7-1-1997

Critical Issues For Understanding Particle Acceleration in Impulsive Solar Flares

James A. Miller  
*University of Alabama - Huntsville*

Peter J. Cargill  
*Naval Research Laboratory*

A. Gordon Emslie  
*University of Alabama - Huntsville*

Gordon D. Holman  
*NASA Goddard Space Flight Center*

Brian R. Dennis  
*NASA Goddard Space Flight Center*

*See next page for additional authors*

Follow this and additional works at: [http://digitalcommons.kennesaw.edu/facpubs](http://digitalcommons.kennesaw.edu/facpubs)

Part of the [Stars, Interstellar Medium and the Galaxy Commons](http://digitalcommons.kennesaw.edu/facpubs), and the [The Sun and the Solar System Commons](http://digitalcommons.kennesaw.edu/facpubs)

**Recommended Citation**


This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Faculty Publications by an authorized administrator of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.
Critical issues for understanding particle acceleration in impulsive solar flares


Abstract. This paper, a review of the present status of existing models for particle acceleration during impulsive solar flares, was inspired by a week-long workshop held in the Fall of 1993 at NASA Goddard Space Flight Center. Recent observations from Yohkoh and the Compton Gamma Ray Observatory, and a reanalysis of older observations from the Solar Maximum Mission, have led to important new results concerning the location, timing, and efficiency of particle acceleration in flares. These are summarized in the first part of the review. Particle acceleration processes are then discussed, with particular emphasis on new developments in stochastic acceleration by magnetohydrodynamic waves and direct electric field acceleration by both sub- and super-Dreicer electric fields. Finally, issues that arise when these mechanisms are incorporated into the large-scale flare structure are considered.

Stochastic and super-Dreicer acceleration may occur either in a single large coronal reconnection site or at multiple “fragmented” energy release sites. Sub-Dreicer acceleration requires a highly filamented coronal current pattern. A particular issue that needs to be confronted by all theories is the apparent need for large magnetic field strengths in the flare energy release region.

1. Introduction

Particle acceleration is a ubiquitous phenomenon at sites throughout the Universe [e.g., Zank and Gaisser, 1992]. An important example occurs in solar flares, which offer a wide range of observations and allow one to probe both electron and ion acceleration. During flares, large amounts of energy, anywhere from \(10^{28}\) to \(10^{34}\) ergs, are released on timescales which vary from a fraction of a second to several tens of minutes [e.g., Švestka, 1976; Priest, 1981; Tandberg-Hanssen and Emslie, 1988]. As will be discussed below, a significant fraction of this released energy is manifested in the form of energetic particles.

Flares are unique in the astrophysical realm for the great diversity of diagnostic data that are available. These data include (1) continuum emission, which spans the dynamic range from radio, to microwaves, soft and hard X rays, and finally gamma rays, which may have energies in excess of 1 GeV; (2) gamma ray line emission at various energies between \(400\) keV and \(8\) MeV; (3) direct charged particle and neutron observations in space; and (4) observations of high-energy neutrons and charged particles by ground-based monitors. The microwave and hard X ray/gamma ray continuum are believed to be the result of gyrosynchrotron emission and bremsstrahlung, respectively, from subrelativistic to relativistic electrons. Lower-frequency radio and soft X ray emission are thought to be plasma radiation and thermal bremsstrahlung, respectively. Soft X ray and EUV spectral lines are also present and are due to the hot thermal plasma. Interactions between accelerated ions with energies \(>1\) MeV nucleon\(^{-1}\) and ambient nuclei yield excited nuclei, neutrons, and positrons, all of which then produce the gamma ray lines. Reactions of relativistic ions with ambient nuclei also produce pions and high-energy neutrons. The pions decay either directly into gamma rays or into ultrarelativistic secondary electrons and positrons, all three of which may contribute to the \(>10\) MeV gamma ray continuum.

Over the past few years, it has been argued [Reames, 1995, and references therein] that flares appear to reside in two broad categories: impulsive and gradual.
inally so named on the basis of the duration of their soft X ray emission [Pallavicini et al., 1977] (and also their hard X ray and gamma ray emission), impulsive events tend to be compact and occur low in the corona, while gradual events occur at greater heights and correlate well with coronal mass ejections (CMEs). However, another major difference between the two classes is the composition of particles that are observed in interplanetary space. There, impulsive events exhibit striking ion abundance enhancements, while gradual events produce accelerated ions with ambient coronal composition [Reames et al., 1994]. The general acceleration scenario that emerges is that all interplanetary particles in gradual events are accelerated by a CME-driven shock, while those in impulsive events are produced by another mechanism(s).

We believe that the most severe constraints on particle acceleration models are imposed during impulsive events, and it is these which we consider. However, a recent refinement [Cliver, 1996] of the two-class picture argues that the same acceleration mechanism(s) is responsible for energizing the particles that remain trapped at the Sun in both impulsive and gradual events and that these trapped particles are similar to those observed in space from impulsive events [see also Mandzhavidze and Ramaty, 1993]. In other words, gradual events possess an impulsive flare “core,” which is responsible for the energetic particles that remain trapped and produce radiation. Some particles from the core also escape (as in pure impulsive events) into space, where they are joined by the particles accelerated by the CME-driven shock. In this case, the relevance of the present paper expands to include those particles in gradual events which were not accelerated by a CME-driven shock.

This review paper grew out of a week-long workshop conducted at NASA Goddard Space Flight Center in the fall of 1993, in which a small group of people working in the field (the authors of this paper) met to discuss issues related to particle acceleration. The paper summarizes the issues discussed there as well as more recent developments in both theory and observations. In section 2, we present a review of the observations and their interpretation. Section 3 reviews the particle acceleration mechanisms which have been proposed to account for these observations. Section 4 examines how these mechanisms might fit into the global constraints of the solar flare geometry. Section 5 summarizes our conclusions and outlines profitable directions for future studies.

2. Review of Pertinent Observations and Their Implications

In this section we present a summary of the essential observations of accelerated particles that any theoretical model of flares must account for. It is convenient to approach this by posing the following questions: (1) To what energies are the particles accelerated? (2) How quickly do they reach these energies? and (3) How many particles are accelerated per second? These questions can be asked of both electrons and ions, which we now consider in the following two subsections. It should be noted that, in addition to particle acceleration, plasma heating also occurs in flares. While this review will not address issues of plasma heating directly, we note that some direct heating will be associated with any of the acceleration mechanisms discussed. The ratio of particle acceleration to direct heating is an important measure of the efficiency of the mechanism.

2.1. Energetic Electrons and Hard X Ray Bursts

A useful paradigm for flares is that they involve the release of magnetic energy in bipolar coronal loops or arcade structures, with the magnetic field connecting photospheric regions of opposite magnetic polarity. A fairly successful model is that much of this energy appears initially as accelerated electrons with energies \( \gtrsim 20 \text{ keV} \). As the accelerated electrons stream from the corona toward and through the chromosphere, they produce hard X ray bremsstrahlung via interactions with ambient protons. However, they concurrently lose far more energy to heating the ambient cooler electrons by Coulomb collisions. Since the electrons remain trapped in the corona and chromosphere and radiate while losing all of their suprathermal energy there, the resulting emission is referred to as thick-target nonthermal bremsstrahlung. (Thin-target bremsstrahlung would arise if the electrons lost only a small fraction of their energy while radiating.)

The chromospheric plasma is heated and then driven upward along the guiding magnetic field lines toward the corona by the large pressure gradient. These flows are referred to as chromospheric ablations or evaporations [e.g., Antonucci et al., 1982; Mariska et al., 1993], and emit relatively long-lived thermal soft (\( \lesssim 20 \text{ keV} \)) X ray emission [e.g., Pallavicini et al., 1977; Doschek et al., 1993]. This picture provides a simple explanation for the close association between the thermal and nonthermal X ray emission [e.g., Wu et al., 1986; Doschek et al., 1996]. In addition, relativistic electrons produce gyrosynchrotron microwave emission as they spiral in the coronal magnetic field. It should be noted that the model is basically one of transport, since the initial electron acceleration is simply assumed ab initio.

This nonthermal thick-target model has explained successfully the observed radiative signatures of flares at a number of wavelengths including optical, \( H\alpha \), EUV, and soft X rays [e.g., Emslie et al., 1981; McClymont and Canfield, 1986; Canfield and Gayley, 1987; Mariska et al., 1989]. Striking support also comes from the observation of simultaneous impulsive soft and hard X ray emission from the chromospheric footpoints of the magnetic structure [Hudson et al., 1994; Sakao, 1994; Masuda, 1994], as would be expected from the interaction of electron beams with the chromosphere.

There are also some problems that should be noted: (1) plasma heating is often observed before the start of the hard X ray emission [Mariska and Zarro, 1991] (and
hence before the acceleration of a significant number of nonthermal electrons), and (2) the model generally predicts more upward moving material than is indicated by the observed blue-shifted component of soft X-ray lines [Li et al., 1991]. However, in the present section we will adopt this nonthermal thick-target model as a working paradigm.

The hard X-ray emission produced by the energetic electrons is widely regarded as the characteristic signature of impulsive flares [Dennis, 1985, 1988; Tandberg-Hanssen and Emslie, 1988]. While microwave and radio emission are also diagnostics of the energetic electrons [e.g., Benz, 1993; Aschwanden et al., 1995a], we can concentrate just on the hard X-ray emission to address the following three basic questions:

**To what energies are electrons accelerated?**

Photons of a given energy are produced principally by electrons of comparable energy. Therefore it can be inferred from observed hard X-ray spectra that electrons with energies well into the relativistic regime ($\gtrsim 100$ keV) exist [Dennis, 1988]. While hard X-ray emission is common, some flares also exhibit gamma-ray emission up to tens of MeV. Processes which contribute to emission above about 1 MeV include electron bremsstrahlung, nuclear deexcitation, and pion decay. These last two result from energetic ions and will be treated in the next section. However, in some flares there is no evidence for the presence of energetic ions, and all of the gamma-ray emission is evidently due to ultra-relativistic electrons. Such flares have been called "electron dominated" [Marschhäuser et al., 1994; Petrosian et al., 1994] and thus signal the acceleration of electrons to tens of MeV. We note that the photon spectra $F(E_\gamma) = KE_\gamma^{-s}$ photons cm$^{-2}$ s$^{-1}$ MeV$^{-1}$ at 1 AU from the Sun, where $E_\gamma$ is the photon energy, can be very hard in the gamma-ray regime, with spectral indices $s$ as low as 1.5.

**How quickly do they reach these energies?**

A precise determination of the acceleration time is complicated by transport from the acceleration region to the interaction region. However, an upper limit on the acceleration time to $\sim 100$ keV can be obtained from hard X-ray time profiles. Observations from the Burst and Transient Source Experiment (BATSE) on the Compton Gamma Ray Observatory (CGRO) have revealed very fine scale structure in the hard X-ray emission from impulsive flares, manifested as spikes in the emission lasting $\approx 400$ ms [Machado et al., 1993]. Particles would thus have to be accelerated to $\sim 100$ keV on such a timescale. Also, Aschwanden et al. [1995b] have reported 10–20 ms delays between two low-energy hard X-ray channels in BATSE. This is consistent with the near-simultaneous acceleration of the particles to both energies, with the delay resulting from the different travel times from a coronal acceleration site.

The acceleration to higher energies ($> 100$ keV) can occur somewhat more slowly. Specifically, during the initial few seconds of the hard X-ray burst, there is sometimes a "high-energy delay," where the flare onset at energies $\gtrsim 150$ keV is delayed by a few seconds relative to the onset at lower energies (e.g., see Bai et al. [1983] and Dulk et al. [1992]; see, however, Kane et al. [1986] for a case when such a delay was absent). The time profiles of the gamma-ray emission from electron-dominated flares can also place an upper limit on the acceleration time. This emission rises and reaches a maximum over a few seconds to about 30 s [Rieger, 1994]. The electron acceleration time to a few tens of MeV must then be no more than a few seconds.

**How many electrons are accelerated per second?**

The number of electrons that escape into space is less than the number that remain trapped on closed magnetic field lines and produce X rays [Ramaty et al., 1993]. We thus consider only the trapped electrons in our discussion.

Above the iron lines at around 7 keV, hard X-ray spectra $F(E_\gamma)$ (photons cm$^{-2}$ s$^{-1}$ MeV$^{-1}$) are smooth continua and are fairly well fitted by power laws $KE_\gamma^{-s}$ with $s$ larger than 2. In a thick-target nonthermal model, the energy-differential rate at which accelerated electrons are produced or injected $N(E)$ (electrons MeV$^{-1}$ s$^{-1}$) is also a power law and steeper than this by a power of roughly one. For a large X-class flare, the flux of X rays above 20 keV at 1 AU can be $\gtrsim 10^4$ photons cm$^{-2}$ s$^{-1}$, and results from an emission area of $\approx 10^{18}$ cm$^2$. The nonthermal model then indicates that $\approx 10^{27}$ electrons s$^{-1}$ were accelerated to energies $> 20$ keV in such a flare. Hence, if the flare lasts $\approx 100$ s, the total number of electrons energized above 20 keV is about $10^{39}$. (We point out that, while these numbers are quite large, they are dwarfed by those from so-called "giant flares," in which the energization rate and total number above 20 keV can be $\approx 10^{39}$ s$^{-1}$ and $10^{41}$, respectively [Kane et al., 1995]. These events, however, are relatively rare, and we do not take them into account in obtaining "typical" numbers for flares.) Given the steepness of the electron energy distribution, the bulk of the energy in nonthermal electrons resides at low energies (20–50 keV). Below $\approx 20$ keV, it becomes harder to distinguish the nonthermal component from a hot thermal component generated by plasma heating. These numbers are appropriate to the entire flare duration, but there is evidence that electron acceleration in impulsive flares occurs in small bursts, which have been termed "energy release fragments (ERFs)" by Machado et al. [1993]. Data obtained with the Hard X-ray Burst Spectrometer (HXRBS) on the Solar Maximum Mission (SMM) have shown spikes of duration $\approx 400$ ms superposed upon the more slowly varying background of hard X rays [Kiplinger et al., 1984]. Employing the nonthermal model for hard X-ray production, they deduced that about $2 \times 10^{34}$ electrons were accelerated to energies greater than 20 keV in one of these spikes. With the aforementioned spike duration, the rate at which electrons are energized above 20 keV is then $\approx 5 \times 10^{34}$ s$^{-1}$. The existence of this spiky structure has been confirmed by observations made with BATSE on CGRO (see above), where the accelerated electron energy content in an ERF is between $10^{26}$ and $10^{27}$ ergs, and with the PHEBUS instrument on...
GRANAT [Vilmer et al., 1996]. In light of all these observations, \( \approx 5 \times 10^{34} \text{electrons s}^{-1} \) need to be energized above 20 keV over \( \approx 400 \text{ms} \) in order to account for an ERF. These ERFs are apparent only in smaller flares, where the number that are firing at any time is sufficiently small for them to be observed separately; in larger events, they presumably blend together to form a smoother hard X ray emission time profile.

