CHAPTER 6

HEATING OF THE UPPER ATMOSPHERE

6.1. Introduction

It was Biermann (1946) and Schwarzschild (1948) who first suggested the heating of the upper atmosphere (the chromosphere and corona) by *sound waves* that are generated from turbulence in the convection zone and then steepen to form *shock waves* as they propagate upwards. Until relatively recently this was universally accepted but now it is thought to be important only for the low chromosphere. This chapter first gives a summary of some energy-balance models that have been proposed for the upper atmosphere (assuming a simple form for the heating), and then it proceeds to discuss the processes that may produce the heating. Qualitatively, it is clear that a source of heat is needed to balance not only the energy radiated away in the chromosphere but also the energy removed by conduction from the temperature maximum. Quantitatively, however, it is still uncertain how the heating varies with altitude, and the detailed nature of the heating mechanism is highly controversial. It is also probable that the heating mechanism in the outer corona is collisionless and so beyond the scope of this book.

According to Withbroe and Noyes (1977), in the chromosphere the heating needed to balance radiation is about 4×10^3 W m^{-2} for quiet regions or coronal holes and 2×10^4 W m^{-2} for active regions. In the corona, the required energy input drops to only 3×10^2 W m^{-2} for quiet regions and 5×10^3 W m^{-2} for active regions. Down at the photosphere, an enormous input of wave flux of between 10^4 and 10^6 W m^{-2} is believed to exist (1 W m$^{-2} \equiv 10^3$ erg cm^{-2} s^{-1}), but it is not clear how much of this reaches higher levels.

Indirect observational support for the classical picture of atmospheric heating by acoustic waves had come from spectral line profiles; these are broadened by the presence of *nonthermal velocities* that increase with height from a few km s^{-1} in the low chromosphere to 25 to 30 km s^{-1} in the transition region and 10 to 30 km s^{-1} in the corona. The nonthermal broadening was thought possibly due to waves that were propagating upwards rather than remaining stationary. Indeed, oscillatory motions with a period of 300 s or shorter had been observed through the photosphere and chromosphere up to the low transition region (but not in the corona). Furthermore, the observations by Deubner (1976) suggested the presence near the temperature minimum of an energy flux of at most 10^5 to 10^6 W m^{-2}, carried by waves with periods 10 to 300 s. Even though this wave flux is radiatively damped in the photosphere, only a small fraction needed to propagate upwards to supply enough energy to heat the chromosphere and corona.

More recently, direct evidence from OSO 8 that only the *low chromosphere may be heated by acoustic waves* has been presented by Athay and White (1977). Mein *et al.*

(1980) and Mein and Schmieder (1981) have summarised other observations. At the temperature minimum a 5-min wave-train is observed, with possibly enough energy to heat the low chromosphere; but, by the time the upper chromosphere is reached, its flux has been reduced and its coherence destroyed (Section 6.3.2). *Heating of the upper chromosphere and corona is therefore probably magnetic in nature* (Section 6.4). Magnetic waves present difficulties because they are both scattered off inhomogeneities and refracted as the Alfvén speed increases with altitude; but, provided enough wave-flux can reach the corona, *short-period (≈ 10 s) Alfvén waves appear to be a viable heating mechanism for most coronal loops* (Section 6.4.2), where the magnetic field is less than about 20 G. However, they are difficult to dissipate in stronger magnetic field regions, such as *X-ray bright points* and *active-region cores*. An alternative heating mechanism that does work in a strong field is *magnetic dissipation in current sheets or filaments* (Section 6.4.4). Here the energy is transferred to the corona by motions of such a low frequency that the magnetic field evolves through a series of *quasi-static configurations*. In the corona the energy then dissipates ohmically in the classical way that has long been proposed for solar flares. A continuous occurrence of such 'mini-flares' is proposed as the means by which the corona is heated generally; but a detailed analysis of the process remains to be carried out.

6.2. Models for Atmospheric Structure

6.2.1. BASIC MODEL

The temperature increases dramatically from the chromosphere up through the narrow transition region to the corona, in a manner that is shown schematically in Figure 6.1. The gradient of the temperature increases from small values in the chromosphere to extremely large values in the transition region and then decreases to zero at the temperature maximum. The inflexion point in the temperature profile corresponds to the place of maximum temperature gradient and is located typically near the base of the transition region ($\approx 2 \times 10^4$ K), whereas the inflexion point in the profile of $T^{7/2}$ (at T_i) gives the maximum heat flux and occurs typically at about 10^6 K.

At each location a thermal equilibrium,

$$C = H - R, \tag{6.1}$$

has often been assumed to hold between some kind of heating (H), radiative losses ($-R$) and a downwards conductive flux ($\mathbf{F}_c = -\kappa_0 T^{5/2} \nabla T$), whose divergence is the conductive loss (C). Below T_i, C is negative, so that conduction deposits heat and the radiation (R) exceeds heating (H), whereas above T_i C is positive and heating dominates radiation.

When Equation (6.1) holds, the temperature structure is determined by the relative sizes of R and H and by the response of C in maintaining the thermal balance. In the *chromosphere*, R and H are both relatively large, while their difference C is small, so that the spatial change in temperature gradient is slow. As the temperature rises, so the radiation (R) increases to a maximum between 10^4 and 10^5 K (see Table 2.2); here R greatly exceeds the heating (H) and is balanced by C. This in turn forces the temperature gradient to increase to large values in the *lower transition region* (which
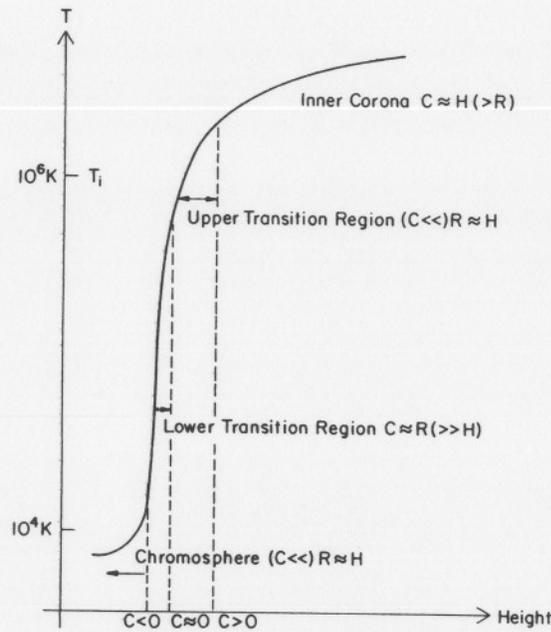
Fig. 6.1. A sketch of the temperature structure in the upper solar atmosphere, indicating the relative roles of conduction (C), radiation (R) and heating (H).

is therefore being heated by conduction from above and cooled by radiation). Above about 10^5 K the radiation falls dramatically and eventually it reaches equality with heating. Then, through the *upper transition region*, the conductive flux $\mathbf{F}_c$ stays relatively constant and at a high value. In the *lower corona* R and H have fallen to much smaller values and heating is balanced mainly by conductive losses, the relatively small value of C implying slow changes of temperature gradient with height. Above the temperature maximum of an open magnetic field region, the energy transport by the solar wind becomes increasingly important and eventually it dominates conduction of heat outwards.

Thus it can be seen that the cause of the extremely steep rise of temperature in the lower transition region is the fact that the *radiation around 10^4 to 10^5 K is so large that it cannot be supplied by mechanical heating* but must be provided by conduction from above. Furthermore, at greater altitudes the energy that is deposited as *heat cannot be radiated away* and so it must be conducted both inwards and outwards from the temperature maximum.

