

Solar Flares on All Scales and Coronal Heating Problem¹

by Markus J. Aschwanden

1 Magnetic Reconnection Topologies

A fundamental aspect in the physical understanding of solar flares is the geometric topology and the dynamics of the magnetic field, which drives magnetic reconnection processes during solar flares. There are three basic topologies of magnetic reconnection geometries between open and closed field lines: The pre-reconnection geometry consists of a pair of (1) open-open, (2) open-closed, or (3) closed-closed magnetic field lines (Fig. 1). If these pairs of pre-reconnection magnetic field lines are coplanar, we have a 2D model, as shown in Fig. 1 (top row, thick dashed lines). The disjoint field lines are brought into contact with each other during the reconnection process (dotted lines in Fig. 1, top row), and then relax into the post-reconnection configuration (shown with solid double lines in Fig. 1 top row).

A standard 2D flare model of the dipolar or open-open type is the Carmichael-Sturrock-Hirayama-Kopp-Pneuman (CSHKP) reconnection model. It starts with a helmet-streamer configuration with two antiparallel magnetic field lines above the cusp of the streamer, where a Y-type reconnection geometry occurs in the cusp, as observed in the famous “candle-flame” flare of 1992-Feb-21 (Tsuneta et al. 1992), which is similar to the 1999-Mar-18 flare shown in Fig. 1 (bottom left). We see that the end product is one closed (postflare) loop (Fig. 1, top left). The observations (Fig. 1, bottom left) show only the lower part with a cusp and postflare loop, but in a vertically symmetric X-type geometry we would expect also an upward reconnected segment that escapes into interplanetary space.

The tripolar type involves three magnetic poles (Fig. 1, top middle), where magnetic reconnection is referred to as “*interchange reconnection*”. Variants of this type of magnetic reconnection in tripolar geometries were also envisioned in the context of emerging-flux models (Heyvaerts et al. 1977) and particularly after the discovery of soft X-ray plasma jets with Yohkoh (Shibata et al. 1992). The observation of long straight soft X-ray jets (e.g., Fig. 1, bottom middle) were taken as evidence of plasma flows along open field lines, a fact that constitutes a flare-like process between a closed and an open field line. The end product of tripolar (open-closed) reconnection is one closed post-reconnection (postflare) loop and one open field line (Fig. 1, top middle), usually associated with a soft X-ray jet.

¹see also chapters 9 and 10 in textbook “*Physics of the Solar Corona. An Introduction*” by Markus J. Aschwanden (PRAXIS Publishing, Chichester, UK, and Springer, New York), and references therein. Electronic version at http://www.lmsal.com/~aschwand/eprints/2004_book/

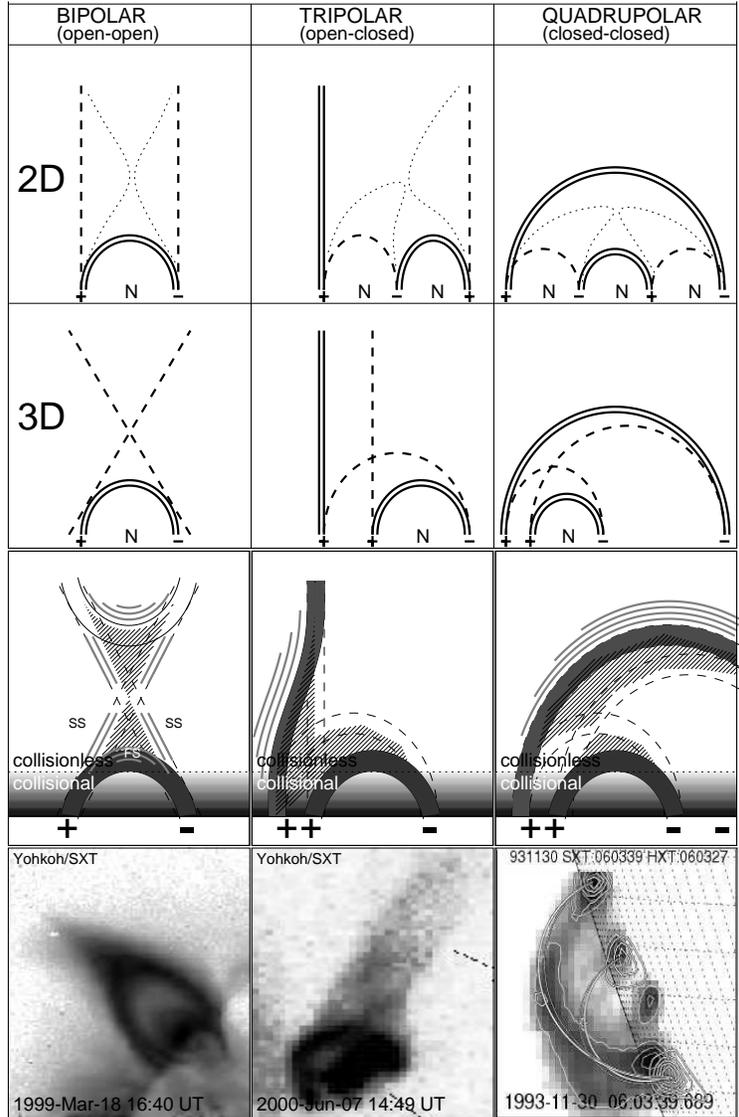


Figure 1: The topology of magnetic reconnection regions is classified into three combinations: bipolar or open-open (left column), tripolar or open-closed (middle column), and quadrupolar or closed-closed field line reconnection (right column). The 2D versions are shown in the top row, with the pre-reconnection field lines marked with dashed lines, during reconnection with dotted lines, and post-reconnection field lines with double solid linestyle. The 3D versions are indicated in the second row, where the pre-reconnection field lines are not coplanar, but located behind each other. The third row indicates the acceleration regions (hatched), the relative densities (greyscale), and upward/sideward directed shocks (grey lines). The bottom row shows flare observations from Yohkoh SXT that correspond to the three different reconnection topologies (Aschwanden 2002, 2005).

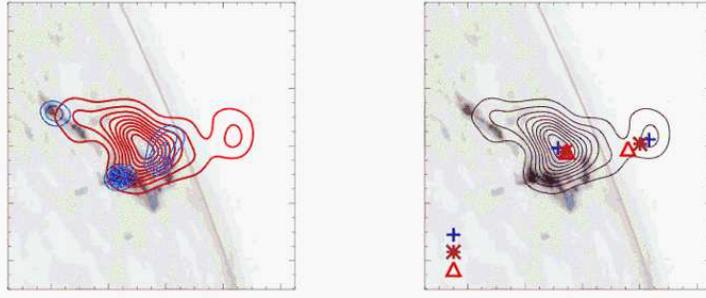


Figure 2: Left: TRACE 1600 Å image of the 2002 April 15, 23:11 UT, flare, overlaid with RHESSI contours of thermal and nonthermal emission. Right: Same image with contours of RHESSI 10-15 keV emission. The symbols indicate the centroid of the coronal sources in the energy bands of 6-8 keV, 10-12 keV, and 16-20 keV for the lower coronal source, and 10-12 keV, 12-14 keV, and 14-16 keV for the upper coronal source. Note the increasing separation of the coronal hard X-ray sources towards lower thermal energies (Sui & Holman 2003).

The quadrupolar type (Fig. 1, top right) is also called interacting-loop model and has been theoretically modeled in terms of magnetic flux transfer between two current-carrying loops (Melrose 1997). Classical examples have been observed with Yohkoh SXT by Hanaoka (1996), Nishio et al. (1997), and modeled in terms of 3D quadrupolar geometries by Aschwanden et al. (1999). The initial situation as well as the end product of quadrupolar reconnection are two closed loops, but the footpoint connectivities between opposite polarities are switched during reconnection. The outcomes are similar in 2D and 3D (Fig. 1, second row), except that the footpoints and loops are not lined up in a single plane in 3D, but can have arbitrary shear angles between the pre-reconnection loops.

Observations usually do not make the pre-reconnection configuration visible, but display the post-reconnection field lines only, because they become filled with dense hot flare plasma by the chromospheric evaporation process, which is easily to detect in soft X-rays, as shown in the examples in Fig. 1 (bottom row). Most solar flare observations are interpreted in terms of one of these basic magnetic topologies, which provide us the approximate location and geometry of particle acceleration regions (hatched areas in 3rd row of Fig. 1) and the likely propagation paths of accelerated electrons along the outgoing magnetic field lines.

1.1 Examples of Bipolar X-Point Reconnection

The standard (Carmichael-Sturrock-Hirayama-Kopp-Pneuman) flare model envisions two oppositely directed magnetic field regions that are stretched in the vertical direction to form a current sheet where they reconnect. After reconnection, the newly-connected field lines form a cusp beneath the X-type reconnection point and relax into a semi-circular, dipolar post-flare loop, which has two conjugate footpoints where the non-thermal hard X-ray emission originates. In a recent unique RHESSI observation (Sui

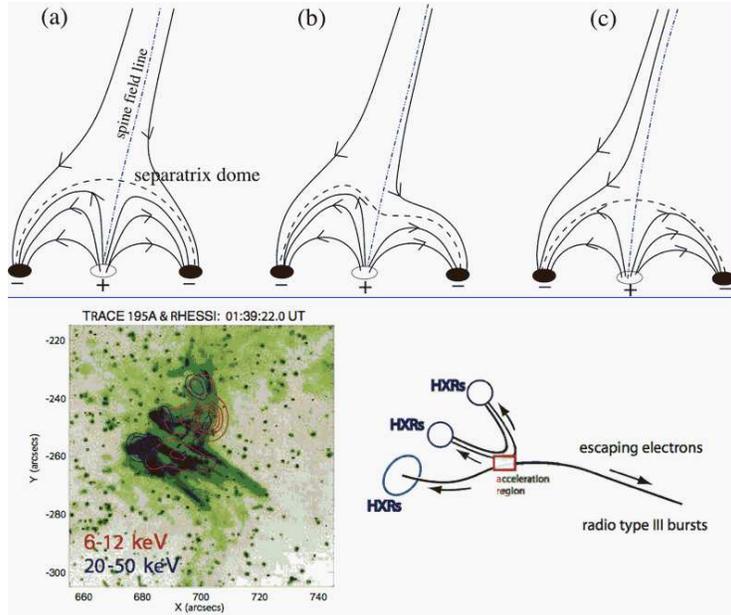


Figure 3: *Top: Magnetic topology inferred for the 1993 May 3, 23:05 UT, flare, suggesting reconnection at a 3D nullpoint where the spine field line intersects the separatrix dome (Fletcher et al. 2001) Bottom: A similar topology is observed in a flare observed with RHESSI and TRACE (Krucker & Hudson 2004).*

& Holman 2003), two coronal hard X-ray sources were observed, symmetrically placed below and above the putative X-point, at locations where the downward and upward outflows from the reconnection region are expected (Fig. 2). Moreover, the separation of the sources increased with lower energies in the thermal range of $\approx 10 - 16$ keV. Since increasing temperatures affect higher photon energies, this particular configuration indicates that thermal hard X-ray emission is observed hottest near the X-point, and progressively cooler with increasing distance from the X-point. This result was interpreted in terms of a current sheet formed above the flare loop location, as expected in the standard flare model. This observation can be considered as the first direct localization of a current sheet in a flare, and thus provides strong support for the standard model.

1.2 Examples of Tripolar Magnetic Reconnection

There is a class of magnetic topologies that involve reconnection between an open field line and a closed field line, which in the simplest case corresponds to a “tripolar” configuration. After reconnection, one footpoint of the closed loop becomes the footpoint of the open field line. In more complex 3D topologies, an isolated polarity is surrounded by a region of opposite polarity, which creates a coronal nullpoint at the

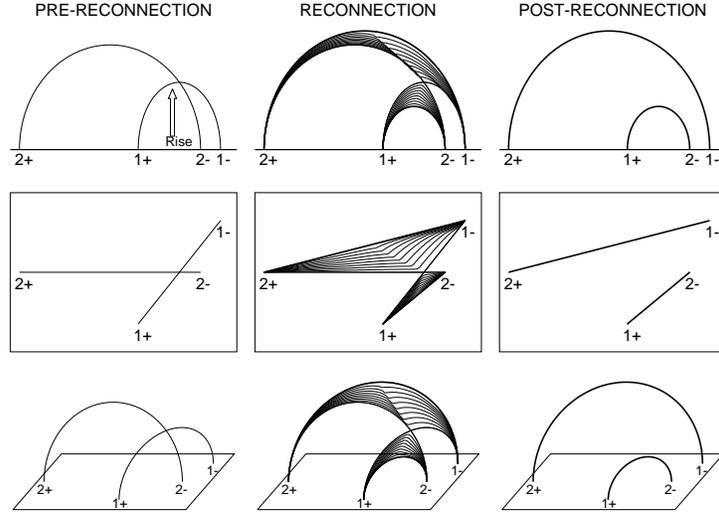


Figure 4: The concept of 3D magnetic reconnection in a quadrupolar geometry is visualized for two semi-circular loops (1 and 2), with initial footpoints (1+, 1−) and (2+, 2−). The three stages of prereconnection (left column), main reconnection phase with subsequent relaxation process (middle column), and final postreconnection phase (right column) are depicted for three different views [i.e., side view (top row), top view (middle row), and perspective view (bottom row)]. Note that all loop shapes in the initial and final phase are represented by circular segments, while the intermediate stages of relaxing field lines are rendered by linear interpolation (Aschwanden et al. 1999c).

intersection of the separatrix dome and the vertical spine field line (Fig. 3 top). Flares that are consistent with such tripolar topologies (in a 2D plane) have been observed with TRACE (Fletcher et al. 2001) and with RHESSI (Krucker & Hudson 2004). A by-product of such tripolar flares is the appearance of linear jets that are detectable in soft X-rays and EUV, as well as escaping electron beams that produce radio type III bursts (Fig. 3 bottom).

1.3 Examples of Quadrupolar Magnetic Reconnection

Some flares clearly show an interaction between two flare loops (or closed-field systems), which can most simply be interpreted as the outcome of a quadrupolar reconnection process. The magnetic configuration corresponds to a 3D reconnection case (Fig. 1, bottom right) that can be represented by a single, common, neutral line for the two interacting flare loops, which is different from the 2D case in Uchida’s model, which has three neutral lines (Fig. 1, top right). The two observed flare loops represent, of course, the postreconnection situation, but the prereconnection topology can be straightforwardly reconstructed by switching the polarities according to the scheme shown in Fig. 4. Thus, magnetic geometry is fully constrained for this type of 3D reconnection and can be reconstructed from the observed postflare loops. A number of

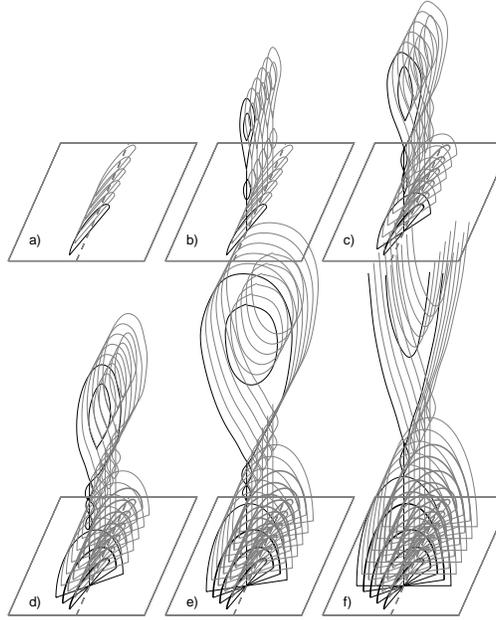


Figure 5: Scenario of the dynamic evolution during the Bastille-Day 2000-Jul-14 flare: (a) low-lying, highly sheared loops above the neutral line first become unstable; (b) after loss of magnetic equilibrium the filament jumps upward and forms a current sheet according to the model by Forbes & Priest (1995). When the current sheet becomes stretched, magnetic islands form and coalescence of islands occurs at locations of enhanced resistivity, initiating particle acceleration and plasma heating; (c) the lowest lying loops relax after reconnection and become filled due to chromospheric evaporation (loops with thick linestyle); (d) reconnection proceeds upward and involves higher lying, less sheared loops; (e) the arcade gradually fills up with filled loops; (f) the last reconnecting loops have no shear and are oriented perpendicular to the neutral line. At some point the filament disconnects completely from the flare arcade and escapes into interplanetary space (Aschwanden 2002b).

flares was found to fit quadrupolar geometry (Hanaoka 1996, 1997; Nishio et al. 1997; Aschwanden et al. 1999c).

A theoretical model for this type of 3D quadrupolar reconnection was developed by Melrose (1997) in terms of two interacting current-carrying loops. A fundamental assumption in Melrose's (1997) model is the conservation of the large-scale currents that flow through coronal loops and close below the photosphere. A consequence of this assumption is that magnetic reconnection processes only redistribute the current paths, while the net current flowing into and out of the corona remains fixed.

Large flares, such as the Bastille-Day 2000-Jul-4 event, clearly reveal a composite structure of over 100 postflare loops, which all represent the remnants of discrete individual reconnection processes. So, the building blocks of large flares may consist of bipolar, tripolar, or quadrupolar reconnection processes, as shown in Fig. 5. Even

simple flares, which display only a single soft X-ray postflare loop in images recorded with low spatial resolution, may consist of multiple reconnection processes.