In general, these electron numbers are obtained by fitting model spectra to data of low spectral resolution (e.g., the Hard X ray Telescope (HXT) on Yohkoh has four channels). To obtain more accurate estimates of the number of energetic electrons, much higher spectral resolution is needed. A 1980 balloon flight using cooled Germanium detectors [Lin et al., 1981] provided data with approximately 2 keV resolution between 15 and \( \approx 200 \text{keV} \) on a relatively small GOES class M6 flare. A discussion of the results is instructive, since they show the wealth of detail available when high spectral resolution is used.

Figure 1 shows the X ray spectra \( F(E_{\gamma}) \) for 15 time intervals spanning the duration of this event [Lin et al., 1981]. When the emission is rising (left panel), the spectra are well fit by power laws of spectral index \( s \approx 3.5 \) below about 100 keV, but show a steepening at higher energies. During the peak of the emission (center panel), the spectra exhibit a strong steepening (spectral index \( \approx 11 \)) below \( \approx 40 \text{keV} \), consistent with thermal bremsstrahlung emission [Emslie et al., 1989], possibly due to a “superhot” \( (\approx 3 \times 10^7 \text{K}) \) thermal plasma [Lin et al., 1981]. The spectra during the decay phase (right panel) are well fit below 40 keV by single temperature thermal spectra with a slowly decreasing temperature. A subsequent reexamination of flare spectra observed using HXRS on SMM has shown that they are also consistent with this same broken power law shape [Winglee et al., 1991; Dulk et al., 1992].

Lin and Johns [1993] obtained directly the spectrum of the accelerated electrons from these data. Their analysis suggests that two hard X ray emitting electron populations exist in this flare: a superhot thermal component of slowly increasing density which emits at low energies \( (\lesssim 30 \text{keV}) \), and a rapidly varying nonthermal component which is responsible for the higher energy X rays, but which also produces the spiky structure at low energies as well. Both components can be integrated to obtain a total energy and particle number [Lin and Johns, 1993], giving the injection rate \( \tilde{N} = \int_{20}^{\infty} dE \tilde{N}(E) \) of electrons with energies \( > 20 \text{keV} \) from the superhot and nonthermal components as \( \approx 4 \times 10^{34} \text{s}^{-1} \) and \( \approx 5 \times 10^{35} \text{s}^{-1} \), respectively. Above 30 keV (where the superhot component is negligible), the rate at which nonthermal electrons are produced is \( \approx 10^{35} \text{s}^{-1} \) over a period of about 150 s.

These numbers have important implications for the energetics of flares, which centers ultimately on the relative efficiency of thick-target bremsstrahlung in producing hard X rays. For example, for keV electrons,
this efficiency is only \(\approx 5 \times 10^{-6}\), so that \(\approx 2 \times 10^6\) ergs of electron kinetic energy are necessary for every erg of X ray bremsstrahlung radiated. For the typical X-class flare involving the acceleration of \(\approx 10^{37}\) electrons s\(^{-1}\), there is a power input in the form of energetic electrons above 20 keV of \(\approx 3 \times 10^{29}\) ergs s\(^{-1}\). Thus, in a flare lasting about 100 s, the total energy in energetic electrons is \(\approx 3 \times 10^{31}\) ergs, or a significant fraction of the total estimated flare energy.

To further appreciate the magnitude of these numbers, consider that a typical magnetic flux loop involved in a flare has an area of \(10^{18}\) cm\(^2\) and a length of \(10^9\) cm. For a density of \(10^{10}\) cm\(^{-3}\) (typical for an active region), the entire loop contains about \(10^{37}\) electrons. This means that, over the flare duration, more electrons must be accelerated than are initially available in the flux tube. Therefore real-time replenishment is a requirement of a viable model. While the large mass reservoir in the chromosphere can easily provide the electrons necessary for replenishment of the acceleration region, some models have electrodynamic constraints that can limit the way in which these electrons can be pulled from the chromosphere. Such constraints can have important implications for the overall structure of the flaring region (see section 4).

In view of these large electron numbers, alternatives to the nonthermal thick-target model have been proposed. The most widely studied of these is a thermal model in which hard X rays are predominantly emitted by electrons in a hot (\(\gtrsim 10^5\) K) coronal plasma. This is substantially more efficient than the nonthermal thick-target model since, on average, the hard X ray emitting electrons do not lose energy to other electrons in the plasma. In this case, the dominant energy loss channel is bremsstrahlung and efficiencies close to unity are possible in theory. The hot plasma expands behind a pair of conduction fronts that propagate at approximately the local ion acoustic speed [Brown et al., 1979] and eventually reach the chromosphere. However, leakage of hot electrons from such a plasma to the chromosphere reduces the efficiency [Brown et al., 1979; Smith and Brown, 1980], and these streaming electrons will produce footpoint hard X ray emission, just as in the thick-target model.

Recent observations from Yohkoh have cast serious doubts on the viability of a purely thermal hard X ray model. Sakao [1994] analyzed the Yohkoh HXT observations of a number of flares that had pairs of hard X ray brightenings on both sides of a magnetic neutral line, presumably corresponding to the footpoints of a bipolar loop. The temporal fluctuations of these footpoints were coincident to \(\approx 0.1\) s. Unless the hot plasma was sited exactly equidistant between the two footpoints, the footpoint brightenings could not be due to the interaction of conduction fronts with the chromosphere. However, simultaneous brightenings could be produced by the aforementioned free-streaming electrons.

An obvious compromise between nonthermal thick-target and thermal hard X ray models is a hybrid model: one involving both heating and acceleration as modes of primary energy release. Holman and Benka [1992] have formulated such a model based on sub-Dreicer electric fields and find that a maximum acceleration rate of \(10^{34}\) electrons s\(^{-1}\) is sufficient to account for the flare discussed by Lin et al. [1981]. This rate is about a factor of 50 lower than that observed above for the purely nonthermal model. However, if the acceleration volume is \(10^{27}\) cm\(^3\), then we still need to accelerate \(10^7\) electrons cm\(^{-3}\) s\(^{-1}\). Assuming that this factor of 50 decrease is applicable to larger flares, we see from the numbers given above for the nonthermal model that typical electron energization rates and total energy contents above 20 keV are \(\approx 2 \times 10^{35}\) s\(^{-1}\) and \(6 \times 10^{29}\) ergs, respectively. Hence this model still requires real-time replenishment of the coronal electron population. Yohkoh observations of Masuda [1994] have also provided some evidence for such a model, with both footpoint and coronal hard X ray sources being present in some limb flares.

### 2.2. Gamma Rays and Energetic Ions

Energetic ions in a solar flare can also be investigated indirectly through the variety of neutral emissions that they produce (see reviews by Chupp [1984] and Ramaty and Murphy [1987]), as well as directly through in situ measurements in space. We thus ask the same three basic questions as in the previous subsection. As was the case there, we assume that the neutral emissions were created in a thick-target interaction region, such as the chromosphere and photosphere. Furthermore, due to the high ion energies involved, the ions are necessarily nonthermal.

**To what energies are ions accelerated?** The most direct answer to this question is offered by the ions that escape from the impulsive solar flare and are observed directly in interplanetary space. Such ions have energies up to 100 MeV nucleon\(^{-1}\) [Reames et al., 1992; Mazur et al., 1992]. However, some impulsive flares (and gradual ones too) possess excess (i.e., above the electron bremsstrahlung continuum) photon emission above about 1 MeV that consists of nuclear radiations. For these flares, ion energies can be probed indirectly using this nuclear excess. A typical gamma ray flare exhibits narrow (\(\approx 100\) keV width) nuclear deexcitation lines between \(\approx 1\) and \(\approx 7\) MeV and a neutron capture (or deuterium formation) line at 2.223 MeV [e.g., Chupp, 1984; Murphy et al., 1991]. A theoretical spectrum similar to what is needed to model a large gamma ray flare is shown in Figure 2. The narrow deexcitation lines result from the interaction of protons and alpha particles having energies between \(\approx 1\) and \(\approx 100\) MeV nucleon\(^{-1}\) with ambient heavier nuclei [Ramaty et al., 1979]. Inverse reactions between energetic heavy nuclei and ambient H and \(^4\)He yield deexcitation lines of width \(\approx 1\) MeV, which, together with many closely spaced and weak narrow lines, constitute broad unresolved features in the gamma ray spectrum. The neutrons which yield the capture line also result from reactions of ions having energies between \(\approx 1\) and \(\approx 100\) MeV nucleon\(^{-1}\) with the...
ambiguously observed in space [Chupp et al., 1982] and are energy $E$ or $K_2$ modified Bessel functions. The Bessel flares. Neutrons between 50 and 500 MeV can be evaluated. They neutron capture line fluences were available, assuming several events [from Ramaty and Lingenfelter, 1995]. The ion and electron spectra incident on a thick-target emission region are power laws of the same spectral index. The dotted line is bremsstrahlung from the electrons, and the solid line is the total gamma ray emission. The principal nuclear deexcitation lines, the neutron capture line, the positron annihilation line, the positronium continuum, and the broad deexcitation lines from $\alpha-\alpha$ reactions are marked. Positrons result from $\beta^+$ decay of radionuclides generated in reactions between incident and ambient ions.

Higher-energy ions are present in some flares. Six gamma ray flares have exhibited a hardening or "bump" in the photon spectrum near $\approx 70$ MeV [see Mandzhavidze and Ramaty, 1993], which is due to mainly neutral pion decay radiation [Murphy et al., 1987]. This pion excess immediately indicates that protons (which dominate pion production) were accelerated above the pion production threshold of $\approx 300$ MeV. Moreover, some of these flares had pion-decay emission up to a few GeV [Akimov et al., 1993; Kanbach et al., 1993], which then pushes the proton energy upper limit to a few GeV. Modeling of pion-decay emission for one flare indicates that a high-energy cutoff of 10 GeV in the proton spectrum is most consistent with the data [Mandzhavidze et al., 1996].

Neutrons are also a signature of very high-energy protons and are generated mostly by protons and alpha particles interacting with ambient H and $^4$He. They usually accompany pion decay radiation in the largest flares. Neutrons between $\approx 50$ and 500 MeV can be directly observed in space [Chupp et al., 1982] and are in turn produced by protons with energies up to $\approx 1$ GeV [e.g., Ramaty and Mandzhavidze, 1994]. The very highest energy ($\approx 1$ GeV) neutrons can be detected by ground-based neutron monitors [e.g., Debrunner et al., 1983], and indicate the presence of protons of roughly the same energy. Hence, while most gamma ray flares exhibit evidence for ions up to 100 MeV nucleon$^{-1}$, some of the largest appear capable of accelerating protons up to at least $\approx 1$ to 10 GeV.

**How quickly do they reach these energies?** As with electrons, a determination of the acceleration time is complicated by transport. An upper limit on the acceleration time to tens of MeV nucleon$^{-1}$ can be obtained from the time profiles of the nuclear deexcitation gamma ray line flux. These light curves rise above background and peak on timescales of $\lesssim 1$ s [Kane et al., 1986] to a few seconds [Forrest, 1983]. The acceleration time upper limit is then approximately equal to this rise time. A determination of acceleration time scales to higher energies requires a higher energy diagnostic, such as pion decay radiation. For example, a comparison of the nuclear deexcitation line and pion radiation time profiles from the June 3, 1982, flare [Forrest et al., 1986; Chupp et al., 1987] indicates that acceleration to $\approx$ GeV nucleon$^{-1}$ energies occurs in $\lesssim 16$ s for this flare [Miller et al., 1987]. Higher time resolution measurements may reduce these upper limits in the future, but, at present, acceleration to MeV nucleon$^{-1}$ energies on timescales of order 1 s must be accounted for in any acceleration model.

**How many ions are accelerated per second?** The number of ions that escape into interplanetary space can be either more or less than the number that remain trapped at the Sun and produce gamma rays [Ramaty et al., 1993]. However, since the trapped number typically exceeds the escaping number [Hua and Lingenfelter, 1987], we consider again only the trapped particles.

The first diagnostic to be used for probing the spectrum of trapped particles above a few MeV nucleon$^{-1}$ was the ratio of the 2.223 MeV neutron capture line fluence to the 4−7 MeV nuclear deexcitation line fluence [Murphy and Ramaty, 1984; Hua and Lingenfelter, 1987]. Since the deexcitation and neutron capture lines are produced by ions in somewhat different energy ranges (the capture line results from relatively higher energy ions), their ratio is a measure of the ion spectral shape in the $\approx 10$−100 MeV nucleon$^{-1}$ range. In the same way, the spectral shape in the $\approx 10$−1000 MeV nucleon$^{-1}$ range can be determined for pion flares by considering the ratio of the 100 MeV fluence to the nuclear deexcitation fluence [Murphy et al., 1987]. The normalization of the ion spectrum is fixed by the magnitude of a particular fluence.

This technique was used by Murphy and Ramaty [1984] to analyze nine flares for which deexcitation and neutron capture line fluences were available, assuming accelerated proton spectra $N(E) = N_0 \frac{E}{E_0} \exp(-E/E_0)$ (protons MeV$^{-1}$) that were either power laws in kinetic energy $E^{-\alpha}$ or $K_2$ modified Bessel functions. The Bessel
functions result from a stochastic acceleration model [Ramaty, 1979] and roll over at high energies while flattening out at lower energies. For both spectral shapes, the typical number of protons above 30 MeV was found to be \( \approx 10^{32} \). In the case of power laws, the spectral index \( \gamma \) was \( \approx 3.5 \), so that the total number of protons above 1 MeV is \( \approx 5 \times 10^{35} \). In the case of Bessel functions, the total number above 1 MeV is somewhat lower, around \( 10^{35} \).

