is the distance between end boundaries (Solovev, 1971; Raadu, 1972; Spicer, 1976; Hood and Priest, 1979). Hence, the larger is $L_0$ the less stabilizing is the effect.

Unfortunately, up to the present, the stability of solar PMFC's has not been treated correctly. To illustrate one fundamental error that has been made in studying the stability of solar magnetic loops, consider the linear diffuse pinch characterized by cylindrical symmetry with the pressure and magnetic field being functions of the radius only. Newcomb (1960) has reduced the ideal MHD stability problem to the study of the ordinary Sturm-Liouville equation,

$$\frac{d}{dr} \left( f \frac{d\xi}{dr} \right) - g \xi = 0 .$$  \hspace{1cm} (18-47)

The functions $f$ and $g$ (not the gravitational acceleration here) depend on the radius through their dependence on equilibrium quantities, and $f$ is singular at various radii at which the perturbation is constant along an equilibrium field. The displacement $\xi$ must satisfy the boundary condition

$$\left. \left( \frac{d\xi}{dr} \right) \right|_{a}^{b} = 0 .$$  \hspace{1cm} (18-48)

Equation (18-48) can be satisfied in a variety of ways. In laboratory devices $\xi(b) = 0$ is required, because a rigid immovable conducting wall exists at a radius $b = R$. However, no such wall exists around solar loops, so that $b = \infty$. Despite this fundamental difference, the boundary condition $b = R$ is the boundary condition that is presently employed in the analysis of loop stability (Anzer, 1968; Raadu, 1972; Van Hoven et al., 1977). In fact, such boundary conditions eliminate free boundary kinks (Shafranov, 1970) and permit only internal kinks, which are not necessarily observable. In addition, the $m = 1$ kink requires that $d\xi/dr$ vanish at $r = 0$ (Newcomb, 1960), and again the analysis of the previous authors has not included this effect, thereby eliminating, a priori, the $m = 1$ kink mode, both internal and external. On the other hand, Sakurai (1976) has treated the boundary conditions properly for prominences, as his illustrations show (Figure 18-2). This point is most important, because the internal kink has a much lower threshold than the external kink. The external kink requires roughly

$$B_\theta^2 \ln \left( \frac{L_0}{a} \right) > B_z^2 ;$$  \hspace{1cm} (18-49)

whereas the internal kink requires (for periodic boundary conditions in $z$ and homogeneity in $\theta$ and $z$)

$$B_\theta > \frac{2\pi a B_z}{L} .$$  \hspace{1cm} (18-50)

Figure 18-2. The development of an external $m = 1$ ideal MHD kink instability in the magnetic field lines (Sakurai, 1976).
where $a$ is the radius of the pinch, $B_\theta$ is the component of $\mathbf{B}$ induced by $J_z$, and $B_z$ is the axial component of $\mathbf{B}$ (see e.g., Hasegawa, 1975).

It would be useful to discuss the kink instability further before proceeding to reconnection, since it is driven by $J_\parallel$ and, as was discussed in The Origins of Resistivity and Electron Runaways, there are microscopic instabilities also driven by currents parallel to $\mathbf{B}$. As shown already, kinks are effectively incompressible perturbations, since $P_0$ does not appear in the driving term; that is, they would exist with or without a finite $\beta$. Since they are incompressible, the kink driving force must provide a torque in the $z$ direction, i.e.,

$$ T_z = \nabla \times \left( \frac{J \times \mathbf{B}}{-\nabla P + \frac{J \cdot \nabla J}{c}} \right) = \frac{(\mathbf{B} \cdot \nabla J - J \cdot \nabla \mathbf{B})_z}{c} . $$

(18-51)

A linearization of Equation (18-51) for perturbations of the form $\xi = \xi(r) \exp[i(k_\parallel r + k_z z)]$ yields

$$ T'_z = \delta J_{z0} \frac{\partial J_{z0}}{\partial r} + ik_\parallel |B_0| \delta J_z , $$

(18-52)

where $\partial J_{z0}/\partial r$ provides the driving force. Generally, the second term is negative and provides a stabilizing force, except when $k_\parallel = (k \cdot B_0)/B_0^2 = 0$, that is, when the perturbation is constant along the equilibrium field, these points correspond to the singular points in $f$ of Equation (18-47). Thus, depending on the magnitude of $\partial J_{z0}/\partial r$ at the point where $k \cdot B_0 = 0$, instability may or may not result. It follows from this that a knowledge of the equilibrium current density profile is of crucial importance if an understanding is to be reached of the stability of solar features, such as loops and prominences in which parallel currents are believed to flow.

If there is no component of $\mathbf{B}$ other than that produced by $J_z$, the system is shearless, since that induced field component $B_\theta$ would have lines of force that were all in the same direction. However, if more than one component of $\mathbf{B}$ exists then $\mathbf{B}$ has shear, because the direction of $\mathbf{B}$ varies from point to point. For a cylindrically symmetrical field, the pitch of the magnetic field is given by

$$ \mu^{-1} = \frac{B_0}{rB_z} ; $$

(18-53)

hence, the shear is just the derivative $d\mu/dr$. However, this shear is related to $dJ_z/dr$ by the integral (Spicer, 1976; Wesson, 1978)

$$ \frac{d\mu}{dr} = \frac{2\pi L}{r^3 B_{z0}} \int r^{r^2} \frac{dJ_z}{dr'} dr' , $$

(18-54)

and it is for this reason that the paradoxical statement is sometimes made that "shear drives kinks" even though, as seen earlier, shear stabilizes kinks. The main point then is that steep current density profiles, that is, profiles with large $dJ_z/dr$, result in large measures of shear, so that a solar PMFC which has large measures of shear should be relatively stable to ideal MHD kink modes as well as to interchange modes. However, the introduction of finite resistivity, which relaxes the MHD constraint of flux conservation, will change the above claim as is seen in the next section.

Reconnection

As noted earlier, there are two means by which the Joule dissipation time,

$$ \tau = \frac{4\pi(\delta \rho)^2}{c^2 \eta} , $$

(18-55)
can be reduced, that is, either by reducing the magnitude of $\delta k$ or by increasing the resistivity. Of course, both can also occur together. In The Origins of Resistivity and Electron Runaways it was noted that the phenomenon called anomalous resistivity involved simultaneously a reduction in $\delta k$ and an increase in $\eta$. However, it was also noted that the $\delta k$ required was on the order of a plasma skin depth ($c/\omega_{pe}$) which, in turn, implied extremely steep magnetic field gradients. It was also argued that such field gradients could only exist in a transient-driven situation if the flare was to be explained by anomalous Joule heating combined with particle acceleration. Here another mechanism will be examined, namely reconnection. Reconnection requires only that $\delta k$ be reduced, but not necessarily to the extremely small values required of anomalous resistivity. In the literature two forms of reconnection are usually treated: reconnection in neutral sheets, and reconnection in sheared magnetic fields. Further, reconnection in neutral sheets is treated as a driven phenomenon, while reconnection in sheared magnetic fields is treated as resulting from an instability developing from a metastable equilibrium called a tearing instability (sometimes called a resistive kink instability), because it tears the magnetic flux surfaces and, thus, violates the conservation of flux constraint of ideal MHD theory. However, intrinsic to reconnection in both neutral sheets and sheared fields is the formation of so-called X-type points. These are regions in which large current densities can be produced without being opposed by $J \times B$ forces, as first shown by Dungey (1953). These regions are therefore ideal for particle acceleration. Indeed, Dungey also showed that such a region is unstable with respect to the growth of the current density. Further, although neutral sheets require X-type points, they are treated in a steady-state driven mode, as if the X-type points already existed; that is, the neutral sheet models ignore the question of how the X-type point is formed in the first place and, therefore, treat the nonlinear steady-state aspect of reconnection. However, something must cause the formation of these X-type points, so that there must exist a phase in which the X-type point is formed. The mechanism generally believed to be the cause of the X-type point is the tearing instability (Furth, Killeen, and Rosenbluth, 1963, hereafter referred to as FKR), as demonstrated by Jaggi (1964), who showed that the tearing instability would set in before the neutral sheet became thin enough for a steady state to be established. Hence, although models will be noted which are treated with one X-type point, this review will concentrate on the physics of the tearing instability, highlighting the differences between neutral sheets and sheared PMFC.

As was noted in The Conservation of Magnetic Flux and Ideal MHD Instabilities, the conservation of magnetic flux represents a constraint on the motion of the PMFC; that is, the topology of flux surfaces remained the same. Allowing for dissipation removes this constraint and the magnetic field structure of the PMFC can change in such a way as to lower the energy of the PMFC leading to instability. The magnetic flux, which is the product of the inductance, $L$, and total current, $I$, was constant in the ideal MHD theory, requiring $I$ to change inversely with any change in $L$. On the other hand, relaxing this constraint allows the flux to change according to

$$\frac{d}{dt} \left( -IL \right) = IR$$

(18-56)

where $R$ is the total resistance. Hence, changes in $I$ and $L$ are allowed which are not in opposition to one another. During reconnection, $dL/dt$ can be very large at the X-type points and can lead to strong inductive electric fields parallel to $B$ or, in the case of neutral sheets, perpendicular to $B$, where $B \rightarrow 0$. Of course, these large inductive electric fields can lead to both intense Joule heating and particle acceleration. It should be noted that $dL/dt = 0$, during pure Joule heating, which is one means of distinguishing ordinary Joule heating, $LdI/dt \neq 0$, and reconnection, $dI/dt (L) \neq 0$. However, before examining in detail the tearing instability and its relationship to reconnection, consider first why reconnection
lowers the energy of a PMFC and why tearing is a very simple and primitive instability.

Useful insight into how reconnection lowers the energy of the PMFC is gained by using a rubberband analogy. In Figure 18-3(a) the field lines are distorted into nested ellipses by some external force; in Figure 18-3(b) the external force has been removed and the field lines become nested circles. As is well known, ellipses have a smaller area than circles with the same circumference. Hence, because of flux conservation the configuration in Figure 18-3(a) has a higher magnetic energy state than that in Figure 18-3(b), and the energy released in going from the configuration in Figure 18-3(a) to that in Figure 18-3(b) is the work required of the external force to maintain the field in its distorted form. Expressed mathematically, it follows that

\[
\Delta \epsilon = \frac{1}{8\pi} \int_{\text{ellipse}} B^2 d^3x - \frac{1}{8\pi} \int_{\text{circle}} B^2 d^3x.
\]

Consider next the situation where the external force is maintained or even increased as illustrated in Figure 18-3(c). In the ideal MHD theory, the nested ellipses would continue to be compressed and the energy state of the system would continue to increase, since there is no way in which the ideal MHD system can relieve itself of the excess energy being stored in it by its compression. However, if one allows for finite dissipation effects, the ideal MHD constraint is relaxed and reconnection can occur, leading to less distorted ellipses or even circles plus an X-type point. Hence, by allowing for finite dissipation, the system was able to find a lower energy state inaccessible in the ideal MHD theory. The system was able to perform this relaxation in energy by altering the shape of its magnetic surfaces—indicated here by the ellipses or circles—something forbidden by MHD theory. The magnetic energy, so released, will appear localized about the X-point and as enhanced Joule heating in the current filaments formed, which are usually called magnetic islands.

![Diagram](image)

Figure 18-3. (a) Flux surfaces in the ideal MHD approximation are distorted into nested ellipses (higher energy state) by the action of an external force in the form of two conducting plates (\(\eta = 0\)) being externally forced together; (b) flux surfaces in the ideal MHD approximation relax back to nested circles (lower energy state) once external force is removed; and (c) distorted flux surfaces reconnect in the presence of finite resistivity and external driving force. A magnetically lower energy state is evident by the appearance of more circular flux surfaces after reconnection.
The next question then is: What force causes the current to bunch into this low energy “magnetic island” configuration? Consider $n$ parallel wires each carrying a current in the $z$ direction. Suppose they are constrained by some external force to remain in the $y = 0$ plane (Figure 18-4). Next, allow a sinusoidal perturbation along the $x$ direction to cause displacements of the wires toward one another. Since parallel currents attract one another, the wires will bunch up to form “islands” of wires. If one now lets $n \rightarrow \infty$ so as to form a continuous current sheet, a similar perturbation will lead to current bunching or filamentation of the current sheet.

So far, the question of why finite dissipation effects are necessary if reconnection is to occur at all has not been considered. The necessity of dissipation can now be demonstrated as follows: Consider Figure 18-5, which illustrates a neutral sheet in which a current is flowing in the positive $z$ direction and is perpendicular to the equilibrium magnetic field. Note that $B_{y0}$ induced by $J_z$ vanishes everywhere along the $x = 0$ line, and, that the current $J_z$ is peaked there. If there exists

![Diagram](image)

Figure 18-4. (a) The geometry and flux surfaces of an array of wires with a parallel current flowing in the $z$ direction within each of them; and (b) the geometry and flux surfaces of the same array of wires after reconnection, where $n \rightarrow \infty$ ($n =$ number of wires) so a "current sheet" is formed.

Provided by the NASA Astrophysics Data System
another component of $\mathbf{B}$ in the $z$ direction, the field is sheared.* Suppose a perturbation $\exp(ik_y y + \gamma t)$ is applied in the $y$ direction that results in a perturbation to the magnetic field in the $x$ direction $\delta B_x$. Using

$$\nabla \times \delta \mathbf{E} = -\frac{1}{c} \frac{\partial \delta \mathbf{B}}{\partial t},$$

(18-57)

it follows that

$$\delta E_z = -\frac{\gamma \delta B_x}{ik_y c}.$$  

(18-58)

$\delta E_z$ represents an induction field generated by the perturbation. Now, from Ohm's law it follows that

$$\eta \delta J_z = \delta E_z + \frac{\delta V_x B_{y0}}{c},$$

(18-59)

but as $x$ approaches zero, $B_{y0}$ approaches zero, so that near $x = 0$

$$\eta \delta J_z \approx \delta E_z = -\frac{\gamma \delta B_x}{ik_y c}.$$  

(18-60)

In the ideal MHD approximation $\eta \to 0$, which implies $\delta B_x \to 0$; that is, no component of $\mathbf{B}$ at right angles to $B_{y0}$ is generated so as to tear the flux surfaces and form $x$ points. One other point emerges from this analysis; namely, that only where $B_{y0}$ goes to zero is $\eta$ important since, generally, the $\delta V_x B_{y0}$ term is far greater in magnitude. This suggests that the mathematics of the tearing instability can be treated by separating the system into those regions in which the ideal MHD theory is valid and those regions in which finite dissipation effects are important. However, before beginning a simple treatment of the tearing instability, it is useful to first consider some of the differences between neutral sheets and sheared magnetic fields.