2 Magnetic Reconnection Processes

The solar corona has dynamic boundary conditions: (1) The solar dynamo in the interior of the Sun constantly generates new magnetic flux from the bottom of the convection zone (i.e., the tachocline) which rises by buoyancy and emerges through the photosphere into the corona; (2) the differential rotation as well as convective motion at the solar surface continuously wrap up the coronal field with every rotation; and (3) the connectivity to the interplanetary field has constantly to break up to avoid excessive magnetic stress. These three dynamic boundary conditions are the essential reasons why the coronal magnetic field is constantly stressed and has to adjust by restructuring the large-scale magnetic field by topological changes, called *magnetic reconnection* processes. Of course, such magnetic restructuring processes occur wherever the magnetic stresses build up (e.g., in the canopy-like divergent field in the transition region, in highly tangled coronal regions in active regions, or at coronal hole boundaries). A classical example is a transequatorial coronal hole that sometimes is observed to rotate almost rigidly during several solar rotations, although the underlying photosphere displays the omnipresent differential rotation (in latitude): The shape of the coronal hole can only be preserved quasi-statically, if the photospheric magnetic field constantly disconnects and reconnects at the eastern and western boundaries. Topological changes in the form of magnetic reconnection always liberate free nonpotential energy, which is converted into heating of plasma, acceleration of particles, and kinematic motion of coronal plasma. Magnetic reconnection processes can occur in a slowly changing quasi-steady way, which may contribute to coronal heating, but more often happen as sudden violent processes that are manifested as *flares* and *coronal mass ejections*. These dynamic processes are the most fascinating plasma processes we can observe in the universe, displaying an extreme richness of highly dynamic phenomena observable in all wavelengths.

2.1 Steady 2D Magnetic Reconnection

Quasi-steady reconnection of magnetic fields enables the coronal plasma to dissipate magnetic energy, a process that has been proposed to yield direct plasma heating of the corona (e.g., Parker 1963a, 1972, 1979, 1983; Sturrock & Uchida 1981; Van Ballegoijen 1986) or to supply direct plasma heating in flares (e.g., Sweet 1958; Parker 1963a; Petschek 1964; Carmichael 1964; Sturrock 1966). This concept represents one of the most fundamental building blocks that has been used in many theoretical models of coronal heating and solar flares, which we outline in the following.

When a new magnetic flux system is pushed towards a pre-existing old magnetic flux system (e.g., as the solar wind runs into the magnetopause at the Earth's bow shock), or as a new emerging flux region pushes through the chromosphere upwards into a pre-existing coronal magnetic field, a new dynamic boundary is formed where the magnetic field can be directed in opposite directions at both sides of the boundary.

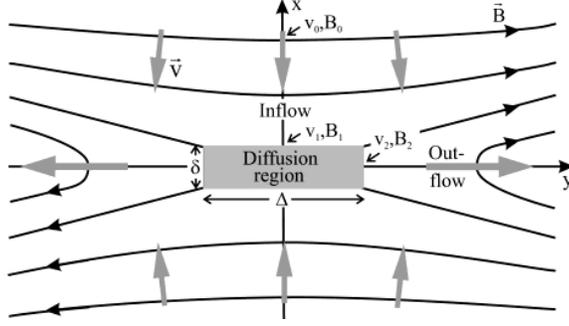


Figure 6: Basic 2D model of a magnetic reconnection process, driven by two oppositely directed inflows (in x-direction), which collide in the diffusion region and create oppositely directed outflows (in y-direction). The central zone with a plasma- β parameter of $\beta > 1$ is called the diffusion region (grey box) (Schindler & Hornig 2001).

The magnetic field has then necessarily to drop to zero at the boundary to allow for a continuous change from a positive to a negative magnetic field strength. Thus the balance between the magnetic and thermal pressure across the *neutral boundary layer*,

$$\frac{B_1^2}{8\pi} + p_1 = p_{nl} = \frac{B_2^2}{8\pi} + p_2, \quad (2.1)$$

yields a higher thermal pressure (p_{nl}) in the neutral layer (where $B = 0$) than on both sides with finite field strengths B_1 and B_2 . In a 1D model we would have an infinite neutral boundary layer. In reality, however, the process of bringing two oppositely directed magnetic flux systems together will always have a finite area of first contact, which limits the extent of the neutral boundary layer and channels outflows to both sides, so that the simplest scenario is a 2D model as shown in Fig. 6, where the lateral inflows (driven by external forces) will create outflows along the neutral line in an equilibrium situation. The plasma- β parameter $\beta = p_{th}/(B_1^2/8\pi)$ becomes larger than unity in the central region (because $B_1 \mapsto 0$), so that the plasma can flow across the magnetic field lines, which is called the *diffusion region*, and is channeled into the outflow regions along the neutral boundary. Outside the diffusion region the plasma- β again drops below unity and the magnetic flux is frozen-in. The highly pointed magnetic field lines in the outflow region experience a high curvature force that tries to smooth out the cusps in the outflow region until a balance between the outward-directed magnetic tension force and the inward-directed magnetic pressure force plus thermal pressure is achieved. This magnetic field line relaxation process is also called the *sling shot effect*, which is the basic conversion mechanism of magnetic into kinetic energy. The stationary outflows are sandwiched between two *standing slow shocks* (which do not propagate). The end result is a thin diffusion region with width δ and length Δ (Fig. 6). The whole process can evolve into a steady-state equilibrium with continuous inflows and outflows, driven by external forces. Since the Lorentz force creates an electric field E_0 in a direction perpendicular to the 2D-plane of the flows (i.e. perpendicular to the image plane of Fig. 6), a current j_{nl} in the neutral layer is

associated with the electric field E_0 according to Ohm's law,

$$E_0 = \frac{1}{c}v_1B_1 = \frac{1}{c}v_2B_2 = \frac{j_{nl}}{\sigma}, \quad (2.2)$$

which is termed the *current sheet* for the diffusion region. The finite resistivity σ requires, strictly speaking, a treatment in the framework of resistive MHD, although the processes outside the diffusion region can be approximated using the ideal MHD equations.

2.2 Sweet–Parker Reconnection Model

There exists no full analytical solution for the steady-state situation of the reconnection geometry shown in Fig. 6 using the full set of resistive MHD equations, but separate analytical solutions for the external (ideal MHD) region and special solutions for the (resistive MHD) diffusion region are available that can be matched with some simplifications. One such solution is the *Sweet–Parker* model (Sweet 1958; Parker 1963a), where it is assumed that the diffusion region is much longer than it is wide, $\Delta \gg \delta$. For steady, compressible flows ($\nabla \cdot \mathbf{v} \neq 0$), it was found that the outflows roughly have Alfvén speeds,

$$v_2 = v_A = \frac{B_2}{\sqrt{4\pi\rho_2}}, \quad (2.3)$$

and that the outflow speed v_2 relates to the inflow speed v_1 reciprocally to the cross sections δ and Δ (according to the continuity equation),

$$\rho_1v_1\Delta = \rho_2v_2\delta, \quad (2.4)$$

and that the *reconnection rate* M_0 , defined as the Mach number ratio of the external inflow speed v_0 to the (Alfvén) outflow speed v_A , is (with the approximation $B_1 \approx B_0$, $v_1 \approx v_0$, and $S_1 \approx S_0$),

$$M_0 = \frac{v_0}{v_A} = \frac{1}{\sqrt{S_0}}. \quad (2.5)$$

The *Lundquist number* S (or *magnetic Reynolds number*) is defined by

$$S = v_AL/\eta, \quad (2.6)$$

analogous to the Reynolds number $R = vL/\eta$ defined for a general fluid velocity v . From Eqs. (2.4–2.6) the following relation follows

$$v_0 = \frac{\eta}{\delta}. \quad (2.7)$$

So, for typical coronal conditions (with a large Lundquist number of $S_0 = R_m \approx 10^8 - 10^{12}$) the reconnection rate is typically $M_0 \approx 10^{-4} - 10^{-6}$, which yields inflow speeds in the order of $v_0 \approx v_AM_0 \approx 1000 \text{ km s}^{-1} \times 10^{-5} \approx 0.01 \text{ km s}^{-1}$ (using Eq. 2.5) and yields extremely thin current sheets with a thickness of $\delta = \Delta(v_A/v_1) \approx \Delta \times 10^{-5}$ (using Eq. 2.4). So, a current sheet with a length of $\Delta \approx 1000 \text{ km}$ would have a thickness of only $\delta \approx 10 \text{ m}$. In typical flares, energies of $\varepsilon_{tot} \approx 10^{28} - 10^{32}$

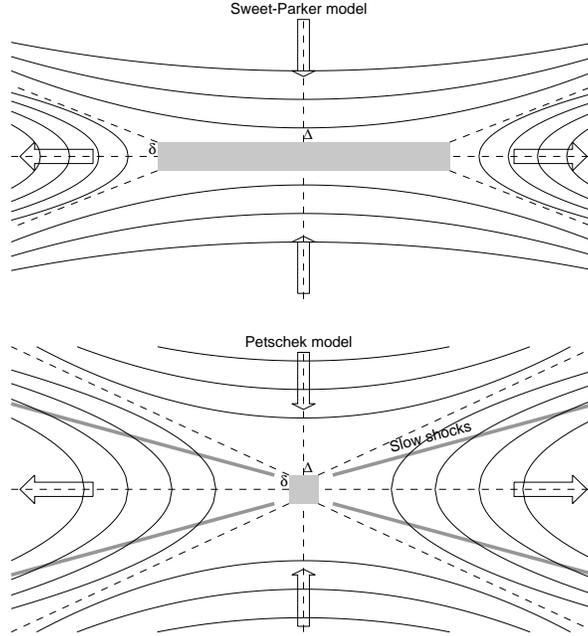


Figure 7: Geometry of the Sweet–Parker (top) and Petschek reconnection model (bottom). The geometry of the diffusion region (grey box) is a long thin sheet ($\Delta \gg \delta$) in the Sweet–Parker model, but much more compact ($\Delta \approx \delta$) in the Petschek model. The Petschek model also considers slow-mode MHD shocks in the outflow region.

erg are dissipated over flare durations of $\Delta t \approx 10 - 10^2$ s, which imply much larger dissipation rates than obtained with the Sweet–Parker current sheet,

$$\frac{d\varepsilon_m}{dt} = \frac{B^2}{8\pi} \frac{dV}{dt} \approx \frac{B^2}{8\pi} L^2 v_0 \approx \frac{B^2}{8\pi} \frac{L^2 v_A}{\sqrt{S_0}} \approx 10^{22} \left(\frac{B}{100 \text{ G}} \right)^2 \left(\frac{L}{1 \text{ Mm}} \right)^2 \left(\frac{v_A}{1 \text{ Mm/s}} \right), \quad (2.8)$$

so the Sweet–Parker reconnection rate is much too slow to explain the magnetic dissipation in solar flare events.

2.3 Petschek Reconnection Model

A much faster reconnection model was proposed by Petschek (1964), which involved reducing the size of the diffusion region to a very compact area ($\Delta \approx \delta$) that is much shorter than the Sweet–Parker current sheet ($\Delta \gg \delta$) (Fig. 7). Summaries of the Petschek model can be found, see example in Priest (1982, p. 351), Jardine (1991), Priest & Forbes (2000, p. 130), Treumann & Baumjohann (1997, p. 148), and Tajima & Shibata (2002, p. 225). Because the length of the current sheet is much shorter, the propagation time through the diffusion region is shorter and the reconnection process becomes faster. However, in a given external area with size L_e comparable with the

length Δ_{SP} of the Sweet–Parker current sheet, a much smaller fraction of the plasma flows through the Petschek diffusion region with size Δ_P , where finite resistivity σ exists and field lines reconnect. Most of the inflowing plasma turns around outside the small diffusion region and *slow-mode shocks* arise where the abrupt flow speed changes from v_1 to v_2 in the outflow region (Fig. 7, bottom). The shock waves represent an obstacle in the flow and thus are the main sites where inflowing magnetic energy is converted into heat and kinetic energy. Simple energy considerations show that inflowing kinetic energy is split up roughly in equal parts into kinetic and thermal energy in the outflowing plasma (Priest & Forbes 2000). Petschek (1964) estimated the maximum flow speed v_e by assuming a magnetic potential field in the inflow region and found that at large distance L_e the external field $B_0(L_e)$ scales logarithmically with distance L_e ,

$$B_0(L_e) = B_0 \left[1 - \frac{4M_0}{\pi} \ln \left(\frac{L_e}{\Delta} \right) \right]. \quad (2.9)$$

Petschek (1964) estimated the maximum reconnection rate M_0 at a distance L_e where the internal magnetic field is half of the external value (i.e., $B_0(L_e) = B_0/2$), which yields using Eq. (2.9),

$$M_0 = \frac{\pi}{8 \ln(L_e/\Delta)} \approx \frac{\pi}{8 \ln(R_{me})}. \quad (2.10)$$

So, the reconnection rate $M_0 = v_0/v_A$ depends only logarithmically on the magnetic Reynolds number $R_{me} = L_e v_{Ae}/\eta$. Therefore, for coronal conditions, where the magnetic Reynolds number is very high (i.e., $R_{me} \approx 10^8 - 10^{12}$), the Petschek reconnection rate is $M_0 \approx 0.01 - 0.02$ according to Eq. (2.10), yielding an inflow speed of $v_0 \approx v_A M_0 \approx 10 - 20 \text{ km s}^{-1}$ for typical coronal Alfvén speeds of $v_A \approx 1000 \text{ km s}^{-1}$. Thus, the Petschek reconnection rate is about three orders of magnitude faster than the Sweet–Parker reconnection rate.

2.4 Unsteady/Bursty 2D Reconnection

When the diffusion region gets too long (such as in the Sweet–Parker model), it becomes unstable to *secondary tearing* (Furth et al. 1963) and an *impulsive bursty regime* of reconnection ensues (Priest 1986; Lee & Fu 1986; Kliem 1995; Priest & Forbes 2000, § 6-7). Such unsteady reconnection modes are very likely to operate in solar flares, because bursty and intermittent pulses (on time scales of seconds to subseconds) have been observed in hard X-ray and radio signatures of particle acceleration during virtually all flares. In the following we describe a few of those unsteady reconnection modes, such as tearing instability (§ 2.5), coalescence instability (§ 2.6), and their combined dynamics (i.e., the regime of *bursty reconnection*, § 2.7). There are also other unsteady reconnection types, such as X-type collapse (Dungey, 1953; Craig & McClymont 1991, 1993; Craig & Watson 1992a; McClymont & Craig 1996; Priest & Forbes, 2000, p. 205), resistive reconnection in 3D (e.g. Schumacher et al. 2000; Priest & Forbes, 2000, p. 230), or collisionless reconnection (e.g. Drake et al. 1997; Haruki & Sakai, 2001a, b). The latter has not yet been applied to solar flares, but has been discovered in the Earth’s magneto-tail (Øieroset et al. 2001).

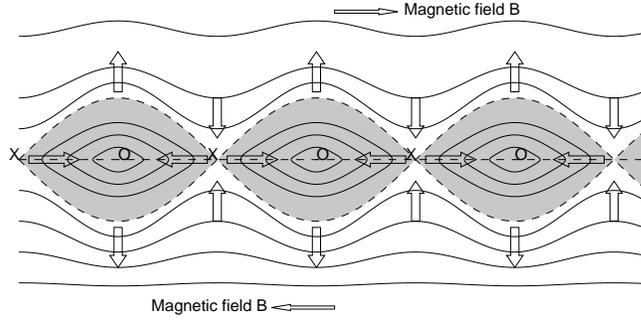


Figure 8: Magnetic island formation by tearing-mode instability in the magnetic reconnection region. Magnetically neutral X and O points are formed at the boundary between regions of an oppositely directed magnetic field, with plasma flow in the directions indicated by the arrows (after Furth et al. 1963).