The number of ions below about 1 MeV cannot be determined from deexcitation and neutron line emission, since the cross sections are zero and the ions therefore have no gamma ray signature. Hence, for an emission duration of about 30 s, the typical rate at which protons are energized above 30 MeV is about \( 3 \times 10^{30} \text{s}^{-1} \), while the energization rate above 1 MeV can range from \( 3 \times 10^{32} \) to \( 2 \times 10^{34} \text{s}^{-1} \).

The energy content of these protons can also be estimated. For the stochastic acceleration spectrum, there is \( \approx 10^{50} \text{ergs} \) in the protons above 1 MeV, while for the power law this content is nearly \( 10^{30} \text{ergs} \). The energy contained in the heavier ions is roughly equal to the energy contained in the protons. The ion energy content is then more than an order of magnitude lower than the energy contained in the electrons. This result has lead to the notion that energetic ions are not the main players in the overall energy budget of flares. However, note that for a flare volume of \( 10^{27} \text{cm}^3 \), the flare must still produce order \( 10^2-10^3 \text{ergs cm}^{-3} \) of accelerated protons, which is much larger than the thermal plasma energy density and still a sizable fraction of the estimated magnetic field energy density.

However, the conclusion that ions are energetically unimportant has changed recently. Using data [Share and Murphy, 1995] from 19 gamma ray flares observed during a 9-year period with the Gamma Ray Spectrometer on SMM, Ramaty et al. [1995] have used the ratio of the 1.63 MeV \(^{20}\text{Ne} \) deexcitation line fluence to the 6.13 MeV \(^{16}\text{O} \) deexcitation line fluence (see Figure 2) to determine energetic ion spectra. This technique works much the same way as the method discussed above and relies on the fact that the lines are produced by ions of different energies: the cross section for the \(^{20}\text{Ne} \) line becomes nonzero above \( \approx 2 \text{MeV nucleon}^{-1} \) and peaks around \( 7 \text{MeV nucleon}^{-1} \), while that for the \(^{16}\text{O} \) line becomes nonzero above \( \approx 7 \text{MeV nucleon}^{-1} \) and peaks around \( 12 \text{MeV nucleon}^{-1} \). These energies are for incident protons; for incident alpha particles, they are somewhat lower. The \(^{20}\text{Ne} \) line is therefore a good diagnostic for energetic ions above \( 1 \text{MeV nucleon}^{-1} \).

The other difference between this and previous studies was the use of an ambient Ne-to-O ratio which is closer to that obtained from EUV and soft X ray line observations. The new ratio is lower and leads to an increased number of ions at low energies.

The observed \(^{20}\text{Ne} \) and \(^{16}\text{O} \) deexcitation line fluences imply that the energetic ion spectra \( N(E) \) are relatively steep power laws (spectral index \( \gamma \approx 4 \)) down to \( \approx 1 \text{MeV nucleon}^{-1} \), with the number of protons above 30 MeV still remaining at about \( 10^{32} \) (Ramaty and Mandzhavidze, private communication, 1996). However, as a result of the steep spectra, the number of protons above 1 MeV now rises to typically \( 3 \times 10^{36} \).

For a 30 s flare duration, the rate at which protons are energized above 1 MeV is then nearly \( 10^{36} \text{s}^{-1} \), and can rival the electron energization rate above a few tens of keV (see previous section). The total ion energy content above 1 MeV nucleon\(^{−1} \) for these flares is shown in Figure 3. While there is significant scatter, a typical energy content is about \( 3 \times 10^{31} \text{ergs} \), more than an order of magnitude above previously derived values. The protons and heavier ions each have approximately the same energy content.

The ion energy for these flares is now comparable to the typical nonthermal electron energy (\( \approx 3 \times 10^{31} \text{ergs} \)) discussed in the previous section and is also comparable to the energy contained in a \( \approx 1 \text{Kg} \) coronal field in a volume of \( \approx 10^{27} \text{cm}^3 \). A case-by-case comparison can also be made for 12 flares for which hard X ray data is also available from the SMM HXRBS. The energy contained in > 20 keV electrons for these 12 flares is shown in Figure 3. Again, while there is significant scatter, a typical electron energy content is about \( 3 \times 10^{31} \text{ergs} \) (consistent with the argument in the previous section). Note that a few flares even have more ion energy than electron energy. Hence, at least for flares with detectable gamma ray emission, there is evidently a near (to within uncertainties in the low-energy cutoffs of the ion and electron energy distributions) equipartition in energy between ions and electrons.

In Table 1, we summarize the above discussion on electrons and ions, and present the average rates \( \dot{N} \) at which particles are energized above a given energy along with the total energy content of the particles. The electron energization rates are for large flares, such as those which have detectable gamma ray emission. In ERFs,
Table 1. Summary of Typical Energization Rates and Total Energy Contents

<table>
<thead>
<tr>
<th>Quantity</th>
<th>ERF, Entire Flare, Entire Flare, Entire Flare, Entire Flare, Nonthermal Model, Hybrid Model, Pre-1995, Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\dot{N}$</td>
<td>$5 \times 10^{24} \text{s}^{-1}$</td>
</tr>
<tr>
<td>$U_p$</td>
<td>$5 \times 10^{26} \text{ergs}$</td>
</tr>
</tbody>
</table>

The quantities $\dot{N}$ and $U_p$ denote, respectively, the energization rate and the total energy content above either 20 keV (for electrons) or 1 MeV (for protons).

The lower limit results from stochastic acceleration proton spectra (specifically $K_2$ Bessel functions), while the upper limit results from power law proton spectra.

the average rate of energization must be sustained for about 400 ms, while in the entire flare it must occur over several tens of seconds. For protons, we present rates and energy contents obtained by both pre-1995 and present calculations.

Finally, ions observed in space yield another valuable diagnostic of the acceleration mechanism: relative abundances. The energetic particles from impulsive flares exhibit dramatic abundance enhancements at energies above about 1 MeV nucleon$^{-1}$ (see reviews by Lin [1987] and Reames [1990]), and specifically have a $^3\text{He}/^4\text{He}$ ratio that ranges between 0.1 and 10. This is a huge increase over the coronal value of about $5 \times 10^{-4}$. In addition, these events are also characterized by (1) enhanced ratios of Ne, Mg, Si, and Fe to C, N, O, and $^4\text{He}$ [Reames et al., 1994]; (2) high charge states of the heavy ions [Mason et al., 1995]; (3) isotopic abundance enhancements ($^{25}\text{Mg}$ to $^{24}\text{Mg}$ and $^{22}\text{Ne}$ to $^{20}\text{Ne}$ [Mason et al., 1994]); and (4) ion spectra that have approximately a species-independent shape [Mason et al., 1994]. Typical impulsive flare and coronal abundance ratios are given in Table 2 (adapted from Miller [1995]).

3. Particle Acceleration Processes

A number of mechanisms have been proposed to account for energetic solar particles. This section focuses exclusively on the kinetic physics of the acceleration process. Section 4 assesses how these kinetic processes can arise in the global solar corona, and what (if any) additional constraints need to be imposed on a given mechanism. We split the acceleration processes up into three broad classes: stochastic acceleration by waves, shock acceleration, and direct electric field (dc) acceleration. The overall properties of these mechanisms as they relate to the data discussed in section 2 are summarized in Table 3, which the reader may find it convenient to refer to throughout this section.

3.1. Stochastic Acceleration

Stochastic acceleration may be broadly defined as any process in which a particle can either gain or lose energy in a short interval of time, but where the particles systematically gain energy over longer times. The most important example of this is acceleration by waves.

Table 2. Ion Abundance Ratios

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Impulsive Flares$^a$</th>
<th>Increase Factor Over Coronal Values</th>
<th>Corona$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{3}\text{He}/^{4}\text{He}$</td>
<td>~1</td>
<td>2000</td>
<td>~ $0.0005$</td>
</tr>
<tr>
<td>$^{4}\text{He}/\text{O}$</td>
<td>~46</td>
<td>—</td>
<td>~55</td>
</tr>
<tr>
<td>C/O</td>
<td>~0.436</td>
<td>—</td>
<td>~0.471</td>
</tr>
<tr>
<td>N/O</td>
<td>~0.153</td>
<td>—</td>
<td>~0.128</td>
</tr>
<tr>
<td>Ne/O</td>
<td>~0.416</td>
<td>2.8</td>
<td>~0.151</td>
</tr>
<tr>
<td>Mg/O</td>
<td>~0.413</td>
<td>2.0</td>
<td>~0.203</td>
</tr>
<tr>
<td>Si/O</td>
<td>~0.405</td>
<td>2.6</td>
<td>~0.155</td>
</tr>
<tr>
<td>Fe/O</td>
<td>~1.234</td>
<td>8.0</td>
<td>~0.155</td>
</tr>
<tr>
<td>H/He</td>
<td>~10</td>
<td>—</td>
<td>~100</td>
</tr>
</tbody>
</table>

$^a$Ratio for ions above ~1 MeV nucleon$^{-1}$.

$^b$Ambient abundances.
Central to understanding stochastic acceleration are the normal modes which may exist in a magnetized plasma. We restrict our attention to cold plasma modes [see Swanson, 1989] and to waves that are discussed in subsections 3.1.1 to 3.1.3. A more general discussion may be found elsewhere [e.g., Stringer, 1968; Formisano and Kennel, 1969; Krauss-Varban et al., 1994]. In a cold hydrogen plasma, there are two important electromagnetic modes which comprise different branches in the $\omega-k$ plane. They are the Alfvén branch, which has a resonance below the hydrogen gyrofrequency $\Omega_H$, and the fast mode (or magnetosonic or whistler) branch, which has a resonance below the electron gyrofrequency $\Omega_e$.

For wave frequency $\omega \ll \Omega_H$, the Alfvén branch has the dispersion relation $\omega = v_A |k_\parallel|$, while the fast mode branch has the dispersion relation $\omega = v_A k$, where $v_A$ is the Alfvén speed and $k$ and $k_\parallel$ are the magnitude of the wavevector $k$ and its field-aligned component, respectively. Low-frequency Alfvén waves propagating obliquely with respect to the ambient magnetic field $B_0$ have a linearly polarized electric field normal to $B_0$ and a linearly polarized magnetic field normal to both $B_0$ and $k$. Low-frequency oblique fast mode waves have a linearly polarized electric field normal to both $B_0$ and $k$ and a linearly polarized magnetic field normal to $k$ and the electric field. The wave magnetic field thus has transverse and compressive components with respect to $B_0$. In each case the electric field can be decomposed into left- and right-handed components. However, for parallel propagation, all waves on the Alfvén branch are left-handed, while all those on the fast mode branch are right-handed.

When the Alfvén branch approaches $\Omega_H$, the phase speed approaches zero and waves in this regime are called $H^+$ electromagnetic ion cyclotron ($H^+$ EMIC) waves. When the fast mode branch passes through $\Omega_H$, the phase speed increases. For $\Omega_H \ll \omega \ll \Omega_e$, the dispersion relation is $\omega = k_\parallel^2 \omega_e/\omega_{pe}^2$, where $\omega_{pe}$ is the electron plasma frequency. Waves in this regime are usually called whistlers. As the frequency increases still further, the phase speed approaches zero and whistlers become electromagnetic electron cyclotron waves. In a multi-ion plasma, the dispersion relation below the various ion cyclotron frequencies becomes more complicated, and we refer the reader to Smith and Brice [1964] or Miller and Viñas [1993] for further details.

In addition to these electromagnetic modes, there are also some electrostatic ones. Lower hybrid waves are readily generated by cross-field ion motion or relative electron-ion drift [Huba, 1985] and have a frequency given by $\omega_{LH}(1+(m_i/m_e)(k_i/k_L)^2)^{1/2}$, where $k_L$ is the perpendicular component of the wavevector, $k_\parallel \ll k_L$, the lower hybrid frequency $\omega_{LH} \approx \omega_{pe}/(1 + \omega_{pe}^2/\Omega_e^2)$, $\omega_{pe}$ is the H plasma frequency, and it is further assumed that $\Omega_H \ll \omega \ll \Omega_e$. Electrostatic ion cyclotron (EIC) waves are also generated by relative electron-ion drift and lie above an ion cyclotron frequency [e.g., Stix, 1992]. Electron plasma (or Langmuir) waves can be generated by streaming electrons and have a dispersion relation $\omega = \omega_{pe}$.

A second key issue for understanding stochastic acceleration by waves is resonant wave-particle interactions. When the wave amplitude is small, stochastic acceleration is a resonant process and occurs when the condition $\omega = -k_\parallel |v_\parallel - \ell \Omega/\gamma = 0$ is satisfied. Here $v_\parallel$ and $\gamma$ are the parallel particle speed and Lorentz factor, $\Omega$ is the cyclotron frequency of the particle, and $x$ is referred to as the frequency mismatch parameter. For harmonic numbers $\ell \neq 0$ (gyroresonance), this equation is a matching condition between the particle's cyclotron frequency and the Doppler-shifted wave frequency in the particle's guiding center frame. It means that the frequency of rotation of the wave electric field is an integer multiple of the frequency of gyration of the particle in that frame and that the sense of rotation of the particle and electric field is the same.

The convention we employ is that $\Omega$ is always positive and the sign of $\ell$ depends upon the sense of rotation of the electric field and the particle in the plasma frame: if both rotate in the same sense (right or left handed) relative to $B_0$, then $\ell > 0$ (normal Doppler resonance); if the sense of rotation is different, then $\ell < 0$ (anomalous Doppler resonance). Hence, when the resonance condition is satisfied, the particle sees an electric field for a sustained length of time and will either be strongly accelerated or decelerated, depending upon the relative phase of the field and the gyromotion. The most effective gyroresonance is $|\ell| = 1$, and $\ell = \pm 1$ is usually referred to as cyclotron resonance. For $\ell = 0$ the resonance condition specifies matching between the parallel components of the wave phase velocity and particle velocity. This resonance is sometimes referred to as the Landau or Cerenkov resonance.