As has been shown, neutral sheets have, by definition, a region in which $B$ goes to zero, but $B$ does not go to zero in a sheared field; instead, only one or two of its components go to zero. How then does one locate such regions? Consider $\mathbf{B}(x)$ as a given sheared field. Next, construct another magnetic field $\mathbf{B}'(x)$ which has exactly the same flux surfaces but without shear, and subtract their difference such that

$$\mathbf{B}'(x) = \mathbf{B}(x) - \mathbf{B}(x).$$  

(18-61)

---

*In a three-dimensional configuration, in which three components of $\mathbf{B}$ are possible, one or two components of $\mathbf{B}$ could vanish and it would be called a sheared field. However, for a neutral sheet to exist, all three components of $\mathbf{B}$ must vanish.
Physically, what has been done is a sheared magnetic field has been taken whose direction is changing with $x$ and it has had subtracted from it another which, although it has the same set of flux surfaces (say circles), has no shear; that is, the current density is constant, so the direction of the field is unchanging with $x$. Hence, there will exist some surface or point in the sheared field where one or more components of $B^*(x)$ vanish. This procedure is equivalent to transforming to a coordinate system in which field lines are straight, so that by looking parallel to $B^*$ one is looking at flux surfaces in which a component of $B^*(x)$ vanishes. To use such a coordinate system is equivalent to using a neutral sheet, but with a component of $B$ parallel to the current such as in Figure 18-5. Figure 18-6 illustrates the procedure as applied to a diffuse pinch.

There are no essential differences between a tearing mode in a neutral sheet and the tearing mode in a sheared field, as long as one can assume incompressibility of the PMFC. This requires consideration of perturbations which last for times longer than the shortest time scale of the PMFC. This means for low $\beta$ PMFC's, one must consider perturbations that last for times longer than a magnetosonic transit time, while for high $\beta$ PMFC's, one must consider perturbations that last for times longer than a sonic transit time.

To treat the tearing mode analytically, the argument follows Drake and Lee (1977) and uses the following system of perturbed equations, referring to Figure 18-7 for orientation:

\[
\delta E = - \frac{\gamma \delta A}{c} - i k \delta \phi , \quad (18-62)
\]

\[
\left[ \frac{\partial^2}{\partial x^2} - k_y^2 \right] \delta A = - \frac{4 \pi \delta J}{c} \quad \text{(Ampere's equation)} , \quad (18-63)
\]

\[
\delta E + \frac{\delta \mathbf{E} \times \mathbf{B}_0}{c} = \eta \delta J \quad \text{(Ohm's law)} , \quad (18-64)
\]

\[
\frac{\partial N_e}{\partial t} + \delta \mathbf{V}_e \cdot \nabla N_{0e} = 0 \quad \text{(electron continuity)} , \quad (18-65)
\]

\[
\frac{\partial N_i}{\partial t} + \delta \mathbf{V}_i \cdot \nabla N_{0i} = 0 \quad \text{(ion continuity)} , \quad (18-66)
\]

\[
\left[ \frac{\partial^2}{\partial x^2} - k_y^2 \right] \delta \phi = - 4 \pi e (\delta N_i - \delta N_e) \quad (18-67)
\]

\[
= 0 \quad \text{(quasi-neutrality)} , \quad (18-67)
\]

Figure 18-6. A diffuse pinch cut open; (a) prior to being laid out flat; (b) $B$ before transformation to $B^*$; (c) $B^*$ after transformation but before a radial reconnection perturbation $\delta B$ is applied (from Bateman, 1978).
and
\[ \nabla \cdot \delta V_{e,i} = 0 \text{ (incompressibility)} \]  \hspace{1cm} (18-68)

where all perturbed quantities have been taken to vary as
\[ \delta f = \delta f(x) e^{i k_y y - \gamma t} . \]

**Ideal MHD Region** \((\eta = 0)\). Referring to Figure 18-7, the ideal MHD region is treated first, where \(\eta\) is taken as zero. Using Equations (18-62) and (18-64), and keeping in mind that \(\eta = 0\) and \(k_z = 0\),

\[ \delta V_x = \frac{\gamma \delta A_z}{B_{0y}} . \]  \hspace{1cm} (18-69)

Notice that \(B_{0z}\) does not appear in Equation (18-77), because the PMFC is incompressible. Next take \(\delta J = \delta x \cdot \nabla J_0\), which yields \(\delta J = \delta v \cdot \nabla J_0 / \gamma\), since \(\delta v = \gamma \delta x\). Using Equations (18-63) and (18-69), it follows that

\[ \left( \frac{\partial^2}{\partial x^2} - k_y^2 \right) \delta A_z = - \frac{4\pi}{c} \frac{\partial J_{0z}}{\partial x} \frac{\delta A_z}{B_{0y}} , \]  \hspace{1cm} (18-70)

since
\[ \delta J_z = \frac{\delta A_z}{B_{0y}} \frac{\partial J_{0z}}{\partial x} . \]

The left-hand side of Equation (18-70) represents the stretching of the field lines, while the right-hand side represents the source of driving energy for the tearing mode. Notice that \(\partial J_{0z} / \partial x\) appears in Equation (18-70) just as in (18-52); this is because these equations are equivalent, which illustrates that the driving energy for both the \(J_z\)-driven ideal MHD and resistive MHD instabilities lies in the physics of the region where \(k_z\) vanishes. Equation (18-70) also demonstrates that only when \(k_y \delta x \ll 1\) can an instability occur, where

\[ \delta l \approx \left( \frac{4\pi}{c} \frac{\partial J_{0z}}{\partial x} \frac{1}{B_{0y}} \right)^{-\frac{1}{2}} ; \]

that is, the wavelength of the perturbation along \(y\) must be greater than the characteristic gradient.

![Figure 18-7](image)

*Figure 18-7. The geometry of a resistive layer split into two spatial regions. Within the hatched area, with a thickness \(\Delta\), resistivity is important, while outside the hatched area the resistivity is taken to be zero.*
scale length of the magnetic field $\delta x$. Also, as a general rule the long wavelength tearing modes can extract more energy from the equilibrium field than the short wavelength modes, because the short wavelength modes bend the field lines more, causing greater restoring forces and subsequently lower saturation amplitudes. On the other hand, the long wavelength modes bend the field less and can draw on a larger volume of free energy than can the short wavelength modes.

**Resistive Layer ($\eta \neq 0$).** From Equation (18-23) one finds the response of the electrons to a perturbed electric field parallel to $B_0$ is given by

$$
\delta V_{\parallel} = \frac{-e \delta E_{\parallel}}{m_e \nu_e} = \frac{-e(\gamma \delta A_{\parallel} - ik_{\parallel} \delta \phi)}{m_e \nu_e} .
$$

Notice that when $k_{\parallel} \rightarrow 0$ only the induction field is important, and that when $\nu_e$ goes to zero, $\delta E_{\parallel}$ must go to zero, or otherwise $\delta V_{\parallel}$ would go to infinity, which is physically inadmissible. Using Equations (18-65) and (18-71),

$$
\delta N_e = \frac{ie \delta E_{\parallel} k_{\parallel} \Omega_{ce} \delta \phi}{m_e \nu_e \gamma} ,
$$

where the reasonable assumption that electron density perturbations perpendicular to $B_0$ are negligible is made, since $\Omega_{ce} \gg \gamma$ (i.e., the instability has a slow growth rate). If the ions are magnetized, the ion response to the time-varying electric field is a polarization drift (Spitzer, 1962),

$$
\delta V_{\perp} = \frac{c^2}{4\pi e V_A^2} \frac{d \delta E_{\perp}}{dt} ,
$$

where $V_A$ is the Alfvén velocity.

Hence, by using Equations (18-66), (18-73), and (18-67)

$$
\delta N_1 = \frac{c^2}{4\pi e V_A^2} \frac{\partial^2 \delta \phi}{\partial x^2} ,
$$

which, on equating with Equation (18-72) (because of the quasineutrality condition in Equation 18-66, which is applicable because of the low frequency of the instability) yields

$$
\frac{\partial^2 \delta \phi}{\partial x^2} - \frac{k_{\parallel}^2 \Omega_{ci} \Omega_{ce} \delta \phi}{\nu_e} \gamma = -i k_{\parallel} \frac{\Omega_{ci} \Omega_{ce} \delta A_{\parallel}}{\nu_e c} .
$$

This result has a number of implications that clearly demonstrates the distinctions between the ideal and resistive MHD instabilities. Notice that the right-hand side of Equation (18-75) represents a source term for $\delta \phi$, and that in the limit $k_{\parallel} \rightarrow 0$ it follows that $\delta \phi \rightarrow 0$ implying that the electrons experience only an induction field $-\gamma \delta A_{\parallel}/c$, where $k_{\parallel} \rightarrow 0$ (see Equation 18-62). Further, one finds from Equation (18-75) that as $\nu_e$ approaches zero ($\eta$ approaches zero),

$$
\frac{\gamma \delta A_{\parallel}}{c} = -i k_{\parallel} \delta \phi .
$$

That is, $\delta E_{\parallel} \rightarrow 0$, which is what is required by the ideal MHD theory. The point at which $\delta E_{\parallel}$ vanishes defines the width of the accelerating channel at which $\delta \phi$ becomes important. One can now understand the role that finite resistivity plays in the tearing process. Figure 18-8 shows that the perturbed parallel induction field $-\gamma \delta A_{\parallel}/c$ causes the electrons to flow with a velocity $\delta V_{\parallel}$. However, because the induction field leads to a charge separation between ions and electrons, a
parallel electrostatic field \(-ik_\parallel \delta \phi\) is also produced which tends to short out the induction field for sufficiently large \(k_\parallel\), so that \(\delta E_\parallel = 0\) and acceleration parallel to \(B\) cannot take place. However, only when \(\eta\) is finite can \(\delta E_\parallel \neq 0\), where \(k_\parallel \rightarrow 0\), so that particle acceleration as well as restructuring of the magnetic surfaces can take place.

In the discussion so far, it has been tacitly assumed that \(\delta A_\parallel\) is constant within the region where \(\delta E_\parallel \neq 0\). This assumption is called the “constant-\(\psi\)” approximation in the literature. However, there are tearing modes that violate this assumption. These different types of modes—those which satisfy the constant-\(\psi\) approximation and those which do not—are referred to as slow tearing modes and fast tearing modes, respectively. It is useful to clarify the differences between these two modes here, because, until recently, all flare models that used tearing modes have invoked slow tearing modes (e.g., Jaggi, 1964; Sturrock, 1966b, 1967). Only recently have the fast modes been applied to the solar flare phenomena (Spicer, 1976, 1977a, 1981a) and to erupting prominences (Spicer, 1979a).

The slow and fast tearing modes are distinguished in both their growth rates and in their nonlinear behavior. Slow modes in the collisionally dominated regime have a growth rate given by

\[
\gamma_s \approx \frac{S^{2/5}}{\tau_R},
\]

(18-77)

where \(\tau_R\) is the resistive skin time given by Equation (18-8),

\[
S = \frac{\tau_R}{\tau_A},
\]

(18-78)

\(\tau_A = \delta \ell / V_A\), and \(V_A\) is the Alfvén velocity. The term, \(S\), is generally referred to as the mag-

![Figure 18-8](image)

*Figure 18-8. Illustrates the physical mechanism responsible for the generation of \(\delta \phi\). The induced field \(-\gamma \delta A_\parallel /c\) generates an electron flow with velocity \(V_e\) as shown. The resulting charge separation results in a parallel electrostatic field \(-ik_\parallel \delta \phi\), which cancels the parallel induction field except for sufficiently small \(k_\parallel\). Hence, the electrons short out the induction field except around \(k_\parallel \rightarrow 0\) (from Drake and Lee, 1977).*

436

Provided by the NASA Astrophysics Data System
nentic Reynolds number, since it is the ratio of the resistive diffusion time and the convection time (the transit time of an Alfvén wave across a layer $\delta \xi$ in thickness) in a magnetized plasma. For $S < 1$, the PMFC is said to be diffusion dominated, while for $S > 1$, it is said to be convection-dominated. For solar plasmas, $S$ is very much less than 1, typically on the order of $10^{11-12}$; solar PMFC's are thus totally convection-dominated.