Typical values for the coronal temperature and transition-region pressure are given in Table 6.1 (from Withbroe and Noyes, 1977) along with estimates for the conductive and radiative losses (C and R) per unit area at different levels. The heating (H) that is required may be obtained by summing the losses (C and R). The energy necessary to heat the corona is typically only a few percent of that needed down in the chromosphere, so a comprehensive model of atmospheric heating would have to

TABLE 6.1.

Energy losses from the upper atmosphere (1 W m^{-2} ≡ 10^3 erg cm^{-2} s^{-1})

	Conduction (W m^{-2})	Radiation (W m^{-2})	Temperature (K)	Pressure (N m^{-2})
Quiet region:				
Lower and middle chromosphere		4×10^3		
Upper chromosphere		3×10^2		2×10^{-2}
Corona.	2×10^2	10^2	$1.1 - 1.6 \times 10^6$	
Coronal hole:				
Lower and middle chromosphere		4×10^3		
Upper chromosphere		3×10^2		7×10^{-3}
Corona	6×10	10	10^6	
Active region:				
Lower and middle chromosphere		2×10^4		
Upper chromosphere		2×10^3		2×10^{-1}
Corona	$10^2 - 10^4$	5×10^3	2.5×10^6	

treat the generation, propagation and dissipation of energy through the whole region to a high degree of accuracy. Table 6.1 also shows the extent to which a coronal hole region is both cooler and less dense than a normal quiet region of the Sun, while an active region is both hotter and denser.

A simple numerical model for the upper atmosphere may be computed by assuming forms for the terms in (6.1). For example, Wragg and Priest (1981b) have adapted an earlier model of McWhirter *et al.* (1975). They solve

$$\frac{d}{dz}\left(\kappa_0 T^{5/2}\frac{dT}{dz}\right) = \chi n_e^2 T^\alpha - H \tag{6.2}$$

for the temperature profile ($T(z)$) as a function of height, and use the standard values for the conduction coefficient (κ_0) (Equation 2.34) and the optically thin radiative-loss coefficients χ and α (Table 2.2). The heating (H) is assumed uniform per unit volume for simplicity, and Equation (6.2) is coupled with the equation of hydrostatic equilibrium for a fully-ionised hydrogen plasma, namely

$$\frac{dp}{dz} = -m_p n_e g, \tag{6.3}$$

where $p = 2n_e k_B T$. Three boundary conditions are needed to solve Equations (6.2) and (6.3), such as the prescription at the base of temperature, temperature gradient and electron density. The resulting profiles depend on one parameter alone, namely H, some examples being shown in Figure 6.2(a). It is particularly interesting to discover how the temperature maximum and the height at which it occurs vary with

(a)

(b)

(c)
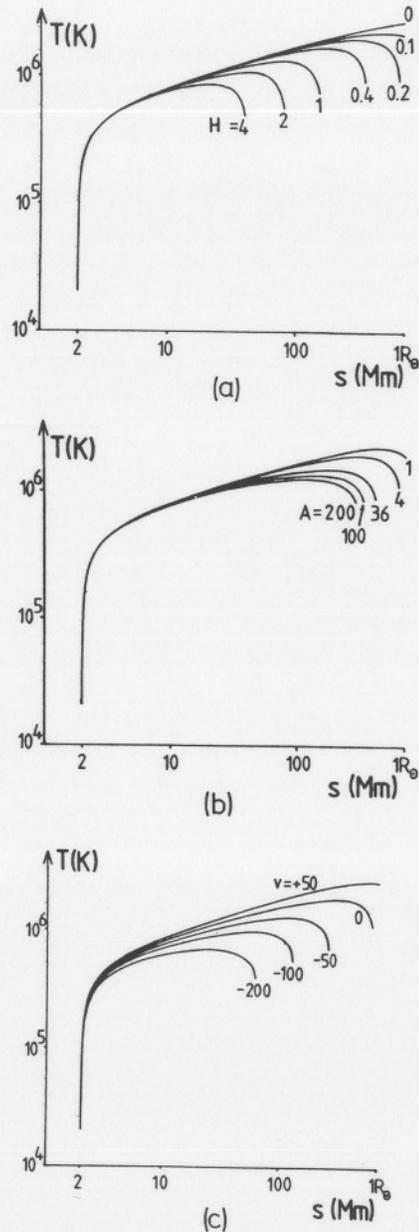
Fig. 6.2. The temperature (T) of a model coronal atmosphere as a function of height s (1 Mm ≡ 10^6 m), showing the effect of varying: (a) the heating strength H, (b) the flux tube divergence for $H = 0.2$, (c) a flow for $H = 0.2$. The heating H is measured in units of the radiation at a temperature of 10^6 K and a density of 5×10^{14} m^{-3}. A is the ratio of the loop area at the summit to that at the base. v is the flow speed in m s^{-1} (+ represents an upflow and − a downflow) (from Wragg and Priest, 1981b).

H. If these models were linked to a coronal hole model, the rapid fall in temperature after the temperature maximum would be replaced by a more gradual decline. In Figures 6.2(b) and 6.2(c) the effects of varying the cross-sectional area and flow speed are also displayed.

6.2.2. MAGNETIC FIELD EFFECTS

It is clear from eclipse or X-ray pictures of the Sun, such as Figures 1.3(d), 1.4 and 1.15, that coronal structure is dominated by the magnetic field. Regions of open field show up as dark *coronal holes*, whereas closed-field regions are seen as bright *coronal loops*. The influence of the magnetic field on coronal plasma is threefold.

(i) *It exerts a force.* The $\mathbf{j} \times \mathbf{B}$ force is able to act inward and so contain plasma with an enhanced pressure in features such as *X-ray bright points*, coronal loops and active regions.

(ii) *It stores energy.* The energy ($B^2/(2\mu)$ per unit volume) that is stored in the magnetic field may provide an extra source of heating, either by allowing *additional wave modes* (Section 4.3) that eventually dissipate or by being released directly by ohmic dissipation (j^2/σ) in regions where electric currents are strong.

(iii) *It channels heat.* The coefficient of thermal conduction ($\kappa_{\parallel}$) along the field is much larger than the coefficient ($\kappa_{\perp}$) across the field, so the magnetic field acts as a 'blanket' and thermally insulates the plasma very effectively. Heat is constrained to flow largely along the field, which means that in the transition region and corona, where conduction is an important means of energy transport, the temperature and density are strongly affected by the structure of the magnetic field. This is one reason why the *coronal structures in eclipse and X-ray pictures probably outline the magnetic field*. A discussion of such coronal loops can be found in Section 6.5.