When a particle is in resonance with a single small-amplitude wave, $v_\parallel$ executes approximate simple harmonic motion about the parallel velocity which exactly satisfies the resonance condition [Karimabadi et al., 1992]. There is no energy gain on average. The amplitude of the oscillation is proportional to the square root of the wave amplitude, and the maximum energy gain is small [see Roberts and Buchsbaum, 1964; Ginet and Heinemann, 1989]. The frequency $\omega_\parallel$ of oscillation, or the bounce frequency, is also proportional to the root of the wave amplitude, and is important for the following reason: If $|x| \leq 2\omega_\parallel$, the particle and wave effectively are in resonance. Hence, the exact resonance condition $x = 0$ does not have to be satisfied in order for a strong wave-particle interaction to occur, which immediately implies that large systematic energy gains in a spectrum of waves are possible.

Consider two neighboring waves, $i$ and $i + 1$, where $i + 1$ will resonate with a particle of higher energy than $i$ will. A particle initially resonant with wave $i$ will periodically gain and lose a small amount of $v_\parallel$. If the gain at some time is large enough to allow it to satisfy $|x| \leq \omega_b,i+1$, where $\omega_b,i+1$ is the bounce frequency for wave $i + 1$, then the particle will resonate with that wave next. After "jumping" from one wave to the next in this manner, the particle will have achieved a net gain in energy. If other waves are present that will resonate with even higher energy particles, the particle will con-
time jumping from resonance to resonance and achieve a maximum energy corresponding to the last resonance present. If the wave spectrum is discrete, then the spacing of waves is critical; if the spectrum is continuous, however, then resonance overlap will automatically occur. Of course, the particle can also move down the resonance ladder, but over long timescales, there is a net gain in energy and stochastic acceleration is the result. This process can be treated by a momentum diffusion equation, and the diffusion coefficients can be calculated using a convenient Hamiltonian approach found in the work of Karimabadi et al. [1992]. For a further discussion of wave-particle resonance, see Karimabadi et al. [1994].

A broadband spectrum of waves is thus typically required in order to stochastically accelerate particles to high energies. The exception is acceleration by resonance overlap in a single large-amplitude wave [Kaney, 1978; Karimabadi et al., 1990]. In this interesting process, a particle resonates with the same wave but through many harmonic numbers, and huge energy gains are possible. However, the importance of such acceleration in flare plasmas has not been considered in detail at the present time. We thus concentrate on acceleration by a spectrum of waves.

3.1.1. Electromagnetic waves: Electrons. A number of different forms of electron acceleration by electromagnetic waves have been considered for flares. The most familiar of these is gyroresonant stochastic acceleration by turbulence with frequencies below \( \Omega_e \).

(For our purposes, turbulence refers simply to a continuous spectrum of randomly phased monochromatic waves.) Alfvén, fast mode, and whistler waves were among the first to be considered [Melrose, 1974]. Electrons can gyroresonate with the first two waves via \( \ell = \pm 1 \), due to the presence of both right- and left-handed electric field components, but \( \ell = +1 \) is most important for whistlers.

Since \( \omega \leq \Omega_H \) for both Alfvén and fast mode waves, we see from the resonance condition that \( \gamma v_{\parallel} \) must be greater than about \( (m_p/m_e) \nu_A \) for electrons to resonate. For a \( \nu_A \) of about 2000 km s\(^{-1}\), this requires electron energies of \( \approx 6 \) MeV. While possibly important for the acceleration of ultrarelativistic electrons, these waves cannot accelerate electrons out of the thermal distribution or through hard X ray producing energies. Whistlers, with \( \Omega_H < \omega < \Omega_e \), yield a resonance requirement of \( \gamma v_{\parallel} \gg (m_p/m_e)^{1/2} \nu_A \). The threshold condition for whistlers is then 20 keV, so that these waves could accelerate hard X ray producing electrons. However, since the threshold is still well above the thermal energy, these waves cannot accelerate electrons directly from the thermal plasma either. Whistlers and Alfvén waves have been used to accelerate deka-keV “seed” electrons to ultrarelativistic energies [Miller and Ramaty, 1987].

However, Steinacker and Miller [1992] and Hamilton and Petrosian [1992] point out that relaxation of the \( \omega \ll \Omega_e \) requirement and the inclusion of higher-frequency waves reduces the energy threshold to values inside the electron distribution. Hamilton and Petrosian [1992] have calculated electron and X ray spectra, the latter of which compare favorably with SMM and Hinotori observations, as well as with the high-resolution spectra of Lin et al. [1981]. Steinacker and Miller [1992] showed that the acceleration times could be reproduced if the whistler turbulence energy density was about 10% of the magnetic field energy density and that acceleration to the highest observed energies could occur if lower-frequency waves on the branch were also included.

The acceleration of electrons from the thermal distribution by gyroresonance thus requires the transfer of spectral energy up the fast mode branch into the whistler and electron cyclotron regimes, since it would appear likely that the initial turbulence exists at low frequencies (\( \omega \ll \Omega_H \)). A cascade of power is one way to achieve this, but such a process has not been investigated and so is speculative at present (for \( \omega > \Omega_H \); in the MHD regime, it is well established that cascading occurs). These whistler acceleration models also have important implications for the overall flare energetics. For a power law spectral density, Steinacker and Miller [1992] showed that for wavelengths shorter than \( \approx 10^6 \) cm, the wave energy density needs to be about 10% the ambient magnetic field energy density in order for the electron acceleration time to be consistent with observations. Hence, if the cascading produces a power law spectral density (which is the case where cascading has been investigated), and if the low-wavenumber cutoff corresponds to about one to one tenth the scale size of the flare (\( \approx 10^6 \) cm), then the total wave energy density exceeds the estimated ambient magnetic field energy density by a significant amount, the exact value depending on the slope of the turbulence spectrum.

One way to avoid the issue of cascading over a large frequency range is to simply use the long-wavelength MHD waves directly. From above, gyroresonance with the transverse electric field is not a viable option, but the \( \ell = 0 \) resonance with the compressive magnetic field component of the fast mode waves is. Using the fast mode dispersion relation, the resonance condition can be written as \( v_{\parallel} = v_A/\eta \), where \( \eta = k_H/k \). Since \( v_A \) is typically much greater than the proton thermal speed \( v_{tp} \) but comparable to the electron thermal speed \( v_{te} \) in a flare plasma (\( B_0 \approx 500 \) G, density \( n \approx 10^{10} \) cm\(^{-3}\), and proton and electron temperatures \( T_p = T_e \approx 3 \times 10^6 \) K), only electrons will be able to interact with the waves. This process is the magnetic equivalent of Landau damping and is called transit-time damping [Lee and Völk, 1975; Fisk, 1976; Achterberg, 1979; Stix, 1992], since the resonance condition can be rewritten to show that the transit time of a particle across a wavelength is equal to the period of the wave. This interaction changes only the parallel energy of a particle, and will lead to anisotropic distributions if ancillary pitch angle scattering is not present.

Miller et al. [1996] have investigated transit time electron acceleration by fast mode waves and found it to be a very efficient mechanism under flare conditions. In this model, low-amplitude fast mode waves are assumed to be generated on very large scales, by, for example, a
large-scale perturbation to the flare magnetic field. The electron acceleration rate is proportional to the mean wavenumber of the spectrum, and the wave damping rate is proportional to the wavenumber, so that both are small initially. Hence there is essentially no damping of the waves and, since Coulomb drag cannot be overcome, no electron acceleration. As the waves cascade to higher wavenumbers, the damping rate increases. The inertial range is the range of wavenumbers where the damping timescale remains larger than the cascading timescale, and the waves can thus cascade relatively uninhibited. The inertial range in this case spans a wide range of wavenumbers and the spectral density therein is a power law. The waves cascade through the inertial range and eventually reach the dissipation range, where transit time damping by electrons with speeds greater than \( v_{te} \) is faster than cascading. The waves are then rapidly damped and these electrons, in turn, are energized out of the tail and to substantially higher energies.

Electron acceleration and wave cascading are described by coupled nonlinear diffusion equations, with particle escape from the acceleration region being neglected. The electron distribution was taken to be isotropic. Sources of sufficiently rapid pitch-angle scattering are Coulomb collisions near \( v_{te} \) and gyroresonance with waves driven unstable by an anisotropic distribution that results from transit time damping. An example of the resulting electron distributions and wave spectral densities is given in Figure 4.

The mechanism is quite robust, and it was found that the generation of 12 ergs cm\(^{-3}\) of fast mode wave turbulence on any scale less than \( \approx 10^8 \) cm and over any time interval less than about a second will yield an acceleration rate above 20 keV that is high enough to account for the hard X ray flux in an ERF. The fast mode waves also accelerate electrons to MeV energies on timescales less than about a second. On timescales of a couple of seconds, electrons are energized to tens to MeV. We also point out that, depending on the nature of the cascading, very hard energy spectra \( (E^{-1.2}) \) can be produced. These spectra are too hard to be consistent with observations, but escape will presumably soften them somewhat. The model has also not been applied to the entire duration of the flare. Here many small discrete injections or a long continuous injection of turbulence will be needed, together with replenishment of the acceleration region. We return to these issues in section 4.

If the amplitude of the MHD waves becomes sufficiently large \( (\delta B/B \sim 1) \), acceleration will no longer be a resonant process, but will proceed according to the classic Fermi mechanism of collisions with scatter-

---

**Figure 4.** Electron energy spectra \( N(E) \) and wave spectral densities \( W_T \) resulting from cascading and transit time damping of fast mode waves. The waves were injected at a wavelength of \( \approx 10^7 \) cm, at a rate of \( \approx 19 \) ergs cm\(^{-3}\) s\(^{-1}\), and over a time of 0.6 s. The ambient electron density was \( 10^9 \) cm\(^{-3}\). (a) and (b) Evolution from \( t = 4 \times 10^5 T_H \) to \( 5 \times 10^5 T_H \). \( N \) and \( W_T \) are shown at times \( t_n = (4 \times 10^5 + 10^4 n) T_H \), for \( n = 0, \ldots, 10 \). (c) and (d) Evolution from \( t = 10^6 T_H \) to \( 3 \times 10^6 T_H \). \( N \) and \( W_T \) are shown at times \( t_n = (10^6 + 2 \times 10^5 n) T_H \), for \( n = 0, \ldots, 10 \). Here \( T_H = \Omega_H^{-1} \approx 2.1 \times 10^{-7} \) s and \( U_B = B_0^2/8\pi \) is the ambient magnetic field energy density. From Figure 4 of Miller et al. [1996].
ing centers. Fermi [1949] pointed out that collisions with moving magnetic scattering centers will lead to a systematic increase in particle energy, and this process was further investigated by Davis [1956], who emphasized its diffusive or stochastic nature, Parker and Tidman [1958], who first applied it to flares, Tverskoi [1967], and Ramaty [1979].

This idea has recently received further attention for electron acceleration. Gisler and Lebons [1990] and Gisler [1992] have shown through Monte Carlo simulations that, in certain instances, Fermi acceleration can be efficient for accelerating electrons out of the background distribution. LaRosa and Moore [1993] and LaRosa et al. [1994] have applied Fermi acceleration to flares and have argued that it can account for the energization of a large fraction of the ambient electrons to \( \approx 25 \text{ keV} \). They assume that during the flare tens or hundreds of elementary flux tubes with radii of order \( 10^6 \text{ cm} \) undergo reconnection and proposed that a shear flow instability in the jets resulting from reconnection produces fast mode waves at similar scales. The wave energy at these large scales then cascades to smaller scales and ultimately to the electrons [LaRosa et al., 1996] through the Fermi mechanism. This model is similar to that of Miller et al. [1996], except that it assumes high levels of turbulence in many small (\( \approx 10^{24} \text{ cm}^3 \)) regions, as opposed to the injection of many packets of low-amplitude turbulence in a single large (\( \approx 10^{27} \text{ cm}^3 \)) region.

The rate of Fermi acceleration of electrons from 0.1 to 20 keV under flare conditions is of order a few tenths of a second once the waves reach wavelengths of about 1 km. However, as with transit time damping, ancillary pitch angle scattering is still required. While this should not be a severe requirement (see above), it remains to be shown that it can occur. The acceleration of electrons to higher energies and the nature of the energy distribution also need to be considered.

Another class of electromagnetic wave acceleration involves high-frequency (\( \omega > \Omega_H \)) waves. Sprangle and Vlahos [1983] and Karimabadi et al. [1987] examined the interaction of electrons with such a wave propagating obliquely with respect to \( B_0 \). However, only a very small fraction (\( < 10^{-3} \)) of the ambient electrons were energized. This is more likely to be a mechanism for type III radio bursts than the large-scale acceleration required for hard X ray bursts [Sprangle and Vlahos, 1983].

### 3.1.2. Electromagnetic waves: Ions

The Alfvén waves are frequently employed for ion acceleration and have been invoked to specifically energize the protons which produce nuclear gamma ray line emission [e.g., Barbosa, 1979; Miller et al., 1990] as well as the ions which escape into interplanetary space [e.g., Möbius et al., 1982; Mazur et al., 1992]. It is found, for example, that turbulence with an energy density \( \approx 10 \text{ ergs cm}^{-3} \) can accelerate protons from suprathermal to GeV nucleon\(^{-1} \) energies on timescales of order 1 to 10 s. It usually has been assumed that \( \omega \) for the resonant waves is \( \ll \Omega_H \), in which case \( |v_{\|}| \gg v_A \) [e.g., Steinacker and Miller, 1992], the first term in the frequency mismatch parameter \( x \) can be neglected, and the diffusion coefficients are simplified. However, this assumption yields an injection problem, since ions in the thermal distribution typically have speeds much less than \( v_A \) and so will be unable to resonate. For example, the threshold kinetic energy is \( \approx (1/2)m_p v_A^2 \) and for a \( v_A \) of about 2000 km s\(^{-1} \) is \( \approx 20 \text{ keV} \). This is much greater than the thermal energy of \( \lesssim 1 \text{ keV} \).