Rutherford (1973) showed that in the nonlinear regime exponential growth of the slow tearing mode decreases to algebraic growth with $dW/dt$ constant when $W > \Delta$, where $\Delta$ is the width of the region within which $\delta A_\parallel \neq 0$ given by (Drake and Lee, 1977)

$$\Delta = \left( \frac{\eta \gamma_s c^2}{4 \pi k^2_{||} V_A^2} \right)^{1/4}, \quad (18-79)$$

$k^2_{||}$ is the derivative of $k_\parallel$ with respect to the inhomogeneous coordinate earlier taken to be $x$, and $W$ is the width of the magnetic island as shown in Figure 18-9. The quantity $k^2_{||}$ is related to the shear of the magnetic field such that $k^2_{||} \simeq k_\parallel x_\parallel/\xi_\parallel$, where $\xi_\parallel$ is the shear length. In a diffuse cylindrical pinch,

$$\xi_\parallel \simeq \left[ \frac{d}{dr} \left( \frac{r B_z}{B_\theta} \right) \right]^{-1} \quad (18-80)$$

However, numerical simulations of fast tearing modes have demonstrated that fast modes

*As a historic note it should be noted that the first nonlinear study of tearing modes was due to Van Hoven and Cross (1973), whose principal aim, interestingly enough, was to demonstrate that the energy release from tearing instabilities was adequate to explain solar flares. Unfortunately, their aim was not achieved, since they considered only Reynolds numbers of $10^2$-$10^3$. It is well known that the nonlinear behavior of high $S$ ($>10^9$) versus low $S$ ($>10^2$) tearing modes is significantly different.

continue to evolve exponentially to very large amplitudes without entering the Rutherford regime. These fast tearing modes have been found to have growth rates $\gamma_F$ which have a range

$$\gamma_F \simeq \frac{S^{1/3}}{\tau_R} \quad \text{to} \quad \frac{S^{2/3}}{\tau_R}, \quad (18-81)$$

and include the $m = 1$ mode in cylindrical geometry (Spicer, 1976; Waddell et al., 1979), long wavelength modes in slab geometry (Furth et al., 1963; Wesson, 1966; Van Hoven and Cross, 1973), and double tearing modes (Schnack and Killeen, 1978). All of these basically long wavelength tearing modes have the common feature that $\delta A_\parallel$ is not constant across the tearing layer, and that they are a driven form of the tearing mode.*

Recently, Pritchett et al. (1979) have shown that the physical differences between the fast and slow tearing modes can be found by comparing the growth time of each mode with the skin time.
of the tearing layer. On using $\Delta$ for $\delta \xi$ in Equation (18-8) this gives for the skin time

$$\tau_{sk} \approx \frac{4\pi\Delta^2}{\eta c^2}.$$  \hspace{1cm} (18-82)

(Note: $\delta \xi$ not $\Delta$ should be used in Equations (18-77) and (18-81) to estimate growth rates, since $\Delta \ll \delta \xi$.) For the fast mode $\gamma^{-1}_{k} \approx \tau_{sk}$, while for the slow mode $\gamma^{-1}_{s} \gg \tau_{sk}$. Physically, this means that the flux perturbations generated by the slow modes have time to communicate across the layer $\Delta$ thick during their growth, but flux perturbations generated by the fast modes do not. This difference in communication time means that $\delta E_{\|}$ does not have time to penetrate into the magnetic islands for the fast modes, so that the volume of plasma undergoing acceleration by $\delta E_{\|}$ remains the same as for the slow modes in their linear regime. On the other hand, $\delta E_{\|}$ for the slow modes does have time to penetrate into the islands, causing $\delta E_{\|}$ to weaken because of the increase in the tearing layer width, that is, $\Delta$ increases to $W$. Thus, since the accelerating volume and $\delta E_{\|}$ for the fast modes are not increasing and decreasing respectively, these modes are unchanged in the nonlinear regime, and the exponential growth must continue.

Figures 18-10 and 18-11 illustrate the $m = 1$ tearing mode in cylindrical geometry and the sheet analog of the double tearing mode, respectively. Note that the currents are antiparallel for the double tearing modes. Figure 18-12 illustrates the flux surfaces and velocity fields found by numerical simulations. It should be emphasized that all of these fast tearing modes are driven, as opposed to the slow tearing modes. For the $m = 1$ tearing mode (usually called a resistive kink), the driver is an $m = 1$ ideal MHD kink which forces the magnetic flux into the X-type point (Figure 18-10). In the case of the double tearing mode, one tearing mode drives the other. The reader should note that $\mathbf{k} \cdot \mathbf{B}_0$ vanishes more than once for the double tearing (or for multiple tearing modes), so that the resulting modes are essentially locked in phase with one another. However, this phase locking is not of crucial importance, and the fast reconnection rates will still occur, although at somewhat reduced rates.

(a) IDEAL KINK MODE

(b) m = 1 TEARING MODE

Figure 18-10. (a) The axial displacement of the flux surfaces due to an $m = 1$ ideal MHD internal kink; and (b) an identical displacement except with finite resistivity thereby allowing the formation of one X-type neutral point. Notice that the flows are such that the ideal $m = 1$ kink is essentially driving new flux into the X-type point (from Drake, 1978).
Figure 18-11. The neutral sheet analog of a double tearing mode. Fluid flows in the directions $V$ and $-V$ toward the two neutral sheets indicated by dashed lines. The field dissipation rate will accelerate until no field is left between the two sheets.

Figure 18-12a. Computed double tearing mode flux surfaces. Singular layers are located in the closed contours, with the neutral points occurring at coordinates $(0.25, -0.5)$ and $(0.75, 0.5)$ (from Schnack and Kileen, 1978).

Figure 18-12b. Illustration of vortex velocity flow during double tearing mode. Use of small arrows indicates direction and magnitude of the flows (from Schnack and Kileen, 1978).
Taking the ratio of $\gamma_F$ and $\gamma_S$ yields

$$f = \frac{\gamma_F}{\gamma_S} \lesssim S^{4/15}.$$  \hspace{1cm} (18-83)

Hence, the fast tearing modes dissipate magnetic energy at a rate $S^{4/15}$ times faster than a slow tearing mode. Since $S$ typically has a magnitude of $10^{11}$ to $10^{12}$, one gets a $10^3$ increase in the dissipation rate for fast modes compared to slow modes, making the fast modes an ideal candidate for a flare mechanism (Spicer, 1976; 1977a; 1981a). The higher is $S$, the more likely it is that the constant-$\psi$ approximation will be violated, making fast tearing modes a likely candidate for the Sun.

The rate at which a tearing mode releases magnetic free energy per unit volume per unit time can be approximated by

$$\frac{dE}{dt} \approx \frac{\gamma B_{0y}^2}{4\pi},$$  \hspace{1cm} (18-84)

which on taking

$$J_{0z} \approx \frac{e B_{0y}}{4\pi \delta \xi},$$

gives

$$\frac{dE}{dt} \approx \eta_{\text{eff}} J_{0z}^2,$$  \hspace{1cm} (18-85)

where

$$\eta_{\text{eff}} \approx \frac{(\delta \xi)^2 4\pi \gamma}{c^2}.$$  \hspace{1cm} (18-86)

It can be demonstrated using linear theory that $\sim 10$ percent of the magnetic energy released goes into kinetic flows, while $\sim 90$ percent goes into internal energy (Furth et al., 1963). Similar results have been found in the nonlinear regime (Van Hoven and Cross, 1973; Schnack and Killeen, 1978). In light of Equation (18-84) one may view reconnection by the tearing mode as causing an increased Joule heating within the tearing layer, although of a kind vastly different from that of the anomalous Joule heating discussed in The Origins of Resistivity and Electron Runaways.

![Figure 18-13. Expanded geometry of a neutral sheet about the X-type neutral point illustrating the inflow ($V_x$) and outflow ($V_{\text{out}}$).](image-url)
Neutral Sheets and Steady-State Reconnection.
The physics of steady-state reconnection and reconnection in neutral sheets has been reviewed frequently in the last few years (e.g., Vasyliunas, 1975; Priest, 1976). Only a brief composite of the many theoretical papers that have treated both steady-state reconnection and neutral sheets is given here. This follows in part Sweet’s review (1969), and in part Priest’s review (1976), making additional points where necessary. The geometry of a current sheet is given in Figure 18-13. Note that there is only one X-type point in the treatment, and that driven inflow of plasmas from both \( x = \pm \infty \) into the \( x = 0 \) plane forces the field toward the \( x = 0 \) plane.

The equations governing the steady-state motion in neutral sheet models are the same as Equations (18-62) to (18-67). Far removed from the neutral sheet, at some point \( x_0 \), fluid flows toward the sheet with a velocity \( V_{x_0}(x_0) \) as specified by the boundary conditions; hence, by Ohm’s law, \( E_{x_0}(x_0) = V_{x_0}(x_0)B_{y_0}(x_0)/c \). Since the electric field is uniform in the steady state, \( \eta_0 J = E_{x_0} \) in the neutral sheet, where \( B_{y_0}(x=0) = 0 \). The thickness of the neutral sheet is such that the net current across the sheet just equals the change in \( B_{y_0} \) across the sheet to yield

\[
\delta \ell \approx \frac{\eta c^2}{4\pi V_{x_0}}.
\]

(18-87)

Assuming there exists a plasma sink at \( \pm \infty \) a more exact solution of the induction equation (Faraday’s equation combined with Ohm’s law) yields a solution for \( B_{y_0} \) given by (Yeh and Axford, 1970)

\[
B_{y_0} = \pm B_0 \left( 1 - \exp \left[ \frac{4\pi V_{x_0} x}{\eta c^2} \right] \right) x \approx 0,
\]

(18-88)

where \( B_0 \) is the field at \( x = \pm \infty \). The characteristic scale length of the magnetic field change in Equation (18-88) is as given by Equation (18-87). According to Equation (18-87), \( V_{x_0} \) determines \( \delta \ell \); however, Sweet (1958a, 1958b) and Parker (1963) attempted to determine \( \delta \ell \) by considering the origin of the plasma sink at the neutral sheet. Parker argued that the fluid flows out of the sheet whose length was \( 2L \) in the \( \pm y \) direction. For incompressible flow, Parker found by using mass conservation that the pressure gradient across the sheet led to flow that left the system at the Alfvén velocity so that

\[
V_{x_0}L = V_A \delta \ell.
\]

(18-89)

Hence,

\[
\delta \ell \approx \frac{V_{x_0}L}{V_A}.
\]

(18-90)

By using Equations (18-89) and (18-87) one finds

\[
\delta \ell = \left( \frac{L\eta c^2}{4\pi V_A} \right)^{\frac{3}{2}},
\]

(18-91)

and

\[
V_{x_0} = \frac{V_A \eta c^2}{4\pi L}^{\frac{3}{2}},
\]

(18-92)

where \( V_A, \eta, \) and \( L \) are assumed given. Petschek (1964) further suggested that \( L \), like \( \delta \ell \), should be determined from Equation (18-89) given \( V_{x_0} \). Petschek argued semiquantitatively that the MFE will be converted to internal plasma energy away
from the neutral X-type point by the presence of slow mode MHD shocks. On the basis of these arguments, Petschek (1964) found that fast reconnection could occur on a time scale \( \tau = \tau_0 \ln(\tau^*_e/\tau^*_A) \), where \( \tau^*_e = 4\pi L^2/\eta c^2 \) and \( \tau^*_A = L/V_A \). The mechanism proposed by Petschek is almost independent of resistivity because of a logarithmic term. Recently Priest and Soward (1976) have put the Petschek mechanism on firmer mathematical grounds.

It is of interest to digress at this point and compare the single X-type point \( m = 1 \) fast tearing mode in cylindrical geometry with the single X-type point steady-state sheet models. On comparing Figures 18-10 and 18-13, the reader will note that the one X-type point steady-state reconnection model has a geometry similar to that of the \( m = 1 \) fast tearing mode if the sheet is rolled into a cylinder so that the points at \( y = \pm L \) meet to form a circle. Here \( 2L \) becomes \( 2\pi r_0 \), where \( r_0 \) is the radius at which \( k \cdot B_0 \) vanishes. Further, the \( m = 1 \) ideal kink mode drives the external flow into the X-type point, and the outflow moves away from the X-type in the \( m = 1 \) fast tearing modes just as in the sheet. Recent numerical work has shown that as \( S \) becomes larger (\( \sim 10^6 \) is the highest value so far simulated) the behavior of the region about the X-type point becomes more and more like Petschek-type reconnection; similar behavior is expected for the other fast tearing modes (Drake, private communication). Hence, it is our feeling that steady-state reconnection models in sheared current sheets should be considered in other geometries, as just discussed.

An alternative means of shortening the reconnection time is to increase the resistivity by turbulence as discussed in The Origins of Resistivity and Electron Runaways. If the current density reaches the threshold values required for onset of the various instabilities listed in Table 18-1, anomalous resistivity will result. Whether \( J_\parallel \) or \( J_\perp \) is the driver will depend on whether the PMFC is taken as sheared or whether the PMFC has a neutral sheet. The question of which \( J_\parallel \)- or \( J_\perp \)-driven mechanism will occur first will depend on how fast \( J_\parallel \) or \( J_\perp \) rises as a function of time. This point will be discussed subsequently, but first the \( J_\parallel \)-driven anomalous resistivity mechanisms associated with neutral sheets will be examined.

Sturrock (1967) and Friedman and Hamberger (1968) suggested that some form of increased turbulence would alter the reconnection rate through the resulting increased resistivity. Sturrock (1967) suggested that an increased diffusion coefficient perpendicular to \( B \) would result if the sheet became turbulent, and that this turbulence would take the Bohm value

\[
D_B = \frac{1}{16} \frac{V_e^2}{\Omega_c \tau_e} \quad (18-93)
\]

It is easy to demonstrate that the maximum diffusion coefficient possible, involving either collisions or collisionless scattering, is \( \sim 8D_B \). This can be seen from the diffusion coefficient perpendicular to \( B \), which is

\[
D_\perp = \frac{V_e^2 \tau_e}{1 + (\Omega_c \tau_e)^2} \quad (18-94)
\]

which has a maximum at \( \Omega_c \tau_e = 1 \); hence, \( D_{\perp,\max} = V_e^2 / 2\Omega_c \). Since the only means of obtaining \( \Omega_c \tau_e \approx 1 \) in a solar plasma is for the plasma to have reduced \( \tau_e \) because of some electrostatic instability, Sturrock was implicitly arguing that the resistivity was also anomalous in the sheet. On the other hand, Friedman and Hamberger (1968) argued that \( J_\perp \) could easily exceed the threshold for the Buneman instability (see Table 18-1) and settle into a marginally stable condition. However, as noted earlier, if the current density is large enough to cause the Buneman instability, then both rapid heating and particle acceleration will occur, with or without reconnection. Hence, the question arises: What is the need for reconnection in a neutral sheet if the current density is high enough to excite the Buneman instability in the first place? Indeed, as
argued by Papadopoulos (1977), it is not possible to achieve any semblance of steady-state anomalous resistivity with the Buneman instability because, as discussed in The Origins of Resistivity and Electron Runaways, the Buneman instability must be driven constantly. Of course other electrostatic instabilities exist that are capable of causing high levels of anomalous resistivity, such as the ion acoustic instability; however, probably the most useful of these is the lower hybrid drift instability which has been recently applied to the Earth's magnetotail (Huba et al., 1977, 1978). This instability has a number of advantages over the instabilities generally applied to neutral sheets in solar physics (e.g., Syrovatskii, 1972; Heyvaerts et al., 1977). Using Table 18-1, which lists the threshold conditions for the lower hybrid drift instability, the reader will note that it is driven by $J_L$, it can be excited by currents far weaker than, say, the ion acoustic instability and occurs even when $T_e/T_i < 1$. The level of turbulent resistivity is also substantial. In addition, the lower hybrid drift instability, which is characterized by wavenumbers $k \cdot B = 0$, is closely related to the modified two-stream instability $k \cdot B \approx 0$ (Gladd, 1976), and, as has been demonstrated by Lampe and Papadopoulos (1977), is capable of accelerating electrons stochastically to very high energies. No application of lower hybrid drift instability or the modified two-stream instability to neutral sheets in the solar atmosphere has taken place as yet, although it should be relatively straightforward by using the results of Huba et al. (1977, 1978).