A model has been developed by Gabriel (1976) (from an earlier attempt of Kopp and Kuperus, 1968) for the atmosphere above a (quiet-region) *supergranule cell*, typically 30 000 km in diameter. In the photosphere, magnetic flux is concentrated at supergranule boundaries by the convective flow and instability (Section 8.7), but higher up in the atmosphere the flux expands until it has become relatively uniform at the corona (Figure 6.3). Thus, images formed in transition-region lines follow the supergranulation pattern, with intensities over supergranule boundaries about a factor of ten higher than those over cell centres, while at the corona the pattern has disappeared (Figure 1.14). Gabriel assumes for simplicity that $\nabla \times \mathbf{B} = \mathbf{0}$, so that the magnetic field is a potential one. He next assumes the plasma to be in thermal equilibrium under a balance

$$\frac{d}{ds}\left(\kappa_{\parallel} A(s) \frac{dT}{ds}\right) = \chi n_e^2 T^\alpha A(s),$$

between only conduction and radiation, where A is a flux-tube cross-sectional area and local heating has been ignored; heat is supposedly deposited at greater heights. This equation is solved along each field line together with the Equation (6.3) of hydrostatic equilibrium, subject to the boundary conditions that the density, temperature and conductive flux be the same on each field line in the corona at an altitude of 30 000 km. A coronal conductive flux of 360 W m^{-2} is found to give closest agree-
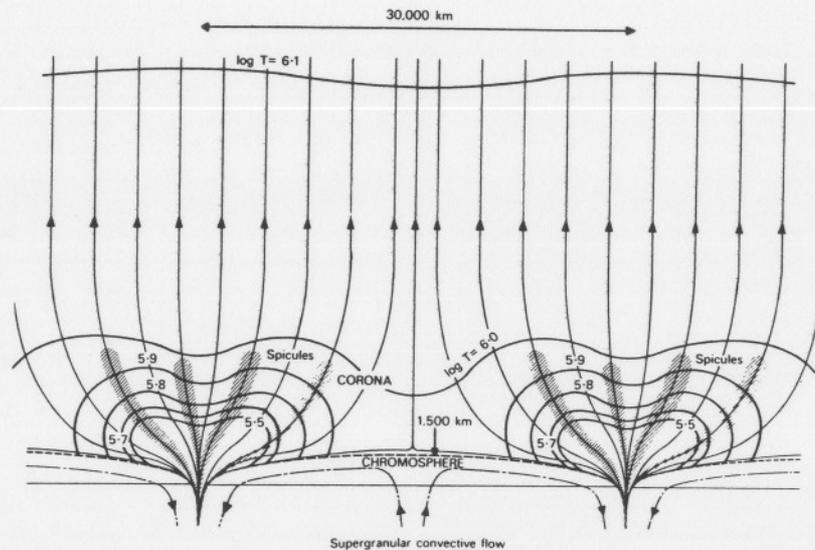
Fig. 6.3. Magnetic field lines and temperature contours for the atmosphere above a supergranule cell in a quiet region (after Gabriel, 1976).

ment with observation for the network width, and the resulting isotherms are sketched in Figure 6.3. By comparison with a plane-parallel model, the effect of the above flux-tube divergence (through $A(s)$) is to increase the temperature gradient at transition-region temperatures and so lower the height at which coronal temperatures are attained. Furthermore, the observed intensities of optically thin lines may be used to derive the differential emission measure ($n_e^2 T \, dh/dT$) as a function of temperature. At temperatures between $10^{5.2}$ and $10^{6.2}$ K, agreement with this emission measure is much better for Gabriel's model than for McWhirter *et al.* (1975)'s previous spherically symmetric models. Below $10^{5.2}$ K there is a need to include heating of amount 2×10^3 W m^{-2}.

Gabriel's model has recently been extended by Athay (1981b), who includes gravitational energy and enthalpy flux but no mechanical heating. With a downflow he obtains good agreement with observations for 3×10^5 K $\leqslant T \leqslant 10^6$ K, and so concludes that there is no need for mechanical heating. In future, there is a need to calculate a wider range of models and to couple the energy balance with a magneto-static force-balance, since the plasma beta is probably of order unity.

6.2.3. ADDITIONAL EFFECTS

Several effects may seriously modify the energy balance in the upper solar atmosphere but they are normally omitted from the models. For example, the waves that may be propagating up from below and heating the atmosphere exert a turbulent pressure, $\langle \rho v^2 \rangle$, which is just the time-average over a wave period of the momentum flux

(ρv^2). This *wave pressure* may well exceed the plasma pressure, and so it needs to be included in calculating the overall hydrostatic equilibrium. It may even overcome gravity and cause the plasma to flow outwards.

Most models of the thermal energy balance assume a static plasma and give a transition-region thickness of only 10 km or so, but in practice the region is probably in a *dynamic state*, with an effective thickness of a few thousand km. For example, material is continually ejected upwards as *spicules* from the chromosphere at supergranule boundaries, and persistent *downflows* of hotter plasma at 5 to 100 km s^{-1} are observed over supergranule boundaries, sunspots and plages. (It may even be the case that spicules are not just superimposed on a static atmosphere, but rather that essentially all of the transition-region emission is from the hot surroundings of spicules.) As cool spicular material moves upwards, it is heated and so extracts energy from its surroundings; when it falls back down, it carries thermal energy with it. Typically, one finds a downward mass flux at 10^5 K over supergranule boundaries of 5×10^{19} m^{-2} s^{-1} for a density of 5×10^{15} m^{-3} and a speed of 10 km s^{-1}. Pneuman and Kopp (1978) estimate that the heat carried down by such motions can exceed the heat transport by conduction, and so they construct a transition-region model based on a balance between heat downflow and radiation. The effect of a downflux (q) on the model of Wragg and Priest (1981b) can be seen in Figure 6.2(c), for which the term $d/dz(5kTq)$ has been added to the right-hand side of Equation (6.2).

6.3. Acoustic Wave Heating

Acoustic waves are believed to be *generated* near the photosphere (Section 4.9.3) and to *steepen* into shock waves at an altitude of a few hundred kilometres. They continue to *propagate* upwards and (according to some models) they may *dissipate* enough energy to balance radiation from the chromosphere (e.g., Schatzman, 1949; Kuperus, 1969; Ulmschneider, 1971, 1974, 1979; Kuperus and Chiuderi, 1976). The models predict that waves with a period of only a few tens of seconds heat the lower chromosphere. At one time, it was thought that 300 s waves could heat the upper chromosphere, but now their flux is thought to be too low (Ulmschneider, 1976). Also, by the time the corona is reached, the uncertainties are very large, since weak shock theory is no longer valid and there is considerable reflection and refraction in the transition region.

6.3.1. STEEPENING

Section 5.1 describes the way in which acoustic waves steepen to form shocks, and Section 5.2 includes a derivation of the jump conditions across a shock front. In a *uniform medium*, sound waves steepen because every part of the wave profile moves with a different speed. The crests possess a higher temperature than the troughs, and so they propagate faster. If c_s is the ambient sound speed and v_1 the velocity amplitude (Section 4.2), the crest of the wave moves with speed $c_s + v_1$, while the trough moves at only $c_s - v_1$. The crest therefore catches up the trough when the relative speed of $2v_1$. Since the trough is initially half a wavelength ($\frac{1}{2}\lambda$) ahead of the crest, the time it takes the trough to be overtaken is just $\lambda/(4v_1)$, and so the distance that a sound wave

can travel at speed c_s before shocking is

$$d = \frac{\lambda c_s}{4v_1},$$ (6.4)

or, in terms of the wave-period ($\tau = \lambda/c_s$),

$$d = \frac{\tau c_s^2}{4v_1}.$$ (6.5)

From this result, it can be seen that short-period waves evolve into shock waves over much smaller distances than long-period waves, which is why the former were thought to heat the low chromosphere and the latter the higher atmosphere.