2.5 Tearing-Mode Instability and Magnetic Island Formation

In current sheet formations, resistive instabilities can occur, where the magnetic field lines can move independently of the plasma due to the non-zero resistivity (opposed to the *frozen-flux theorem* for zero resistivity). In magnetic reconnection regions with high magnetic Reynolds numbers ($R_m = \tau_d/\tau_A$), where the outward diffusion (on a time scale of $\tau_d = l^2/\eta$, with $2l$ the width of the current sheet and $\eta = (\nu\sigma)^{-1}$ the magnetic diffusivity) is much larger than the Alfvén transit time $\tau_A = l/v_A$ (i.e., $\tau_d \gg \tau_A$), three different types of resistive instabilities can occur: *gravitational*, *rippling*, and *tearing mode* (Furth et al. 1963). Essentially, an Alfvén disturbance can trigger an instability before it can be stabilized by magnetic diffusion, when $\tau_d \gg \tau_A$ (i.e., for large Reynolds numbers $R_m = \tau_d/\tau_A$). The tearing mode, which has a wavelength greater than the width of the sheet ($kl < 1$), has a growth time τ_G^{tear} of

$$\tau_G^{tear} = [(kl)^2 \tau_d^3 \tau_A^2]^{1/5}, \quad (2.11)$$

for wave numbers in the approximate range $(\tau_A/\tau_d)^{1/4} < kl < 1$ (e.g., see derivations in Furth et al. 1963; Priest 1982, p. 272; White 1983; and Sturrock 1994, p. 272). Thus, the mode with the longest wavelength has the fastest growth rate,

$$\tau_{G,min}^{tear} = [\tau_d \tau_A]^{1/2}. \quad (2.12)$$

Tearing mode produces magnetic islands in 2D (see Fig. 8) or magnetic fluxropes in 2.5D, respectively. These structures saturate in the nonlinear phase of the tearing mode (if coalescence is not permitted) and their subsequent diffusion at the diffusive time scale τ_d is extremely slow (since $R_m \gg 1$ in the corona). The energy release of tearing-mode instability occurs during the process of island formation. Tearing modes have been applied to solar flares in a number of theoretical studies (e.g. Sturrock 1966; Heyvaerts et al. 1977; Spicer 1977a, b, 1981a; Somov & Verneta, 1989; Kliem 1990), and

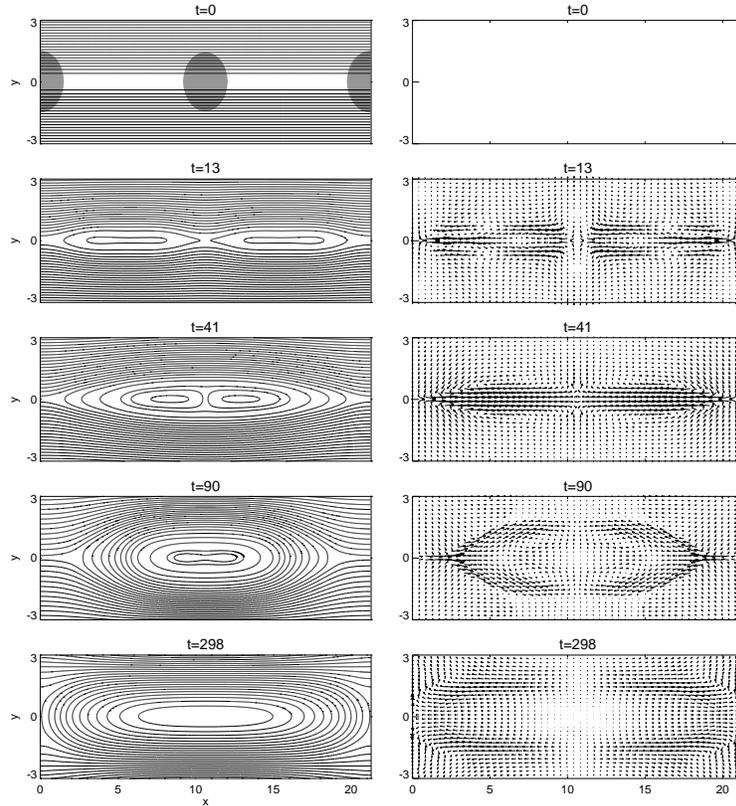


Figure 9: MHD simulation of the coalescence instability for a Lundquist number of $S = 1000$ and a plasma- β of 0.1. The magnetic field is shown in left-hand panels, the velocity field in the right-hand panels. Initial resistivity perturbation is shown shaded (Schumacher & Kliem 1997a).

numerical MHD simulations have been performed (Biskamp & Welter 1989). Kliem (1995) estimated the growth time of tearing mode for coronal conditions ($n_e = 10^{10} \text{ cm}^{-3}$, $T = 2.5 \times 10^6 \text{ K}$, $B = 200 \text{ G}$, with smallest current sheet half-widths of $l \approx 7 \times 10^3 \text{ cm}$), which yields $\tau_G^{tear} \approx 0.4 \text{ s}$. This time scale is comparable with the duration of elementary time structures observed in the form of hard X-ray pulses and radio type III bursts. Because tearing mode has a threshold current density orders of magnitude below the threshold of kinetic current-driven instabilities, it will occur first. Continued shearing and tearing may reduce the width of the current sheet until the threshold of kinetic instability is reached (Kliem 1995).

2.6 Coalescence Instability

While tearing mode leads to filamentation of the current sheet, the resulting filaments are not stable in a dynamic environment. If two neighboring filaments approach each other and there is still non-zero resistivity, they enter another instability, the *coales-*

cence instability, which merges the two magnetic islands into a single one (Pritchett & Wu 1979; Longcope & Strauss 1994; Haruki & Sakai 2001a, b). An example of an MHD simulation is shown in Fig. 9 (Schumacher & Kliem 1997a). Coalescence instability completes the collapse in sections of the current sheet, initiated by tearing-mode instability, and thus releases the main part of the free energy in the current sheet (Leboeuf et al. 1982). There is no complete analytical description of coalescence instability, but numerical MHD simulations (Pritchett & Wu, 1979; Biskamp & Welter 1979, 1989; Leboeuf et al. 1982; Tajima et al. 1982, 1987; Schumacher & Kliem 1997a) show that the evolution consists of two phases: first the pairing of current filaments as an ideal MHD process, and then a resistive phase of pair-wise reconnection between the approaching filaments. The characteristic time scale of the ideal phase is essentially the Alfvén transit time through the distance l_{coal} between the approaching current filaments,

$$\tau_{coal} = \frac{1}{q_{coal}} \frac{l_{coal}}{v_A}, \quad q_{coal} = \frac{u_{coal}}{v_A} \approx 0.1 - 1 \quad (2.13)$$

where u_{coal} is the velocity of the approaching filaments. For coronal conditions (say $n_e = 10^{10} \text{ cm}^{-3}$, $B = 200 \text{ G}$, $l_{coal} = 1000 \text{ km}$) we estimate coalescence times of $\tau_{coal} \approx 0.2 - 2.0 \text{ s}$, which is again the typical time for the observed modulation of hard-X ray pulses and type III electron beams in flares.

2.7 Dynamic Current Sheet and Bursty Reconnection

In praxis, the two previously described processes of tearing instability and coalescence instability occur iteratively, leading to a scenario of *dynamic current sheet evolution*, also known as *impulsive bursty reconnection* (Leboeuf et al. 1982; Priest 1985a; Tajima et al. 1987; Kliem 1988, 1995). A long current sheet is first subject to tearing that creates many filaments, while rapid coalescence clusters and then combines groups of closely spaced filaments, which are once again unstable to secondary tearing, to secondary coalescence, and so forth. MHD simulations reproduce this iterative chain of successive tearing and coalescence events (Malara et al. 1992; Kliem et al. 2000). An example of such a numerical simulation from the study of Kliem et al. (2000) is shown in Fig. 9 (magnetic field evolution). Let us review three key studies (Tajima et al. 1987; Karpen et al. 1995; Kliem et al. 2000), where numerical MHD simulations of this process have been applied to solar flares.

Tajima et al. (1987) performed numerical MHD simulations of the nonlinear coalescence instability between current-carrying loops and derived an analytical model of the temporal evolution of electromagnetic fields [see also two comprehensive reviews on this subject by Sakai & Ohsawa (1987) and Sakai & De Jager (1996), and references therein]. This nonlinear system evolves into an oscillatory relaxation dynamics, driven by the interplay of the $\mathbf{j} \times \mathbf{B}$ force and the hydrodynamic pressure response, which was modeled analytically by Sakai & Ohsawa (1987). The oscillatory behavior is very appealing, because it provides a possible explanation for the numerous quasi-periodic time structures observed in radio and hard X-rays during flares. An oscillatory regime of fast reconnection has also been found from other theoretical work on current instabilities in current sheets (Smith 1977) and X-point relaxation (Craig & McClymont 1991, 1993).

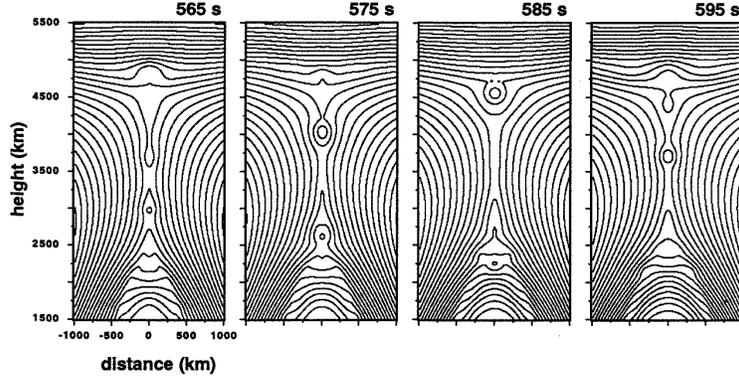


Figure 10: Magnetic field lines near the reconnection region at four different times (565, 575, 585, 595 s) during a strong-shear MHD simulation by Karpen et al. (1995). Note the tearing along the vertical current sheet (first frame), which forms two magnetic islands (second frame), which are ejected from the sheet and merge with the flux systems above or below the sheet (third frame), followed by another tearing plus magnetic island formation (fourth frame). (Karpen et al. 1995).

Karpen et al. (1995) performed 2.5-dimensional numerical MHD simulations of shear-driven magnetic reconnection in a double arcade with quadrupolar magnetic topology. For strong shear, the initial X-point was found to lengthen upward into a current sheet, that reconnects gradually for a while but then begins to undergo multiple tearing. Several magnetic islands develop in sequence, move towards the ends of the sheet, and disappear through reconnection with the overlying and underlying field (Fig. 10). A second study with similar quadrupolar configuration was performed, but with asymmetric shear in dipoles with markedly unequal field strengths (Karpen et al. 1998). Similar intermittency was found in the shear-driven magnetic reconnection process, and the simulations moreover show that each dissipated magnetic island leaves a footprint in the form of fine filaments in the overlying separatrix layer (Fig. 11). This dynamic behavior is essentially identical to the pattern of repeated tearing and coalescence first investigated by Leboeuf et al. (1982) and dubbed *impulsive bursty reconnection* by Priest (1985b). In Fig. 11 there are also some other dynamic processes present: (a) a thin region along the slowly rising inner separatrix is compressed; (b) a downflow with $v \approx 30 \text{ km s}^{-1}$; (c) this is followed by an upflow along the same field lines. Although these simulations by Karpen et al. (1995, 1998) are carried out using parameters corresponding to chromospheric conditions, it demonstrates that magnetic reconnection in sheared flare arcades occurs in a bursty and intermittent mode, and not in a quasi-stationary Sweet–Parker or Petschek mode. The physical origin of this intermittent reconnection dynamics is most essential to understanding the rapidly varying time structures of accelerated particles.

A recent work on *impulsive bursty reconnection* applied to solar flares was carried out by Kliem et al. (2000). Fig. 12 shows the evolution of tearing, magnetic island formations, magnetic island coalescence, secondary tearing, and so forth. Tearing and

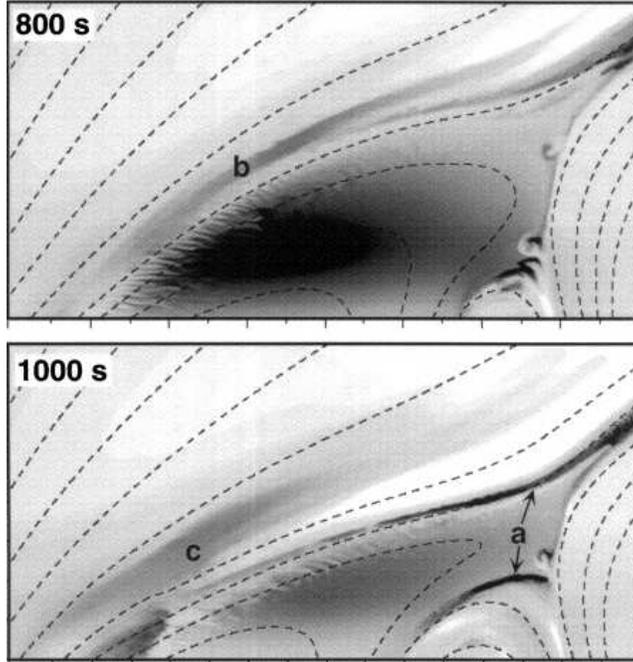


Figure 11: Mass density difference ratio (greyscale) and projection of magnetic field lines into the image plane (dashed lines) at 800 s and 1000 s in the vicinity of the reconnection region, during an MHD simulation of a sheared arcade. The location **a** corresponds to a thin compressed region along the slowly rising inner separatrix, **b** to a narrow downflow falling outside of the left outer separatrix, and **c** indicates a broader upflow that follows along the same field lines (Karpen et al. 1998).

coalescence in the bursty magnetic reconnection mode also modulates particle acceleration on time scales that are observed in radio and hard X-rays, and is more consistent with flare observations than steady reconnection scenarios. The iterative processes of tearing and coalescence may repeat down to microscopic scales (of the ion Larmor radius or the ion inertial length), producing a *fractal current sheet* (Shibata & Tanuma 2001). A similar concept is that of MHD turbulent cascading, which leads to similar high fragmentation at the smallest spatial scales, called *turbulent reconnection* (e.g., Kim & Diamond 2001; Matthaeus 2001b) and applied to flares (Moore et al. 1995; Somov & Oreshina 2000). The two concepts of fractal current sheets and turbulent reconnection could possibly be discriminated observationally from the frequency distribution of time scales, since fractal processes are scale-free and generally produce powerlaw distributions, while turbulent processes are controlled by incoherent random processes that generally produce exponential distributions (Fig. 22).

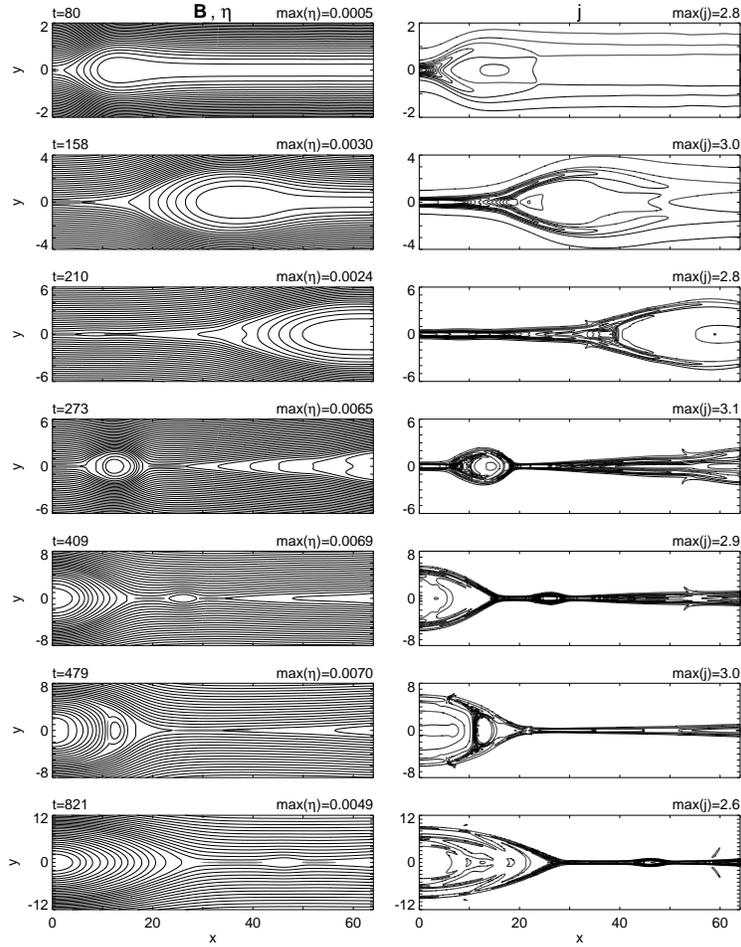


Figure 12: 2D MHD simulation of dynamic magnetic reconnection, showing the magnetic field (left panels) and current density (right panels). Regions with anomalous resistivity are shown shaded in the magnetic field plot (at $y=0$) (Kliem et al. 2000).

3 Flare/CME Models

In this section we discuss the most eminent physical models for flare and CME processes, which all involve magnetic reconnection in some form. What distinguishes the different flare models are mainly the initial magnetic topologies, which are prone to specific instabilities or drivers. This section covers mainly the theoretical aspects of flare models, while supporting observations are compiled elsewhere (e.g., § 10.6 of Aschwanden 2004). Theoretical reviews on flare/CME models can be found in Brown & Smith (1980), Melrose (1993), Shibata (1998), Priest (2000), Forbes (2000b, 2001), Klimchuk (2001), Low (1999b, 2001b), Priest & Forbes (2002), or in textbook chapters

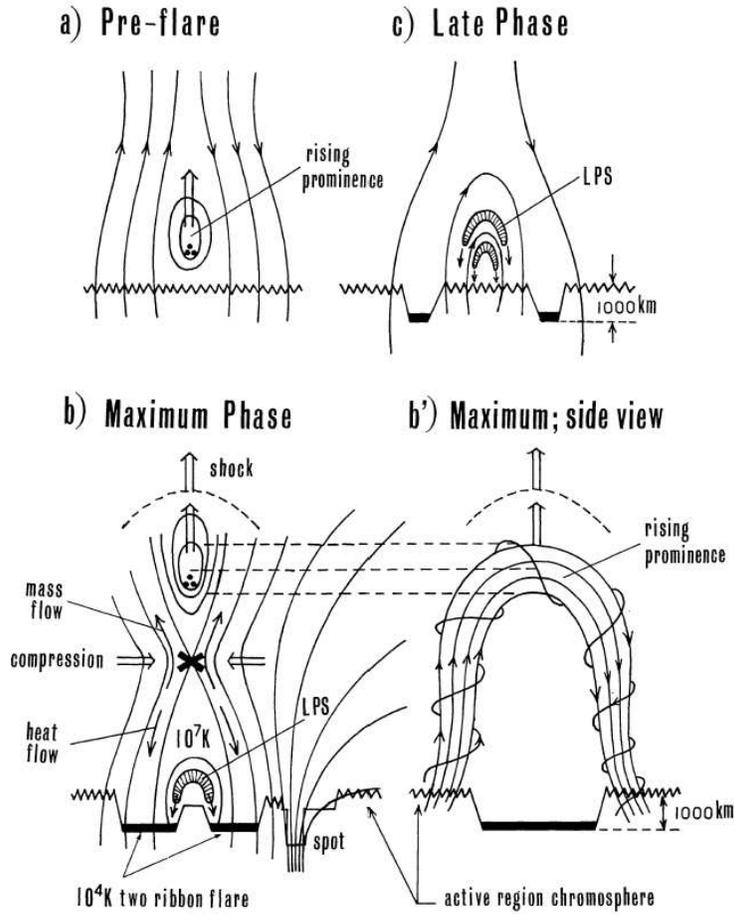


Figure 13: Temporal evolution of a flare according to the model of Hirayama (1974), which starts from a rising prominence (a), triggers X-point reconnection beneath an erupting prominence (b), shown in sideview (b'), and ends with the draining of chromospheric evaporated, hot plasma from the flare loops (c) (Hirayama 1974).