**Double Layers**

A "double layer" is defined as consisting of two equal but oppositely charged, essentially parallel but not necessarily plane, space charge layers (Block, 1977). The potential, electric field and the space charge density vary qualitatively within the layer as shown in Figure 18-14. A double layer is believed to occur when a large potential difference is applied to a finite length plasma, but the potential difference, rather than being felt over the entire length of the system, is concentrated in a shocklike localized region. There are essentially four conditions which must be fulfilled for a double layer to occur (Block, 1977):

1. The potential difference $\phi_0$ through the layer must satisfy $|\phi_0| \gtrsim kT_e/e$.
2. The electric field must be much stronger inside the double layer than outside, so that the integrated positive and negative charges nearly cancel one another.
3. Quasineutrality is locally violated in both charge layers.
4. The collisional mean free path must be much greater than the double layer thick-
ness. This follows because the formation of a double layer is a collisionless phenomenon.

The reader will note that for a double layer to be useful in explaining a solar flare, the potential difference \( \phi_0 \) must be much greater than \( kT_e/e \), otherwise the energy gain by an electron traversing a double layer with \( \phi_0 \gg kT_e/e \) will not be appreciably greater than the thermal energy of the plasma. Hence, one requires \( \phi \gg kT_e/e \) which implies that a strong double layer is required. It was shown by Goertz and Joyce (1975) that a strong double layer requires a current density given by

\[
|J| = N_e eV_{Te}, \quad (18-95)
\]

which is the current density required to excite the current-driven Buneman instability.

As was noted in The Origins of Resistivity and Electron Runaways, to be able to use current driven microinstabilities, the system—external driver and instability—must be at marginal stability throughout the duration of the flare, or otherwise, the instability will saturate and switch off permanently. A similar argument is applicable to double layers since the double layer must exist for a \( \Delta t \) characteristic of a flare, if it is to be responsible for the flare. Hence, the external driving voltage difference, or equivalently, the electric field, must exist for a time \( \Delta t \). A double layer, as well as anomalous resistivity \( J_{\parallel} \)-driven systems are essentially systems in which a large-scale circuit carrying a current and driven by a generator has a dissipative load within it, which is the double layer or region of anomalous resistivity (i.e., a current interrupter, Alfvén and Carlqvist, 1967). For this reason the whole current needs analysis to understand the nature of the generator and the nature of the load. Hence, the double layer and \( J_{\parallel} \)-driven anomalous resistivity systems require special boundary conditions, and the question arises of whether these boundary conditions can be maintained for the duration of the flare.

A further constraint on the double layer is that since it leads primarily to particle acceleration, it must be able to accelerate enough electrons to explain the flare. As shown by Hubbard and Joyce (1979), the thickness of a double layer is roughly

\[
L_{DL} \sim 6 \left( \frac{e\phi_0}{kT_e} \right)^{1/2} \lambda_{De}. \quad (18-96)
\]

Combining Equations (18-40), (18-41), and (18-96) with \( L_{DL} = D \) yields

\[
N_T \sim \frac{N_e \Delta t \left( e \right)^{1/2} \phi_0 A}{\left( m_e \right)^{1/2}}. \quad (18-97)
\]

Hence,

\[
e\phi \sim \left( \frac{N_T}{N_e \Delta t A} \right)^2 m_e. \quad (18-98)
\]

is necessary to accelerate \( N_T \) particles during a solar flare, where \( A \) is the cross-sectional area of the double layer. This result will be used in the section on Flare Models. The reader should consult the reviews by Block (1972, 1975, and 1978) and Goertz (1979) for more details concerning double layers.

**Which Mechanism Occurs First?**

As the reader has undoubtedly noticed, for either \( J_{\parallel} \) or \( J_{\perp} \) there exists more than one mechanism that could be excited (see Table 18-2). The question naturally arises: Which mechanism for a given value of \( J_{\parallel} \) or \( J_{\perp} \) will be excited first, and which mechanism will dominate the dissipation
of \(J_\parallel\) or \(J_\perp\)? It is relatively straightforward to show that the tearing instability has the lowest threshold and should be excited first, either in sheared fields or in neutral sheets. However, this is not the whole story. An additional piece of information is needed—specifically, the rate at which the current density rises—if one is to decide which mechanism is to be excited first. The reason for this is as follows: Suppose, for example, the Sun were somehow to cause either \(J_\parallel\) or \(J_\perp\) to rise in a time less than a tearing mode growth time (called a strongly driven current as opposed to a weakly driven current). The tearing mode would not have time to develop and dissipate the current before the current density increases to magnitudes sufficient to excite, say, double layers in the case of \(J_\parallel\) or the lower hybrid drift instability in the case of \(J_\perp\). Hence, the rise time of the current density is an observable means of distinguishing, in principle, possible flare mechanisms (Spicer, 1979b). For mechanisms such as double layers and anomalous Joule heating, the \(J_\parallel\) required implies magnetic flux changes comparable to >10^{18} \text{ m}^2 \text{ s}^{-1} are necessary (Spicer, 1981a). Further, note that if the current is driven strongly and causes, say, anomalous resistivity for a period comparable to a tearing mode growth time, but with allowance taken for an anomalous resistivity, tearing can occur. Which mechanism will dominate the system energetically will depend on the parameters involved. For weakly driven currents only tearing and ideal MHD kinks in \(J_\parallel\) systems are expected, at least initially, and, in \(J_\perp\)-driven systems, only tearing modes followed by steady-state reconnection are expected as discussed by Sweet (1958a, 1958b), Parker (1963), or Petschek (1964).

A Classification of Flare Mechanisms

Throughout this section, various mechanisms proposed over the years to explain solar flares have been reviewed. In particular, it should now be apparent that the source of driving energy for each mechanism was either due to \(J_\parallel\) or to \(J_\perp\). There is a convenient classification scheme that allows one to identify the mechanisms that would be important at a particular altitude in the solar atmosphere and what kind of PMFC's can be expected to have a particular type of mechanism operating within it. Table 18-2 tabulates the mechanisms with their respective source of free energy, as well as their obvious advantages and disadvantages. Also indicated in Table 18-2 is that \(J_\parallel\)-driven mechanisms will generally be important in closed PMFC's, while \(J_\perp\)-driven mechanisms will generally be important in open PMFC's. In PMFC's where \(|J_\parallel| \approx |J_\perp|\), one cannot be as definitive. The association of closed PMFC's with \(J_\parallel\) and of open PMFC's with \(J_\perp\) follows from the fact that \(\beta \ll 1\) is required for \(J_\parallel\) to dominate the energetics of the PMFC, which in turn implies that the magnetic field pressure is much greater than the gas pressure, further implying that the PMFC will be closed. On the other hand, \(\beta \gg 1\) is required for \(J_\perp\) to dominate the PMFC, which in turn implies that the gas pressure is much greater than the magnetic pressure, further implying that the PMFC will be open. However, most PMFC's are not equilibrium PMFC's but are closer to a steady state or dynamic equilibrium, because there are other body forces in the solar atmosphere, particularly in the photosphere and in the high corona. Consequently, the above arguments are not exact but they do allow one to be more specific as to what type of mechanism can be expected in what kind of PMFC's and what its driver will be. Later in Flare Models this classification scheme will be used to classify various flare models according to their mechanisms. This eliminates the need for reviewing and criticizing in detail all the flare models ever proposed. Instead, attention can be focused on the advantages and disadvantages of a given mechanism, which in turn reflects on the models that use that mechanism.

The above classification scheme also leads naturally to some comments about energy storage prior to flares in the solar atmosphere. Since the driving energy of a \(J_\perp\)-driven mechanism is the pressure gradient and not the magnetic field, flare models that utilize \(J_\perp\) mechanisms, such as neutral sheets, must, of necessity, require an additional source of driving energy. This

Provided by the NASA Astrophysics Data System
Table 18-2
Mechanisms Classified According to Whether the Driver is $J_{\parallel}$ or $J_{\perp}$

<table>
<thead>
<tr>
<th>Driver</th>
<th>Mechanism</th>
<th>Requirements and Comments</th>
<th>Advantages</th>
<th>Possible Disadvantages*</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_{\parallel}$</td>
<td>Anomalous Joule Heating</td>
<td>Strongly driven currents</td>
<td>Extremely rapid energy release both heating and copious runaways expected.</td>
<td>Requires transient external driver.</td>
</tr>
<tr>
<td>$\beta&lt;&lt;1$ Closed PMFC</td>
<td>Double Layers</td>
<td>Strongly driven currents (energy cannot be stored in situ) Transient external driver</td>
<td>Primarily a particle acceleration mechanism.</td>
<td>Requires transient external driver.</td>
</tr>
<tr>
<td>$J_{\perp}$</td>
<td>Tearing Instabilities</td>
<td>Weakly driven currents (energy can be stored in situ) Transient external driver</td>
<td>Can occur in any sheared PMFC. Heating and particle acceleration.</td>
<td>No obvious disadvantages (but see models).</td>
</tr>
<tr>
<td>$\beta&gt;&gt;1$ Open PMFC</td>
<td>Reconnection in Neutral Sheets</td>
<td>Weakly driven currents (energy cannot be stored in situ) Transient external driver</td>
<td>Heating and particle acceleration.</td>
<td>Slow energy release except in Petschek regime. Requires $B$ to vanish over large surface. Requires transient external driver.</td>
</tr>
<tr>
<td>$J_{\perp}$</td>
<td>Anomalous Joule Heating in Neutral Sheets</td>
<td>Strongly driven currents required (energy cannot be stored in situ) Transient external driver</td>
<td>Rapid heating and particle acceleration.</td>
<td>Requires transient external driver. Requires $B$ to vanish over large surface.</td>
</tr>
</tbody>
</table>

*These disadvantages are of a subjective nature; they may or may not be true disadvantages.
Table 18-2 (continued)
Mechanisms Classified According to Whether the Driver is $J_\parallel$ or $J_\perp$

<table>
<thead>
<tr>
<th>Driver</th>
<th>Mechanism</th>
<th>Model</th>
<th>Geometry</th>
<th>Specific Comments</th>
<th>General Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_\parallel$</td>
<td>Double Layer</td>
<td>Current interruption</td>
<td>Loop</td>
<td>Current strongly driven; primarily a nonthermal model.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Alfvén and Carlqvist, 1967)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anomalous Joule Heating</td>
<td>Revised current interruption</td>
<td>(Smith and Priest, 1972; Spicer, 1981a)</td>
<td>Loop</td>
<td>Current strongly driven; primarily a thermal model.</td>
<td>All mechanisms in principle will work in any low $\beta$ PMFC.</td>
</tr>
<tr>
<td>Fast and Slow Tearing Mode</td>
<td>Tearing mode model</td>
<td>(Spicer, 1976; 1977a, 1977b, 1978; 1981a; Colgate, 1978)</td>
<td>Loop</td>
<td>Current weakly driven; primarily a thermal model.</td>
<td>Possible problem with Type III radio bursts.</td>
</tr>
<tr>
<td>Anomalous Joule Heating and Tearing Instabilities</td>
<td>Hybrid model of revised current interruption and tearing mode models</td>
<td>(Spicer, 1981a)</td>
<td>Loop</td>
<td>Current strongly driven; primarily a thermal model. High field shear required.</td>
<td></td>
</tr>
<tr>
<td>Reconnection in Neutral sheets (by Sweet-Parker and Slow Tearing)</td>
<td>Gold-Hoyle (1960)</td>
<td>Interacting loops Multiple loops</td>
<td></td>
<td>Too qualitative.</td>
<td>Utilize interchange instability to speed up reconnection.</td>
</tr>
</tbody>
</table>
Table 18-2 (continued)
Mechanisms Classified According to Whether the Driver is $J_\parallel$ or $J_\perp$

<table>
<thead>
<tr>
<th>Driver</th>
<th>Mechanism</th>
<th>Model</th>
<th>Geometry</th>
<th>Specific Comments</th>
<th>General Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sweet (1958a, 1958b)</td>
<td></td>
<td>Sheet formed between two distinct flux systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kopp and Pneuman (1976)</td>
<td></td>
<td>Inverted Y PMFC.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J_\perp$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reconnection in</td>
<td>Carmichael (1964); Sturrock</td>
<td>Emerging Flux Model</td>
<td>Sheet between two PMFC's one of which is emerging.</td>
<td>Difficulty with driver.</td>
<td></td>
</tr>
<tr>
<td>Petschek, Sweet-</td>
<td>Emerging Flux Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parker, Slow</td>
<td>Priest and Heyvaerts, (1974);</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tearing, and</td>
<td>Canfield et al. (1974);</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anomalous Resistivity</td>
<td>Heyvaerts et al. (1977);</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tur and Priest (1978); and Krivsky</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Syrovatskii (1966, 1969)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Somov and Syrovatskii (1977)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
follows because it is logically inconsistent for a pressure gradient to steepen on its own against a balancing magnetic field. Only an additional source of free energy can cause such steepening. This is the reason why the neutral sheet mechanism requires convective flow external to the sheet [i.e., finite \( V_{x0} \) in Equation (18-87)] to force new magnetic flux into the reconnecting region. Hence, the neutral sheet mechanism is, in effect, a mechanism that dissipates the convective flow energy of a large remote plasma volume, whereas the magnetic field only serves as an intermediate medium that allows this flow energy through the localized neutral sheet mechanism to be concentrated into kinetic energy. Thus, unless there is some unknown mechanism preventing continuous dissipation of the magnetic flux piling up outside the neutral sheet, preflare storage of magnetic energy around a neutral sheet cannot occur. In the absence of any such mechanism, the neutral sheet mechanism must be a driven mechanism, and must be driven for the duration of the flare. Thus, to identify a neutral sheet as the flare mechanism one must not only detect the neutral sheet, but also identify its external driver such as emerging flux (Canfield et al., 1974; Heyvaerts et al., 1977; Tur and Priest, 1978). Note also that anomalous Joule heating mechanisms driven by \( J_\parallel \) are similar to the neutral sheet mechanisms in that they require an external driver.