For a *vertically stratified atmosphere*, rather than a uniform medium, the distance for shock formation is greatly reduced because the wave-amplitude increases rapidly with altitude. The density for an isothermal atmosphere decreases with height like $\rho(z) \sim e^{-z/\Lambda}$, where Λ is the scale-height. With no dissipation, the total *wave energy* is proportional to $\frac{1}{2}\rho v_1^2$; this remains constant as the wave propagates up, and so the wave amplitude increases as

$$v_1(z) \sim e^{z/(2\Lambda)}.$$ (6.6)

Thus, for example, taking a scale-height of 100 km, a wave starting in the low chromosphere would find its initial amplitude of only 0.2 km s^{-1}, say, grow to the sound speed (about 7.5 km s^{-1}) at an altitude of 1000 km. The distance required for a shock to form in the stratified medium becomes

$$d = 2\Lambda \log_e \left(1 + \frac{\tau c_s^2}{2(\gamma + 1)\Lambda v_1} \right)$$

in place of Equation (6.5), according to Stein and Leibacher (1974). For instance, setting $\gamma = 5/3$, $c_s = 6$ km s^{-1}, $\Lambda = 130$ km, a wave with an initial amplitude (v_1) of 0.6 km s^{-1} at the base of the chromosphere would develop into a shock at an altitude of 500 km if its period were 10 s or 800 km for a period of 30 s. This makes it most unlikely that the heating of the chromosphere is due to damping (by thermal conduction) of small-amplitude linear waves, since they develop into shocks too quickly.

The height of shock formation is in practice governed by the extent to which the wave is damped (as well as the height of generation, the initial amplitude and the wave-period). A wave can dissipate its energy through viscous, thermal or radiative losses, the last being the most important up to a height of about 1000 km (Osterbrock, 1961). Ulmschneider (1971) has estimated the radiative damping time to increase from 30 s at $h = 0$ in the photosphere to 750 s at $h = 400$ km and thence to decrease to 300 s at $h = 1000$ km, for the Harvard–Smithsonian Reference Atmosphere. His calculations demonstrate that sound waves with a period less than 100 s can propagate into the chromosphere and steepen into shocks below a height of 1000 km. Later, Ulmschneider and Kalkofen (1977) found that the effect of radiative damping is to make the height of shock formation independent of period.

6.3.2. Propagation and Dissipation

Three approximate theories have been developed to describe the propagation of a shock wave through an inhomogeneous atmosphere. *Geometrical acoustics* (Jeffrey

and Taniuti, 1964) makes the assumption that the shock energy remains constant, although a correction to include the dissipation can be added easily when the atmosphere has weak gradients. The *Chisnell–Witham method* (Witham, 1974) replaces an inhomogeneous atmosphere by a series of homogeneous layers. It neglects the reflected waves that are generated at each interface as the shock passes through, and also it does not include the effect of dissipation on the shock strength. However, a more useful theory for obtaining the dissipation from *periodically generated shocks* is that of Brinkley and Kirkwood, as developed by Schatzman (1949), Osterbrock (1961) and Ulmschneider (1971). It has been described by Bray and Loughhead (1974) and is outlined below.

Consider a train of shock waves that have developed from waves of frequency ν, so that the time interval between successive shocks is ν^{-1}. The flux of energy transmitted per second by the shocks at an altitude z_1 is the work done by the pressure, namely $F(z) = \nu\int(p - p_1)v \, dt / \int dt$, where the integrals are performed over one period; the front of the shock and its rear are denoted by subscripts 1 and 2, respectively. The energy flux may be rewritten

$$F(z) = \nu(p_2 - p_1)v_2 t_0 / \tau$$ (6.7)

in terms of a characteristic time (t_0) that is assumed (for self-similarity) to be independent of height. t_0 depends on the shock profile and is taken as $\tau/12$ by Ulmschneider (1971), where τ is the duration of the shock pulse. The jump conditions (5.14) to (5.16) may be used to write the energy flux (6.7) in terms of the fractional compression $\bar{\eta} = (\rho_2 - \rho_1)/\rho_1$. For *weak shocks* ($\rho_2 \approx \rho_1$), it becomes

$$F(z) = \nu\rho_1(z)c_{s1}(z)^3\bar{\eta}(z)^2/12 \,.$$ (6.8)

Now, following the passage of a shock, some energy is used in returning the plasma to its initial state. Schatzman suggested that it first expands adiabatically to its original pressure and then cools by radiation to its original density, ready for the passage of the next shock. For weak shocks, the resulting rate of energy dissipation is

$$\frac{-\nu\gamma(\gamma + 1)p_1\bar{\eta}^3}{12} \equiv \frac{dF}{dz}.$$ (6.9)

Since this is proportional to the wave frequency, short-period waves will dissipate faster than long-period ones. The corresponding rate of decrease of the peak pressure is

$$\frac{dp_2}{dz} = \frac{dp_1}{dz} - \frac{\nu\gamma^2(\gamma + 1)p_1^2\bar{\eta}^4}{12F},$$

which, when solved simultaneously with Equation (6.9), leads to a decrease of both p_2 and F with altitude like $z^{-1/2}$. An alternative procedure is to differentiate Equation (6.8) with respect to z and equate it to Equation (6.9). The resulting differential equation for the shock strength ($\bar{\eta}(z)$) has been solved numerically by Ulmschneider (1971) for a model atmosphere, including the additional effect of shock refraction away from the vertical. He deduces the dissipation as a function of height for a range of periods, and he finds agreement between this energy dissipation and the radiative loss in the *lower chromosphere* only if the shock period lies between 10 and 30 s. The

necessary shock strengths $(\bar{\eta})$ lie between 0.1 and 0.5, so that the weak shock approximation is justified. In order of magnitude, the distance it takes a shock to dissipate significantly, the so-called *shock damping-length*, is

$$d \approx \frac{F}{dF/dz} \approx \frac{c_s t_0}{\bar{\eta}} \qquad (6.10)$$

from Equations (6.7) and (6.9).

It may be noted that, in order to use the general energy equation (2.30) in modelling, one needs to insert a form for the heating. When heating is from weak acoustic shocks, it is given by Equation (6.9) in terms of the shock strength $(\bar{\eta}(z))$. If, furthermore, the shock *Mach number* $(M_1(z))$ is known as a function of altitude, the heating may be deduced from

$$H = \tfrac{2}{3}v\gamma(\gamma + 1)p_1(M_1 - 1)^3, \qquad (6.11)$$

and so, for constant M_1, it is proportional to the plasma pressure. This differs from the heating forms that are sometimes used in static models, namely heating uniform or proportional to density. There is always a definite relation between $\bar{\eta}$ and M: in general, it can be obtained by integrating Equations (6.8) and (6.9), but for weak shocks $\bar{\eta} = (4(M_1 - 1)/(\gamma + 1))$.

Short-period acoustic waves (10 to 50 s) are, according to the above theory, expected to develop into weak shocks within a few hundred kilometres of the chromospheric base and to heat the low chromosphere. This heating mechanism has recently been put on a firmer foundation by some numerical calculations of Ulmschneider and Kalkofen (1977) with an improved hydrodynamic code including radiative damping. The height of shock formation agrees with the position of the temperature minimum (namely 500 km for the Harvard–Smithsonian Reference Atmosphere) provided the wave period lies between 25 and 45 s and the initial acoustic flux is 3 to 6×10^4 W m^{-2}. Furthermore, the acoustic flux at the height of shock formation agrees with the chromospheric radiation provided the period is less than 35 s and the initial flux is 2 to 6×10^4 W m^{-2}.