It was shown by Miller [1991] and Smith and Brecht [1993] that nonlinear Landau damping [Lee and Vlk, 1973] of the Alfvén waves can lead to significant and rapid proton heating and will energize a number of protons above this threshold. A spectrum of Alfvén waves (with \( \omega \ll \Omega_H \)) therefore can accelerate protons from thermal to ultrarelativistic energies through a combination of nonlinear and linear wave–particle interactions. Miller and Ramaty [1992] made a rough estimate of the overall efficiency of this process and Smith and Miller [1995], in a more detailed study, found that steady-state levels of \( \sim 1 \text{ erg cm}^{-3} \) of turbulence will accelerate the required number of protons in the pre-1995 scenario (see Table 1). The model has not been investigated in light of the latest observational requirements but is likely to still be viable using higher levels of turbulence. We note that Miller and Ramaty [1992] also considered nonlinear Landau damping in a multispecies plasma. They showed that the heating rate for an ion species is proportional to its mass and pointed out that this process may lead to element enhancements in the energetic particles.

While quite efficient, nonlinear Landau damping is actually not essential for the energization of protons out of the thermal distribution. Higher-frequency waves on the Alfvén branch are able to accelerate protons with energy well inside the thermal distribution, and this section of the dispersion relation is naturally populated by a cascade of wave energy from low frequencies [e.g., Zhou and Matthaeus, 1990; Verma, 1994]. This scenario thus employs only cyclotron resonance throughout the entire energization process and was first proposed by Eichler [1979] and subsequently elaborated upon by Miller and Roberts [1993].

Alfvén waves are assumed to be generated at large wavelengths by either reconnection [LaRosa et al., 1994] or large-scale perturbations to the magnetic field. Unable to resonate with protons, the waves cascade on short timescales to larger \( k \). As \( k \) increases, they are able to cyclotron resonate with progressively lower energy protons, but damping remains negligible since the particles are initially confined to thermal energies. An inertial range thus results. Ultimately, however, the waves will encounter a large number of protons in the tail and be strongly damped. Cascading will cease as a result of the rapid energy flow into the tail protons and a dissipation range will form. The tail protons, in turn, will be energized out of the thermal distribution by the high-\( k \) waves and then accelerated to much higher energies by the lower-\( k \) waves already present in the spectral density.
Miller and Roberts [1995] have explored this model with a quasi-linear code that treats acceleration and cascading with coupled nonlinear diffusion equations. Proton escape is treated with a leaky-box loss term. An example of the resulting proton distributions and wave spectral densities is shown in Figure 5. They find that the injection of $\approx 400 \text{ergs cm}^{-3}$ of Alfvén waves at any scale and over any time likely to be encountered in a flare results in the acceleration of $> 3 \times 10^5$ protons cm$^{-3}$ to energies $> 30 \text{MeV}$ on timescales $\approx 1 \text{s}$. For a volume $\approx 10^{27} \text{cm}^3$, the total number of $> 30 \text{MeV}$ protons is consistent with that inferred from gamma ray line emission. However, the proton spectra at present are not consistent with those needed to account for gamma ray line emission and can range from being either too soft to very hard ($E^{-1.5}$). A better treatment of escape may remedy this but has not yet been explored.

The explanation of ion abundance enhancements has been an active area of research, and several theories have been advanced over the years. For cold or very low energy ions, waves with $\omega \approx \Omega_i$ are required for resonance, where $\Omega_i$ is the ion cyclotron frequency. Along with the observed selectivity of the ion acceleration mechanism (especially with respect to $^{3}\text{He}$ and $^{4}\text{He}$), this strongly suggests that gyroresonance with plasma waves of frequency close to the cyclotron frequency of the enhanced ion is responsible. This is the idea be-

![Figure 5](image-url)

**Figure 5.** Proton spectrum $N(E)$ and wave spectral density $W_T$ resulting from cascading and cyclotron damping of Alfvén waves. The waves were injected at a wavelength of $\approx 10^7 \text{cm}$, at a rate of about $100 \text{ergs cm}^{-3} \text{s}^{-1}$, and over a time of $2 \text{s}$. The ambient proton density was $10^{10} \text{cm}^{-3}$. (a) $N$ at times $t_n = n(5 \times 10^4T_H)$, for $n = 0, \ldots, 10$. The leftmost curve is the spectrum for $n \leq 7$, and the remaining curves, from left to right, are the spectra for $n = 8, 9,$ and $10$. (b) Spectral densities at the same times. (c) Proton spectrum at times $t_n = n(9.5 \times 10^5T_H) + 5 \times 10^5T_H$, for $n = 0, \ldots, 10$. (d) Spectral densities at the same times. $T_H$ is the same as in Figure 4. From Figures 1 and 2 of Miller and Roberts [1995].
hind the model of Fisk [1978], in which $^3$He is energized by cyclotron resonance with $^4$He$^+$ EIC waves in the vicinity of the $^3$He cyclotron frequency. These waves are driven unstable by a relative electron-ion drift (i.e., a current) when the drift speed exceeds a critical value. However, this model requires an ambient $^4$He/He ratio of about 20–30% in order for this critical drift speed to be less than that for the excitation of other modes. These other modes could destroy or severely alter the drifting distribution and thus prevent the desired waves from being excited. In a similar model, Zhang [1995] energizes $^3$He by $\epsilon = +2$ gyroresonance with $^4$He$^+$ EIC waves above the $^3$He cyclotron frequency, and avoids the requirement of a high $^4$He/H ratio. Both models produce $^3$He heating, and another process is needed to achieve energies beyond a few tens of keV nucleon$^{-1}$. \cite{rygopoulous, varvougts, and winglee} have alternative theories, but all possess difficulties [see \cite{miller and viñas}].

The most attractive theory for the $^3$He/$^4$He ratio is that of Temerin and Roth [1992], who proposed that a bump-on-tail electron distribution is present in flares and excites $^4$He$^+$ EMIC waves around the $^3$He cyclotron frequency. The waves then resonate with and accelerate $^3$He to tens of MeV nucleon$^{-1}$ but not other ions, thus leading to the large $^3$He/$^4$He ratio. They justified the electron beam hypothesis by analogy with the aurora. Specifically, since impulsive events are rich in 2–100 keV electrons, and since electron beams are observed in the aurora and thought to be responsible \cite{temerin and lysak} for the $^4$He$^+$ EMIC waves comprising the $300 \text{ Hz ELF hiss}$ \cite{gurnett and frank}, it is likely the same instability and waves will be present in solar flares. Temerin and Roth’s idea has been elaborated upon by Miller and Viñas \cite{miller and viñas} and Miller et al. \cite{miller et al.}, who performed a linear Vlasov stability analysis of the beam, and calculated $^3$He and Fe distributions \cite{litvinenko}.

The electron beam also excites Alfvén waves around the Ne, Mg, Si, and Fe cyclotron frequencies and can preferentially accelerate these ions and lead to their enhancement over C, N, O, and $^4$He \cite{miller et al.}. The beams required for this mechanism to work are quite strong, with an energy content of around 20 ergs cm$^{-3}$ and a current density of $\approx 10^4 \text{ A m}^{-2}$. It is intriguing, though, that this current density is consistent with that implied by hard X-ray emission. However, Ne, Mg, Si, and Fe are accelerated to at most a few MeV nucleon$^{-1}$, and so the highest observed energies are not attained. A second acceleration mechanism, responsible for energization beyond a few MeV nucleon$^{-1}$, is thus required. This mechanism would also presumably accelerate C, N, O, $^4$He, and H as well at high energies. It is likely that all ions are accelerated by the same mechanism (which does not depend strongly on the charge-to-mass ratio) at high energies, for if this were not the case, it is difficult to see how the shape of the spectra would be species independent (see section 2.2).

The cascading Alfvén wave model fits in naturally with the observed heavy ion enhancements. As the waves cascade to higher frequency, they will first encounter Fe. Because of the low Fe abundance, the waves will not be completely damped and will continue to cascade up to the Ne, Mg, and Si group. Again, these ions will be accelerated and the waves will continue cascading. The waves then encounter $^4$He, C, N, and O. In light of the diminishing wave power above each group of ions, one would expect Fe to be enhanced the most, followed by Ne, Mg, and Si (relative to $^4$He, C, N, and O). Enhancements consistent with those observed have been obtained by Miller and Reames \cite{miller and reames}, who employ a quasi-linear code that simultaneously solves all ion diffusion equations and the wave equation. However, the model has not been fully explored yet and it appears that the parameter ranges in which the model works are restrictive.

3.1.3. Electrostatic waves. Langmuir waves are very effective for accelerating electrons, and were investigated extensively in the 1970s \cite{melrose}. However, a problem that has never been solved is the source of the Langmuir turbulence. In most cases, these waves are supposedly generated by a suprathermal beam of electrons ($v > v_{th}$). However, such a beam is what one is trying to produce in the first place, so that the nature of this mechanism reduces to somewhat of a “chicken and egg” problem.

Another form of electrostatic turbulence is that composed of lower hybrid waves. Relative drifts between electrons and ions are unstable to a wide range of plasma instabilities, but one with a low threshold (relative drift $< \text{ion thermal speed}$) results in the generation of lower hybrid waves \cite{papadopoulos1}. It was noted by Lampe and Papadopoulos \cite{lampe and papadopoulos} that in this instability the waves could undergo a nonlinear frequency shift, relocating the wave power toward the tail of the electron distribution function. They then argued that electrons could be diffusively accelerated out of the thermal pool into a tail. The energy gains are easily enough to account for the hard X rays below 50 keV \cite{benz and smith}, but it is unclear whether the highest-energy electrons can be produced. A problem with this mechanism is that only a small number of electrons are accelerated. Lampe and Papadopoulos \cite{lampe and papadopoulos} and Vlahos et al. \cite{vlahos et al.} estimate between 10$^{-5}$% to a few times 10$^{-3}$% of the ambient distribution. Such numbers of energetic electrons may account for weak radiation bursts, such as seen in radio emission \cite{vlahos et al.}, but it is unclear whether the highest-energy electrons can be produced. A problem with this mechanism is that only a small number of electrons are accelerated. Lampe and Papadopoulos \cite{lampe and papadopoulos} and Vlahos et al. \cite{vlahos et al.} estimate between 10$^{-5}$% to a few times 10$^{-3}$% of the ambient distribution. Such numbers of energetic electrons may account for weak radiation bursts, such as seen in radio emission \cite{vlahos et al.}, but it is unclear whether the highest-energy electrons can be produced. A problem with this mechanism is that only a small number of electrons are accelerated.
the conditions for shock formation in the corona are severe, and McClements et al. [1990, 1993] use a large ring number density corresponding to a strong shock, thus worsening the formation problem. Secondly, they assume that the ion ring is formed throughout the coronal region of a flare, but this is unlikely to be valid when the accelerating agent is a shock or reconnection site; rather one probably has many energization sites scattered randomly throughout the corona. This will greatly decrease the overall efficiency of the electron acceleration.

3.2. Shock Acceleration

Shocks have been invoked as a highly efficient acceleration mechanism in many areas of space physics and astrophysics. In particular, they can produce very high energy cosmic rays (e.g., see papers in Zank and Gaisser [1992]), and so can readily account for flare energies, provided certain conditions are met. One of the main conditions is that the shock forms in the first place; this is discussed in section 4.1

Shock acceleration is generally split into 2 types: drift and diffusive. Drift acceleration involves particles moving along the shock front, gaining energy from the shock electric field. Electrons behave approximately adiabatically [Wu, 1984; Krauss-Varban et al., 1989], since their Larmor radius is much smaller than the characteristic scale of the shock front. Electron drift acceleration is fast (a few $\Omega^{-1} m$), which in solar applications is $\ll 1$ s, but its effectiveness is limited in two ways. First, once the particle has gained energy, it tends to escape along the upstream magnetic field and, in the absence of scattering (from, for example, whistler turbulence), will not return to the shock. Second, Wu [1984] and Krauss-Varban et al. [1989] showed that acceleration took place only when the direction of propagation of the shock gets to within 1 or 2 deg of being perpendicular to the upstream magnetic field. Unfortunately, in this regime, only a few electrons ($\ll 1\%$) are accelerated. While electron drift acceleration can successfully account for a number of observations at the Earth's bow shock [Krauss-Varban and Burgess, 1991] and may be important in type II bursts [Holman and Pesses, 1983], it is effective in far too restrictive a regime to be considered seriously in flares. In drift acceleration, the ion energy gain is also very limited in the absence of upstream turbulence [Decker, 1988]. However, inclusion of upstream turbulence confines the particle to the vicinity of the shock. Test particle simulations of Decker and Vlahos [1986] have shown that in this case ion energies of a few MeV can be obtained.

Diffusive shock acceleration is similar to stochastic acceleration in that particles undergo a systematic energy gain by interacting with moving scattering centers. The difference is that since the scattering centers are moving toward each other in the rest frame of the shock, there is a first-order energy gain with each interaction so that the acceleration is much faster. Only fast mode shocks are of interest since the scattering centers do not converge in slow shocks [Isenberg, 1986]. Once again, an injection energy is needed for the process to be effective. For ion acceleration [e.g., Blandford and Eichler, 1987, and references therein], injection is relatively simple. Numerical simulations have shown that heated shocked plasma readily leaks into the upstream region and generates low-frequency electromagnetic waves which in turn resonantly scatter the ions [e.g., Quest, 1988; Kucharek and Scholer, 1991]. High energies can be readily attained (100 MeV in $< 1$ s [Ellison and Ramaty, 1985]) since the turbulence generated by the particles keeps them from escaping from the shocks [e.g., Lee, 1982]. In addition, the acceleration is prompt (a few hundred $\Omega^{-1} m$, which is $\ll 1$ s), so that shock acceleration must be viewed as a viable process for ion acceleration in flares. Note, however, that the issues of abundance anomalies have not been addressed.

For electron acceleration at shocks, many of the problems discussed in the context of stochastic acceleration arise, particularly the injection problem. In order to resonate with the turbulent Alfvén waves associated with the shock, the electrons need to be relativistic. Whistlers have a lower threshold energy (section 3.1.1) but need to be located near the electron cyclotron frequency for this threshold to lie near thermal energies. While the largest-amplitude waves at the Earth's bow shock reside at lower frequencies, whistler turbulence has also been reported extensively [Gurnett, 1985]. The implications of this for electron acceleration have not been explored. Finally, we note that shocks have one important difference from plasma wave turbulence: namely, in the shock itself there are dc electric fields [e.g., Goodrich and Scudder, 1984], which could directly produce energetic particles that are subsequently injected into a second acceleration process.