Neutral sheets cannot store energy before a flare, but sheared PMFC can, since it is the addition of shear that stabilizes against ideal MHD instabilities, but drives reconnection by tearing instabilities once enough shear is accumulated in the PMFC. Such behavior is common in laboratory devices such as reverse field pinches, where strong shear stabilizes the ideal MHD instabilities but results in fast tearing modes if the shear level is not carefully monitored (Christiansen and Roberts, 1978; Schnack and Killeen, 1979). Thus, preflare storage of magnetic energy in sheared PMFC’s can occur and can possibly be detected by searching for evidence of the increase of \( J_\parallel \) before a flare (cf. Flare Models). However, as discussed in this section, other \( J_\parallel \)-driven mechanisms such as anomalous Joule heating and double layers can occur only if \( J_\parallel \) is driven externally and if \( J_\parallel \) rises in a time less than a tearing or kink growth time. Such mechanisms are also incompatible with preflare storage of magnetic energy in situ.

**FLARE MODELS**

In this section, several topics related to the concept of a flare model are discussed. These topics include the following: requirements of a flare model, triggering mechanisms if applicable, and particle acceleration and heating. A brief description of the various flare models is offered, along with a classification of them under \( J_\parallel \) and \( J_\perp \). Flare models which fall outside of this classification scheme will also be discussed. Frequent references will be made to Chapter 7, which reviews flare observations and the constraints imposed on flare theories by these observations.

**Requirements of a Flare Model**

The objective here is to summarize the generally accepted requirements of a flare model. However, in doing so, the authors will take the liberty of injecting some of their own feelings on the question, and they wish to emphasize to the reader that the question of flare model requirements is still a subjective topic, and one that the theorist should consider cautiously.

**Energy Storage.** The generally accepted belief is that the entire energy \( \epsilon_F \) (ergs) of a flare is stored above the photosphere and probably above the chromosphere. This belief is primarily rooted in the undisputed fact, first demonstrated by Parker (1957), that the overwhelmingly dominant form of energy stored above the photosphere is magnetic; in the context of Basic Physical Processes the PMFC’s have low \( \beta \) values above the photosphere until reaching altitudes above the transition zone, except in the plages and the chromospheric network. Since a potential field contains no magnetic free energy, current carrying PMFC’s are assumed to be the source of flare energy, and it is further assumed that \( \epsilon_F \) is stored in these currents above the photosphere.
Consider this point in more detail. Notice that one obvious conclusion inherent to the premise that all of \( \epsilon_F \) is stored above the photosphere is that externally driven flare mechanisms such as neutral sheet or double layer models are eliminated a priori. Indeed, one of the most interesting neutral sheet models yet proposed—that of the emerging flux model or its alternative the emerging sheared loop model—is eliminated unless one is prepared to accept the hypothesis that the flare has its origin in an external driver. In fact, what could be more reasonable than to attribute photospheric or subphotospheric energy sources as the primary cause of flares in view of the fact that the photosphere is “wagging” the corona and not the inverse? However, this is not the whole story, for implicit to both the emerging flux model (e.g., Heyvaerts et al., 1977) and the sheared loop model (Spicer, 1976, 1977a) is the requirement that the initial release of energy caused by the respective models be able to propagate outward and thereby destabilize other nonpotential PMFC’s such as prominences. This has the effect of greatly increasing the volume in which the flare energy is released while simultaneously expanding on the available magnetic free energy supply. One can now envision three possible flare scenarios: (1) All of \( \epsilon_F \) is stored in situ above the photosphere in some equilibrium or slowly evolving steady state and then released. (2) An externally driven mechanism such as a double layer causes the total \( \epsilon_F \) to be released from some given volume. (3) An externally driven mechanism releases some fraction of \( \epsilon_F \), which in turn propagates outward, expanding both the available magnetic free energy supply and flare volume by destabilizing other PMFC’s. The choice between (1) and (3) will be determined by what fraction of \( \epsilon_F \) is required to excite neighboring nonpotential PMFC’s, since the external driver can be viewed simply as a trigger if the fraction is small, whereas it must be considered as the primary flare mechanism if it is large. Another point to consider is the energy density of the magnetic free energy. If the energy density of the PMFC is low, the first scenario will require large volumes to accommodate \( \epsilon_F \). If, however, the energy density of the PMFC is high before its release, the first scenario will require a mechanism to cause the magnetic free energy to accumulate before it is released. This is not readily accomplished, since a PMFC will tend to expand to minimize its energy density with respect to its surroundings unless the energy accumulation rate exceeds the expansion rate.

In the case of scenarios (2) and (3), the preflare magnetic energy density must be reasonably high if the release of this energy is to either result in the whole of \( \epsilon_F \) being released in a time characteristic of a flare, or result in a disturbance sufficient to excite the neighboring nonpotential PMFC’s into releasing their energy. For driven systems, high magnetic free energy densities are, by definition, required; otherwise the thresholds necessary for their occurrence will not be reached.

**Energy Release.** The magnitude of \( \epsilon_T \), released in a given flare model is the flare model requirement usually offered as evidence that a given model is workable. The general approach is to take a given magnitude for \( B \) and calculate the total magnetic energy stored in the volume of some PMFC, that is,

\[
\epsilon_T = \int \frac{B^2}{8\pi} \, d^3x ,
\]

and then to demonstrate that \( \epsilon_T \) is released in some time characteristic of a flare. The authors feel this approach is unacceptable because it does not allow for the fact that no flare mechanism will ever dissipate 100 percent of \( \epsilon_T \) but rather some fraction, since \( \epsilon_T \) has both potential and nonpotential components. This point is dramatically illustrated by the fact that the loop PMFC causing the compact loop flare still exists after the flare; that is, the total field was not dissipated. What this implies is that \( \epsilon_T > \epsilon_F \), if the \( \epsilon_F \) is to be stored above the photosphere before the flare. Of course, this condition does not apply to driven
flare mechanisms, since the flare energy supply is external to the locus of the flare energy release.

The point concerning the available magnetic free energy can be expanded further by noting that, if a flare mechanism is to heat a plasma it must overwhelm the various energy sinks. To see this, note that in the total energy density of the plasma within the PMFC in steady state before the flare is roughly $NkT + \tau P_{\text{loss}}$ and that a source of energy $E$ per unit volume is required corresponding to the atmospheric heating mechanism, which maintains the energy density at its steady-state value, such that

$$E = NkT + \tau P_{\text{loss}}, \quad (18-99)$$

where $P_{\text{loss}}$ represents the power density lost by radiation, thermal conduction, and convection, during a period of time $\tau$, which corresponds to the cooling time of the most efficient cooling mechanism, and $NkT$ is just the total internal energy density of the plasma. Thus, to heat the plasma in a time $\tau'$ by some mechanism which generates an input power density $P_{\text{in}}$, requires

$$(\epsilon_0 \tau' P_{\text{in}} + E) > (NkT + \tau' P_{\text{loss}}), \quad (18-100)$$

where $\epsilon_0$ represents the efficiency of the coupling of the heating mechanism to the plasma. Since, in general, $\epsilon_0 \tau' P_{\text{in}}$ is much greater than $E$ except during the cooling period, and since $E > NkT$, Equation (18-100) reduces to

$$\epsilon_0 \tau' P_{\text{in}} > \tau' P_{\text{loss}}. \quad (18-101)$$

In the gradual phase of the flare, Equation (18-100) becomes

$$\epsilon \tau'' P_{\text{in}} \leq NkT + \tau'' P_{\text{loss}} - E, \quad (18-102)$$

where $\tau''$ is the lifetime of the gradual phase of the flare, and where $E$ has been included again in Equation (18-102). Including $E$ again follows because there is no a priori reason to assume that $E$ is shut off during the flare. It is noteworthy that an efficiency factor is associated with the coupling of the heating mechanism to the plasma, one that is not usually included in flare modeling analysis. This is more significant than it seems, since no mechanism heats a plasma with 100 percent efficiency; lower efficiencies must be allowed for in the model.

As noted above, $P_{\text{loss}}$ includes radiation, thermal conduction, and convective losses. It is of interest to examine each of these mechanisms for a given flare mechanism. Assume that $\epsilon_0 \tau' P_{\text{in}}$ is sufficient to cause heating so that the temperature rises. As is well known (Cox and Tucker, 1969) the radiation losses in a solar plasma decrease with temperature from $\sim 10^5$ to $\sim 10^7$ K, at which temperature they increase again for constant density conditions. On the other hand, thermal conductive losses parallel to $\mathbf{B}$ increase rapidly with temperature as $T^{7/2}$, while convective losses will only be important if any large pressure gradients exist. One might ask: Under what circumstances are thermal conduction and convection parallel to $\mathbf{B}$ unimportant, and what flare mechanisms operate under these circumstances? The answer to the first part of this question is that thermal conduction and convection parallel to $\mathbf{B}$ become unimportant when the flare mechanism causes heating over a sufficiently large temperature scale length $\delta L_\parallel$ parallel to $\mathbf{B}$. This is a result of the fact that the cooling time by thermal conduction parallel to $\mathbf{B}$ scales as $(\delta L_\parallel)^2$, and the cooling time by convection parallel to $\mathbf{B}$ scales roughly as $\delta L_\parallel$. 

451

Provided by the NASA Astrophysics Data System
Thus, to find a flare mechanism that satisfies this constraint, it requires only that it be a long wavelength heating mechanism parallel to B, that is, the wavelength parallel to B must be comparable to the length of the PMFC. By definition such mechanisms are fast tearing modes and neutral sheets. Double layers do not satisfy this requirement while anomalous Joule heating may, if it occurs along a sufficiently large distance parallel to B. What this requirement implies is that long wavelength heating mechanisms are more efficient plasma heating mechanisms than short wavelength mechanisms, since they need only overwhelm the radiation losses within the scale length $\delta L_\parallel$. End losses by conduction and convection must be considered also, but they should expect them to be important only if the ratio of the area of the ends to be heated volume is large, which is not likely for the mechanisms cited. Thermal conduction losses perpendicular to B will be negligible even if the conduction coefficient takes the Bohm value. In low $\beta$ PMFC’s convective losses perpendicular to B are negligible, since $\nabla P \approx 0$, except perhaps during the initial heating phase, while for high $\beta$ PMFC’s convection can be very important. However, convection losses will be important in neutral sheets, only within the sheet itself where $\beta$ is large, and most of the energy will probably go into macroscopic fluid turbulence and fluid flows.

**Particle Acceleration and Heating.** Particle acceleration is perhaps the most controversial aspect of solar flare theory, apart from the question of which mechanism is the cause of the flare. The fact that particle acceleration is treated as a subject area independent of the primary flare energy release mechanism (Ramaty et al., 1980)* is paradoxical since every mechanism discussed so far accelerates particles. The really relevant question is: Can any of the mechanisms so far proposed accelerate the numbers of electrons and ions required by the observations? The phrase “required by the observations” contains the origins of the paradoxical situation noted above, since solar astronomy is an observational as opposed to an experimental science. Such a situation leaves the theoretician and data analyst in a position in which different interpretations of the same data are fully consistent with that data. A classic example of this situation is the proof of Brown (1974) that a multithermal plasma can mimic any observed nonthermal collisional bremsstrahlung spectrum. One could go even further and state that some multithermal plasma can mimic any observed single nonthermal spectrum. For example, a multibeta-Maxwellian (thermal plus superthermal) distribution can mimic a power law distribution.

The solution of the integral equations involved is inherently ambiguous, and interpretation is made more difficult by the presence of noise in the data (Craig and Brown, 1976; Brown, 1978). In the case of solar hard X-ray bursts this lack of decisive data leaves open at least two interpretations, one of which leads to the so-called “electron number problem” (Hoyng et al., 1976; Brown and Melrose, 1977) when the bremsstrahlung X-ray spectrum, which usually “fits” a power law spectrum (over about one decade in energy), is interpreted as being caused by a purely nonthermal power law electron distribution colliding with a cold dense target. The analysis of the impulsive hard X-ray data from the ESRO TD 1A satellite (Hoyng et al., 1976) illustrates the so-called “thick target interpretation” problem (e.g., Brown, 1971, 1975; Kane, 1974). These authors show that a thick target interpretation requires about $10^{36}$ 25 keV electrons per second be accelerated for approximately $10^4$ to $10^5$ seconds, implying that $10^{37}$ to $10^{39}$ 25 keV electrons are necessary to explain medium-large, impulsive hard X-ray bursts. This further implies $10^{30}$ to $10^{32}$ ergs is carried by the streaming 25 keV electrons. Ignoring the question of return currents (Hoyng et al., 1976; Brown and Melrose, 1977) associated with such an electron stream, the thick target interpretation is a theoretical dilemma since one can either accept the thick target interpretation as correct and ponder the question of how any mechanism can accelerate $10^{36}$ 25 keV electrons per second.

---

*This review discussed only the first stage acceleration phase and not the second phase.
or one can challenge the thick target interpretation. Keeping in mind that there is nothing sacred about power laws, and that any multidistribution of thermal and nonthermal electrons (or one or the other) can mimic them, an alternative simple interpretation of the power law spectrum is a multithermal plasma interpretation (Chubb, 1972; Brown, 1974, 1975). This interpretation is diametrically opposed to the nonthermal thick target interpretation in that it assumes that the 25 keV X-ray flux has its origins in the multithermal plasma, that is, in a relaxed electron distribution. It has the additional strength that far fewer electrons are necessary to generate the same X-ray spectrum than in a nonthermal target interpretation. This follows because the nonthermal electrons in a nonthermal source lose their energy primarily in collisions with the cold ambient electrons. In a confined thermal source, on the other hand, the only losses are the bremsstrahlung radiation and the collisional energy losses with the ions, if the electrons and ions have not equilibrated. The electron numbers in these two cases are $10^{-5}$ and $10^{-3}$ of those for a thick target interpretation. On the other hand, because of lack of confinement in practice, conductive losses reduce these improvement factors to $10^{1}$ to $10^{2}$. For detailed discussions of the strengths of multithermal versus nonthermal interpretation the reader should consult (Brown, Melrose, and Spicer, 1979; Smith and Lilliequist, 1979).