(Svestka 1976, § 6; Priest 1982, § 10; Priest & Forbes 2000, § 11; Tajima & Shibata 2002, § 3.3).

3.1 The Standard 2D Flare Model

Although not all flares can be explained by a single model, it is justified to establish a standard model that fits most of the observations and has a well-understood theoretical foundation. The most widely accepted standard model for flares is the 2D magnetic reconnection model that evolved from the concepts of Carmichael (1964), Sturrock (1966), Hirayama (1974), Kopp & Pneuman (1976), called the *CSHKP model* accord-

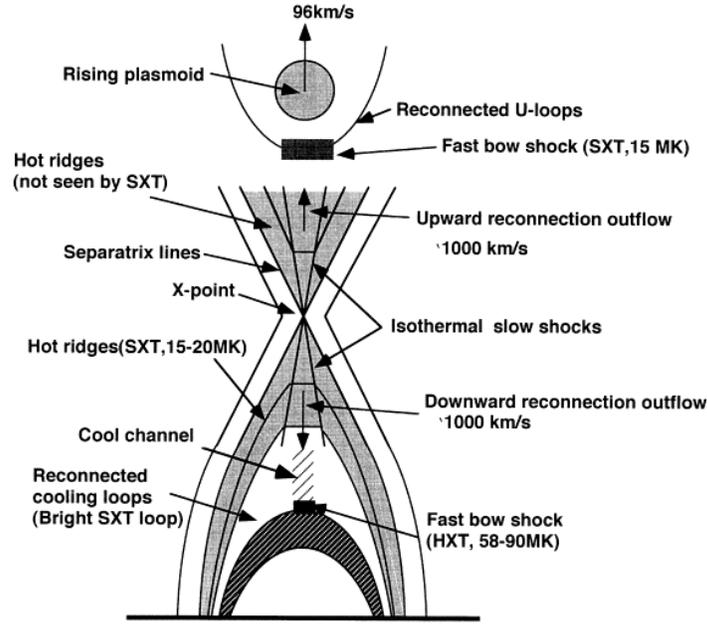


Figure 14: Elaborate version of the standard 2D X-type reconnection model that also includes the slow and fast shocks in the outflow region, the upward-ejected plasmoid, and the locations of the soft X-ray bright flare loops (Tsuneta 1997).

ing to the initials of these five authors. This has been further elaborated by Tsuneta (1996a; 1997) and Shibata (1995) based on the modeling of Yohkoh observations.

The initial driver of the flare process is a rising prominence above the neutral line in a flare-prone active region (Fig. 13a). The rising filament stretches a current sheet above the neutral line, which is prone to Sweet–Parker or Petschek reconnection (Fig. 13b). In the model of Sturrock (1966), a helmet streamer configuration was assumed to exist at the beginning of a flare, where the tearing-mode instability (induced by footpoint shearing) near the Y-type reconnection point triggers a flare, accelerating particles in a downward direction and producing shock waves and plasmoid ejection in an upward direction. Hirayama (1974) explains the preflare process as a rising prominence above a neutral line (between oppositely directed open magnetic field lines), which carries an electric current parallel to the neutral line and induces a magnetic collapse on both sides of the current sheet after eruption of the prominence. The magnetic collapse is accompanied by lateral inflow of plasma into the opposite sides of the current sheets. The X-type reconnection region is assumed to be the location of major magnetic energy dissipation, which heats the local coronal plasma and accelerates non-thermal particles. These two processes produce thermal conduction fronts and precipitating particles which both heat the chromospheric footpoints of the newly reconnected field lines. As a result of this impulsive heating, chromospheric plasma evaporates (or ablates) and fills the newly reconnected field lines with overdense heated plasma, which produces soft X-ray-emitting flare loops with temperatures of $T_e \approx 10 - 40$ MK

and densities of $n_e \approx 10^{10} - 10^{12} \text{ cm}^{-3}$. Once the flare loops cool down by thermal conduction and radiative loss, they also become detectable in EUV ($T_e \approx 1 - 2 \text{ MK}$) and $\text{H}\alpha$ ($T_e \approx 10^4 - 10^5 \text{ K}$). Kopp & Pneuman (1976) refined this scenario further and predicted a continuous rise of the Y-type reconnection point, due to the rising prominence. As a consequence, the newly reconnected field lines beneath the X or Y-type reconnection point have an increasingly larger height and wider footpoint separation. Tsuneta (1996a; 1997) and Shibata (1995) elaborated on the temperature structure, upward-ejected plasmoids, slow shocks, and fast shocks in the outflow region of the X-type reconnection geometry (Fig. 14). The heated plasma in the reconnection outflow produces hot ridges ($T \approx 15 - 20 \text{ MK}$) along the separatrices with the slow shocks, sandwiching the denser soft X-ray flare loops that occupy the newly reconnected relaxing field lines, which are filled with chromospheric evaporated plasma. The fast shocks in the reconnection outflows collide with the previously reconnected field lines and may produce hot thermal (as well as nonthermal) hard X-ray sources above the flare looptops. Numerical hydrodynamic simulations of this model reproduce heat conduction fronts and slow-mode shocks (Yokoyama & Shibata 1997) and chromospheric evaporation (Magara et al. 1996; Yokoyama & Shibata 1998; 2001)

This model is essentially a 2D model that describes the evolution in a vertical plane, while evolution along the third dimension (in the direction of the neutral line) can be independently repeated for multiple flare loops (where footpoints extend to a double ribbon) or can be stopped (in the case of a single-loop flare). It is likely that the extension in the third dimension is not continuous (in the form of a giant 2D Sweet–Parker current sheet), but rather highly fragmented into temporary magnetic islands (due to tearing-mode and coalescence instabilities, see § 2.5 and 2.6). Numerical simulations of enhanced resistivity in the current sheet enables the fast reconnection regime (Magara & Shibata 1999) that is required to explain the observed fast (subsecond) time structures. This model fits a lot of the observational features in hard X-rays, soft X-rays, $\text{H}\alpha$, and radio wavelengths, provides a physical mechanism to explain self-consistently the processes of filament eruption, magnetic reconnection, and coronal mass ejection, but does not specify what drives the initial magnetic system to become unstable. This model fits single-loop and double-ribbon arcade geometries, but is not appropriate for quadrupolar flare loop interactions and 3D nullpoint topologies.

3.2 The Emerging Flux Model

The most decisive criterion to judge the relevance of a particular flare model is the driver mechanism that dictates the magnetic evolution, the loss of stability, and subsequent magnetic reconnection process. While the driver is a rising filament/prominence in the Kopp & Pneuman (1976) model, the process of flux emergence has been considered as a driver in the model of Heyvaerts et al. (1977). The model of Heyvaerts et al. (1977) consists of three phases: (1) a preflare heating phase where a new magnetic flux emerges beneath the flare filament and continuously reconnects and heats the current sheet between the old and new flux; (2) the impulsive phase starts when the heated current sheet loses equilibrium at a critical height and turbulent electrical resistivity causes the current sheet rapidly to expand, accelerating particles and triggering chromospheric evaporation; and (3) the main phase where the current sheet reaches a

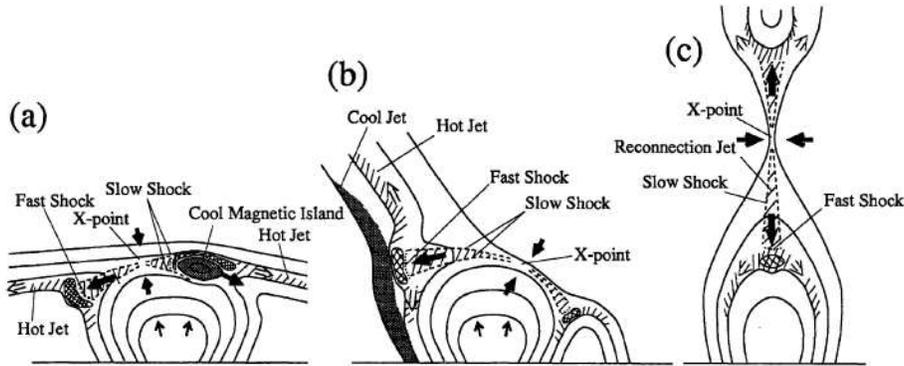


Figure 15: X-type reconnection scenarios for three different orientations of the external magnetic field: horizontal (a), oblique (b), and vertical (c). The two versions (a) and (b) represent geometries used in the emerging flux model of Heyvaerts et al. (1977), while version (c) corresponds to the Kopp & Pneuman (1976) model (Yokoyama & Shibata 1996).

new steady state with marginal reconnection [Fig. 15(a) and (b)]. A requirement of this model is the pre-existence of a stable current sheet (with very low resistivity) for periods of the order of a day or more. However, numerical simulations indicate that the current sheets reconnect almost as quickly as they are formed (Forbes & Priest 1984; Shibata et al. 1990). It is therefore believed that the Heyvaerts model can only apply to small flares (Priest & Forbes 2000).

The geometry of the pre-existing magnetic field is assumed to have a horizontal or oblique angle. A consequence of this geometry is the expulsion of two oppositely directed plasma jets during the impulsive phase [for a horizontal orientation see Fig. 15(a)], or a single jet in the upward direction [for an oblique orientation see Fig. 15(b)]. This model was further elaborated in terms of reconnection outflow characteristics by Shibata et al. (1996c), inspired by the numerous plasma jets that have been observed with Yohkoh/SXT (e.g., Shibata et al. 1992a, 1994a, 1996a, b). The initial driver in Shibata's emerging flux model is the nonlinear evolution of the magnetic buoyancy (Parker) instability simulated by Shibata et al. (1989b). This instability was applied to the reconnection between the emerging flux and the overlying coronal field, leading to formation and ejection of magnetic islands or plasmoids (Shibata et al. 1992b). Further numerical hydrodynamic simulations succeeded in modeling coronal X-ray jets and $H\alpha$ surges (Yokoyama & Shibata 1995, 1996). The locations of the slow-mode and fast-mode shocks of reconnection outflows are indicated in Fig. 15(c). Shibata et al. (1994a, 1996a, b, c) distinguishes between hot and cool jet structures, where the hot jets emerge from the reconnection region, while the cooler jets result from chromospheric evaporation into open field lines. Yokoyama & Shibata (1995, 1996) pointed out that the hot jet ejected from the current sheet region is not the reconnection jet itself, but a secondary jet accelerated by the enhanced gas pressure behind the fast shock, which prevents a direct escape of the primary reconnection jet.

There are other variants of reconnection-driven jet models. The model of Priest

et al. (1994) also produces two-sided, soft X-ray jets, but the drivers are converging flows at the footpoint, which is motivated by the observed correlation with magnetic flux cancellation, while the driver in the model of Shibata et al. (1994b, 1996a, b, c) is flux emergence caused by the upward-pushing Parker instability. The model of Karpen et al. (1995, 1998; Figs. 11) can also produce jets, but is driven by shearing motion at the footpoints, which drives magnetic reconnection in a quadrupolar geometry. Thus, the production of plasma jets is a common characteristic of many flare models, which provides a useful diagnostic of the geometric orientation and spatial location of the involved magnetic reconnection regions (Fig. 15).

3.3 The Equilibrium Loss Model

The driver in the Kopp–Pneuman flare model (§ 3.1) is a rising filament, but the magnetic pre-evolution that leads up to flare instability is not quantified in the various concepts of the CSHKP models. The driver in the Heyvaerts model is emerging flux (§ 3.2), but the onset of flare instability is not quantified in terms of a magnetic field evolution. Another criticism of the latter model is that stable current sheets are unlikely to exist over extended periods of time (as numerical simulations demonstrate), which implies that free magnetic energy has to be stored in the form of field-aligned currents (i.e., force-free fields; Forbes 1996). An evolutionary model that starts with a stable (force-free) magnetic field configuration, then applies converging flows as a continuous driver, and demonstrates how the (force-free) evolution passes a critical point where the system becomes unstable and triggers the rise of a filament, has been developed by Forbes & Priest (1995) in 2D. The initial situation of the magnetic field is shown in Fig. 16(b), where the magnetic field is quantified by the 2D equilibrium of a fluxrope at a stationary height, described by the Grad–Shafranov equation. The two footpoints of the field lines that envelop the fluxrope are then driven closer together, while the system evolves through a series of equilibrium solutions. The height h of the fluxrope as a function of the separation half-distance λ is shown in Fig. 16(a), which monotonically decreases while the source separation is made smaller from $\lambda = 4 \mapsto 1$. Once the source separation passes the critical point at $\lambda = 1$, the fluxrope enters a loss of equilibrium and jumps in height (from $h = 1$ to $h \approx 5$), while forming a current sheet beneath [Fig. 16(d)]. In *ideal* MHD, the rising fluxrope would stop at a higher equilibrium position, because the tension force associated with the current sheet is always strong enough to prevent the fluxrope from escaping (Lin & Forbes 2000). If there is some resistivity, magnetic reconnection is enabled, and even a fairly small reconnection rate is sufficient to allow the fluxrope to escape (Lin & Forbes 2000). Magnetic reconnection in the current sheet releases most ($\approx 95\%$) of the magnetic energy that has been built up from the initial force-free configuration by the converging motion of the footpoints before the loss of equilibrium. This model is formulated fully analytically and yields reasonable amounts of released energies, suitable to explain flares and CMEs. Although this analytical model is restricted to 2D (with a fluxrope that is not anchored at both ends), it demonstrates quantitatively how a loss of magnetic equilibrium leads to a rapid energy release, which probably also takes place in more complex 3D configurations. The question is whether the driver in terms of converging flows is realistic, because typically observed photospheric flows are in the order of ≈ 1

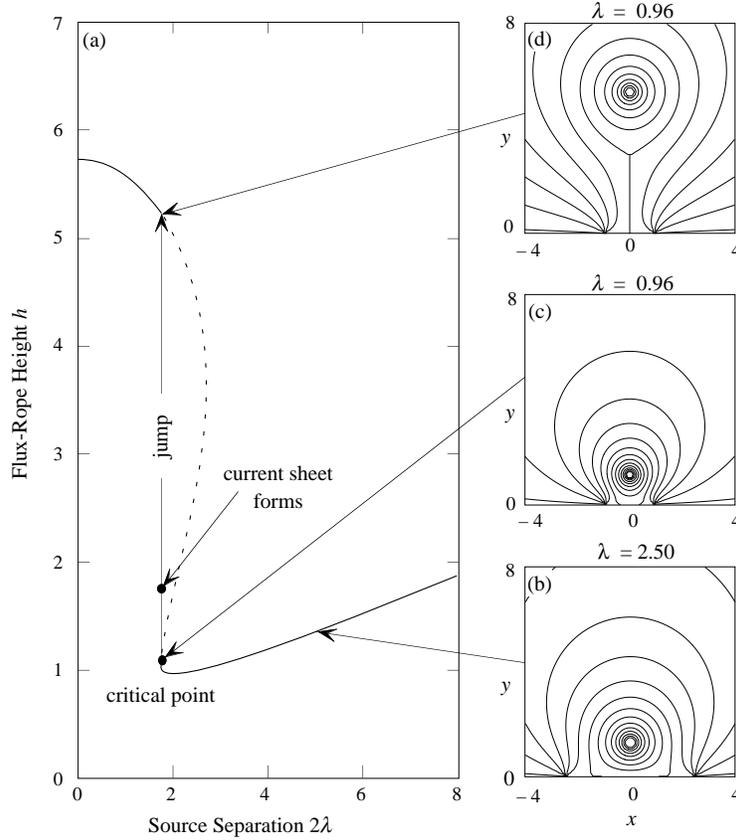


Figure 16: Flare dynamics in the model of Forbes & Priest (1995), inferred from the ideal MHD evolution of a 2D arcade containing an unshielded fluxrope (a)–(c). The fluxrope arcade jumps upwards when the two photospheric field sources are pushed too close to one another. (d) The vertical current sheet is subject to magnetic reconnection if enhanced or anomalous resistivity occurs (Forbes & Priest 1995).

km s^{-1} , which could be too slow or may be randomly oriented. Also, shear flows with subsequent tearing instability have been found to be important drivers of flares, which would require a generalization of Forbes’s model to 3D. Numerical 3D simulations of a similar dipolar configuration driven by converging flows have been performed by Birn et al. (2000).