Longer-period waves were once expected to increase in amplitude as radiative damping becomes less effective above 200 km and to develop into *strong shocks* with $\bar{\eta} > 1$ in the upper chromosphere. (Indeed, the shocks may well be strong lower in the atmosphere, so that weak-shock theory is invalid and gives no more than a qualitative picture.) Self-consistent models for heating the upper chromosphere and corona by strong shocks have been computed by, for instance, Kuperus (1965) and Ulmschneider (1971) assuming hydrostatic equilibrium and a thermal balance between conduction, radiation and heating. However, observations from the OSO 8 satellite (Athay and White, 1977) give strong evidence that 300-s waves have insufficient energy to heat the upper chromosphere, because their amplitudes have been reduced either by scattering from chromospheric inhomogeneities such as spicules and fibrils or by refraction in the region of rapidly increasing sound speed (see Section 6.4.1). In the low transition region at a temperature of 10^5 K the observed fluctuations are largely aperiodic and the shocks are very weak with an energy flux of only 10 W m^{-2}. Thus, although *short-period acoustic shock waves may well heat the low chromosphere*, some form of *magnetic heating is probably needed for the upper chromosphere and corona* (i.e., resonant absorption or tearing instability).

6.4. Magnetic Heating

The importance of the magnetic field in heating the solar atmosphere is being increasingly recognised, and one expects this subject to receive a great deal of attention in the future. At first, it was believed that the magnetic field had the secondary role of just enhancing the heating in active regions and in the network. But the discoveries of *intense kilogauss fields* (Sections 1.3.2B and 8.7) at *supergranulation boundaries* in the photosphere, and of coronal loops and bright points in soft X-ray photographs, have emphasised the dominance of the magnetic field in these regions too. Furthermore, the probability that insufficient acoustic flux reaches the upper chromosphere and corona has led to the suggestion that, even for the 'quiet' corona, the dominant heating mechanism is due to the magnetic flux that spreads upwards from supergranulation boundaries. The heating of both quiet and active regions would then be caused by the same mechanism, with active regions receiving more heat simply because the concentration of intense flux elements is higher. Two of the possibilities for magnetic heating are *magnetic waves* and direct *magnetic dissipation*. The actual cause of heating is the same in both cases, namely ohmic or viscous dissipation in small-scale regions, but the means of producing the regions is different, so that in one case the current sheets are propagating (i.e., magnetic shocks) and in the other case they are non-propagating.

Magnetic field disturbances are generated by the motion of the footpoints of field lines in the photosphere. For instance, the *intense fields* at supergranulation boundaries are continually being buffeted by granulation with a period of roughly 5 min and a scale size of about 1000 km. Also, the footpoints are being shuffled around on a supergranular time-scale of many hours. A general photospheric disturbance produces waves of several types (such as fast and slow magnetoacoustic waves and Alfvén waves) that may propagate upwards. The *magnetoacoustic waves* steepen into shocks and dissipate in a similar manner to pure acoustic waves, with ohmic heating providing extra dissipation. However, *Alfvén waves* dissipate much less readily. They require a nonlinear interaction to produce magnetoacoustic waves that subsequently relinquish their energy. Areas in need of further study are the way in which the generation of wave flux is affected by the magnetic field, the nonlinear coupling of the different wave modes and the propagation of magnetic waves in an inhomogeneous medium.

The classical linear treatment for the propagation of a photospheric disturbance as a wave through a uniform or slowly varying medium becomes inadequate in four situations:

(i). when the initial disturbance is so large that nonlinear effects are important;

(ii) when the source of the disturbance is closer than a few wavelengths, so that a wave-train has not developed;

(iii) when the disturbance is so slow that its wavelength exceeds a scale-height, i.e., $\lambda > \Lambda$, in which case the ambient medium cannot be considered slowly-varying;

(iv) when the footpoints of magnetic field lines move more slowly than the Alfvén travel-time (τ_A); in other words

$$\tau > \tau_A = \frac{L}{v_A}, \qquad (6.12)$$

or, equivalently, the wavelength exceeds the dimension (L) of the configuration:

$$\lambda > L. \tag{6.13}$$

In case (iv) the magnetic configuration evolves passively through a series of equilibria, set up by the relatively fast propagation of magnetic waves. However, the new equilibria can contain *current sheets*, which then dissipate magnetic energy ohmically and allow a *reconnection of the magnetic field lines* in the process. This has long been proposed as a mechanism for releasing magnetic energy as heat and kinetic energy in solar flares (see Chapter 10), but it may be taking place more often on a smaller scale throughout the corona. Such mechanisms for heating plasma in a magnetic field by waves or current-sheet dissipation are sketched in the sections that follow, and they are also described in the reviews by Heyvaerts and Schatzman (1980), Chiuderi (1981), Kuperus *et al.* (1981) and Hollweg (1981a).

6.4.1. PROPAGATION AND DISSIPATION OF MAGNETIC WAVES

In a classic paper, Osterbrock (1961) analysed the effect of the magnetic field on the generation, propagation and dissipation of the waves that may heat the chromosphere and corona. The analysis is limited to low-amplitude waves, with a wavelength much smaller than a scale-height. For quiet regions, where the mean magnetic field may be 2 G, upward-moving sound waves become increasingly magnetohydrodynamic in character as the fall-off in density makes the Alfvén speed increase; indeed, above a height of about 2000 km the Alfvén speed (v_A) exceeds the sound speed (c_s). For *plage regions*, where the mean field strength is typically 50 G, the magnetic field dominates even more and may cause the enhanced heating that is observed. Furthermore, the recent discovery of *kilogauss fields* at supergranulation boundaries (even in 'quiet' regions) makes the magnetic field less homogeneous than Osterbrock supposed and increases its effect on wave propagation.

It will be remembered that *fast magnetoacoustic waves* can propagate in any direction, with a phase speed that varies from the maximum of c_s and v_A along the magnetic field to $(c_s^2 + v_A^2)^{1/2}$ across the field. Thus, in a region where $v_A \gg c_s$, propagation is at v_A equally in all directions. The *slow mode*, by comparison, can transmit energy only in directions that are close to that of the magnetic field, the speed of propagation being the smaller of c_s and v_A. The third type of wave, namely the Alfvén mode, possesses a group velocity of v_A along the field and produces no change in either density or pressure.

Osterbrock points out that, in regions where the magnetic field strength is below the equipartition value of a few hundred Gauss, most of the wave energy that is generated by isotropic turbulence in the convective zone takes the form of fast-mode waves. Much smaller fluxes of slow-mode or Alfvén waves are generated, and so he proposes the *dissipation of fast shocks* as the dominant heating mechanism in the chromosphere.

The propagation and dissipation of *fast waves* differs quantitatively from that of ordinary sound waves in several respects, according to Osterbrock. Consider, for example, the refraction of waves as they propagate upwards in the stratified solar atmosphere. Since the acoustic speed (c_s) increases with height (from, say, 10 km s^{-1}

at the photosphere to 200 km s^{-1} in the corona), sound waves find their directions of propagation rotate away from the vertical as they progress upwards, in much the same way that light is refracted away from the normal as it crosses from water to air. However, the effect on the Alfvén speed (v_A) of the fall-off in plasma density away from the solar surface is to make the Alfvén speed increase even more rapidly with altitude (from typically 10 km s^{-1} at the photosphere to 10^3 km s^{-1} in the corona). Thus the refraction of fast-mode waves as they propagate upwards is even more pronounced than that of sound waves, so that even less energy is likely to reach coronal layers. Osterbrock calculated the *ray paths* in a vertically stratified plasma as follows. As the phase speed ($v_p(h)$) increases with height h (with the wave frequency remaining constant), so the wavelength increases. But, by Snell's Law, the horizontal component of $\mathbf{k}$ must remain constant, and this makes $\theta(h)$, which is the inclination to the vertical of the propagation direction, increase in such a way that $\sin \theta(h)/v_p(h) = $ constant. The resulting ray path ($h(x)$) taken by the wave in this approximation follows from integrating $dh/dx = \tan \theta$.