With these two interpretations in mind, one can ask what flare mechanism is consistent with which interpretation of the X-ray bremsstrahlung. Consider the nonthermal target interpretation first. Of all the possible mechanisms, only the $J_{||}$-driven double layer mechanism can come close to satisfying the thick target interpretation, since it is primarily a particle acceleration mechanism. One can use Equation (18-98) to estimate the energy of the electrons accelerated through the double layer. To achieve this one first requires the cross-sectional area of the double layer. Since the threshold condition for a current to form a double layer is $V_{D} \geq V_{Te}$, and since it must be maintained at this value for the duration of the impulsive phase, the area is just $2\pi \delta r$, where $\delta r$

is the width in which the current flows. Using Ampere's equation yields

$$\delta r = \left( \frac{c}{\alpha_{pe}} \right) \delta B^{1/2} ,$$  \hspace{1cm} (18-103)

where

$$\delta B = \frac{8 \pi N_{e} k T_{e}}{\delta B_{p}^{2}}$$

and $\delta B_{p}$ is the change in the field component induced by $J_{p}$ across the layer $\delta r$ thick. Hence, Equation (18-98) becomes

$$e \phi \approx 5 \times 10^{32} \left[ \frac{N_{T}}{N_{e}^{2} \Delta t} \right]^{2} \beta , \text{ keV} .$$  \hspace{1cm} (18-104)

By using $N_{T} \approx 10^{9}$, $N_{e} \approx 5 \times 10^{11}$ cm$^{-3}$, $\Delta t \approx 100$ s, $r \approx 5 \times 10^{8}$ cm, and $\beta \approx 10^{-3}$, one finds $e \phi \approx 4 \times 10^{8}$ keV an energy totally unacceptable observationally. Alternatively, one can use Equation (18-97), employing similar arguments for the area, to obtain

$$N_{T} \approx 4 \times 10^{15} N_{e}^{1/2} r \Delta t \left( \frac{\phi}{\beta} \right)^{1/2} \hspace{1cm} (\phi \text{ in keV}) .$$  \hspace{1cm} (18-105)

By taking $e \phi \approx 25$ keV and the values for $N_{e}$, $\Delta t$, and $r$ previously used, one gets $N_{T} \approx 8 \times 10^{33}$ electrons, which is about 1000 times less than the necessary number. Notice that a lower $N_{e}$ will
reduce \( N_p \), while \( N_e \) will increase linearly with \( \Delta t \) and \( r \), but even for a judicious selection of these parameters it seems scarcely possible to obtain agreement between theory and observation. In addition, the authors feel the double layer mechanism has theoretical problems, if it is to explain the electron numbers required of the target interpretation.

Turning to the thermal interpretation, it is obvious that reconnection and anomalous Joule dissipation mechanisms, both \( J_\parallel \) and \( J_\perp \), result in primarily thermal plasmas with nonthermal tails. Previous estimates by Spicer (1981a) of the temperatures expected from \( J_\parallel \)-driven anomalous Joule dissipation and fast collisional tearing modes indicate electron temperatures of roughly 10 keV for both. Similar results can be expected for neutral sheets (cf. Heyvaerts et al., 1977). The total number of particles accelerated by the anomalous Joule dissipation mechanism and by fast collisional tearing modes have also been estimated by Spicer (1981a), using Equation (18-38) with both the anomalous and the classical collisional frequencies. The numbers obtained still fall short of the numbers required of the thick target approximation, indicating that only weak nonthermal tails are to be expected. Hence, both anomalous Joule dissipation and fast tearing modes are inconsistent with the nonthermal thick target approximation but consistent with the thermal approximations. There is no currently proposed energy release mechanism that is directly consistent with the nonthermal interpretation of hard X-ray bursts. Hence, if the observations require acceptence of the thick target approximation in the future either the flare mechanisms so far discussed must be abandoned, or some additional acceleration mechanism (or mechanisms) must be at work. Langmuir wave acceleration is one such mechanism (Benz, 1977; Melrose, 1978; Hoyng, 1977a, 1977b; Smith, 1977a, 1977b; Hoyng et al., 1980).

**Mass Ejections.** As discussed in Chapter 7, only a small fraction of flares exhibit mass ejections, and it constitutes a large fraction of the total flare energy released when it does occur. However, the question arises whether the energy associated with mass ejections is really a result of the primary energy release mechanism or is it an indirect result of the triggering of another mechanism, for example, a prominence as discussed under energy storage. If the former possibility is correct, the primary energy release mechanism must provide a pressure pulse constituting a major fraction of the energy release during a flare. Placing theoretical constraints on the primary energy release mechanism which could cause mass ejections is, at this time, a purely subjective issue, since the available observations do not allow one to distinguish between the two choices mentioned above. Nevertheless, the authors feel that if the mass ejections observed are not a result of the triggering by the primary energy release mechanism of other mechanisms into activity, then the energetics of mass ejections can place strong constraints on the primary energy release mechanism by a combination of high power/volume output from the mechanism (local constraint) and a high integrated power output from the mechanism (global constraint).

**Miscellaneous Flare Model Requirements.** Probably the most obvious flare model requirement is that the PMFC of the model be consistent with the observations. Thus, one requirement is that the flare model employ a mechanism that will work in simple bipolar magnetic regions, since observations clearly demonstrate that some flares occur in such regions (Svestka, 1976; Sturrock et al., 1980). Thus, the model must use not only a mechanism that will operate in a bipolar magnetic region but it also must start with (or produce) a PMFC consistent with the observed PMFC, which appears to be a loop in the case of compact flares and appears to be multiple loops for large double ribbon flares (i.e., assuming that EUV structures delineate field lines).

Other obvious requirements (all of which are discussed in Chapter 7) are that a model needs to be able to explain homologous flares, sympathetic flares, temperature minimum heating, the close relationship between high photospheric shear and flares, the relationship of emerging flux with flares, and the simultaneity (or near simul-
taneity) of hard X-ray, microwave, EUV, and Type III bursts. Theoretically, the existence of homologous flares implies that the basic PMFC in such cases is not destroyed by the flare, which is consistent with the points made in the sections on Energy Storage and Energy Release. It also suggests a high induction circuit model, where the total current, $I$, is maintained or replaced after the flare mechanism dissipates the driving energy. Since a high induction circuit model has a high $L/R$ time, the flare mechanism in such a situation must be associated with an $L^{-1}(dL/dt)^{-1}$ time. This means that either the flare mechanism itself generates changes in the inductance, or that $\mathbf{v} \times \mathbf{B}$ forces in the photosphere or convection zone provide such changes. In the former case, tearing modes cause such changes, while in the latter the twisting of the field lines by $\mathbf{v} \times \mathbf{B}$ forces will produce $dL/dt$. (Some laboratory devices such as Tokamaks possess a high $L/R$ time, which is maintained by external circuits, and tearing modes are observed to occur in such devices, giving rise to what are known as negative voltage spikes which are very large in magnitude, e.g. Bateman, 1978.) Sympathetic flares, if they exist (Švestka, 1976), argue in favor of preflare storage of energy. Thus, a model that uses a driven mechanism such as double layers, must also explain why homologous flares exist, if indeed they do exist.

Heating of the temperature minimum region (Machado et al., 1978) is a recent development that needs further exploration. Any model that uses a transient $J_{||}$ driver may be able to explain temperature minimum heating by Joule heating, since the classical resistivity at the temperature minimum is dominated by electron neutral scattering of current carrying electrons. It is trivial to demonstrate that the current densities required to heat the temperature minimum during a flare are roughly $10^6$ statamps/cm$^2$ which implies mean magnetic fields of about $10^4$ gauss distributed over a total area of $\sim 10^{16}$ cm$^2$. Since these fields would necessarily be transient this may not be unacceptable. Alternatively, if $\int J_{\perp} \cdot dS \ll |J| S$ (Emslie and Machado, 1979) Joule heating may still be viable. Such a situation will occur if a set of antiparallel currents were concentrated in very small areas within a typical element of spatial resolution. Then it is possible to raise the Joule heating greatly in a local region, since this is $\int \eta F dS$, without increasing the total current $\int J_{||} dS$. Such multiple reversals, if they exist, would be susceptible to fast multiple tearing modes and would favor a $J_{||}$-driven flare mechanism.

The frequent occurrence of flares in magnetic field configurations with high photospheric field shear (Švestka, 1976) needs to be reconciled with a given model. Such shear (not to be confused with magnetic shear defined in Ideal MHD Instabilities) results from a displacement of the footpoints of one side of a PMFC with respect to the other as measured from the so-called neutral line. Since magnetic shear would be expected if field lines at different altitudes traversing the neutral had different directions with respect to one another because of photospheric shear, tearing instabilities driven by $J_{||}$ can be expected.

Emerging flux appears to be a sufficient condition for a flare, although not a necessary one. Nevertheless, models must make allowance for emerging flux. Present models have only considered emerging flux as a driver when it interacts directly with preexisting PMFC's to excite a flare. However, emerging flux may represent a manifestation of a more general phenomenon occurring in the photosphere or convection zone. For example, the emerging flux might induce electric fields within other preexisting structures and excite flares, rather than directly interacting with the preexisting PMFC's.

The observed simultaneity of electromagnetic bursts represents a problem for any model. Such simultaneity can be accomplished by a variety of means, for example, by electron beams or by long wavelength mechanisms with high characteristic velocity, that heat the PMFC over a significant fraction of its length.

The escape of high energy electrons to produce Type III bursts can present difficulties for some models which utilize closed rather than open PMFC's. However, the argument can be turned around to make difficulties for open PMFC's because most flares do not cause Type III bursts.
suggesting that a closed PMFC is more common. Nevertheless, the escape mechanism must be explained and it must work within a time short enough to maintain the observed simultaneity (or near simultaneity) with other bursts.

**Flare Triggers**

Historically there has always been the belief (cf. Sweet, 1969) that a flare requires some external perturbation, or trigger, to excite it. This belief has no observational justification at the present time. However, if one accepts the contention that the flare energy is stored in situ above the photosphere before a flare, the need arises for a trigger mechanism (or mechanisms). On the other hand, mechanisms which require external drivers do not need triggers, since no more is necessary than to switch on the driver; for example, in neutral sheet models, mass flow or emerging flux must appear and exist for the duration of the energy release. This situation is physically distinct from a trigger, since once the perturbation pushes the system “over the brink” the system is on its own. In this regard, Sturrock (1966a) has noted that there are essentially two types of instability onsets: explosive and non-explosive. The explosive onset occurs when a system is near marginal stability and linearly stable to infinitesimal perturbations but nonlinearly unstable to finite perturbations. On the other hand, the nonexplosive onset occurs when a system is linearly unstable to infinitesimal perturbations, Sturrock (1966a) further argues that an instability must result from an explosive onset if it is to be viable as a flare mechanism. This point of view can be understood by recognizing that, for preflare storage of energy to occur, energy must accumulate in the PMFC in such a way that it is not continuously dissipated because of instabilities excited by infinitesimal perturbations, but rather the energy must accumulate up to some magnitude beyond which a finite perturbation can drive it to instability, which in turn releases the accumulated energy. The questions thus arise: What kinds of instabilities require explosive excitation, and how can energy be accumulated without continuous dissipation? Both of these questions are difficult to answer, but one can gain some insight into them. Consider the latter question first. Aside from instabilities, a PMFC has two important means by which it can lower its magnetic energy density: either by classical Joule dissipation or by expansion. It is easy to show that the classical Joule dissipation time will generally be very large, because the resistivity is so small in the solar atmosphere (except at the temperature minimum), making classical Joule dissipation ineffective for preventing magnetic energy accumulation. However, plasma expansion can lead to serious difficulties if the rate of energy accumulation is less than the rate at which plasma expansion reduces it. All PMFC’s have a tendency to expand in an open system such as an exponential atmosphere. Hence, one must either prevent the expansion or increase the energy accumulation rate beyond the expansion rate. In laboratory devices, expansion (which defeats plasma containment for fusion) is prevented by the use of conducting walls and externally imposed magnetic field configurations. Hence, in laboratory devices magnetic energy accumulates, and it is this accumulation of magnetic energy that causes the violent phenomena commonly observed in these devices, such as kinks or tearing modes. One must therefore raise the question: How is a PMFC in the solar atmosphere prevented from expanding? There does not appear to be an easy answer to this question, although it is obvious that additional magnetic fields prevent such expansion in the solar atmosphere if they are nearly potential to begin with and have little expansion energy intrinsic to them, or, alternatively, if they contain additional materials in the magnetic field region. Hence, for preflare energy storage it is not useful to use the term equilibrium configuration in its precise sense, but rather one must assume from the outset that any PMFC considered must be evolving on a time scale far longer than any instabilities expected in it. In this context, some recent work has assumed that the preflare PMFC is slowly evolving, accumulating energy until the system shows indications of losing its quasiequilibrium (Jockers, 1976; Low,
1977; Anzer, 1978; Birn et al., 1978; Heyvaerts et al., 1979; Priest and Milne, 1979; Hood and Priest, 1980). The principle involved is related to bifurcation theory, since, as the relevant parameters are varied, there is a point of bifurcation at which the equilibrium can either go unstable, or go to a more stable regime. Usually the point of bifurcation separates a class of stable equilibria from a class of unstable equilibria. Though bifurcation theory is on firm ground theoretically, its application to solar flare theory should be viewed with caution, because all treatments to date have imposed some constraint from the outset, such as a symmetry on the PMFC. This has the unfortunate result of reducing the degrees of freedom that the PMFC has available to lower its energy. For example, suppose the condition of axial symmetry on a PMFC such as a prominence were imposed, while simultaneously assuming it to be force-free. Can one then believe the argument that after twisting the PMFC, it will erupt at some stage into an open PMFC, simply because the force-free PMFC has a higher magnetic field energy density than does the open PMFC? The answer may, in reality, be yes, but on the basis of arguments from bifurcation theory with the above constraint the answer is no! By imposing axial symmetry, all ideal and resistive MHD instabilities with $|m| \geq 1$ have been suppressed. In addition, by imposing the force-free condition, the $m = 0$ mode has been eliminated, since the $m = 0$ mode is easily stabilized in a force-free system ($\beta = 0$) (Newcomb, 1960). Hence, by not being careful, it is possible to eliminate a priori just about every degree of freedom the PMFC had to lower its energy, which makes any conclusions such as those mentioned above fallacious.