The analytical model of Forbes & Priest (1995) predicts a specific height evolution of the fluxrope $h(t)$, which grows initially as $h(t) \propto t^{5/2}$, or $v(t) \propto t^{3/2}$, and reaches an asymptotic constant speed of order $v_{term} \approx 1500 \text{ km s}^{-1}$. The solutions of the height $h(t)$, velocity $v(t)$, dissipated energy $dW(t)/dt$, electric field $E_0(t)$, Alfvén speed $v_A(t)$, and reconnection speed $v_R(t)$ are shown in Fig. 17, calculated for a model with initial half-separation $\lambda_0 = 50 \text{ Mm}$, fluxrope length $L = 100 \text{ Mm}$, fluxrope

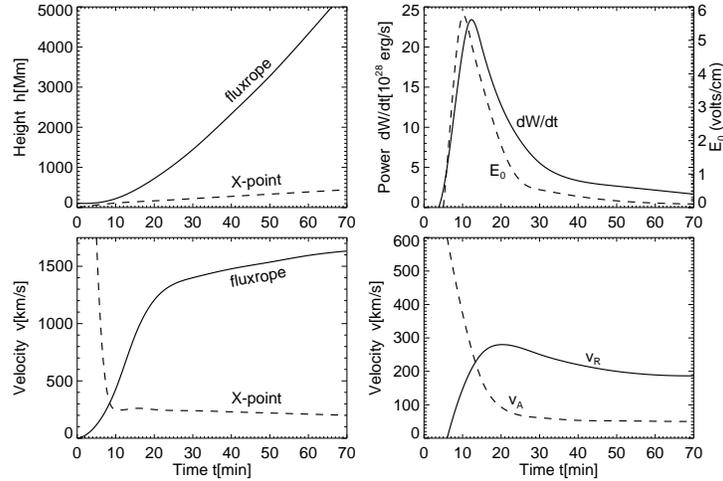


Figure 17: Evolution of height $h(t)$ (top left) and velocities $v(t)$ (bottom left) of fluxrope and magnetic reconnection X-point in the equilibrium loss model of Forbes & Priest (1995). The upper limit of the dissipated energy dW/dt in the current sheet and the generated electric field E_0 at the X-point are shown (top right), as well as the reconnection speed $v_R(t)$ and the ambient Alfvén speed (bottom right). See parameters in text (Priest & Forbes 2000).

radius $a_0 = 0.4\lambda_0$, density $n_e = 5 \times 10^{10} \text{ cm}^{-3}$, and magnetic field $B_0 = 100 \text{ G}$. This calculation represents an upper limit, in the case of an unlimited reconnection rate, so that all magnetic energy goes into the kinetic energy of the upward-accelerated fluxrope. The upward motion of the unstable fluxrope with associated reconnection also predicts a *shrinkage* of flare loops, characterized by the height ratio of the cusp at the beginning of reconnection to the height (of the relaxed dipolar field line) in the postflare phase, which was found to be 20% and 33% in two flares (Forbes & Acton 1996).

3.4 2D Quadrupolar Flare Model

Among 2D models, which we classified in Fig. 1 into *bipolar* (e.g., Kopp–Pneuman § 3.1; Priest–Forbes § 3.3), *tripolar* (e.g., Heyvaerts et al. § 3.2), and *quadrupolar* ones, we describe here a representative of the latter category (namely, the *quadrupolar photospheric source model*), which was first proposed by Uchida (1980), and later developed further by Uchida et al. (1998a, b) and Hirose et al. (2001). The initial configuration consists of two parallel arcades (as shown in Fig. 1, top right), which altogether requires three parallel neutral lines. Formation of such double arcades with current sheets inbetween have been inferred from neighboring active regions (Sakurai & Uchida 1977) and from polar crown filaments with arcades (Uchida et al. 1996). As in the Forbes–Priest (1995) model, the principal driver is a converging flow pattern that pushes the two arcades together. The X-point above the middle neutral line supports a dark filament. The two flanking arcades that suspend the filament might be partially ob-

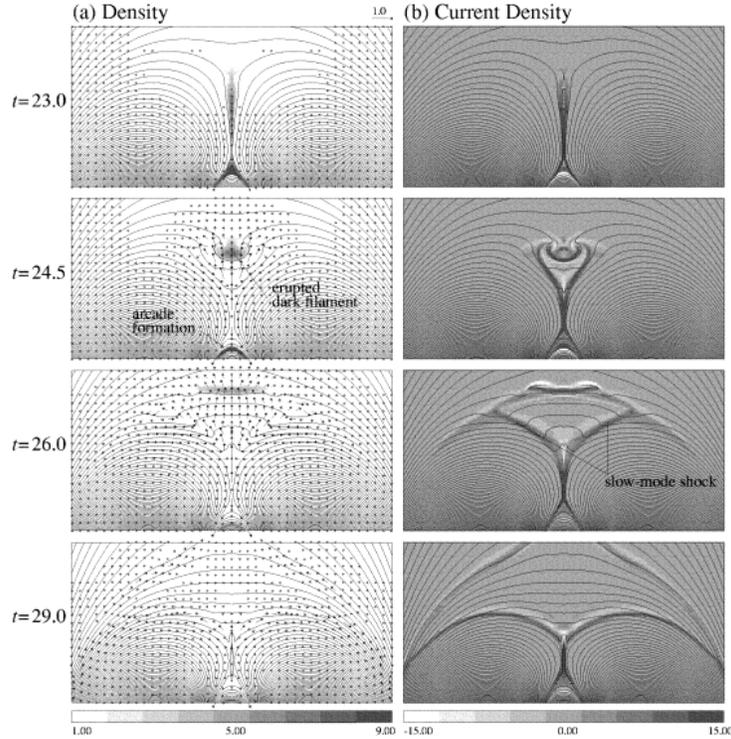


Figure 18: Dynamical evolution of the plasma density (left) and current density (right) in the quadrupolar magnetic reconnection model of Uchida (1980), simulated with a 2.5D MHD code. Note that the erupted dark filament transforms into a CME with two slow-mode shocks at both sides (Hirose et al. 2001).

servable as the “barbs” of the filament channel. While the two arcades push together, the dark filament transforms into a thin vertical current sheet, which at some point becomes unstable due to tearing-mode instability, triggering anomalous resistivity and fast reconnection. The dark filament with helical field lines is accelerated upward in the expanding field structure with a rounded shape and transforms into a CME. The reconnected field lines below the X-point shrink and form the postflare arcade. Numerical simulations of the driven reconnection in quadrupolar arcades or between interacting loops have been performed by Rickard & Priest (1994), Sakai et al. (1995), Sakai & De Jager (1996), and for the quadrupolar model of Uchida specifically by Hirose et al. (2001), see Fig. 18 for an example.

There are several motivations for this 2D quadrupolar model. The Kopp–Pneuman (1976) model cannot explain the magnetic field in the dark filaments seen from the side, because the direction of the magnetic field at the lower side is opposite to what is expected from the polarity of photospheric sources (Leroy et al. 1984). A proposed solution was to introduce a fluxrope (Kuperus & Raadu 1974), which reverses the polarity at the lower dip of the prominence. Therefore, the involvement of a fluxrope-

like filament is an essential element in a flare model, which naturally evolves in the Forbes–Priest (1995) and quadrupolar model of Uchida (1980). Further it is argued that the 2D quadrupolar model of Uchida (1980) solves the energy problem to open up the field and to accelerate the filament to escape velocity (Uchida et al. 1996; 1998a, b). In Sturrock’s (1966) model, the erupting filament is required to have an energy equal to or greater than that of the flare itself, since the open-field configuration (which is the final state in Sturrock’s 1966 model) is conjectured to be the state of maximum energy (Aly 1984; Sturrock 1991). In the numerical simulation of Hirose et al. (2001) it is actually found that the major part of the stored magnetic energy is converted into kinetic energy carried away by the CME (containing the erupted dark filament), while only a minor part is left for heating of the associated arcade flare.

3.5 The Magnetic Breakout Model

A further development of the 2D quadrupolar model of Uchida (1980) is the so-called *magnetic breakout* model of Antiochos et al. (1999b) and Aulanier et al. (2000b), which involves the same initial quadrupolar magnetic configuration, but undergoes an asymmetric evolution with the opening up of the magnetic field on one side. The asymmetric evolution is driven by footpoint shearing of one side arcade, where reconnection between the sheared arcade and the neighboring (unsheared) flux systems triggers an eruption. In this magnetic breakout model, reconnection removes the unsheared field above the low-lying, sheared core flux near the neutral line, which then allows the field above the core flux to open up (Antiochos et al. 1999b). Thus, this model addresses the same energy problem as Uchida’s model: How very low-lying magnetic field lines can open up (down to the photospheric level) into an open-field configuration during the eruption. Moreover, the eruption is solely driven by free magnetic energy stored in a closed, sheared arcade. It circumvents the Aly–Sturrock energy limit by allowing external, disconnected magnetic flux from a neighboring sheared arcade (which is not accounted for in the “closed-topology” model of Aly and Sturrock) to assist in the opening-up process. Thus, a key point of the magnetic breakout model is the interaction of a multi-flux system (e.g., in a quadrupolar double arcade). It has the same initial configuration as Uchida’s model, but is driven by asymmetric shear.

The magnetic topology of the magnetic breakout model has been applied to the Bastille-Day flare by Aulanier et al. (2000b), who found a more complex 3D topology than the 2D quadrupolar model of Antiochos et al. (1999b). Aulanier et al. (2000b) actually identified a magnetic nullpoint in the corona above the flare arcade which was connected with a “spine” field line to a photospheric location where the flare brightens up first. The other side of the coronal nullpoint sits on a dome-like “fan” surface, which encloses the main flare arcade. This topology can be considered as one of the many possible 3D reconnection scenarios in which the magnetic breakout model can be realized. Aulanier et al. (2000b) suggest a more general definition: “A *magnetic breakout* is the opening of initially low-lying sheared fields, triggered by reconnection at a nullpoint that is located high in the corona and that defines a separatrix enclosing the sheared fields”. This represents a generalization of the 2.5D version (Fig. 19) into 3D reconnection topologies. Obviously, observations are crucial to pinning down the involved magnetic configurations, which are now becoming available increasingly

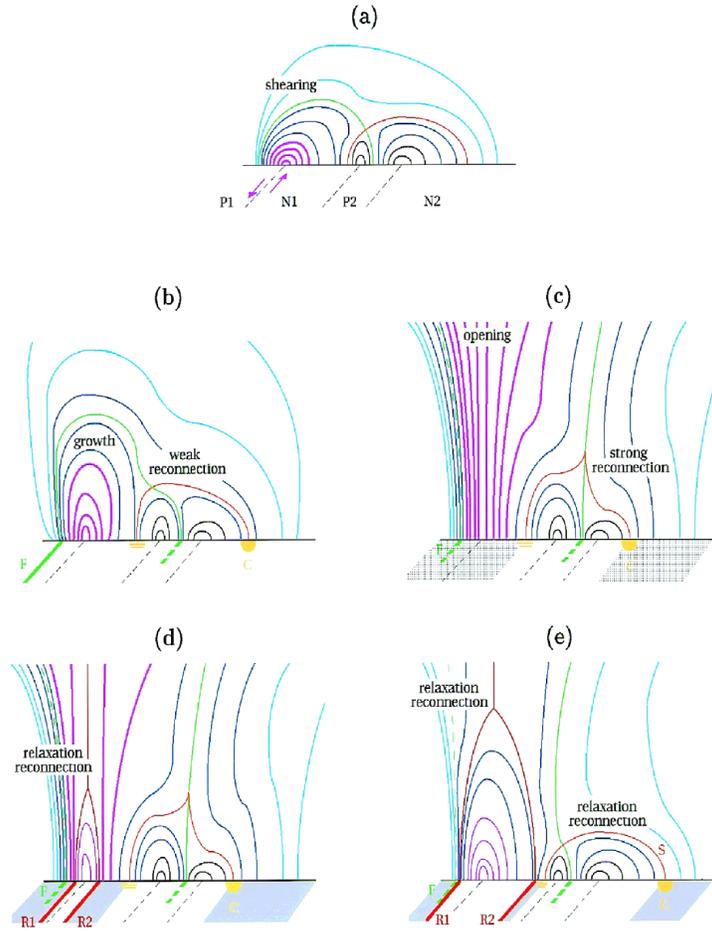


Figure 19: 2.5D version of the magnetic breakout model: (a) The initial quadrupolar potential state, with shear applied on both sides of the neutral line P1/N1; (b) the shear triggers some weak reconnection at the coronal nullpoint; (c) fast reconnection at the nullpoint leads to opening up of the field; (d) relaxation reconnection in the opening field lines, forming footpoint ribbons and flare loops; (e) ongoing formation of postflare loops and reconnection at the null. The Kopp–Pneuman (1976) model is a special case in which the magnetic breakout does not occur [eliminating phase (c)] (Aulanier et al. 2000b).

clearly from TRACE postflare loop observations. The reconstruction of the preflare configuration, which is necessary to track down the reconnection process, however, is hampered by the unavailability of high-resolution observations at the much higher flare temperatures ($T \approx 10 - 40$ MK) during the impulsive flare phase. Nevertheless, the 3D reconnection topology could be reconstructed for some cases, clearly showing evidence for 3D reconnection involving a separatrix dome (Fig. 20; Fletcher et al. 2001).

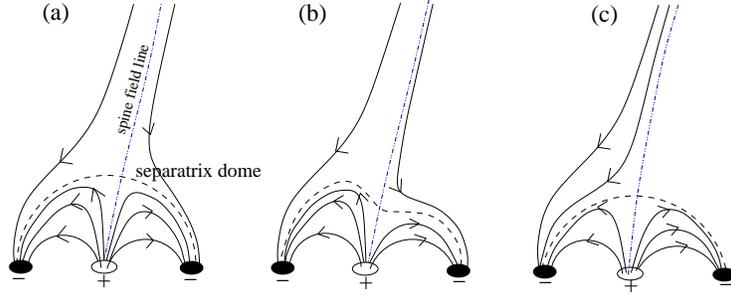


Figure 20: Magnetic field topology inferred in the 1993-May-3, 23:05 UT, flare by Fletcher et al. (2001). The sequence shows a 2D representation of the reconnection process via a separator dome (Fletcher et al. 2001)

3.6 3D Quadrupolar Flare Models

Some flares clearly show an interaction between two flare loops (or closed-field systems), which can most simply be interpreted as the outcome of a quadrupolar reconnection process. The magnetic configuration corresponds to a 3D reconnection case (Fig. 1, bottom right) that can be represented by a single, common, neutral line for the two interacting flare loops, which is different from the 2D case in Uchida’s model, which has three neutral lines (Fig. 1, top right). The two observed flare loops represent, of course, the postreconnection situation, but the prereconnection topology can be straightforwardly reconstructed by switching the polarities according to the scheme shown in Fig. 4. Thus, magnetic geometry is fully constrained for this type of 3D reconnection and can be reconstructed from the observed postflare loops. A number of flares was found to fit quadrupolar geometry (Hanaoka 1996, 1997; Nishio et al. 1997; Aschwanden et al. 1999c).

This 3D quadrupolar reconnection model only describes the interaction between two closed loops, which was found to match closely the observed topology of some flares. Obviously the model does not include any open field lines and thus cannot explain the simultaneous rise of a filament and expulsion of a CME, which may occur in a detached magnetic field domain above the interacting loop system. However, a key aspect of this model is that it relates the currents of the prereconnection to the postreconnection field lines in a highly sheared configuration. It also quantifies the efficiency of the reconnection rate as a function of shear angles. Interestingly, most of the relevant observations indicate shear angles between the reconnecting field lines which range from near-parallel to near-perpendicular (Aschwanden et al. 1999b), rather than being anti-parallel as expected in the standard Kopp–Pneuman flare scenario. This observational fact, however, does not violate the Petschek reconnection model, since the reconnection rate can still operate at small angles θ , with efficiency scaling as $\propto [\sin(\theta/2)]^{1/2}$ (Soward 1982).

Similar 3D quadrupolar reconnection models are also described in Somov et al. (1998), Somov (2000, § 16.5.2), and Kusano (2002), where loops that are sheared along the central neutral line of a flare arcade reconnect with the overlying less sheared

Table 1: Classification of flare/CME Models according to the driver mechanisms and dimensionality of magnetic reconnection geometry.

<i>Driver mechanism:</i>	<i>2D models:</i>	<i>3D models:</i>
Rising filament or prominence $v_z (h \gg h_{Ph})$	X-type reconnection (Hirayama 1974) (Kopp & Pneuman 1976)	
Photospheric flux emergence $v_z (h = h_{Ph})$	Emerging flux model (Heyvaerts et al. 1977)	Quadrupolar flux transfer (Melrose 1995)
Photospheric converging flows $+v_x, -v_x \perp NL$	Equilibrium loss model (Forbes & Priest 1995) Quadrupolar double-arcade (Uchida 1980)	
Photospheric shear motion $+v_y, -v_y \parallel NL$		Tearing-mode instability (Sturrock 1966) Magnetic breakout model (Antiochos et al. 1999b) Sheared loops inside arcade (Somov et al. 1998)

arcade. Alternatively, 3D quadrupolar reconnection in large flares could also be driven by emerging current loops (Mandrini et al. 1993). Such 3D quadrupolar configurations are particularly suitable to explain double-ribbon flares, but it could not yet be decided observationally whether the primary driver of this type of reconnection is a rising filament (Kopp–Pneuman 1976) or the shear along the neutral line (Sturrock 1966; Somov et al. 1998).

3.7 Unification of Flare Models

In Table 1 we sort the previously discussed flare/CME models according to the driver mechanisms and dimensionality of magnetic reconnection geometry. There are essentially two locations of drivers: (1) above the flare site (in the form of a rising filament, prominence, or plasmoid); and (2) below the flare site (in the form of photospheric emergence, convergence flows, or shear flows). The three photospheric drivers can essentially be discriminated by their directions: (1) flux emergence corresponds to a flow in the vertical direction (v_z); (2) convergence flows are counter-directed perpendicular to the neutral line ($+v_x, -v_x \perp NL$); and (3) shear flows are counter-directed parallel to the neutral line ($+v_y, -v_y \parallel NL$). The classification in Table 1 also shows that 2D models can only be constructed when the driver force is in the 2D plane of the loops (e.g., converging flows in the the x-direction or emergence in the z-direction), while a driver force perpendicular to the 2D loop plane (e.g., shear in the y-direction) requires 3D models. Table 1 is by no means a complete list of flare/CME models; in principle there could be for every type of driver at least one 3D model, and moreover multiple models could be conceived for any combination of multiple loops (open or closed, and arcades). The 2D models are probably all idealized approximations, but more accurate

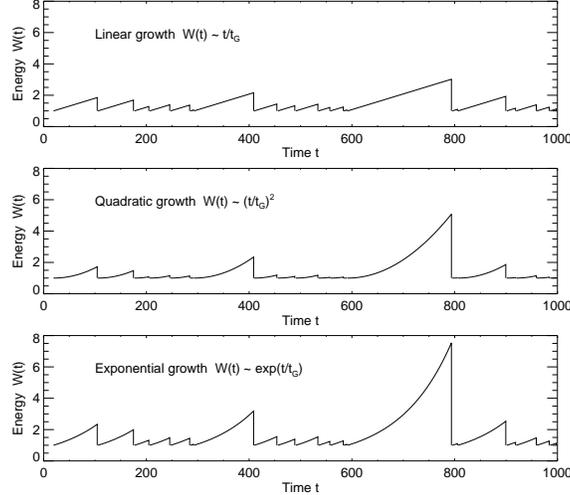


Figure 21: Time evolution of energy build-up interrupted by random times: linear (top), quadratic (middle), and exponential growth (bottom). Note that linear growth produces exponential distributions of saturation energies, while nonlinear growth produces powerlaw-like distributions, see Fig. 22.

future observations might require generalizations of each one to a 3D version. Nevertheless, the classification in Table 1 indicates that at least models with the same driver mechanism could be unified into a 3D model. In future we might even distill a single unified flare/CME model by combining all the important drivers.