3.3. The dc Electric fields

Perhaps the most direct way to accelerate particles is by a large-scale quasi-static electric field. Most work in this area focuses on electrons, which we consider first. In addition to the force due to the electric field, an electron also experiences a Coulomb drag force from the other electrons in the distribution. It is the interplay between these two forces that govern whether or not an electron is accelerated out of the bulk distribution. As the speed of an electron increases, the drag force increases, until reaching a maximum at the electron thermal speed $v_{te}$. Above the electron thermal speed, this drag force decreases with increasing electron speed. The value of the electric field $E$ where the drag force at the thermal speed equals the electric field force is called the Dreicer field $E_D$ [Dreicer, 1960] and is given by $E_D = (e/4\pi\epsilon_0)(\omega_{pe}/v_{te})^2\ln\Lambda V m^{-1}$. Here $\ln\Lambda$ is the Coulomb logarithm, $e$ is the electron charge magnitude, and all quantities are in SI units.

This simple picture is modified somewhat by Coulomb pitch angle scattering and electron/ion collisions [Fuchs et al., 1986]. Neglecting these, for $E > E_D$ the electric force exceeds the drag force on all electrons, which will then be freely accelerated to higher energies. Such fields are called super-Dreicer. For $E < E_D$, there exists a critical velocity $v_{ci}$, below which the drag force
overcomes the electric force. Above $v_c$, the situation is reversed. Electrons with speeds $< v_c$ will then be heated, while those with speeds $> v_c$ will be freely accelerated.

For solar flare acceleration, models which employ both super- and sub-Dreicer fields have been proposed. The most advanced model in the former category is that of Litvinenko [1996b] (see also Martens [1988]). The geometry of this model is that of a large reconnecting current sheet above a bipolar magnetic loop or arcade. The sheet has a height ($x$) and length ($z$) of $\approx 10^9$ cm, a width ($y$) of a hundred meters, and contains an electric field along the length of the sheet whose strength is $\approx 10$ V cm$^{-1}$. This is several orders of magnitude higher than the Dreicer field ($\approx 10^{-4}$ V cm$^{-1}$), but is a reasonable $\vec{V} \times \vec{B}$ field for quasi-steady magnetic reconnection in the corona. The magnetic field in the sheet has a constant longitudinal $B_z$ component along the electric field, a reconnecting $B_y$ component normal to the electric field and parallel to the height of the sheet, and a transverse $B_x$ component normal to the electric field and parallel to the width of the sheet. The geometry is similar to that found in the Earth's magnetotail when the interplanetary magnetic field has an east/west component, except that $B_y$ in the magnetotail corresponds to $B_z$ in Litvinenko's model and vice versa. The longitudinal component ($\approx 100$ G) is much larger than the transverse component ($\approx 1$ G). Particles will therefore be magnetized mostly along the direction of the electric field and be able to gain large energies.

However, the energy corresponding to the potential drop along the length of the sheet ($\approx 10$ GeV) will not be realized as a result of the finite $B_y$. This component slightly magnetizes particles in the transverse direction and causes them to escape from the sheet over distances of order $10^4$ cm. Typical maximum energies are then about 100 keV. After escaping, particles can follow the magnetic field lines down to the chromosphere and generate hard X rays there. While limiting the particle energy, this rapid transverse escape prevents the current in the sheet from reaching values where the self-induced magnetic field would exceed observational limits (see section 4.2).

Litvinenko's model can yield electron energies and fluxes consistent with hard X ray observations, with few problems associated with replenishment of the acceleration region (see section 4.2). It also employs a simple geometry which seems to correspond nicely to flares with cusp structures, such as have been observed by Yohkoh [e.g., Masuda, 1994]. However, the nature of the predicted electron spectra has not been investigated yet. Also, while electron energies above 10 MeV are possible in light of the total potential drop, it is unknown how many electrons will achieve this energy before escaping (using the present ratio of magnetic field components). Decreasing the transverse component will increase this number but may lead to currents not consistent with the magnetic field.

Sub-Dreicer acceleration has been considered in detail for several years, with most of this work being applied to laboratory plasmas [Kulsrud et al., 1973] (see review by Kneepf and Spong [1979]). Electron distribution functions have been numerically calculated for a variety of $\epsilon = E/iD$ values, taking into account both electron/electron and electron/ion Coulomb energy loss rates and pitch angle scattering [e.g., Wiley et al., 1980; Fuchs et al., 1986]. Qualitatively, electrons above about $v_c$ are drawn out to higher energies and form a relatively flat distribution in parallel velocity space. Coulomb collisions pitch angle scatter the particles and thus increase the effective perpendicular temperature of the distribution.

Application of sub-Dreicer field acceleration to flares has been carried out by Holman [1985], Tsuneta [1985], and Benka and Holman [1994]. The work of Benka and Holman [1994] employed a simplified method for calculating the electron distribution function: they assume electron isotropy and solve a continuity equation that includes a loss term proportional to a power of momentum, subject to the boundary condition that the solution match one obtained by Fuchs et al. [1986] in the high-momentum regime. The resulting electron spectra are in general complicated functions of momentum and are then used to calculate X ray emission using the hybrid model discussed in section 2.1.

The X ray spectra derived from such a model have been compared with the high spectral resolution data of Lin et al. [1981]. Using an electron distribution consisting of a thermal component and a nonthermal tail of runaways, Holman and Benka [1992] obtained fits to a spectrum from early in the flare and to one from later in the flare. Figure 6 shows the results of this comparison. Each spectral fit uses five parameters: the electron temperature $T_e$ and emission measure of the thermal plasma, the critical velocity $v_c$, the maximum energy attained by a particle with initial velocity $v_c$, and the area of the thick-target interaction region. The potential drop and $\epsilon$ can be calculated from these parameters. Assuming a length $L$ for the scale of the potential drop, the density and electric field in the accelerating region can also be derived. Some of these parameters are given in the Figure 6 caption. Using the derived electron spectra, they found that $10^{33}$--$10^{34}$ electrons s$^{-1}$ needed to be accelerated in the electric field model. As mentioned in section 2.1, this is more than an order of magnitude less than that required by using a purely nonthermal model, the difference due to the fact that much of the hard X ray emission is produced by the more efficient thermal process.

Benka and Holman [1994] have carried out a more extensive analysis, decomposing the X ray light curve into gradual and spike components, as suggested by the work of Lin and Schwartz [1987]. In summary, they found that the observed hard X ray emission can be produced by fields with $\epsilon = 0.01$--0.1, classical resistivity (i.e., no enhanced resistivity from wave turbulence), and with a density of $\approx 10^{12}$ cm$^{-3}$ in the acceleration region. The spectral fits are consistent with a hot plasma contributing to the hard X ray emission early in the flare, before a distinct thermal component was evident.
at lower energies. This plasma has a higher temperature (up to $10^6$ K) and lower emission measure than the $\sim 3-4 \times 10^7$ K component observed later in the flare. This would occur if, early in the flare, heating were confined to a volume in the immediate vicinity of the current channel (a region which contains the electric field and thus a current, in addition to the runaway electrons), while later in the flare this energy was distributed over a larger volume. (The derived densities are also somewhat higher later in the flare.) However, only the gradual component produced observable heating of the plasma. They also found that $\epsilon$ did not vary systematically throughout the flare if the length of the current channels was assumed constant but that $\epsilon$ did decrease, primarily as a result of the change in temperature.

Typically, the electric fields in this model are of order $10^{-5}$ V cm$^{-1}$ and the lengths are about $3 \times 10^9$ cm, yielding maximum energies around 10 to 100 keV. Hence, while able to accelerate electrons to hard X ray producing energies, sub-Dreicer fields cannot energize electrons to the tens of MeV as needed in some flares. This situation can be corrected by invoking anomalous resistivity, which amounts to saying that Coulomb collisions are negligible compared with the scattering rate that results from resonance with waves. If this rate is assumed to be very high, the electric field required to accelerate a thermal electron (the effective Dreicer field) will also

---

**Figure 6.** Comparison of the spectra observed from the June 27, 1980, flare with the model predictions of Holman and Benka [1992]. (top) Early in the flare, before the emission peak. The dotted and dashed lines are the nonthermal and thermal bremsstrahlung contributions, respectively, and the solid line is the total. The best fit parameters are $T_e = 10^8$ K, $\epsilon = 0.13$, $L = 3 \times 10^9$ cm, $E = 7.7 \times 10^{-8}$ V cm$^{-1}$, and density $n = 9.8 \times 10^{10}$ cm$^{-3}$. (bottom) Latter in the flare, after the emission peak. The best fit parameters are $T_e = 3.6 \times 10^7$ K, $\epsilon = 0.054$, $L = 3 \times 10^9$ cm, $E = 9.9 \times 10^{-6}$ V cm$^{-1}$, and $n = 1.1 \times 10^{11}$ cm$^{-3}$. From Holman and Benka [1992].
become much larger than that for the usual Coulomb collision case. Hence, the electric field could be large ($\approx 10^{-2}$ V cm$^{-1}$, say) and still be sub-(effective) Dreicer. The problem now is generating suitable waves, in the face of high Landau damping.

The model also predicts a relationship between the rate of plasma heating $Q$ (erg s$^{-1}$) and the rate at which electrons are accelerated out of the thermal plasma $N$ (electrons s$^{-1}$). Expressions for both [Holman et al., 1989] demonstrate that $Q$ and $N$ are determined by the same physical parameters: any changes in these parameters drive the heating and electron acceleration in the same direction. Since the hard X ray emission is proportional to the instantaneous accelerated electron flux $N$, and the time rate of change of the soft X ray emission is most closely related to the heating rate $Q$, the model predicts that the hard X ray flux should follow the time derivative of the soft X ray emission. Neupert [1968] pointed out a resemblance between the observed soft X ray time profile and the integral of the microwave profile (and noted that microwaves and hard X rays are produced by electrons of similar energy). Dennis [1991] and Dennis and Zarro [1993] subsequently noted a similarity between the hard X ray time profile and the time derivative of the soft X ray profile. Benka and Holman [1992] have also successfully applied the electric field model to a representative microwave spectrum from the Owens Valley Radio Observatory.

This simple treatment of runaway acceleration does not include some potentially important plasma physics. As a result of the directed nature of electric field acceleration, the distribution function is anisotropic, with an excess of energy in the longitudinal direction. This makes it subject to velocity space instabilities, in particular the anomalous Doppler resonance (ADR) instability. Here lower hybrid waves are driven unstable by tail electrons above some critical velocity [Liu et al., 1977; Fuchs et al., 1988]. Electrons below this critical velocity resonate with waves where Landau damping by electrons in the thermal bulk prevents a net growth. Simulations by Moghaddem-Taaheri et al. [1985] showed that this instability leads to pitch angle scattering of the electrons into $v_{\perp}$ space, with the consequent limitation of the parallel energy. Moghaddem-Taaheri and Goertz [1990] (see also Holman et al. [1982]) also pointed out that the pitch angle scattering will lead to a significant enhancement in the level of synchrotron emission. The ADR instability can also greatly increase the runaway rate [An et al., 1982] above $v_{c}$.

Electric fields can also cause ions to run away. The number of ions above $v_{te}$ is negligible unless $T_{i} \gg T_{e}$. As a result, ions are not as readily available as electrons to be freely accelerated. The collisional drag force on the ions, however, unlike that on the electrons, has a minimum near 0.1$v_{te}$. If the force exerted on the ions by the electric field exceeds this minimum drag force, ions will be pulled out of the thermal distribution. For $E < E_{D}$, the ions are limited by electron drag to velocities between 0.1$v_{te}$ and $v_{ce}$. Protons drift in the direction of the electric field, while heavier ions are dragged in the opposite direction along with the drifting electrons [Harrison, 1960; Gurevich, 1961; Farth and Rutherford, 1972]. In a field with $E > E_{D}$, ions may be directly accelerated to higher velocities.

The threshold for the generation of suprathermal protons in the solar corona is $E \approx 0.5E_{D}$, and for He ions it is $E \approx 0.2E_{D}$ [Holman, 1995]. Higher Z ions, such as $^{56}$Fe$^{+26}$, have a threshold as low as $E \approx 0.08E_{D}$. (Higher values of $E/E_{D}$ are required, however, to provide a sufficient number of particles.) Since the electron thermal velocity exceeds $10^{5}$ cm s$^{-1}$ at flare temperatures of $10^{7}$ K and higher, the ions attain velocities $\approx 10^{5} - 10^{6}$ cm s$^{-1}$, and energies $\approx 10^{-4}$ keV nucleon$^{-1}$. Subsequently these ions may be accelerated by Alfvén waves through stochastic acceleration. Some of the higher Z ions have enhanced abundances relative to hydrogen in the accelerated particles. The difficulty with this process, however, is that the specific observed ion abundance enhancements noted in section 2.2 are not achievable in a simple way.

Another form of acceleration by direct electric fields involves double layers. A double layer may be defined as consisting of two equal but oppositely charged, essentially parallel but not necessarily plane, space charge layers [Block, 1978]. Present theory of double layers is split into two different types, corresponding to strong or weak double layers. Strong double layers have a large potential drop ($\gg k_{B}T_{e}$, where $k_{B}$ is Boltzmann’s constant) and are believed to be formed during the nonlinear evolution of the Buneman instability [Smith, 1985]. Volwerk and Kuijpers [1994] have suggested that such structures can be the source of observed emission in the megahertz and gigahertz bands. Weak double layers have a potential drop only of order $k_{B}T_{e}$ and arise during the evolution of an ion-acoustic instability. Clearly, a single weak double layer is of little interest to flares, but numerical simulations [Sato and Okuda, 1981; Barnes et al., 1985] indicate that for a strong magnetic field (i.e. $\Omega_{A} > \omega_{pe}$), sequences of weak double layers can form along a magnetic field line. Khan [1988] (see also Haerendal [1988]) has suggested that electrons can be accelerated to high energies by moving through many such double layers in succession.