The damping, prior to shocking, of upward-propagating *magnetic waves* may also differ from that of acoustic waves. It is caused by *ohmic* as well as *viscous dissipation* and, near the temperature-minimum region, by *ambipolar diffusion* too. For a wavelength λ, the time-scale for *ohmic dissipation* is simply $\tau_d = \lambda^2/\eta$, where η is the magnetic diffusivity, and so the distance that a wave can travel at the Alfvén speed, say, before it dissipates (the so-called *damping length*) is

$$L_d = v_A \tau_d = \frac{v_A \lambda^2}{\eta}.$$

In terms of the wave frequency (ω), this may be written $L_d = v_A^3/(\eta \omega^2)$, from which it follows that dissipation is smallest for waves at the lowest frequency and in regions of the highest magnetic field strength (and so largest Alfvén speed). Osterbrock found that in the *quiet Sun* a *weak field* of 2 G would imply strong damping of Alfvén waves or slow-mode waves (with a frequency $\omega = 1.2 \times 10^{-2}$ Hz) at the low chromosphere. However, the fast-mode waves suffer negligible attenuation, since they are basically acoustic in the photosphere and therefore subject to only viscous dissipation, which is less effective than ohmic dissipation. In *plage regions*, where the field strength exceeds 50 G, Alfvén waves are able to propagate up through the chromosphere with negligible damping and so contribute to the heating above, provided an efficient dissipation mechanism exists there.

There are two effects which make *fast magnetoacoustic waves steepen into shock waves more slowly* than sound waves. When $v_A \gg c_s$, the distance that a fast wave needs to travel before shocking in a uniform medium is $d = \tau v_A^2/(4v_1)$, by analogy with Equation (6.5), and so it is larger than for sound waves. Furthermore, in a stratified isothermal atmosphere the density falls off as $\rho(z) \sim e^{-z/\Lambda}$, and the wave energy flux is proportional to $(\frac{1}{2}\rho v_1^2)v_A$, where v_A is the group velocity. Since this flux remains constant as the wave propagates up and the wave speed (v_A) is proportional to $\rho^{-1/2}$, the wave velocity amplitude increases as $v_1(z) \sim \rho^{-1/4} \sim e^{z/(4\Lambda)}$, which represents a much slower rate than that given by Equation (6.6) for sound waves.

Osterbrock extended the Brinkley–Kirkwood analysis for weak acoustic-shock dissipation to fast-mode shocks in a straight-forward manner. One effect of the

magnetic field is to introduce extra terms into the expression (6.7) for the energy flux, due to the additional work done by the magnetic tension and pressure. The resulting *damping length* for fast shocks is just $d \approx v_f t_0/\bar{\eta}$, in which the sound speed (c_s) in Equation (6.10) has been replaced by the fast-mode speed (v_f). For a magnetic field of 2 G, Osterbrock found the dissipation peaks at heights of 1000 to 2000 km and is sufficient to account for heating of the low chromosphere. By the time the corona is reached, however, the calculated direct fast-mode flux is far too small to provide the necessary dissipation, and so he suggested that heating of the upper chromosphere may be by slow-mode waves that are generated from the interaction of the fast-mode shocks.

The heating of a coronal loop by fast magnetoacoustic waves of high frequency ($\approx 3\,\mathrm{s}^{-1}$) has been suggested by Habbal *et al.* (1979). Such waves propagate across the magnetic field and are refracted into regions of low Alfvén speed, where they suffer significant collisionless damping if β is larger than about 0.1. The required wave flux for such (as yet unobserved) waves is about $100\,\mathrm{W\,m}^{-2}$ at the coronal base. Provided that the wavelength is much smaller than the length-scale for Alfvén speed variations, the ray paths ($\mathbf{r}(t)$) satisfy

$$\frac{d\mathbf{r}}{dt} = \frac{\partial\omega}{\partial\mathbf{k}}, \qquad \frac{d\mathbf{k}}{dt} = -\frac{\partial\omega}{\partial\mathbf{r}},$$

where the dispersion relation is $\omega = kv_A$ when $\beta \ll 1$. The authors extend Osterbrock's analysis for a plane-parallel atmosphere by considering propagation in a dipole magnetic field that contains isothermal plasma in hydrostatic equilibrium. Sample results are presented for a large loop that reaches a height of half a solar radius and contains plasma at about twice the ambient density. If β is large enough in the loop, the required dissipation occurs in a loop-shaped region coming up from the magnetic loop foot-points and passing below its summit. However, the effect of gravity is to make these waves evanescent, and so the coronal amplitudes are significant only for large horizontal wavelengths; furthermore, for such long waves a non-local analysis is required. Zweibel (1980) has evaluated the damping of fast waves in more detail and considered the energy balance between radiation and heating alone; she finds the equilibrium is unstable thermally (Section 7.5.7) to the formation of cool filaments parallel to the field. Furthermore, Hollweg (1981b) has pointed out two difficulties with fast modes: they are efficiently reflected at the transition region and so may never reach the corona; also, for values of ω and horizontal wavenumber corresponding to observed photospheric motions, fast waves should be evanescent.

6.4.2. NONLINEAR COUPLING OF ALFVÉN WAVES

The corona may possibly be heated by *Alfvén waves*. Osterbrock considered such low magnetic field strengths that the Alfvén waves would be strongly damped in the photosphere and so would need to be generated higher up by interactions of magnetoacoustic waves. However, it is now thought likely that a large flux of Alfvén waves is generated at *supergranulation boundaries*, where the presence of *intense kilogauss fields* allows them to penetrate the photosphere. They then propagate way

up into the corona with hardly any attenuation at all. In fact, the damping of Alfvén waves for fields larger than 10 to 100 G is so small that the problem is to explain how they give up their energy before propagating away, either into the solar wind along open field lines or back down to the photosphere along closed fields.

An Alfvén wave is likely to dissipate in practice because of its *nonlinear interaction* with either the non-uniform ambient field or another Alfvén wave (Wentzel, 1974). The method used to analyse the *effect of magnetic field inhomogeneity* depends on the wave-period $\tau = \lambda/v_A$. For a magnetic field of 10 G and a density of $10^{16}\,\mathrm{m}^{-3}$, characteristic of the active-region corona, the Alfvén speed (v_A) is about $300\,\mathrm{km\,s}^{-1}$ according to Equation (2.48b). Wavelengths (λ) of 100 000 km, comparable with the coronal scale-height, then correspond to wave-periods of 5 min. For periods much less than 5 min, the wavelength is much smaller than a scale-height and so the Alfvén waves propagate in a 'slowly varying' medium. As they propagate around a bend in the magnetic field much of their energy is converted to fast-mode waves (provided the waves are not well collimated along the field). They in turn decay rapidly and would *brighten the bends in a magnetic configuration* such as a coronal loop or the region above a supergranule cell. For wave periods of order or greater than 5 min, there are large variations in the ambient medium over a wavelength and the 'slowly-varying' approximation fails. Significant dissipation is still expected, but the details have not yet been worked out.