4 Flare Statistics and Frequency Distributions

4.1 Theory of Frequency Distributions

The statistics of energies in the form of frequency distributions became an important tool for studying nonlinear dissipative events. A *frequency distribution* is a function that describes the occurrence rate of events as a function of their size, usually plotted as a histogram of the logarithmic number $\log(N)$ versus the logarithmic size $\log(S)$, where the size S could be a length scale l , an area A , a spatial volume V , or a volume-integrated energy E . The two most common functional forms of such frequency distributions are the exponential and the powerlaw function. We will demonstrate that an exponential distribution results from linear or incoherent processes, while a powerlaw distribution results from nonlinear or coherent processes. The latter function has therefore been established as the hallmark of nonlinear dissipative systems. A powerlaw function has no characteristic spatial scale, in contrast to an exponential function, which has an e-folding scale length. The size range over which a powerlaw function applies is called the *inertial range*. We will see that this inertial range extends over more than 8 orders of magnitude in energy for solar flares and nanoflares. Nonlinear dissipative systems, which are constantly driven by some random energy input evolve

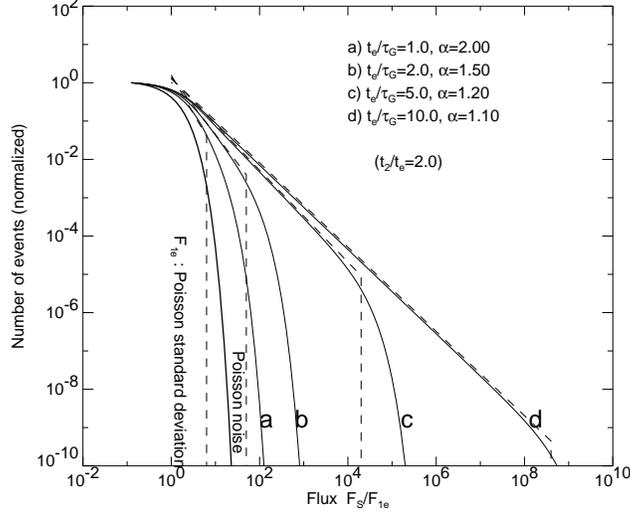


Figure 22: Theoretical frequency distributions $N(W_S)$ for 4 different ratios of the mean saturation times t_{S_e} to the growth time τ_G : (a) $t_{S_e}/\tau_G = 1.0$ (linear case), (b) $t_{S_e}/\tau_G = 2.0$, (c) $t_{S_e}/\tau_G = 5.0$, (d) $t_{S_e}/\tau_G = 10.0$ (nonlinear cases). Note that the frequency distribution evolves from an exponential to a powerlaw distribution the higher the nonlinear saturation time is (Aschwanden et al. 1998a).

into a critical state that is maintained as a powerlaw distribution. The fluctuations of the input does not change the powerlaw slope of the dissipated energy events that make up the output, but are just adjusted by a scale-invariant number factor and by a slow shift of the upper cutoff of the distribution. The maintenance of an invariant powerlaw slope is also called *self-organized criticality* and is a property that is inherent to nonlinear dissipative systems. The principle of self-organized criticality has been first applied to solar flare phenomena by Lu & Hamilton (1991).

We can build a simple mathematical model of a nonlinear dissipative system just by two rules: (1) energy is dissipated in random time intervals, and (2) energy builds up with a nonlinear power as a function of time. So, let us consider linear and nonlinear time evolutions (e.g., a quadratic and an exponential function) for the build-up of energy $W(t)$, see Fig. 21,

$$W(t) = W_1 \times \begin{cases} (t/\tau_G) & \text{linear} \\ (t/\tau_G)^2 & \text{quadratic} \\ \exp(t/\tau_G) & \text{exponential} \end{cases}, \quad (4.1)$$

where τ_G represents an exponential growth time. If we let each process grow to randomly distributed saturation times $t = t_S$, we will obtain a distribution of saturation energies $W_S = W(t = t_S)$. The distribution of random times t_S obeys Poisson statistics and can be approximated with an exponential distribution with an e-folding time constant t_{S_e} (in the tail $t_S \gtrsim t_{S_e}$),

$$N(t_S)dt_S = N_0 \exp\left(-\frac{t_S}{t_{S_e}}\right)dt_S \quad (4.2)$$

where N_0 is a normalization constant. With these two definitions (Eqs. 4.1–4.2) we can derive the frequency distribution of dissipated energies $N(W_S)$ by substituting the saturation times t_S with the energy variable $W_S(t_S)$ in the distribution of saturation times $N(t_S)$ in Eq. (4.2),

$$N(W_S)dW_S = N[t_S(W_S)] \left| \frac{dt_S}{dW_S} \right| dW_S . \quad (4.3)$$

So we have to invert the energy evolution time profile $W_S(t_S)$ (Eq. 4.1),

$$t_S(W_S) = \tau_G \times \begin{cases} (W_S/W_1) & \text{linear} \\ (W_S/W_1)^{1/2} & \text{quadratic} \\ \ln(W_S/W_1) & \text{exponential} \end{cases} , \quad (4.4)$$

and to calculate the derivatives of the inversions, dt_S/dW_S ,

$$\left(\frac{dt_S}{dW_S} \right) = \tau_G \times \begin{cases} (1/W_1) & \text{linear} \\ (1/2W_1)(W_1/W_S)^{1/2} & \text{quadratic} \\ (1/W_S) & \text{exponential} \end{cases} , \quad (4.5)$$

which then can be plugged into Eq. (4.3) to yield the frequency distributions of energies:

$$N(W_S) \propto \begin{cases} \exp[-(\tau_G/t_{Se})(W_S/W_1)] & \text{linear} \\ \exp[-(\tau_G/t_{Se})(W_S/W_1)^{1/2}] \times W_S^{-1/2} & \text{quadratic} \\ W_S^{-(1+\tau_G/t_{Se})} & \text{exponential} \end{cases} . \quad (4.6)$$

The resulting frequency distribution for linear growth is an exponential distribution, similar to the exponential distribution of saturation times. This is trivial, because the energy W_S is proportional to the saturation time t_S for linear growth. For exponential growth, however, the resulting frequency distribution becomes a powerlaw (Eq. 4.6; Fig. 22) with an index

$$\alpha_W = \left(1 + \frac{\tau_G}{t_{Se}} \right) , \quad (4.7)$$

(Rosner & Vaiana 1978; Aschwanden et al. 1998a). So the powerlaw slope is determined by the ratio of the exponential growth time τ_G of the nonlinear energy evolution and by the e-folding saturation time t_{Se} of the random distribution of saturation times, with the limit of $\alpha_W \geq 1$ for $t_{Se} \gg \tau_G$ (Fig. 23). The linear growth case can also be mimicked by the exponential model for $t_{Se} \ll \tau_G$. We illustrate the relation between the time profiles of the energy evolution $W(t_S)$ and the distribution of saturation energies in Figs. 21 to 23. The theoretical relation (4.7) for the powerlaw slope gives us a diagnostic as to whether the underlying nonlinear dissipative process is incoherent (if $\alpha_W \gg 1$) or coherent with a high amplification factor (if $\alpha_W \gtrsim 1$).

4.2 Frequency Distributions of Flare Parameters

We started out to specify the relation between different physical parameters in flares ($n_e, T_e, \varepsilon_{th}$) as a function of the spatial scale l . Therefore, once we have measured the

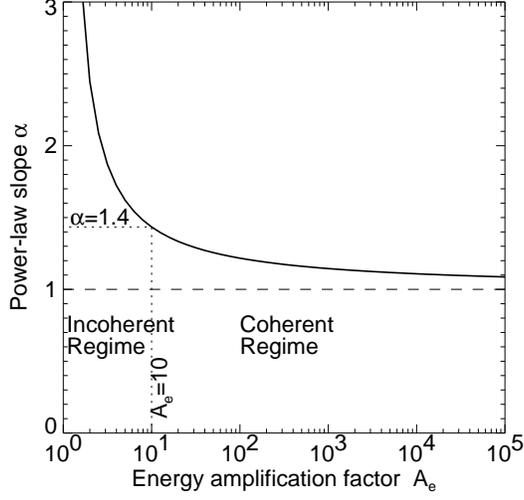


Figure 23: Dependence of the powerlaw index α_W of the frequency distributions $N(W_S) \propto W_S^{-\alpha}$ on the ratio of the saturation time t_{Se} to the exponential growth time τ_G (amplification factor). Note that the distribution becomes the flatter the higher the amplification factor (Aschwanden et al. 1998a).

distribution of spatial scales, which might be characterized by a powerlaw distribution,

$$N(l)dl \propto l^{-\alpha_l} dl \quad 1 \leq \alpha_l \leq 3, \quad (4.8)$$

and know the relation of a parameter $y(l)$ as a function of this independent variable l , for example,

$$y(l) \propto l^D \begin{cases} V(l) \propto l^{D_3} & \text{volume} \\ n_e(l) \propto l^{D_n} & \text{density} \\ T_e(l) \propto l^{D_T} & \text{temperature} \\ \varepsilon_{th}(l) \propto l^{D_E} = l^{D_n+D_T+D_3} & \text{thermal energy} \end{cases} \quad (4.9)$$

we can predict the frequency distributions $N(y)$ of these parameters in the same way as we did in Eqs. (4.3–4.6). We need only to calculate the inversion of $y(l)$ and its derivative, which for powerlaw functions is straightforward,

$$l(y) \propto y^{1/D}, \quad \frac{dl}{dy} \propto y^{1/D-1} \quad (4.10)$$

and yields the desired frequency distributions,

$$N(y) dy = N[l(y)] \frac{dl}{dy} dy = l(y)^{-\alpha_l} \frac{dy}{dl} dy = y^{-[1+(\alpha_l-1)/D]} dy \quad (4.11)$$

So we obtain again powerlaw distributions with the following slopes,

$$N(y)dy \propto y^{-\alpha_y} dy, \quad \alpha_y = \left(1 + \frac{\alpha_l - 1}{D}\right). \quad (4.12)$$

For the RTV model ($n \propto l^3, T \propto l^2$) we predict the following powerlaw slopes α_y of the frequency distributions $N(y)$ of various physical parameters y ,

<u>RTV model</u>	<u>Volume model</u>	
$\alpha_l = 2.5 \pm 0.5$	$\alpha_l = 2.5 \pm 0.5$	
$D_3 = 2.5 \pm 0.5$	$D_3 = 2.5 \pm 0.5$	
$D_n = 3.0$	$D_n = 0.0$	
$D_T = 2.0$	$D_T = 0.0$	
$D_E = D_n + D_t + D_3 = 7.5 \pm 0.5$	$D_E = 2.5 \pm 0.5$	(4.13)
$\alpha_V = 1 + \frac{\alpha_l - 1}{D_3} = 1.67 \pm 0.33$	$\alpha_V = 1.67 \pm 0.30$	
$\alpha_n = 1 + \frac{\alpha_l - 1}{D_n} = 1.50 \pm 0.17$...	
$\alpha_T = 1 + \frac{\alpha_l - 1}{D_T} = 1.75 \pm 0.25$...	
$\alpha_E = 1 + \frac{\alpha_l - 1}{D_E} = 1.21 \pm 0.08$	$\alpha_E = 1.67 \pm 0.33$	

Thus we predict a powerlaw distribution of energies with a slope of $\alpha_E = 1.21 \pm 0.08$ for the RTV model, where the error bars include only the propagation errors of the fractal dimension $D_3 = 2.5 \pm 0.5$ and the length distribution $\alpha_l = 2.5 \pm 0.5$. Let us define also an alternative model, the so-called *Volume model*, where the thermal energy is directly proportional to the volume, without any dependence on the density and temperature. Such a model may be representative if statistics is done in a subset of the data, say in a narrowband filter with a small temperature range and with a flux threshold (which restricts the range of detected densities). In such a restricted data subset we would predict a slope of $\alpha_E = 1.67 \pm 0.33$. Real data with some temperature range and some moderate flux range, of course, could produce any value between these two cases, $1.21 \leq \alpha_E \leq 1.67$. Powerlaw indices in the range of 1.1 – 1.64 have also been derived from magnetic braiding and twisting models (Zirker & Cleveland 1993a,b).

4.3 Energy Budget of Flare-like Events

The frequency distributions specify the number of events $N(W_i)$ in an energy bin $[W_i, W_{i+1}]$. If we want to know the total energy budget over some range that is bracketed by the minimum W_1 and maximum W_2 , we have to integrate the energy powerlaw distribution,

$$\begin{aligned}
 W_{tot} &= \int_{W_1}^{W_2} N(W) W dW = \int_{W_1}^{W_2} N_1 W_1 \left(\frac{W}{W_1} \right)^{1-\alpha_W} dW = \\
 &= N_1 W_1^2 \begin{cases} \frac{1}{(2-\alpha_W)} \left[\left(\frac{W_2}{W_1} \right)^{2-\alpha_W} - 1 \right] & \text{if } \alpha_W \neq 2, \\ [\ln(W_2/W_1)] & \text{if } \alpha_W = 2, \end{cases} \quad (4.14)
 \end{aligned}$$

From this expression we see immediately that the integral is dominated by the upper limit W_2 for flat powerlaw indices $\alpha_W < 2$, and by the lower limit W_1 in the case of steep powerlaw indices $\alpha_W > 2$. This implies that nanoflares are important for coronal heating if the frequency distribution of their energy has a slope steeper than 2, a necessary condition that was pointed out by Hudson (1991a).

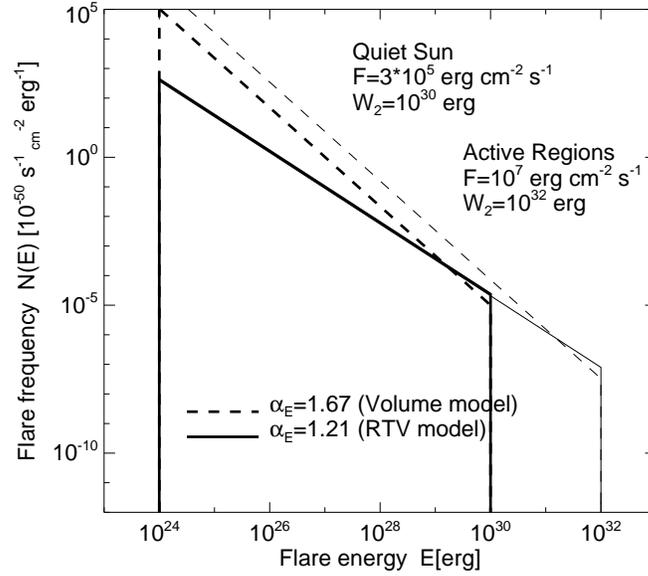


Figure 24: Predicted frequency distributions that fulfill the coronal heating requirement, for quiet Sun regions ($F = 3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$) and active regions ($F = 10^7 \text{ erg cm}^{-2} \text{ s}^{-1}$), according to the RTV model (with a powerlaw slope of $\alpha_E = 1.21$ and the volume model ($\alpha_E = 1.67$), see text.