This picture has some problems. First, one needs to account for the generation of the double layer. While the ion acoustic instability is widely believed to give rise to double layers, it does require hot electrons ($T_{e} \gg T_{i}$), so that some kind of electron preheating is needed. Perhaps Joule heating from an electric field can perform this. In this case, however, an electric field must have been already present, and since this field is also able to accelerate electrons, the need for subsequent double layer formation is not clear. Transit time damping of MHD fast mode waves is also a possibility, but again the turbulence is able to accelerate electrons from thermal to relativistic energies, and double layers are not needed. Second, the current needed for an ion acoustic instability restricts the width of the layers to a few
meters [Huba, 1985; Papadopoulos, 1979], so that a very high degree of current filamentation is needed, as is also the case in the sub-Dreicer acceleration model.

4. Global Considerations and Particle Acceleration

Another issue of importance in understanding particle acceleration is how the small scale kinetic processes discussed in section 3 are incorporated into the large scale coronal structures present during flares. Observations have long indicated that radiation in flares comes from large volumes, with scales often \( > 10^9 \) cm. In addition, the magnetic energy in a large volume of the corona must participate in the flare. One thus has to account for the connection between acceleration physics, occurring on kinetic scales, with the large-scale coronal structure. This has been approached in two ways. One is to treat the global structure in a very crude way, such as will be discussed in a moment. The second is to model the complete flare process numerically using codes that can simulate both kinetic and global processes. This has been attempted by Wingler et al. [1991] using an electrostatic particle code, but such an approach suffers from a compression of spatial and temporal scales of many orders of magnitude. For example, in these simulations, the extent of the coronal region of the flare is \( \approx 10^3 \) electron Debye lengths, or a few meters. In this review, we concentrate on the former class of modeling.

It is also important to determine how the current associated with the accelerated particles modifies the coronal magnetic field. It has long been recognized that some form of return current must exist; otherwise, the accelerated particles would generate a coronal field many orders of magnitude larger than that which is actually believed to exist, and the acceleration region would be depleted of particles in less time than the flare duration. As we shall see, in some models these problems can readily be dealt with by the presence of a cospatial return current (i.e., a bulk flow of electrons from the chromosphere to the corona, spatially coincident with the accelerated particles, of a magnitude sufficient to yield no net electric current). However, other models have geometrical constraints that forbid a cospatial return current. In these cases, other means must be found to minimize the influence of the accelerated electrons on the coronal magnetic field.

We now address the global ramifications of the three classes of acceleration models (stochastic acceleration by electromagnetic waves, collisionless shocks, and dc electric fields) in more detail.

4.1. Stochastic Acceleration and Shocks

It is convenient to consider shock and stochastic acceleration within a specific flare scenario, namely energy release by magnetic reconnection. A common feature of reconnection models is that shocks and high-speed plasma jets (of order the Alfvén speed based on the reconnecting field component) may be produced [e.g., Parker, 1963; Petschek, 1964; Vasyliunas, 1975; Forbes and Priest, 1987]. Shocks are part of the structure of the reconnecting fields in some models [Petschek, 1964] (although these are slow shocks, which are not very effective at accelerating particles), they can be generated by intense plasma heating associated with reconnection [e.g., Cargill et al., 1988], or can form when a super-Alfvénic plasma jet runs into the neighboring plasma and field [Forbes, 1986]. The jets can produce the long-wavelength waves needed in the MHD acceleration models discussed in the previous section through either a shear flow instability [Roberts et al., 1992] or through the interaction of a sub-Alfvénic jet with the plasma. Thus it is clear that coronal magnetic reconnection has the means to generate regions of particle acceleration.

A major issue with shocks is the question of their formation. Fast mode shocks are generally formed rapidly when they propagate either perpendicular or parallel to the ambient magnetic field, typically taking at most a few hundred \( 1 \) [Cargill, 1991]. However, the question in flares is whether the plasma is ever given a big enough “kick” to form a shock. Formation could occur in two ways. First, a super-Alfvénic jet such as mentioned above will form a standing shock if it interacts with neighboring plasma and field. However, it should be noted that in reconnection the jet speed is approximately Alfvénic with respect to the reconnecting field component, whereas to form a fast shock the jet must be super-Alfvénic with respect to the total field. Since the reconnecting field is likely to be a small fraction of the total field, it may not be possible to form a shock in this way [Forbes et al., 1989]. Second, shocks can form due to intense plasma heating, as was discussed by Cargill et al. [1988]. Strong heating is also a feature of magnetic reconnection, but shock formation requires that locally the plasma \( \beta \) be \( \gg 1 \). For preflare densities (\( \approx 10^{10} \) cm\(^{-3} \)) and magnetic fields (300 G), we require a temperature of at least \( 5 \times 10^6 \) K. This is not only hard to imagine, but presents seemingly impossible constraints on particle confinement mechanisms and is inconsistent with all hard X ray observations to date. It is thus our feeling that stochastic acceleration is more promising that shock acceleration in the context of coronal reconnection.

Broadly speaking, three global reconnection scenarios could arise. The first is when magnetic reconnection occurs at the top of a large coronal arcade structure, as was suggested many years ago by Carmichael [1964], Sturrock [1968], Kopp and Pneuman [1976], Cargill and Priest [1983], and Forbes et al. [1989]. In these models, magnetic reconnection proceeds vertically, with a series of loops being energized. While this model was originally developed for eruptive flares and their associated postflare loops, Yohkoh observations of Masuda [1994] and Tsuneta [1996] have provided pictures of soft X ray coronal sources in a number of flares that were not obviously associated with eruptions but had a cusp-like
shape expected from this model. In addition, Masuda [1994] showed that coronal hard X ray sources were also present in these structures, indicative of particle energization there.

In this scenario, MHD turbulence is generated from the jets below the reconnection site [Tsuneta, 1995]. Particles are accelerated there, so that some of the hard X rays come from this coronal source, and the rest comes from the footpoints as the electrons that have escaped from the footpoints hit the chromosphere. The minimum volume required for stochastic acceleration is not known at present. Current models do not include replenishment from return currents (see below) and employ volumes of $10^{24}$ to $10^{27}$ cm$^3$, which could be lower for larger levels of turbulence. Replenishment will lower the volume (for a given level of waves), but the specific amount still needs to be determined in order to place this scenario on more firm footing. Masuda [1994] has also argued that fast shocks may also exist in this region.

A second reconnection scenario involves the interaction of large scale coronal loops. This was originally proposed by Gold and Hoyle [1960], and was revived more recently by Sakai and collaborators (see Sakai and Ohsawa [1987] and Koide and Sakai [1994, and references therein]) using both magnetohydrodynamic and particle simulations. Recent Yohkoh observations have also indicated that this may be a credible scenario [DeJager et al., 1995; Inda-Koide et al., 1995]. In such a model, reconnection would again lead to pairs of plasma jets, giving rise to turbulence and subsequent acceleration.

A third invocation of coronal reconnection comes from models of fragmented flare energy release [Benz, 1985; Lu and Hamilton, 1991; Vlahos et al., 1995]. In this case the flare is comprised of many small elemental bursts, distributed randomly throughout the corona. Shock formation and development of turbulent cascades can again occur at these numerous locations. An interesting additional feature of such a model was first noted by Anastasiadis and Vlahos [1991] in the context of shock acceleration, but could equally well apply to stochastic acceleration. The difference from the first two (single-site) reconnection models is that once particles leave one region of fragmented energy release, they can be "captured" by another, and accelerated further. Anastasiadis and Vlahos [1991] gave a simple demonstration of such a model in the context of shock drift acceleration and showed that significant additional energy gains did indeed result from the interactions with multiple shocks. We suggest here that a similar process can occur when stochastic acceleration is operating.

However, the fragmented energy release model requires the spontaneous and approximately simultaneous creation and "firing" of many small energy release sites, a process that has not yet been modeled using the equations of either MHD or kinetic plasma physics. The most encouraging approach began with the work of Lu and Hamilton [1991], who noted that the magnetized solar corona could behave as a self-organized critical system, continually adjusting to reduce the free energy in such a way as to allow a whole continuum of energy releases from large flares down to microflares. This intriguing idea has unfortunately received little more in the way of rigorous justification.

Whichever of these reconnection scenarios occurs, some common facts will hold. Turbulent and shock acceleration are likely to occur over an extensive coronal volume, much larger than the volume of the reconnection current channel. Thus a particle undergoing turbulent acceleration will wander three dimensionally through the corona as it picks up energy from a continual interactions with different waves. This three dimensionality has some important implications. It is most unlikely that local buildups of charge or current can arise in the acceleration region. The former is avoided because there are no large-scale electric fields to inhibit the flow of electrons, say, into a region that happens to have too many protons. Hence quasi-neutrality will hold. Current buildup is avoided since cospatial return currents can exist in the absence of a dc electric field. As the particles stream away from the turbulence region, the associated electrostatic and inductive fields [van den Oord, 1990] will draw an immediate cospatial return current that permits them to travel to the chromosphere, subject to the usual caveats about beam/return current instabilities. Thus turbulent acceleration does not suffer from any significant global electromagnetic constraints.

4.2. The dc Acceleration

As in section 3, it is convenient to discuss the integration of super- and sub-Dreicer dc field acceleration models into the global corona separately. Super-Dreicer acceleration can occur at a single reconnecting current sheet [Litvinenko, 1996b; Martens, 1988; Martens and Young, 1990], such as that at the top of a magnetic arcade (see section 4.1). The high electric field ensures that electrons drop through an appropriate potential before escaping from the sides of the sheet on a distance much smaller than the overall sheet length. While limiting the maximum energy, rapid transverse escape prevents the current in the sheet from exceeding the value implied by Ampere's law and the assumed magnetic field (which makes a single current channel geometry possible). Once these electrons escape from the side, they can travel along the magnetic field lines toward the chromosphere. Replenishment of the current sheet can be accomplished by either bulk plasma reconnection inflows or flows upward from the chromosphere. Note that these later flows can be cospatial with the downward moving accelerated electrons, since there is no electric field in this region of space. The other two reconnection scenarios discussed above can also admit super-Dreicer fields, but the details of replenishment have not been investigated.

The sub-Dreicer model cannot involve a cospatial return current [e.g., Holman, 1985; Benka and Holman, 1994]. The requirement that the self-induced magnetic field associated with the accelerated electrons be less than the coronal field (100–1000 G) must then be introduced. To see how this constrains the model, consider that the typical flare discussed in section 2.1 produced a...
current I due to the accelerated electrons of $1.6 \times 10^{18}$ A. For a flare footpoint area of $10^{18}$ cm$^2$ and a coronal magnetic field of 500 G, we find from Ampere’s law that the radius l of a cylindrical current channel carrying these electrons cannot be greater than a few meters. This implies that the flaring corona must be filamented in such a way that neighboring current channels have oppositely directed electric fields, so that the self-induced magnetic fields due to the accelerated electrons cancel each other out. For the above parameters, $\approx 10^{12}$ cylindrical current channels are needed. If the currents are in sheets rather than cylinders, $\approx 10^4$–$10^6$ sheets are required [Holman, 1985]. This model also implies that particles are accelerated toward both footpoints in approximately equal number.

An important issue is whether this filamentation can exist in the corona before the flare begins, or is a direct result of the flare itself. The first possibility can be dismissed on the grounds that, for typical preflare parameters, the magnetic diffusion timescale $\tau_d = \mu_0 I^2/\eta$ (where $\eta$ is the resistivity) is less than a minute for classical [Spitzer, 1962] resistivities. Therefore it is difficult to see how such structures could be formed, or could persist, over the many hours prior to a flare. In addition, all these channels would be required to accelerate particles almost simultaneously. It is unclear how this could be orchestrated.

The implications of filamentation resulting from the flare or flare onset can be assessed by a simple application of Faraday’s law and have been discussed by Emslie and Hénoux [1995]. The total magnetic energy content of the cylinder is $W = (\mu_0/16\pi)L^2$, where $L$ is the length between the footpoints of the current channel and we have approximated the self-inductance of the channel by $(\mu_0/8\pi)L$. To produce this energy over a time $\tau$ requires a voltage $V = W/I = W/\pi l^2 J$, where $l$ is the current channel radius. If we write $\dot{W} = W/\tau$, then the maximum value of $l$ for a given voltage is $\approx (16\pi V/\mu_0 J)^{1/2}$. The current “turn-on” time $\tau$ can be estimated from the risetime of hard X ray bursts. This timescale can lie between tens of milliseconds and a few seconds (section 2.1), so we set $\tau = 1$s as a reasonable estimate. An upper limit on the applied voltage is imposed by the maximum energy of electrons accelerated by the direct electric field, which we take to be of order 300 kV. We again find $l$ to be a few meters. A similar value for $l$ is obtained from the analysis of a current sheet geometry. This limit is only slightly more stringent than that deduced from Ampère’s law above, and implies the existence of roughly the same number of current channels or sheets.

There are mechanisms that can produce current filamentation. One is the superheating instability, which is claimed [Heyvaerts, 1974; Bodo et al., 1991] to generate very fine scale structures, with scales less than 1 km. This instability is based on the increase in electrical conductivity with temperature, so that a positive (say) perturbation in the temperature $T$ gives rise to an increased Ohmic heating rate $\sigma E^2$. The instability thus clearly works only for voltage-driven systems; for current-driven systems the Ohmic heating rate is given by $J^2/\sigma$ and thus decreases with an increase in $T$. A positive perturbation in $T$ thus leads to reduced Ohmic heating, stabilizing the system to the perturbation. Furthermore, it is not established that oppositely directed current channels form.

Current filamentation at subkilometer scale lengths can also occur from kinetic effects. Wingee et al. [1988] have shown that the injection of an electron beam into an ambient plasma can lead to filamentation of the total current if the beam is injected over too large an area. More recent work [Winglee et al., 1991] has shown how a current sheet model for flare energy release can lead to arrays of oppositely directed currents. In each case, it is not clear whether the saturated state has a sub-Dreicer field associated with it. In any case, given that the required fine-scale filamentation is present, the concept of anywhere from $10^4$ to $10^{12}$ separate current channels (i.e., acceleration regions) fits in with the fragmented energy release concept discussed in section 4.1.