The question of preflare storage and triggering is also complicated by the fact that, for those mechanisms in which preflare energy storage is, in principle, possible, the stability analysis currently being used has only global validity for PMFC's with high degrees of symmetry and homogeneity, which solar PMFC’s definitely lack. Hence, many of the results discussed so far are only locally valid and, as is well known, local stability tells one nothing about global stability. For example, consider the kink or tearing instabilities, both of which are global modes.

As discussed earlier, one condition for the occurrence of these global modes with high symmetry and homogeneity is that $k \cdot B$ vanish. This requires that the PMFC be homogeneous in two spatial directions. However, such is not the case in solar PMFC's, which are intrinsically three-dimensional. Hence, the condition that $k \cdot B$ vanish can, at best, be a local condition in a realistic three-dimensional situation, whereas the global condition will be made up of differential operators in those directions previously assumed to be homogeneous. This will have the result that threshold conditions usually advanced for stability are no longer valid and must be used with caution.

A further problem associated with flare triggers and preflare storage is that, while the assumption of force-free PMFC's is reasonable, with the external stresses being accommodated by boundary conditions, the question of how plasma transport processes control parameters such as current density gradients is not accommodated in standard analyses. For example, it is well known that currents tend to flow in regions of low resistivity while avoiding regions of high resistivity, so that one would expect currents to concentrate in regions of high temperature when classical resistivity applies. However, no allowance is made for such inhomogeneous transport effects on the global structure of PMFC's in standard force-free equilibrium analysis (see, for example, Ferraro and Plumpton, 1966). The force-free equilibria usually advocated are thus of questionable validity for solar PMFC's, because even a cursory examination of the observations illustrates that there are strong temperature and density inhomogeneities in solar PMFC's; for example, a prominence has a hot transition sheath (Schmahl, 1979).

Many instabilities have been suggested as triggers for flares, but perhaps the instability most widely believed to have some validity is the so-called thermal instability (Parker, 1953; Field, 1965). Thermal instabilities have their origin in the behavior of the radiative losses with temperature; for example, if the radiative losses decrease with temperature as they do for solar
Abundances at temperatures of $10^5$ K and the thermal conductivity is somehow prevented from removing any excess input of energy, then the temperature of the plasma will increase until either the radiative losses increase with temperature again, or until plasma expansion cools the plasma at a rate faster than the external heating rate. However, the concept of thermal stability has a more general applicability than stated above, and it is well worth examining some of the resulting physical effects which can occur in solar PMFC’s. As an illustration, consider the superheating instability (Kadomtsev, 1966), which arises as a result of the fact that Joule heating increases the temperature, $T$, and reduces the resistivity, $\eta$, so that the current density, $J$, increases, which can increase the Joule heating and so on, providing that radiative and conductive losses are overcome. The growth rate for the heating is just

$$\gamma = -\kappa_0^2 \beta - \kappa_1^2 - \nu_r + \nu_q \left( \frac{k_1^2 - k_1^2}{k_1^2} \right),$$  
(18-106)

where

$$\nu_r = \frac{2}{3N_e} \frac{dQ_r}{dT_0}, \quad \nu_q = \frac{2J_0^2}{3N_e T_e \sigma_0} \frac{d\ln \sigma_0}{d\ln T_e}.$$

This corresponds to steep temperature gradients on either side of the surface where $k \cdot B = 0$ for sheared fields and neutral sheets; hence, it may provide a stabilizing effect. Suppose that $Q_r = 0$, $k_0^2 \rightarrow 0$, and the term $\kappa_1^2$, is ignored. It follows that $\gamma = -\nu_r$. Hence, radiative instability can occur if $dQ_r/dT_0 < 0$, which occurs for solar plasmas (cf. McWhirter, et al., 1975), Sweet (1969) and Kahler and Kreplin (1970) were the first to point out its possible importance in flares.

Next assume isobaric perturbations and that there exists an external heating function $C$ that is a function of $T$ only. This implies that

$$\gamma = -\nu_r + \frac{2}{3N_e} \frac{dC}{dT_0}. \quad (18-107)$$

Hence, if $(2/3N_e)(dC/dT_0) > -\nu_r$ or if $\nu_r < 0$, instability can result. However, for solar plasmas $Q_r$ has both local maxima and minima, implying that unless $(2/3N_e)(dC/dT_0) >> -\nu_r$ always for $\nu_r > 0$ the system can become stabilized around one of the minima, though it will still be unstable to small but finite perturbations. In that case, one should expect $\gamma = 0$ (marginal stability) to determine the marginal temperature.

Returning to the superheating instability (i.e., $C$ is determined by Joule heating), it should be noted that if $k_0^2 \rightarrow 0$ instability results when

$$\frac{2J_0^2}{3N_e T_e \sigma_0} \frac{d\ln \sigma_0}{d\ln T_e} > -\nu_r, \quad (18-108)$$

$d\ln \sigma_0/d\ln T_0 > 0$, which it is when the resistivity is classical. The superheating instability will then result in highly elongated hot current filaments parallel to $B$. In sheared fields these filaments will be elongated and localized around $k \cdot B = 0$. Various authors have proposed the superheating instability as a flare trigger in the context of neutral sheets (Coppi and Friedland, 1971;
Heyvaerts, 1974a) and in sheared fields (Spicer, 1976, 1977a), while nonlinear studies by Sokolov et al. (1977) have demonstrated its applicability to solar conditions.

The discussion is so far only locally valid, since all studies to date have made the WKB approximation. That is, the assumption is made that the wavelength of the perturbation is shorter than the characteristic temperature scale length. Such an approximation ignores the fact that the temperature has steep gradients in various parts of a given solar PMFC. Hence, these arguments are not globally valid at present, but they do indicate some of the basic phenomena expected.

In summary it should be evident that the question of the flare triggering and the associated problem of preflare energy storage are still in the early development stage. However, it should also be evident that these two problems are probably the most fundamental issues in solar flare theory for those who believe in preflare storage of energy, aside from the primary energy release mechanism itself.

Flare Models

The intent in this review is not to review each model in detail, but rather to classify each model according to the mechanism it uses and its source of driving energy. The authors feel that this is the most appropriate approach since the only basic difference between different models that use a given mechanism and driver is the geometry of the models and in what form the driver manifests itself. Another reason for this approach is that most models, except some very recent models, are highly approximate. They attempt to mimic in a very qualitative fashion the observations which have been made up to the time the given model was prepared. If the reader will examine the original papers for each of these models, he will find that, generally speaking, each model followed but did not lead observations. In this sense, very few models have ever made any useful predictions which permit a valid observational test of the model. In cases where predictions are made, the predictions are often of a nondefinitive nature, in that other models using the same mechanism and driver would also make the same predictions. Finally, one finds predictions which require observations on some unsolvable scale. The current vogue seems to be for loop flare models (Alfvén and Carlqvist, 1967; Spicer, 1976, 1977a, 1981a; Colgate, 1978). It might be noted for historical accuracy that the Alfvén-Carlqvist model was proposed long before it was evident from the data that some flares, so-called compact flares, consisted of one or more loops; so, in this single instance, theory led experiment!

It was noted earlier that other mechanisms have been proposed which fall outside the classification scheme given here. There are basically three: the bottling up of Alfvén wave energy (Pneuman, 1967; Piddington, 1973, 1974), the release of trapped energetic protons and electrons from high altitude mirror configurations (Elliot, 1964, 1968), and adiabatic heating (Crannell et al., 1978; Mätzler et al., 1978). The latter was actually proposed for local high temperature effects rather than as a flare mechanism. Of the three, the most difficult to access theoretically is the Alfvén wave model, because it has not progressed much beyond the hypothesis stage; namely, it has not been demonstrated that Alfvén wave energy can be trapped and stored in the manner envisaged by Pneuman (1967), because there are many processes which can lead to the decay of Alfvén waves. Consequently, Alfvén wave trapping may not occur (cf. Wentzel, 1974, 1977).

The release of trapped energetic protons (200 MeV) and electrons into the conjugate mirrors of a trapping configuration is a very appealing idea, if it can be demonstrated observationally that such particles are indeed trapped before a flare with sufficient numbers. A modification of this idea (Spicer, 1976) is that a shock driven by a pressure pulse or by an erupting prominence can couple to preexisting mirror configurations and drive the various crossfield instabilities listed in Table 18-1. These, in turn, can lead to highly anisotropic ion and electron distributions, which will then precipitate into the mirror conjugates.
This hypothesis eliminates the problem of preflare storage, but the problem of the origins of the shock still remain. Nevertheless, the hypothesis, originated by Elliot (1964, 1968) and modified by Spicer (1976), has promise for the second stage acceleration phase of major double ribbon flares and may help explain the impulsive electromagnetic bursts generated during that phase.

The adiabatic heating hypothesis (Crannell et al., 1978; Mätzler et al., 1978) was introduced because a certain select class of hard X-ray bursts showed a vaguely reversible time profile, which is sometimes indicative of adiabatic heating and cooling. These bursts were also shown to be fully consistent spectrally with a single thermal source. The idea envisaged by these authors was that the heating and cooling of the plasma occurred by adiabatic heating from compression and adiabatic cooling by expansion, instead of heating by a dissipative, nonadiabatic mechanism and cooling by, say, conduction. The only means of compressing a plasma adiabatically in the solar atmosphere is by radial or axial magnetic compression. The X-ray bursts observed had a time scale far exceeding the collisional relaxation time of the plasma. Under these conditions, it can be shown that radial adiabatic compression by an external magnetic field of a plasma with a weak trapped magnetic field leads to the relation (Glasstone and Lovberg, 1960)

\[ T = T_0 \left( \frac{B}{B_0} \right)^{4/5} \]

(18-109)

where \( T_0 \) and \( B_0 \) are the initial temperature and field, respectively. Hence, to radially heat a plasma adiabatically from, say, 0.5 keV to 40 keV one requires about 240-fold increase in \( B \). If the initial internal field is comparable to the initial external \( B \), even larger increases are required. Axial compression yields the relation

\[ T = T_0 \left( \frac{V_0}{V} \right)^2 = T_0 \left( \frac{V_0}{V} \right)^2 \]

(18-110)

which requires a factor of \( \sim 9 \) reduction in the axial scale size of the containing PMFC, where \( V_0 \) is the initial axial scale length of the system, \( V_0 \) is the initial volume, and one has assumed no loss of particles from the mirror field so that \( TV^2 \) is constant. Radial compression and expansion implies that the emission measure varies as \( T^{3/2} \), while for axial compression and expansion the emission measure varies at \( T^{1/2} \).

Next consider the question of the initial field strength versus the final field strength after compression. Starting with a 0.5 keV plasma with a density of \( 10^9 \) as deduced by Crannell et al. (1978) and Mätzler et al. (1978), a minimum start field of at least 5 gauss is required for plasma containment. Hence, the final field must be \( \sim 1.2 \times 10^3 \) gauss if radial compression is to work. Further, since the emission measure obtained by these authors was \( \sim 5 \times 10^{44} \), they require a volume of at least \( 5 \times 10^{26} \) cm\(^3\), which means the emitting volume has a characteristic scale of about \( 10^9 \) cm. Fields of \( 1.2 \times 10^3 \) gauss over scales of \( 10^9 \) at altitudes of \( 10^9 \) cm are difficult to reconcile with observations. Also, as noted by Brown and Smith (1979) there is no obvious reason why an adiabatic process must be symmetrical during its heating and cooling phases, as assumed by the authors of the model. Furthermore, even the observed reversibility is unconvincing since the ratios of rise time to fall time for X-ray bursts range from \( \lesssim 1/3 \) up to 3 (Crannell et al., 1978) and the reversibility is not confirmed in otherwise similar events seen by other observers. Also, most adiabatic heating theory ignores nonadiabatic cooling mechanisms such as radiation, which will alter the functional behavior of temperature versus total emission measures noted above.

The flare models that use magnetic free energy are all basically compatible with the primary flare model requirements as they are presently understood. These are considered next. The essence of many of the flare models currently
in vogue is discussed. However, since many of these flare models are similar in many respects, the discussion will focus on describing that model in a class of models which is most complete, while citing references to the others.

Models That Utilize $J_\parallel$-Driven Mechanisms

These models fall into two basic categories: current interruption models and tearing mode models. They are considered in turn.

Current Interruption Model. This model, first proposed by Alfvén and Carlqvist (1967), was the first model to emphasize the loop structure recently detected by the Skylab ATM instruments as the basic structure of small compact flares. It was also the first model to emphasize the importance of viewing the flare energy release as a form of current dissipation whatever the proposed mechanism. Alfvén and Carlqvist (1967) also made one other very important contribution aside from their model, by emphasizing the necessity of treating the whole current system. They emphasized that the problem of dissipating current is not just a local problem but a global problem. Even though one may have found a mechanism that dissipates the magnetic free energy fast enough to explain the flare, it remains necessary to demonstrate that it is possible to replenish the magnetic free energy sufficient rate. This rate must be sufficient to maintain the critical parameters at the magnitudes necessary to sustain the flare mechanisms throughout the duration of the flare. In remarks made earlier, this replenishment mechanism was called the external driver.

The current interruption model presently has three versions: (1) a double layer model, (2) an anomalous Joule heating model, and (3) a hybrid model. The first of these was qualitatively developed by Alfvén and Carlqvist (1967) who argued that, if the current density within a magnetic loop reached a critical value characteristic of the value necessary to excite the ion acoustic or Buneman instabilities (Smith and Priest, 1972), then the development of very high electric fields might occur to maintain the original current across the regions where the resistivity became anomalous. These regions are presently called double layers, as discussed in the section on Double Layers. Note also from Table 18-2 that double layers require strong external drivers and that they can operate only in a transient situation. The second version of this model can be traced to the arguments made by Smith and Priest (1972) that the most likely energy release mechanism to occur in this situation would be simply anomalous Joule heating. However, Smith and Priest rejected this version of the model because of some arguments that have since been questioned. A recent treatment by Spicer (1981a) has demonstrated that the anomalous Joule heating version of the current interruption model might work if the strong external driver required is present. The hybrid version of the current interruption is based on maintaining the current density at marginal stability for a period of time longer than an MHD or tearing mode growth time, if the current density is sufficient to generate anomalous resistivity, as noted by Spicer (1981a). Hence, both $J_\parallel$-driven electrostatic instabilities, as well as $J_\parallel$-driven MHD instabilities will occur in such a model, making it difficult to distinguish between the competing $J_\parallel$-driven mechanisms.