Let us calculate some practical cases that are relevant for the coronal heating problem. The energy requirements for coronal heating are given in Table 3. For the quiet Sun we need a heating rate of $F_{QS} \approx 3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$, and for active regions we need $F_{AR} \approx 10^7 \text{ erg cm}^{-2} \text{ s}^{-1}$. We know that flares occur only in active regions and could have a maximum energy up to $W_{2,AR} \approx 10^{32} \text{ erg}$ (Fig. 24), while the largest microflares occurring in the quiet Sun are the so-called X-ray bright points, which have energies up to $W_{2,QS} \approx 10^{30} \text{ erg}$. For the lower energy limit we take the smallest nanoflares that have been observed so far, which have energies of $W_{1,QS} \approx W_{1,AR} \approx 10^{24} \text{ erg}$. With these values we obtain with (Eq. 4.14) the following rates,

$$N_1 = \frac{W_{tot}}{W_1^2} \frac{1}{(2 - \alpha_W)} \frac{1}{\left(\frac{W_2}{W_1}\right)^{2 - \alpha_W} - 1} = \begin{cases} 4.3 \times 10^0 & \text{QS : (RTV model)} \\ 1.0 \times 10^3 & \text{QS : (Volume model)} \\ 1.4 \times 10^2 & \text{AR : (RTV model)} \\ 3.5 \times 10^4 & \text{AR : (Volume model)} \end{cases} \quad (4.15)$$

We visualize these four distributions in Fig. 24. Because the powerlaw slopes are all below the critical value of $\alpha_E = 2$, the total energy is dominated by the upper energy cutoff W_2 , so that the lower energy cutoff has almost no effect, since $W_1 \ll W_2$. Consequently, the integrated total energy is also not very sensitive to the exact value of the powerlaw slope. In essence, the occurrence rate at the high energy cutoff $N(W_2)$

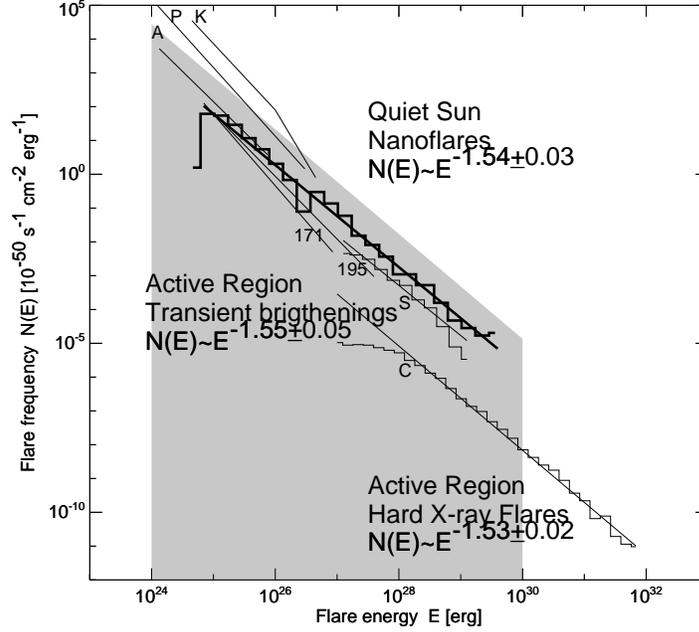


Figure 25: Compilation of frequency distributions of thermal energies from nanoflare statistics in the quiet Sun, active region transient brightenings, and hard X-ray flares. The labels indicate the following studies: K=Krucker & Benz (1998), Benz & Krucker (2002); P=Parnell & Jupp (2000) (corrected for an erroneous factor of 100 in the original paper); A=Aschwanden et al. (2000c); S=Shimizu (1995); C=Crosby et al. (1993), and 171, 195=Aschwanden & Parnell (2002). The overall slope of the synthesized nanoflare distribution, $N(E) \propto E^{-1.54 \pm 0.03}$, is similar to that of transient brightenings and hard X-ray flares. The grey area indicates the coronal heating requirement of $F = 3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$ for quiet Sun regions. Note that the observed distribution of nanoflares falls short of the theoretical requirement by a factor of 10 in occurrence rate or a factor of ≈ 3 in energy.

is the most decisive parameter determining the energy budget.

4.4 Measurements of Frequency Distributions

A compilation of some recent frequency distributions of nanoflare energies is shown in Fig. 25, which all have a powerlaw slope of approximately $\alpha_E \approx 1.55$. In the same figure we also show the energy distribution of the coronal heating requirement for the quiet Sun (grey area in Fig. 25), for the same powerlaw slope and the parameters: $F = W_{tot} = 3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$, $W_1 = 10^{24} \text{ erg}$, $W_2 = 10^{30} \text{ erg}$, $\alpha_E = 1.55$. We see that the observed nanoflare distribution lies about a factor of 10 below the theoretical occurrence rate, or shifted to the left by about a factor of ≈ 3 in energy. Now we have to be aware that the thermal energy is calculated based on the radiation we detect in EUV and soft X-rays, so it characterizes only the energy equivalent to the radiative losses,

Table 2: Frequency distributions of small-scale phenomena observed in quiet Sun regions.

Phenomenon	Number of events N	Powerlaw slope α_E	Energy range E_1, E_2 10^{24} [erg]	Total flux F [erg cm $^{-2}$ s $^{-1}$]
EUV transients, EIT, 171+195 ¹	233	2.45 ± 0.15	10 – 300	0.7×10^5
EUV transients, EIT 195 ²	228	1.35 ± 0.20	1 – 100	...
EUV transients, EIT 195 ³	277	1.45 ± 0.20	10 – 100	...
Nanoflares, TRACE, 171+195 ⁴	5131	2.48 ± 0.11	0.3 – 60	0.2×10^5
Nanoflares, TRACE+SXT ⁵	281	1.53 ± 0.02	10 – 10^6	0.5×10^5
Blinkers, CDS, O V ⁶	790	1.34 ± 0.08	0.01 – 0.3	...
Explosive ev., SUMER C III ⁷	3403	2.8 ± 0.1	0.05 – 2	0.45×10^5
Explosive ev., SUMER Ne IV ⁷	2505	2.8 ± 0.1	0.6 – 10	0.16×10^5
Explosive ev., SUMER O VI ⁷	5531	3.3 ± 0.4	0.1 – 2	0.79×10^5
Explosive ev., SUMER Ne VIII ⁷	2907	2.8 ± 0.5	0.06 – 1	0.03×10^5

¹ Krucker & Benz (1998); ² Berghmans et al. (1998); ³ Berghmans & Clette (1999); ⁴ Parnell & Jupp (2000) [corrected for a factor of 100 in original paper]; ⁵ Aschwanden et al. (2000b); ⁶ Brkovic et al. (2001); ⁷ Winebarger et al. (2002).

while it does not include energy losses due to conduction to the chromosphere or the solar wind flux. The radiative losses in the quiet Sun alone are indeed about a factor of 3 lower than the total coronal energy losses (i.e., $F = 1 \times 10^5$ erg cm $^{-2}$ s $^{-1}$). So we can conclude that the detected radiation of the EUV and SXT nanoflares roughly corresponds to a third of the total coronal heating requirement in quiet Sun regions, which covers approximately the radiative losses. Because there are many uncertainties involved in the quantification of observed frequency distributions, this result still needs to be corroborated. If this result holds up, it has the important consequence that we have localized the coronal heating sources in the form of detectable nanoflares in EUV and soft X-rays with a sufficient rate, and thus we do not need to invoke invisible energy sources such as heating by Alfvén waves to explain the radiation of the heated plasma, at least not in quiet Sun regions. In coronal holes, the total energy losses are much higher due to the solar wind fluxes, where heating by Alfvén waves is probably required in addition to nanoflare heating.

In Table 2 we compile frequency distributions of small-scale phenomena that have been reported from the quiet Sun and calculate their total energy flux F based on the observed energy ranges $[W_1, W_2]$, powerlaw slopes α_E , and rate $N_1 = N(W_1)$. EUV transients, nanoflares and microflares generally are found in the energy range of $W \approx 10^{24} - 10^{26}$ erg and the integrated flux over the entire frequency distribution lies in the range of $F \approx (0.5 \pm 0.2) 10^5$ erg $^{-1}$ cm $^{-2}$ s $^{-1}$, which makes up about one to two-thirds of the total heating requirement of the quiet corona, roughly covering the radiative losses in the quiet Sun corona. This corroborates our finding in Fig. 25. A similar flux was also measured for explosive events in C III, Ne IV, and O VI (Winebarger et al. 2002), which fits into the picture that explosive events and nanoflares are probably controlled by the same physical process as a magnetic reconnection process in the

transition region, which is manifested with comparable amounts of thermal plasma inside the transition region (as detected in the cooler EUV lines in C III, Ne IV, and O VI) as well as in the lower corona (in the hotter EUV lines of Fe IX/X and Fe XII). Other phenomena such as blinkers carry several orders of magnitude less energy ($\varepsilon_{th} \approx 10^{22} - 3 \times 10^{23}$; Brkovic et al. 2001), and thus seem to be energetically less important for coronal heating.

There are some significant variations in the powerlaw slope of the frequency distributions, ranging from as low as $\alpha_E \approx 1.34$ up to $\alpha_E \lesssim 2.6$ (Table 2). Our theoretical RTV model predicts a slope in the range of $1.21 \leq \alpha_E \leq 1.67$ (Fig. 24), depending on the sampling over a broad or narrow temperature and flux range. There are a number of systematic effects in the data analysis and modeling of the thermal energy that affect the resulting powerlaw slope, such as: (1) event definition and discrimination, (2) sampling completeness, (3) observing cadence and exposure times, (4) pattern recognition algorithm, (5) geometric, density, and thermal energy model, (6) line-of-sight integration, (7) extrapolation in undetected energy ranges, (8) wavelength and filter bias, (9) fitting procedure of frequency distributions, and (10) error estimates of powerlaw slopes. Technical details about these issues are discussed and compared in a number of papers (e.g., Aschwanden & Parnell 2002; Aschwanden & Charbonneau 2002; Benz & Krucker 2002; Parnell 2002a, 2002b; Berghmans 2002). The main lesson is that extrapolation of the powerlaw to unobserved energies that are many orders of magnitude smaller than the observed energy ranges remains questionable. The integrated energy flux over the observed energy range is less susceptible to the powerlaw slope, because the total flux of the sum of all measured events is conserved, regardless how the fine structure is subdivided into discrete subevents. Fortunately, since the total energy of the observed nanoflare distributions is commensurable with the radiative losses, there is no need to extrapolate the distribution to unobserved energy ranges, and thus the question whether the powerlaw slope is below or above the critical value of 2 is not decisive for the heating budget. Another lesson is the completeness of temperature coverage, which generally requires coordinated multi-wavelength observations. For instance, a statistical analysis of coronal bright points with EIT 195 Å revealed that bright points cover only about 1.4% of the quiet Sun area, and their radiation accounts for about 5% of the quiet Sun radiation (Zhang et al. 2001), while the multi-wavelength data sets reproduce almost all of the quiet Sun flux.

5 The Coronal Heating Problem

When Bengt Edlén and Walter Grotrian identified Fe IX and Ca XIV lines in the solar spectrum (Edlén 1943), a coronal temperature of $T \approx 1$ MK was inferred from the formation temperature of these highly ionized atoms, for the first time. A profound consequence of this measurement is the implication that the corona then consists of a fully ionized hydrogen plasma. Comparing this coronal temperature with the photospheric temperature of 6000 K (or down to 4800 K in sunspots), we are confronted with the puzzle of how the 200 times hotter coronal temperature can be maintained, the so-called *coronal heating problem*. Of course, there is also a *chromospheric heating problem* and a *solar wind heating problem* (Hollweg 1985). If only thermal conduc-

tion was at work, the temperature in the corona should steadily drop down from the chromospheric value with increasing distance, according to the second law of thermodynamics. Moreover, since we have radiative losses by EUV emission, the corona would just cool off in matter of hours to days, if the plasma temperature could not be maintained continuously by some heating source. In this section we will specify the energy requirement for coronal heating, review a fair number of theoretical models that provide coronal heating mechanisms, and scrutinize them with observational tests if possible. However, all we have available for observational testing are mostly measurements of basic physical parameters, such as density, temperatures, and flow speeds, while theoretical heating models are expressed in parameters that are often not directly measurable in the corona, such as the magnetic field strength, azimuthal field components, nonpotential fields, currents, resistivity, viscosity, turbulence, waves, etc. However, the detection of MHD waves in the corona by TRACE and EIT, the spectroscopic measurements of line widths by SUMER, and the ion temperature anisotropy measurements with UVCS opened up powerful new tools that promise to narrow down the number of viable coronal heating mechanisms in the near future.

The coronal heating problem has been narrowed down by substantial progress in theoretical modeling with MHD codes, new high-resolution imaging with the SXT, EIT and TRACE telescopes, and with more sophisticated data analysis using automated pattern recognition codes. The total energy losses in the solar corona range from $F = 3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$ in quiet Sun regions to $F \approx 10^7 \text{ erg cm}^{-2} \text{ s}^{-1}$ in active regions. Theoretical models of coronal heating mechanisms include the two main groups of DC and AC models, which involve as a primary energy source chromospheric footpoint motion or upward leaking Alfvén waves, which are dissipated in the corona by magnetic reconnection, current cascades, MHD turbulence, Alfvén resonance, resonant absorption, or phase mixing. There is also strong observational evidence for solar wind heating by cyclotron resonance, while velocity filtration seems not to be consistent with EUV data. Progress in theoretical models has mainly been made by abandoning homogeneous fluxtubes, but instead including gravitational scale heights and more realistic models of the transition region, and taking advantage of numerical simulations with 3D MHD codes. From the observational side we can now unify many coronal small-scale phenomena with flare-like characteristics, subdivided into microflares (in soft X-rays) and nanoflares (in EUV) solely by their energy content. Scaling laws of the physical parameters corroborate the unification of nanoflares, microflares, and flares; they provide a physical basis to understand the frequency distributions of their parameters and allow estimation of their energy budget for coronal heating. Synthesized data sets of microflares and nanoflares in EUV and soft X-rays have established that these impulsive small-scale phenomena match the radiative loss of the average quiet Sun corona, which points to small-scale magnetic reconnection processes in the transition region and lower corona as primary heating sources of the corona.

5.1 Heating Energy Requirement

We start to analyze the coronal heating problem by inquiring first about the energy requirements. A coronal heating source E_H has to balance at least the two major loss terms of radiative loss E_R and thermal conduction E_C , as we specified in the energy

equation for a hydrostatic corona,

$$E_H(\mathbf{x}) - E_R(\mathbf{x}) - E_C(\mathbf{x}) = 0, \quad (5.1)$$

where each of the terms represents an energy rate per volume and time unit ($\text{erg cm}^{-3} \text{s}^{-1}$), and depends on the spatial location \mathbf{x} . Because the corona is very inhomogeneous, the heating requirement varies by several orders of magnitude depending on the location. Because of the highly organized structuring by the magnetic field (due to the low plasma- β parameter in the corona), neighboring structures are fully isolated and can have large gradients in the heating rate requirement, while field-aligned conduction will smooth out temperature differences so that an energy balance is warranted along magnetic field lines. We can therefore specify the heating requirement for each magnetically isolated structure separately (e.g., a loop or an open fluxtube in a coronal hole), and consider only the field-aligned space coordinate s in each energy equation, as we did for the energy equation of a single loop,

$$E_H(s) - E_R(s) - E_C(s) = 0. \quad (5.2)$$

Parameterizing the dependence of the heating rate on the space coordinate s with an exponential function (i.e., with a base heating rate E_{H0} and heating scale length s_H), we derived scaling laws for coronal loops in hydrostatic energy balance, which are known as *RTV laws* for the special case of uniform heating without gravity, and have been generalized by Serio et al. (1981) for nonuniform heating and gravity. It is instructional to express the RTV law as a function of the loop density n_e and loop half length L , which we obtain by inserting the pressure from the ideal gas law, $p_0 = 2n_e k_B T_{max}$,

$$T_{max} \approx 10^{-3} (n_e L)^{1/2} \quad (5.3)$$

$$E_{H0} \approx 2 \times 10^{-17} n_e^{7/4} L^{-1/4} \quad (5.4)$$

This form of the RTV law tells us that the heating rate depends most strongly on the density, $E_{H0} \propto n_e^{7/4}$, and very weakly on the loop length L . Actually, we can retrieve essentially the same scaling law using a much simpler argument, considering only radiative loss, which is essentially proportional to the squared density,

$$E_{H0} \approx E_R = n_e^2 \Lambda(T) \approx 10^{-22} n_e^2 \quad (\text{erg cm}^{-3} \text{s}^{-1}) \quad (5.5)$$

where the radiative loss function can be approximated by a constant $\Lambda(T) \approx 10^{-22}$ [$\text{erg cm}^{-3} \text{s}^{-1}$] in the temperature range of $T \approx 0.5 - 3$ MK that characterizes most parts of the corona. This gives us a very simple guiding rule: the coronal heating rate requirement is essentially determined by the squared density. The rule (Eq. 5.5) gives us the following estimates: in coronal holes the base density is typically $n_e \approx 10^8 \text{ cm}^{-3}$ and the heating rate requirement is thus $E_{H0} \approx 10^{-6} (\text{erg cm}^{-3} \text{s}^{-1})$.

Since the heating flux is quickly distributed along a magnetic field line, we can just specify a heating rate per unit area at the coronal base, by integrating the volume heating rate in the vertical direction. For hydrostatic structures, we can integrate the heating rate in the vertical direction simply by multiplying it with the density scale

Table 3: Chromospheric and coronal energy losses, in units of ($\text{erg cm}^{-2} \text{s}^{-1}$) (Withbroe & Noyes 1977).

Parameter	Coronal hole	Quiet Sun	Active region
Transition layer pressure [dyn cm^{-2}]	7×10^{-2}	2×10^{-1}	2
Coronal temperature [K], at $r \approx 1.1R_{\odot}$	10^6	1.5×10^6	2.5×10^6
Coronal energy losses [$\text{erg cm}^{-2} \text{s}^{-1}$]			
– Conductive flux F_C	6×10^4	2×10^5	$10^5 - 10^7$
– Radiative flux F_R	10^4	10^5	5×10^6
– Solar wind flux F_W	7×10^5	$\lesssim 5 \times 10^4$	($< 10^5$)
– Total corona loss $F_C + F_R + F_W$	8×10^5	3×10^5	10^7
Chromospheric radiative losses [$\text{erg cm}^{-2} \text{s}^{-1}$]			
– Low chromosphere	2×10^6	2×10^6	$\gtrsim 10^7$
– Middle chromosphere	2×10^6	2×10^6	10^7
– Upper chromosphere	3×10^5	3×10^5	2×10^6
– Total chromospheric loss	4×10^6	4×10^6	2×10^7
Solar wind mass loss [$\text{g cm}^{-2} \text{s}^{-1}$]	2×10^{-10}	$\lesssim 2 \times 10^{-11}$	($< 4 \times 10^{-11}$)

height λ_T , which is proportional to the temperature. We denote the heating flux per unit area with the symbol F_{H0} (also called *Poynting flux*),

$$F_{H0} = E_{H0} \lambda_T \approx 5 \times 10^3 \left(\frac{n_e}{10^8 \text{ cm}^{-3}} \right)^2 \left(\frac{T}{1 \text{ MK}} \right) \quad [\text{erg cm}^{-2} \text{s}^{-1}] \quad (5.6)$$

Thus for a coronal hole, with $n_e = 10^8 \text{ cm}^{-3}$ and $T = 1.0 \text{ MK}$, we estimate a required heating flux of $F_{H0} = 5 \times 10^3 \text{ erg cm}^{-2} \text{s}^{-1}$, and in an active region with a typical loop base density of $n_e = 2.0 \times 10^9 \text{ cm}^{-3}$ and $T = 2.5 \text{ MK}$, we estimate $F_{H0} \approx 5 \times 10^6 \text{ (erg cm}^{-2} \text{s}^{-1})$. Thus the heating rate requirement varies by about 3 orders of magnitude between the two places.