Another issue for the sub-Dreicer acceleration model is current closure. Recently, Emslie and Hénoux [1995] have proposed a model which relies on the partially ionized nature of the chromosphere to address these concerns. Incoming electrons on one field line are absorbed onto ambient protons as part of a recombination process, while spontaneous ionization of hydrogen atoms on adjacent field lines supplies the electrons for an oppositely-directed current channel. This process effectively transfers electrons across field lines; a cross-field current carried by protons, and an oppositely directed flow of hydrogen atoms provide the necessary continuity for the proton and hydrogen atom population. These complementary processes, occurring simultaneously at opposite footpoints of a magnetic flux tube, provide a natural mechanism for closure of the entire current system (see Figure 1 of Emslie and Hénoux [1995]).

4.3. The Question of Efficiency

In section 2.1 we noted that electrons with energies > 20 keV contained perhaps a few times $10^{31}$ ergs. The results of Ramaty et al. [1995] and Ramaty and Mandzhavidze (private communication, 1996) suggest that about this much energy can go into the ions above 1 MeV nucleon$^{-1}$. If we assume the canonical coronal volume of $10^{27}$ cm$^3$ (although the actual volume involved may be more or less), then the total accelerated particle energy is the equivalent to the dissipation of a coronal field with a strength of several hundred gauss. Smaller flares may require smaller field strengths, but one also needs to account for a smaller volume too. Note that we assume a 100% conversion efficiency of magnetic to particle energy! The following points need to be considered [see Sudan and Spicer, 1996; Caryil, 1996].

1. Is all of this magnetic energy accessible for dissipation? At least part of the coronal field is potential and will not be dissipated. Parker [1983], in a discussion of coronal heating mechanisms, suggests that the nonpotential component may be only 25% of the total field, thus $\approx 6\%$ of the energy.

2. How is the energy in a large volume of coronal field channeled into the dissipation regions? What fraction...
of the available field energy cannot be accessed, perhaps for topological reasons?

3. When the magnetic energy is dissipated (into plasma flows, heating, or particles), what fraction ends up in particles? Early speculation by Smith [1980] suggested < 10%.

Our present understanding of flares does not permit us to make even vaguely qualitative guesses about these numbers. However, let us assume that each part is 50% efficient (a generous estimate). Then the whole acceleration process is \((0.5)^3 \approx 10\%\) efficient. So, in order to account for the observed radiation from energetic particles, one may be forced to postulate at least kilogauss coronal field strengths. Such a field strength is problematic for several reasons. Observations suggest that the field is 1–2 kG in the photosphere, and it is generally accepted that the coronal field is less than that in the photosphere. Also, a kG field in a local region of the corona could not be confined, since its magnetic pressure would lead to its expansion and subsequent weakening.

An alternative may lie in the work of Lites, Leka, and collaborators [Lites et al., 1995; Leka et al., 1996] who have presented recent observations of the emergence of already twisted magnetic field through the photosphere. In such a scenario, there would be no need to store energy in the corona, since the electric currents responsible for the flare would have been generated below the photosphere (see also McClymont and Fisher [1989] and Sudan and Spicer [1996] for similar suggestions). The flare would then proceed as these intense magnetic fields expanded and interacted with the pre-existing coronal structure. These observations are only beginning to be interpreted and their relation to flares needs to be clearly established.

5. Summary

Flare observations and our ability to model physical processes in magnetized plasmas have developed enough that it now makes sense to strive toward a comprehensive model for impulsive flare particle acceleration. Although the observational data and our knowledge of plasma processes are still not extensive enough to settle upon one (or more) acceleration mechanism(s), we have been able to identify a number of issues that must be addressed by a successful model of flare particle acceleration.

1. The model must be capable of accelerating electrons and ions to energies in excess of 100 keV and 100 MeV, respectively, in order to account for hard X ray and gamma ray line emission. It should also allow the possibility of energizing electrons to about 10 MeV and protons to about 1 GeV, in order to account for the less common ultrarelativistic electron bremsstrahlung and pion radiation.

2. The model must be able to accelerate electrons and ions to the lower energies in less than 1 s and to the higher energies over several seconds.

3. For a large flare, the model must produce at least \(2 \times 10^{37}\) electrons s\(^{-1}\) (hybrid model), and possibly as many as \(10^{37}\) electrons s\(^{-1}\) (nonthermal model), above 20 keV and over a period of 10–100 s. It must also produce \(\approx 3 \times 10^{30}\) protons s\(^{-1}\) above 30 MeV and \(\approx 10^{35}\) protons s\(^{-1}\) above 1 MeV over the same time.

4. The model must provide electron and ion distributions that are consistent with (i.e., can be used to successfully fit) high-resolution X ray and nuclear gamma ray line spectra, respectively.

5. The model must reproduce the observed enhancements of \(^3\)He, Ne, Mg, Si, and Fe relative to \(^4\)He, C, N, and O.

6. The model must describe how the accelerated electrons and ions are pulled out of the thermal plasma.

7. The model must describe the relationship between electron acceleration and heating and, in particular, provide the observed relationships between the evolution of hot plasma and accelerated particles.

8. The model should describe the relationship between electron and ion acceleration, and in particular it should account for the simultaneity to within \(\pm 1\) s of the acceleration of these two particle species.

9. It should be clear how the local acceleration model can be incorporated into the large scale coronal structure, as observed by Yohkoh for example.

Table 3 summarizes the results of this paper for the three main acceleration processes: stochastic acceleration by MHD waves, sub- and super-Dreicer dc electric fields, and shocks. The top 13 rows deal with the properties discussed above. Each of the mechanisms has successes and failures. For example, none can account for the enhancement of \(^3\)He in flares; this requires a separate process. All can account for the observed acceleration times. With the possible exception of shock electron acceleration, particles can be extracted from the thermal plasma in each case.

A major failing of both the dc electric field models is their inability to produce energetic protons above a few MeV. The sub-Dreicer model cannot produce the most energetic electrons either, due to the finite length of the small electric field. However, at low energies, the sub-Dreicer model is the only one that has been able to reproduce measured hard X ray spectra.

The turbulence model has fewer glaring failings at this time. It can produce both high and low energy particles, but no detailed comparison with spectra has been carried out. Such comparisons are as much an issue of transport as anything else. The model may also unify ion and electron acceleration, since both fast mode and shear Alfvén waves are likely to be produced together. Shocks models are less developed, and their viability for the production of low energy electrons is a major unresolved issue.

The final three rows of the table discuss some geometrical constraints on each model. We find that the models separate out quite readily into those that are associated with large-scale coronal reconnection, and those that are not. Sub-Dreicer acceleration requires what we call hyperfilamentation of the coronal current. This is not a requirement in the other models. Instead, they can be associated with large scale (arcade or fragmented) reconnection.
Table 3. Summary of Acceleration Models

<table>
<thead>
<tr>
<th>Observation</th>
<th>Sub-Dreicer ε&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Super-Dreicer ε&lt;sup&gt;b&lt;/sup&gt;</th>
<th>MHD Turbulence&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Shocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>~ 100 keV electrons</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>~ 10 MeV electrons</td>
<td>no</td>
<td>?</td>
<td>yes</td>
<td>?</td>
</tr>
<tr>
<td>~ 100 MeV protons</td>
<td>no</td>
<td>?</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>~ 1 GeV protons</td>
<td>no</td>
<td>?</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>~ 1 s acceleration time&lt;sup&gt;e&lt;/sup&gt;</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>&lt; 100 keV electron flux</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>?</td>
</tr>
<tr>
<td>&lt; 100 MeV proton flux</td>
<td>no</td>
<td>?</td>
<td>yes</td>
<td>?</td>
</tr>
<tr>
<td>electron distribution</td>
<td>yes</td>
<td>?</td>
<td>yes</td>
<td>?</td>
</tr>
<tr>
<td>proton distribution</td>
<td>no</td>
<td>?</td>
<td>yes</td>
<td>?</td>
</tr>
<tr>
<td>³He enhancement&lt;sup&gt;f&lt;/sup&gt;</td>
<td>no</td>
<td>no</td>
<td>yes&lt;sup&gt;g&lt;/sup&gt;</td>
<td>no</td>
</tr>
<tr>
<td>Heavy ion enhancement</td>
<td>no</td>
<td>no</td>
<td>yes&lt;sup&gt;h&lt;/sup&gt;</td>
<td>no</td>
</tr>
<tr>
<td>Electron acceleration</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>?</td>
</tr>
<tr>
<td>from thermal plasma</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Ion acceleration</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Hyperfilamentation essential</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Current complexity</td>
<td>high</td>
<td>low</td>
<td>low</td>
<td>medium/low</td>
</tr>
<tr>
<td>Strength of link to</td>
<td>weak</td>
<td>good</td>
<td>good</td>
<td>?</td>
</tr>
<tr>
<td>large-scale reconnection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Based on the work of Holman et al. as discussed in text.
<sup>b</sup> Based on the work of Litvinenko as discussed in text.
<sup>c</sup> Based on the work of Miller et al. as discussed in text.
<sup>d</sup> But does not appear promising.
<sup>e</sup> To those energies given in the above 4 rows accessible by the mechanism.
<sup>f</sup> However, Hamilton and Petrosian [1992] have shown that whistlers can produce correct distributions.
<sup>g</sup> None of these models will directly give the ³He enhancement. However, if any generate a bump-on-tail electron distribution, then the model of Temerin and Roth [1992] and Miller and Viñas [1993] may be applicable.
<sup>h</sup> But the parameters are restrictive.

5.1. Recommendations for Future Observational Studies

In conclusion, we outline briefly future observational needs to help address the problem of what causes a flare.

1. The hard X ray observations of Lin et al. [1981] have demonstrated clearly the value of high spectral resolution. However, only one flare was observed. Such observations need to be repeated over a wide range of flare sizes and over energy ranges from a few keV up to 10 MeV. In addition, CGRO has demonstrated the need for high temporal resolution and Yohkoh has demonstrated the need for high spatial resolution in the hard X ray energy bands. A future instrument must have at least the same resolution.

2. High-resolution spectra of the emission above ≈ 400 keV will allow many lines associated with the accelerated ions to be resolved for the first time. The shape of the lines is a sensitive indicator of the angular distribution of the interacting ions [Wernitz et al., 1990; Murphy et al., 1991]. This allows their degree of anisotropy to be measured, which is a useful diagnostic of transport. An understanding of transport is a necessary step between the acceleration model and the observable radiations that one is making detailed comparisons with. The relative intensities of different lines provides unique information on the spectrum of the ions and on their abundances, both in the accelerated beam and in the target.

Imaging at gamma ray energies will allow the spatial distribution of the emission from accelerated ions to be studied. They could allow the fraction of gamma rays that originate in the corona to be determined for different types of flares when observed near the limb. Comparisons of X ray and gamma ray images would provide fascinating information on the spatial relationship between accelerated ions and electrons.

3. It is essential that observations of high energy particles be placed in the context of both the preflare corona and the less energetic flaring plasma. Thus good observations of a variety of UV and X ray lines are required with high spatial and temporal resolution (1-2 arc secs and 1 s). By measuring line profiles, information on mass motions will also be obtained. This will enable an assessment to be reached of whether such flows are consistent with large scale coronal magnetic reconnection, a common flare model, and could represent a fundamental breakthrough in our understanding of flares.

4. In view of the possible importance of low energy protons (10 keV to 1 MeV), it is imperative that diagnostics be developed to detect their presence. One possibility for detecting their presence would be through charge-exchange emission in the red wing of the hydro-
gen Lyman α line [Orrall and Zirker, 1976; Canfield and Chang, 1985]. Energetic helium could similarly be detected in the red wing of He 304 Å [Peter et al., 1990]. Such emission has not been successfully detected on the Sun, although it does appear to have been detected during a flare on the flare star AU Mic [Woodgate et al., 1992]. Another possible means would be the detection of p-γ lines around 8 MeV due to the resonant capture of ≈ 500 keV protons on ambient carbon [MacKinnon, 1989]. This possibility is perhaps the most direct, but it should be noted that the fluences from these lines are very low.

Acknowledgments. We thank Reuven Ramaty and Natalie Mandzhavidze for important discussions, a careful reading of the manuscript, and permission to use the unpublished electron data in Figure 3; two anonymous referees for their many comments that have much improved this paper; and Judith Karpen for also carefully reading the paper. J.A.M. was supported by the NASA Space Physics Theory, Heliospheric Physics, and Solar Physics Programs. P.J.C. was supported by the Office of Naval Research. A.G.E. was supported by the NASA Solar Physics Program and by the NSF Solar-Terrestrial Physics Division. T.N.L. was supported by a NASA Joint Venture (JOVE) grant. The Editor thanks I. S. Veselovsky and another referee for their assistance in evaluating this paper.

References


Cargill, P. J., Do electrons or protons dominate particle acceleration in solar flares?, Eos Trans. AGU, 77(37), 353, 1996.


Huba, J. D., Anomalous transport in current sheets, in *Unstable Current Systems and Plasma Instabilities in Astro-


Sakao, T., Characteristics of solar flare hard X-ray sources as revealed with the hard X-ray telescope aboard the Yohkoh satellite, Ph.D. Thesis, Univ. of Tokyo, Tokyo, Japan, 1994.


Tsuneta, S., Heating and acceleration processes in hot ther-


S. G. Benka, American Institute of Physics, College Park, MD 20740.

P. J. Cargill, Space and Atmospheric Physics, The Blackett Laboratory, Imperial College, London SW7 2BZ, United Kingdom. (e-mail: p.cargill@ic.ac.uk)


T. N. LaRosa, Department of Biological and Physical Sciences, Kennesaw State University, 1000 Chastain Rd, Kennesaw, GA 30144.

J. A. Miller and A. G. Emslie, Department of Physics, The University of Alabama in Huntsville, Huntsville, AL 35899. (e-mail: miller@mpingo.uah.edu; emslieg@email.uah.edu)

S. Tsuneta, Institute of Astronomy, University of Tokyo, Osawa 2-21-1, Mitaka, 181 Tokyo, Japan.

R. Winglee, Geophysics Program, University of Washington, Seattle, WA 98195.

(Received December 11, 1996; revised March 24, 1997; accepted March 26, 1997.)