The *nonlinear interaction of magnetohydrodynamic waves* has been treated in detail by Kaburaki and Uchida (1971), Chiu and Wentzel (1972), Uchida and Kaburaki (1974). When the magnetic field is so weak that $v_A < c_s$, it is found that two Alfvén waves travelling in opposite directions along a magnetic field line can couple nonlinearly to give an acoustic wave, which in turn dissipates relatively quickly. Suppose that the frequency (ω) and wavenumber (k) of the two Alfvén waves and the acoustic wave are denoted by subscripts 0, 1, 2, respectively, so that

$$\omega_0 = v_A k_0, \qquad \omega_1 = v_A k_1, \qquad \omega_2 = c_s k_2, \tag{6.14}$$

where all the frequencies and wavenumbers are assumed positive. If a coupling of the two incident waves is to occur, the resulting acoustic wave must possess a frequency

$$\omega_2 = \omega_0 + \omega_1 \tag{6.15}$$

and wavenumber

$$k_2 = k_0 - k_1, \tag{6.16}$$

the minus sign resulting from the fact that the two Alfvén waves are propagating in opposite directions. After substituting for the wavenumbers from Equation (6.14) and eliminating ω_2 between Equations (6.15) and (6.16), we find that the two Alfvén waves can interact in this way only if their frequencies are in the ratio $\omega_1/\omega_0 = (c_s - v_A)/(c_s + v_A)$. Furthermore, the resulting acoustic frequency is $\omega_2 = 2\omega_0 c_s/(c_s + v_A)$.

In regions of *strong magnetic field* such that $v_A > c_s$, one Alfvén wave (ω_0, k_0) can decay into another Alfvén wave (ω_1, k_1) travelling in the opposite direction together with a sound wave (ω_2, k_2) travelling in the same direction. The interaction takes place

provided the selection rules $\omega_1 + \omega_2 = \omega_0$, $-k_1 + k_2 = k_0$, are obeyed, giving Alfvén and acoustic frequencies of

$$\omega_1 = \omega_0 \frac{v_A - c_s}{v_A + c_s}, \qquad \omega_2 = \frac{2\omega_0 c_s}{v_A + c_s}.$$

The resulting Alfvén wave has a frequency smaller than the original one and it can in turn decay to another lower-frequency Alfvén wave plus an acoustic wave. The cascade continues until all the Alfvénic energy has been converted to acoustic waves that dissipate rapidly.

Wentzel (1974, 1977) has calculated the rates at which the waves interact and has estimated the heating. He finds a significant production of acoustic waves from Alfvén waves in regions where v_A/c_s lies between 1/30 and 30. Dissipation occurs over distances comparable with coronal loop lengths, and it is strongest for wave periods of about a minute and at locations where the magnetic field is greatest. For an Alfvén wave of velocity amplitude v_1, the wave energy flux is given by

$$F = \tfrac{1}{2}\rho v_1^2 v_A, \tag{6.17}$$

where ρ is the ambient plasma density. Furthermore, the dissipation length for waves of wavenumber k generating a density perturbation ρ_1 can be written in order of magnitude as $d = \rho/(k\rho_1)$. For sound waves $\rho_1/\rho \approx v_1/c_s$, which gives Equation (6.4), but Alfvén waves produce much smaller density changes, namely $\rho_1/\rho \approx 2\pi(v_1/v_A)^2$. (The factor 2π is appropriate to the case when oppositely travelling Alfvén waves interact to give sound waves.) The dissipation length then becomes

$$d = \frac{\tau v_A}{(2\pi)^2}\left(\frac{v_A}{v_1}\right)^2, \tag{6.18}$$

in terms of the wave-period (τ).

Consider now whether Alfvén waves of period 10 s, say, may produce the required heating for *coronal loops*. For *interconnecting loops* or *quiet-region loops* let us adopt values for the necessary wave flux of 300 W m^{-2} and for the magnetic field of 12 G. Then Equations (6.17) and (6.18) imply a wave amplitude of 20 km s^{-1} and a dissipation length of 200 000 km, which is comparable with the half-length of such loops. For weak *active-region loops* a wave flux of 5000 W m^{-2} and a magnetic field of 20 G give a rather large wave amplitude of 60 km s^{-1} and a dissipation length of 110 000 km, again comparable with the half-length. We thus conclude that *short-period ($\approx$ 10 s) Alfvén waves provide a viable means of heating coronal loops outside or on the edge of active regions.* Deep within active regions, the magnetic field (B) is stronger, and so the dissipation length (proportional to B^4 for constant F) is too long. In this case longer-period waves or magnetic field dissipation may be the answer.

When equal fluxes of Alfvén waves propagate up the two legs of a loop, the heating due to their interaction is concentrated near the summit. Wentzel (1976, 1978) has extended the above discussion to include asymmetric heating due to unequal fluxes and also wave reflection at the transition region, which tends to equalise the fluxes along a coronal loop and to disorder the waves below the transition region.

Hollweg (1979, 1981a, b) has considered heating by Alfvén waves in some detail. Besides being tractable mathematically, the advantage of these waves is that they propagate so easily without becoming evanescent or being internally reflected. Also,

the energy propagates along the magnetic field, so heating the strongly magnetic regions, and they are observed to dominate solar wind fluctuations at 1 AU. Hollweg shows that the propagation of axisymmetric twists near the axis of a vertical flux tube obeys

$$\left(v_A^2 \frac{\partial^2}{\partial s^2} + \omega^2\right)(B_0^{1/2}\,\delta v) = 0, \qquad i\omega\,\delta B = B_0^{1/2}\frac{\partial}{\partial s}(B_0^{1/2}\,\delta v),$$

where ω is the frequency and s is the distance along the ambient field (B_0). He solves these equations for the perturbed velocity (δv) and magnetic field (δB) in a model atmosphere, in response to imposed photospheric motions. Two cases of wave propagation are considered, namely in the *open field* of a coronal hole and in the *closed field* of an *active-region* loop.

In the open region, *long-period* waves (with a period (τ) in excess of 10 min) possess an energy flux of typically 10 W m^{-2}, and they may drive the solar wind (Sections 12.3 and 12.4.2). *Short-period* waves (10s $< \tau <$ 5 min) have an energy flux of 10^3 to 10^4 W m^{-2}. They may drive spicules, and their energy is enough to account for chromospheric and coronal heating. The problem used to be how to damp them, but now there appear to be several viable dissipation mechanisms, although their details still need to be worked out. *Joule damping* is important in the middle chromosphere at high frequencies (Section 6.4.1). *Nonlinear damping* may occur by local velocity shears inducing *Kelvin-Helmholtz instability* (Section 7.5.4) or by local magnetic shears driving *tearing modes* (Section 7.5.5). *Nonlinear coupling* to fast and slow modes (which damp efficiently) takes place especially in the chromosphere, because the wave pressure is large (unlike the photosphere) and the waves are nonlinear (unlike the corona). Finally, linear *geometric wave coupling* turns Alfvén waves into fast waves (in the low-β corona), as they refract or propagate around curved field lines. Solutions to the *nonlinear* equations for propagation up into a realistic atmosphere by Hollweg *et al.* (1982) show that Alfvén waves can steepen into fast shocks in the chromosphere, provided their periods are smaller than a few minutes and the photospheric velocity amplitudes are of order 1 km s^{-1} (or greater). They suggest that such waves can drive upward flows (spicules) and can heat the upper chromosphere and corona.