Another conclusion we can immediately draw about the heating function is that the height dependence of the heating has roughly to follow the hydrostatic equilibrium. The heating scale height s_H required in hydrostatic equilibrium is therefore half of the density scale height λ_T , because the radiative loss scales with the squared density, $E_H(h) = E_{H0} \exp(-h/s_H) \propto E_R(h) \propto n_e(h)^2 \approx [n_0 \exp(-h/\lambda_T)]^2$,

$$s_H \approx \frac{\lambda_T}{2} \approx 23 \left(\frac{T}{1 \text{ MK}} \right) \quad [\text{Mm}]. \quad (5.7)$$

This simple theoretical prediction, assuming that radiative loss is the dominant loss component in the coronal part of loops, is also confirmed by hydrostatic modeling of 40 loops observed with TRACE, where including the effect of thermal conduction yielded only slightly smaller values (i.e., $s_H = 17 \pm 6 \text{ Mm}$, Aschwanden et al. 2000d).

The spatial variation of the coronal heating requirement is illustrated in Fig. 26, where we deconvolved the mean coronal base density n_{e0} and differential emission measure distribution $dEM(T)/dT$ in 36 different sectors of the corona from Yohkoh SXT two-filter measurements (Aschwanden & Acton 2001), and determined the heating requirement F_{H0} for these 36 different sectors, finding $5 \times 10^3 \lesssim F_{H0} < 1 \times 10^4$

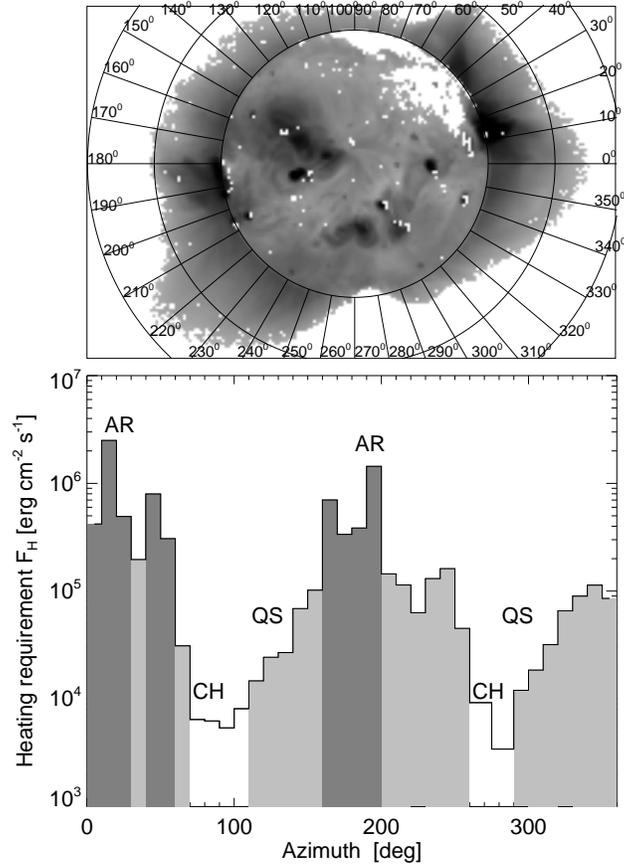


Figure 26: Composite soft X-ray image of the Sun observed on 1992 Aug 26 with Yohkoh (*top panel*). The histogram shows the heating rate requirement (*bottom panel*) in the 36 azimuthal sectors around the Sun. The labels indicate the locations of active regions (AR; dark grey), quiet Sun regions (QS; light grey), and coronal holes (CH; white) (Aschwanden 2001b).

($\text{erg cm}^{-2} \text{s}^{-1}$) in coronal holes, $1 \times 10^4 \lesssim F_{H0} < 2 \times 10^5$ ($\text{erg cm}^{-2} \text{s}^{-1}$) in quiet Sun regions, and $2 \times 10^5 \lesssim F_{H0} < 2 \times 10^6$ ($\text{erg cm}^{-2} \text{s}^{-1}$) in active regions. These measurements agree with the radiative losses found in other observations (e.g., Jordan 1976; Withbroe & Noyes 1977; see Table 9.1).

So, we have a quite specific perception of the heating requirement in the solar corona. The simplest rule is the dependence on the squared electron density, $F_{H0} \propto n_e^2$, which is also proportional to the optically thin emission measure in EUV and soft X-rays, and thus to the observed flux. This sounds trivial, that the heating rate is directly proportional to the observed brightness, if we associate radiation as the major loss, but it would not be true for optically thick radiation, where the observed brightness temperature is lower than the actual electron temperature. A direct consequence of the squared density dependence is that most of the heating is required in the lowest half

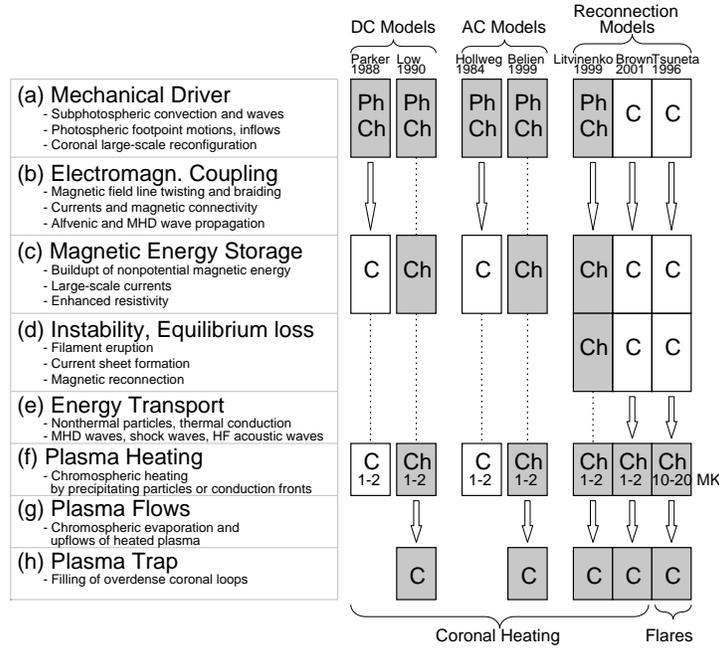


Figure 27: The process of coronal heating can be broken down into 8 subprocesses (a–h). Theoretical models include different subsets of these subprocesses, but only models that include the last step and can parameterize the physical parameters of the heated plasma can be compared with observations. The right side of the diagram shows a flow chart for the major heating models (with a typical representative listed at the top). Boxes mark physical steps that are part of the model, arrows mark transport processes between different locations, and dotted lines mark co-spatial locations. The boxes are colored in grey if the physical process takes place in a high-density region (Ph=photosphere, Ch=chromosphere, overdense coronal loops) and appear white for low-density regions (C=coronal background plasma) (Aschwanden 2001b).

density scale height. When we ask what the dependence of the heating rate is on the temperature, the RTV law (Eqs. 5.3 and 5.4) predicts a dependence with the three-and-a-half power, $F_{H0} \propto T^{3.5}$. Thus a soft X-ray-bright loop with a typical temperature of $T = 3$ MK needs about 50 times more heating flux than an EUV-bright loop with $T = 1$ MK.

In Table 3 we list the energy losses in the corona and chromosphere for comparison, given separately for coronal holes, quiet Sun regions, and active regions. We see that the radiative losses are fully comparable with the conductive losses (within a factor of 2) in the quiet Sun and active regions. Only in coronal holes, radiative loss is substantially less than the losses by thermal conduction and the solar wind flux, because of the low density. So, we can summarize that the minimum heating requirement at any place on the solar surface is $P_{H0} \gtrsim 3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$, mostly needed in the lowest half density scale height, and the heating requirement increases up to two orders of magnitude in dense loops in active regions, roughly scaling with the squared density.

Table 4: Coronal heating models (adapted from Mandrini et al. 2000).

Physical process	References
<i>1. DC stressing and reconnection models:</i>	
– Stress-induced reconnection	Sturrock & Uchida (1981) Parker (1983, 1988); Berger (1991, 1993) Galsgaard & Nordlund (1997)
– Stress-induced current cascade	Van Ballegooijen (1986) Hendrix et al. (1996) Galsgaard & Nordlund (1996) Gudiksen & Nordlund (2002)
– Stress-induced turbulence	Heyvaerts & Priest (1992) Einaudi et al. (1996a,b) Inverarity & Priest (1995a) Dmitruk & Gomez (1997) Milano et al. (1997, 1999); Aly & Amari (1997)
<i>2. AC wave heating models:</i>	
– Alfvénic resonance	Hollweg (1985, 1991)
– Resonant absorption	Ionson (1978, 1982, 1983), Mok (1987) Davila (1987), Poedts et al. (1989) Goossens et al. (1992, 1995) Steinolfson & Davila (1993) Ofman & Davila (1994); Ofman et al. (1994, 1995) Erdélyi & Goossens (1994, 1995, 1996) Halberstadt & Goedbloed (1995a,b) Ruderman et al. (1997) Bélien et al. (1999)
– Phase mixing	Heyvaerts & Priest (1983) Parker (1991); Poedts et al. (1997) De Moortel et al. (1999, 2000a)
– Current layers	Galsgaard & Nordlund (1996)
– MHD turbulence	Inverarity & Priest (1995b) Matthaeus et al. (1999) Dmitruk et al. (2001, 2002)
– Cyclotron resonance	Hollweg (1986), Hollweg & Johnson (1988) Isenberg (1990), Cranmer et al. (1999a) Tu & Marsch (1997, 2001a,b) Marsch & Tu (1997a,b,2001)
<i>3. Acoustic heating:</i>	
– Acoustic waves	Schatzman (1949) Kuperus, Ionson, & Spicer (1981)
<i>4. Chromospheric reconnection:</i>	
	Litvinenko (1999a) Longcope & Kankelborg (1999) Furusawa & Sakai (2000) Sakai et al. (2000a,b, 2001a,b) Brown et al. (2000) Tarbell et al. (1999) Ryutova et al. (2001) Sturrock (1999)
<i>5. Velocity filtration:</i>	
	Scudder (1992a,b; 1994)

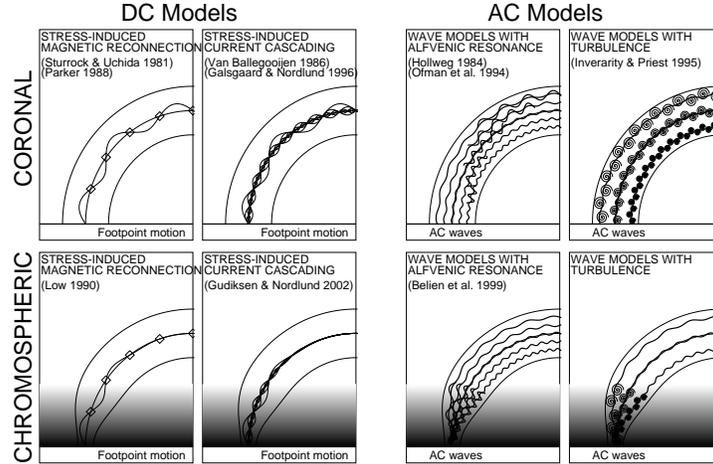


Figure 28: Categories of DC (left panels) and AC models (right panels), subdivided into coronal (top row) and chromospheric versions (bottom row). The greytones demarcate high-density regions (chromosphere and transition region) (Aschwanden 2002b).

5.2 Overview of Coronal Heating Models

In Table 4 we categorize theoretical models of coronal heating processes into 5 groups, according to the main underlying or driving physical processes. It became customary to classify coronal heating models into *DC* (*Direct Current*) and *AC* (*Alternating Current*) types, which characterize the electromechanic coronal response to the photospheric driver that provides the ultimate energy source for heating. Magnetic disturbances propagate from the photosphere to the corona with the Alfvén speed v_A . If the photospheric driver, say random motion of magnetic field line footpoints, changes the boundary condition on time scales much longer than the Alfvén transit time along a coronal loop, the loop can adjust to the changing boundary condition in a quasi-static way, and thus the coronal currents are almost direct ones, which defines the *DC models*. On the other hand, if the photospheric driver changes faster than a coronal loop can adjust to (e.g., by damping and dissipation of incident Alfvén waves), the coronal loop sees an alternating current, which is the characteristic of *AC models*. For each of the two model groups there are a number of variants of how the currents are dissipated, either by Ohmic dissipation, magnetic reconnection, current cascading, and viscous turbulence in the case of DC models, or by Alfvénic resonance, resonance absorption, phase mixing, current layer formation, and turbulence in the case of AC models. As an alternative to current dissipation, some heating could also be produced by compressional waves (i.e., by acoustic waves or shocks). Finally, a completely different physical mechanism is that of velocity filtration, which is based on the influence of the gravitational potential field in the corona on a postulated non-Maxwellian chromospheric velocity distribution.

Let us have a look at the compatibility and completeness of coronal heating models. In most of the theoretical models, coronal heating is a multi-stage process, which can

be conceptually organized in a scheme with 8 steps, as illustrated in Fig. 27: the initial energy comes from a mechanical driver (a), which has an electromagnetic coupling (b) to the location of magnetic energy storage (c). At some point, a magnetic instability and loss of equilibrium (d) occurs, with possible energy transport (e), before plasma heating (f) starts. The resulting overpressure forces plasma flows (g), which become trapped (h) in coronal loops, where they are eventually observed. Various coronal heating models cover only an incomplete subset of these steps, so that these concepts first have to be combined with specific geometric and physically quantified models of coronal structures before they can be applied or fitted to observations. Therefore, observational tests of theoretical heating models are still in their infancy.

An aspect of over-riding importance for modeling coronal heating is the treatment of a realistic chromospheric and transition region boundary. This is visualized in Fig. 28 for some standard models. Early versions of coronal heating models usually approximate a coronal loop with a uniform fluxtube (Fig. 28, top row), which produces a more or less uniform energy dissipation for stressing of magnetic field lines and has rather large dissipation lengths for Alfvén waves. In other words, these highly idealized models produce an almost uniform heating function that stands in stark contrast to the observations. Recent, more realistic, models include gravity and the density and temperature structure of the chromosphere/transition region at the lower boundary (Fig. 28, bottom row), which changes the resulting heating function drastically. Typically, the heating rate is much more concentrated near the footpoints, because of stronger stressing in the canopy-like magnetic field in the transition region, or due to vertical gradients in the density and Alfvén velocity caused by gravitational stratification.

The consideration of the transition region in coronal heating models also plays a crucial role for all models that involve magnetic reconnection. Essentially, the transition region is a dividing line between collisional (chromospheric) and collisionless (coronal) regimes, as illustrated in Fig. 1. Magnetic reconnection in collisionless regimes leads, besides plasma heating, to particle acceleration, which in turn, can efficiently contribute to chromospheric plasma heating (e.g., by chromospheric evaporation or thermal conduction fronts, as known for flares). The very same process is also believed to be responsible for heating of the quiet corona to some extent, as the nonthermal signatures of nanoflares in the quiet Sun suggest. However, if the same magnetic reconnection process happens inside the chromosphere, no particles can be accelerated because their collision time is shorter than their escape time out of the chromosphere. So, no secondary heating via accelerated particles is possible for reconnection processes in collisional plasmas. Therefore, the location of the magnetic reconnection region with respect to the transition region (above or below) is extremely decisive for the efficiency of coronal plasma heating.

Solar Flares on All Scales and Coronal Heating Problem

<u>1. Magnetic Reconnection Topologies</u>	1
1.1 Examples of Bipolar X-Point Reconnection	3
1.2 Examples of Tripolar Magnetic Reconnection	4
1.3 Examples of Quadrupolar Magnetic Reconnection	5
<u>2. Magnetic Reconnection Processes</u>	7
2.1 Steady 2D Magnetic Reconnection	7
2.2 Sweet–Parker Reconnection Model	9
2.3 Petschek Reconnection Model	10
2.4 Unsteady/Bursty 2D Reconnection	11
2.5 Tearing-Mode Instability and Magnetic Island Formation	12
2.6 Coalescence Instability	13
2.7 Dynamic Current Sheet and Bursty Reconnection	14
<u>3. Flare/CME Models</u>	17
3.1 The Standard 2D Flare Model	18
3.2 The Emerging Flux Model	20
3.3 The Equilibrium Loss Model	22
3.4 2D Quadrupolar Flare Model	24
3.5 The Magnetic Breakout Model	26
3.6 3D Quadrupolar Flare Models	28
3.7 Unification of Flare Models	29
<u>4. Flare Statistics and Frequency Distributions</u>	30
4.1 Theory of Frequency Distributions	30
4.2 Frequency Distributions of Flare Parameters	32
4.3 Energy Budget of Flare-like Events	34
4.4 Measurements of Frequency Distributions	36
<u>5. The Coronal Heating Problem</u>	38
5.1 Heating Energy Requirement	39
5.2 Overview of Coronal Heating Models	45