The current interruption model in any of the three forms noted requires a strong transient external current driver, and this transient driver must initially raise the current density in a time shorter than an MHD kink or tearing mode growth time (Which Mechanism Occurs First?). Since the current interruption model utilizes the fact that the total current circuit has a high inductance, the external driver must have its source in convection, that is in the $\mathbf{v} \times \mathbf{B}$ forces located in the photosphere. This of course implies an inductance term, $L$, in the circuit equation with a time constant shorter than the $L/R$ time. While these physical requirements are incompatible with the prevailing opinion that the total flare energy is stored in situ in the solar atmosphere before the flare, there is presently no objective justification for discounting the possibility of a transient external driver.
Tearing Mode Model (TMM). This model, first introduced by Spicer (1976, 1977a), has undergone a number of revisions and improvements since its initial introduction (Spicer, 1978, 1981a, 1981b). The tearing mode model, like the current interruption model, utilizes a current carrying loop (though the geometry of the environment is not crucial to either model). However, it differs fundamentally from the current interruption model in that the $J_\parallel$-driven mechanisms used are the MHD kink and tearing mode instabilities. The tearing mode model was the first model to utilize the fast tearing modes. In addition, the importance of nonlinear mode coupling between different tearing modes, and the possibility of overlapping of resonances was first introduced into flare theory within the context of this model. Recent simulations (Waddell et al., 1979; Hicks et al., 1979) of mode coupling in Tokamak PMFC's have confirmed the argument (Spicer, 1976, 1977a) that nonlinear mode coupling of tearing modes should result in explosive reconnection. This supports the view that such mode coupling could be important in solar flares. Further, the importance of multiple tearing modes for certain types of current density profiles has been demonstrated (Spicer, 1976, 1977a, 1981a, 1981b).

The value of the tearing mode model is that it is capable of explaining many aspects of the impulsive phase of the flare without requiring anomalous resistivity and the associated high magnetic field gradients. In addition, the model is rich enough in possibilities to explain a number of phenomena common to flares, such as millisecond microwave burst structure (Slottje, 1978) and quasiperiodic bursts. Since the model was developed to explain the Skylab observations, it is presently consistent with many observations of compact flares.

Colgate (1978) has attempted to explain the gamma ray observations of solar flares, using the tearing mode model, by arguing that it is the ions which must carry the current, rather than the electrons as in the original tearing mode model. The reader should note that the Colgate refers to the tearing mode instability as a filamentation instability, which is true; however, the term filamentation instability is a generic term applying to a broad variety of instabilities that cause the current to filament (e.g., Cap, 1976). Colgate has also concentrated on the secondary processes associated with the tearing mode model, such as X-ray radiation. In doing so, he rejects nonthermal electron beam (i.e., target) interpretations of hard X-ray bursts on the grounds that the magnetic self-energy of the beam would be enormous. However, his treatment ignores return currents, and others have argued that this effectively negates his position on the magnetic self-energy (Hoyng et al., 1976; Brown and Melrose, 1977). Van Hoven (1979) has repurposed the tearing mode model from the above viewpoint.

The TMM can work (1) if $J_\parallel$ is externally driven, such as in the current interruption model, but with much weaker field gradients being allowed, or it can work (2) if the loop is emerging from beneath the photosphere and a skin current forms (which is expected as demonstrated by Parker, 1974) or (3) it can develop from a quasiequilibrium PMFC which contains enough shear energy to drive the flare (Spicer, 1977a, 1981a). As far as the latter possibility is concerned, the loop must be flat in the sense that it has a large radius of curvature; otherwise, expansion can become important, making preflare energy storage difficult.

Before proceeding to the $J_\parallel$-driven models, a number of points should be emphasized concerning the $J_\parallel$-driven models: First, they can occur in any PMFC geometry in which $\beta << 1$; that is, loops are not necessarily the only geometry in which they will work. For example, any force-free PMFC will do. Second, these models are, at present, applicable only to compact loop flares, although there is some evidence that the large double ribbon flares may consist of flat multiple loops flaring in a domino fashion. Third, all of the arguments concerning the $J_\parallel$-driven mechanisms can be applied directly to prominences with kinks (Sakurai, 1976) and tearing modes (Spicer, 1979a).
Models That Use $J_\perp$-Driven Mechanisms

Because of the multitude of $J_\perp$-driven flare models only two $J_\perp$ flare models will be considered here. The first of these models can be called the helmet streamer model, proposed by Carmichael (1964) and developed semiquantitatively by Sturrock (1966b, 1967, 1972, 1974). The second model, developed by a series of authors (Priest and Heyvaerts, 1974; Canfield et al., 1974; Heyvaerts et al., 1977; Tur and Priest, 1978), is referred to here as the emerging flux model. These two models are considered because the former model illustrates some of the difficulties with neutral sheet models and the latter illustrates how these difficulties may be overcome.

The Helmet Streamer Model. This model is essentially a neutral sheet placed on top of a closed PMFC (Figure 18-15); that is, a PMFC is assumed which has both a high $\beta (\gg 1)$ and low $\beta (\ll 1)$ part, as well as an intermediate $\beta$ regime. The argument goes that the neutral sheet can reconnect by tearing modes to give rise to the flare. Physically, there are a number of shortcomings associated with such a model: (1) As noted by Priest (1976), the available magnetic free energy in such a model, once dissipated, cannot be replenished automatically. Since the available magnetic free energy is determined by the pressure gradient across the neutral sheet, there must be an external driver to convect in new magnetic flux, or otherwise, only the magnetic free energy associated with $\nabla P$ will be dissipated. (2) Since $\beta \gg 1$ in an open PMFC, it is necessary for the effective plasma pressure to exceed the magnetic field pressure. However, the typical plasma pressure where helmet streamer PMFC's occur is, at most, 1.0 ergs cm$^{-3}$; hence, $B$ is at most $(8\pi)^{1/2} \simeq 5$ gauss. Such a PMFC is not expected to exist low in the solar atmosphere in regions capable of flaring. There the field is expected to be $\gtrsim 10^3$ gauss.

An alternate argument is that a closed configuration ($\beta \ll 1$) erupts into an open ($\beta \gg 1$) configuration, at which time reconnection occurs, as was proposed by Barnes and Sturrock (1972). Their argument is that when a force-
the PMFC to the outside. Since the PMFC has $\beta << 1$, the pressure difference is determined by the weak, albeit finite pressure gradient. From this it is easy to demonstrate that the force-free field will expand with an acceleration which is, at most, roughly $c_s^2/R$, where $R$ is the radius of curvature of the field lines and $c_s$ the sound velocity. If the closed PMFC is strongly out of equilibrium, the maximum acceleration for expansion is roughly $V_A^2/R$. A PMFC will only be strongly out of equilibrium if it is suddenly introduced into a region where its own total pressure greatly differs from the background pressure, as can occur with emerging flux or if the PMFC were to become unstable. Since it has been assumed from the outset that the PMFC is force-free, the former possibility is eliminated a priori, so that only the latter possibility is to be considered if the force-free PMFC is to erupt in a transient manner. Otherwise, the expansion would be very slow. Since the PMFC is, by assumption, a force-free one, the source of driving energy is $J_{\parallel}$, not $J_{\perp}$. This means that some combination of $J_{\parallel}$-type instabilities must be excited to increase $\beta$ so that the force-free PMFC will open. In other words, the flare would be the result of the opening process and not the reclosing of the field by reconnection at some later time, and the relevant mechanisms and models would be those discussed under $J_{\parallel}$ models.

Emerging Flux Model. The emerging flux model (Priest andHeyvaerts, 1974; Canfield et al., 1974; Heyvaerts et al., 1977; Tur and Priest, 1978) was developed to make use of observational evidence that emerging magnetic flux appears to be a sufficient condition for the occurrence of a flare. However, the model is limited to neutral sheet mechanisms. In addition, the emerging flux model clearly demonstrates the importance of considering an external driver, the emerging flux, in the flare process, and, in particular, the need for an external driver for $J_{\perp}$-driven models. The basic idea behind the model is that a neutral sheet develops between the newly emerging flux and the preexisting flux, and various $J_{\perp}$-driven mechanisms are subsequently excited if the speed of the emerging flux is great enough. It is further argued that the emerging flux model acts, in reality, as a trigger to excite preexisting neutral sheets within the preexisting PMFC in which the emerging flux couples. It is the opinion of the authors that, of all the $J_{\parallel}$ models, the emerging flux model is the most reasonable and useful as applied to neutral sheet mechanisms.

In comparing all of the flare models it is difficult to objectively assess which is closest to describing the actual flare process or which is the most useful in furthering our understanding. In detail, the tearing mode model and emerging flux models are the most comprehensive and quantitative. However, even though more detail and a greater degree of quantitative prediction is presently being achieved, none of the models is at the stage which would allow detailed predictions or comparisons with data. It can therefore be safely concluded that flare modeling, while probably already using the correct mechanism or mechanisms, is still in its infancy.

Summary and Conclusions

The authors have endeavored to present to the reader a different picture of solar flare theory from that portrayed in most reviews. The picture drawn was the view a plasma physicist might take in attempting to identify which of the many possible mechanisms could explain the rapid conversion of magnetic free energy into the various forms of energy observed during a flare. The observations are currently insufficient in both quantity and quality, (though one hopes data from the Solar Maximum Mission will alleviate things somewhat), and theories are not yet sufficiently predictive to identify the dominant flare mechanism. The authors have therefore tried to sidestep many controversial questions which invariably arise among theoreticians and observers who favor a particular idea or model.

With this in mind, it is of interest to discuss what kinds of theoretical research might be pursued, so as to better identify which of the mechanisms discussed could possibly be the cause of the flare, its driver, and the geometry in which it occurs. Again mechanisms rather than models are emphasized since the principal
differences among models that use a given mechanism are their geometries and their drivers. Note that, in particular, one difference among the different mechanisms is that some mechanisms require very strongly driven, transient currents while others do not. Recently, it has been shown (Spicer, 1979b) that a preflare transient buildup of current can, in principle, be distinguished from a slow buildup in current. The measurable (cf. Spicer, 1979b) time interval between the formation of tails in an electron distribution and the onset of Joule heating should therefore be a means of distinguishing a transient versus a slow buildup of current, and should thus allow one to eliminate various possible flare mechanisms. Related to this effect is the fact that all of the mechanisms that require strongly driven transient currents result in plasma microturbulence, which, in turn, will result in plasma radiation at radio wavelengths. Such a diagnostic has been recently discussed by Smith and Spicer (1979). This diagnostic has additional importance in that it is one of the few known means of distinguishing between a thermal and a nonthermal electron distribution, thereby yielding additional information concerning the numbers of nonthermal electrons necessary to produce the plasma radiation. A second diagnostic is the observation of inner shell ($K_{\alpha}$) transitions, which require the presence of exciting fast particles in a plasma cool enough for the inner shells to be populated, as would occur in a thick target model. In turn, a determination of the fraction of nonthermal electrons produced by a given flare mechanism will allow one to narrow the choice of mechanisms. Other possible diagnostics would compare the heating rate of a given mechanism (which would have to be very sensitive to the observable parameters for this to work) with the observed rate. The above illustrates a number of possible theoretical diagnostics that have been proposed but that need much further development. There are undoubtedly others to be found.

To distinguish external drivers, one needs to look at how an external driver can be produced in the solar atmosphere, and where the mechanism will manifest itself. For example, it is possible to produce a $J_{\parallel}$ driver only in the photosphere or convection zone, since only very weak body forces exist in the corona, which, in addition, is highly ionized. This makes the production of $J_{\parallel}$ by, say, neutral winds (Heyvaerts, 1974b) impossible in the corona but very probable in the photosphere. Next, consider a $J_{\perp}$ driver. To utilize a particular $J_{\perp}$ mechanism that releases its energy in the lower corona requires that one identify a possible mechanism to force new magnetic flux into a neutral sheet, which either already exists in the lower corona or is about to be formed there. Such a mechanism is the emerging flux, which is the basis of the emerging flux model.

Another helpful means of identifying the possible flare mechanism is to identify the simplest magnetic region in which a flare is known to occur and the geometry of the PMFC in which the flare energy has its origins. As noted under miscellaneous flare requirements the simplest magnetic region appears to be bipolar with a closed PMFC. Such a set of circumstances supports a $J_{\perp}$-driven mechanism if no total nulls in $B$ are observed, when magnetographs with higher spatial and temporal resolution become available.

In summary, it is evident that more theoretical effort needs to be directed at understanding how different flare mechanisms would manifest themselves observationally so that differences can be tabulated and compared to data. Such an effort will clearly require more detailed calculations both analytically and numerically. Once the broad features of each mechanism are more fully understood and have survived the gauntlet of observational tests, then and only then should detailed models begin to take form, with numerical simulations of the models leading the way.

We are grateful to Professor Derek Tidman and Dr. James Ionson for reading the manuscript and making useful comments. In addition, we thank Dr. Richard Hubbard for discussions on the phenomenon of double layers. Our special thanks go to Lori Kreger for her excellent work in typing the various drafts of this manuscript as well as her patience with us.
REFERENCES


Block, L.P., 1972, Cosmic Electrodynamics 3, 349.


Dungey, J.W., 1953, Phil. Mag. 44, 725.


Gold, T., Hoyle, F., 1960, Monthly Notices RAS 120, 89.


Schnack, D., Killeen, J., 1979, Nuclear Fusion 19, 877.


Spicer, D.S., 1979b, Solar Phys. 64, 121.


469


Sweet, P.A., 1958b, Nuova Cim. Suppl. 8 (ser. 10) 188.


Ware, A.A., 1965, Culham Lab Rept. CLM-M53.