In a closed loop, *resonant frequencies* appear at multiples of $v_A/(2L)$, where L is the coronal length of the loop. For example, a short loop with $L = 20\,000$ km, $B = 100$ G and $n = 10^{16}$ m^{-3} has resonant periods of 20 s, 10 s, 7 s .... The resonances occur because of reflections off the transition regions at the ends of the loop. They act like windows, which allow a large energy flux (typically 1.5×10^4 W m^{-2}) to pass unimpeded up into the corona, rather than being reflected off the steep Alfvén-speed gradient. Leroy (1980) has used methods from optics to analyse the reflection of Alfvén waves propagating up a vertical magnetic field ($B_0\hat{z}$) containing an isothermal plasma, for which the perturbed field (B_{1x}) satisfies $\partial^2 B_{1x}/\partial t^2 = (B_0^2/\mu)\partial/\partial z (\rho_0(z)^{-1}\,\partial B_{1x}/\partial z)$. He finds that in a 1 G field, waves with periods of less than an hour can reach the corona unreflected, but in a 3000 G field only those with periods lower than 1 s can propagate to the corona. This suggests that only short-period waves are able to make use of intense flux tubes to reach the corona.

6.4.3. Resonant absorption of Alfvén waves

As pointed out in Section 4.10.1, when the ambient medium is *nonuniform* a continuous spectrum of Alfvén waves may exist. The *resonant absorption* of such waves at singular surfaces in the plasma has been suggested as a means of heating in laboratory plasma devices and has been analysed by Grossman and Tataronis (1973) and Hasegawa and Chen (1974). It may also provide a mechanism for absorbing energy in the solar corona.

Consider a *force-free flux tube* with magnetic field components $\mathbf{B}_0 = (0, B_{0\phi}(R), B_{0z}(R))$, and suppose there is a wave-like disturbance of the form $f_1(R, \phi, z, t) = f_1(R) \exp{(i(\omega t - m\phi - kz))}$. Then, by putting $p_0 \equiv 0$ in Equation (4.64), we can see that the radial velocity perturbation $(v_{1R}(R))$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}R}\left(\frac{(\rho_0\omega^2 - (\mathbf{k}\cdot\mathbf{B}_0)^2/\mu)B^2/\mu}{(\rho_0\omega^2 - (m^2/R^2 + k^2)B^2/\mu)R}\frac{\mathrm{d}}{\mathrm{d}R}(Rv_{1R})\right) + Fv_{1R} = 0,$$

where $F(R)$ is given by Equation (4.65) and $\mathbf{k}\cdot\mathbf{B}_0 = kB_{0z} + mR^{-1}B_{0\phi}$.

When $m^2/R^2 \gg \rho_0\omega^2 - k^2B^2/\mu$ this simplifies to the equation

$$\frac{\mathrm{d}}{\mathrm{d}R}\left((\rho_0\omega^2 - (\mathbf{k}\cdot\mathbf{B}_0)^2/\mu)R\frac{\mathrm{d}}{\mathrm{d}R}(Rv_{1R})\right) + m^2Fv_{1R} = 0$$

for Alfvén waves alone.

Consider also a *unidirectional field* and a plasma pressure which vary with x

$$\mathbf{B}_0 = B_0(x)\hat{\mathbf{z}}, \qquad p_0 = p_0(x),$$

and suppose the disturbances behave like

$$f_1(x, y, z, t) = f_1(x)\exp{(i(\omega t - k_y y - k_z z))}.$$

Then, according to Equation (4.60), the perturbation equation for the transverse velocity component (v_{1x}) in the limit when k_y is large enough becomes,

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\varepsilon(x)\frac{\mathrm{d}v_{1x}}{\mathrm{d}x}\right) - k_y^2\varepsilon(x)v_{1x} = 0,$$

for Alfvén waves alone, where $\varepsilon(x) = \rho_0(x)\omega^2 - k_z^2 B_0(x)^2/\mu$.

Now, suppose that the footpoints of either the force-free flux tube or the unidirectional field are forced to vibrate at a given frequency (ω). In general, the local Alfvén frequency $\omega_A = \mathbf{k}\cdot\mathbf{B}_0/(\mu\rho_0)^{1/2}$ is not uniform but varies with R (or x). If there exists a radius R^* (or distance x^*) at which $\omega_A = \omega$, the coefficient of the second-order term in the above differential equations vanishes; the equations will therefore be singular at that point and so a *singular surface* (or *resonant absorption sheath*) will form there. The radial velocity component v_{1R} (or v_{1x}) possesses a logarithmic singularity, while $v_{1\phi}$ (or v_{1y}) has a hyperbolic singularity and the plasma energy

becomes infinite. However, analytic continuation of v_{1R} (or v_{1x}) through the singularity shows that *energy is continuously accumulating at the singular surface, and so it may heat the plasma there*. As the energy builds up, so the width of the surface decreases until a steady state is reached where the energy flux is dissipated ohmically (or viscously).

The thickness of the singular layer is typically 10 ion-Larmor radii, but heating may possibly occur over a much larger region if the Alfvén wave is first converted linearly to a *kinetic Alfvén wave*. It has a dispersion relation

$$\omega^2 = k_\parallel^2 v_A^2(1 + k_\perp^2 \rho_i^2),$$

where $k_\parallel$ and $k_\perp$ are wavenumbers parallel and perpendicular to the magnetic field and ρ_i is the ion Larmor radius. Dissipation of the wave is by kinetic effects such as Landau damping. The above ideas have been examined by Ionson (1978) as a mechanism for heating *coronal loops*. He suggests that 5-min chromospheric oscillations shake the footpoints of a loop and cause *Alfvénic surface waves* (Section 4.10.1) to propagate upwards along the surface of the loop. They in turn couple to the kinetic Alfvén waves, which dissipate in an extremely thin sheath only 1 km thick.

Resonant absorption of more general magnetoacoustic waves is described by Equation (4.59), which possesses resonance points where $\omega_A(x) = \omega$ or $\omega_T(x) = \omega$ and cut-off points where $\omega_+(x) = \omega$ or $\omega_-(x) = \omega$. At the resonance points, v_{1x} possesses a logarithmic singularity and energy is resonantly absorbed as short wavelength oscillations build up until they are limited by viscous or ohmic dissipation. At the cut-off points, the wave becomes evanescent and most of the energy is reflected while some can tunnel through. Thus, as a magnetoacoustic wave propagates in the solar atmosphere with a certain frequency ω, so the values of $\omega_T, \omega_-, \omega_A, \omega_+$ will vary and its nature will change. When $\omega_+ < \omega$ or $\omega_T < \omega < \omega_-$ it propagates, but when $\omega_- < \omega < \omega_+$ or $\omega < \omega_T$ it is evanescent. The wave is evanescent both sides of the Alfvén resonance, but it changes from being propagating to evanescent at the cusp resonance, which is therefore likely to have more energy being fed into it.

6.4.4. Magnetic field dissipation

When photospheric motions are *sufficiently slow* and the wavelength *sufficiently long* that the conditions (6.12) and (6.13) hold, a wave description ceases to be helpful. Instead the coronal magnetic configuration *evolves passively through a series of equilibria*, which store energy in excess of potential. This energy has come originally from the photospheric motion. The electric currents associated with such large-scale equilibria produce ohmic heating, which is, however, completely negligible, since the coronal conductivity (σ) is so large. This is true even when allowance is made for the fact that the corona is probably in a permanent state of weak turbulence, with an anomalous electrical conductivity (σ^*) that is a factor of a hundred or so lower than the classical value (Benz, private communication). The only way that magnetic field (i.e., ohmic) dissipation can produce the necessary coronal heating is for the magnetic field changes and accompanying electric currents to be concentrated in
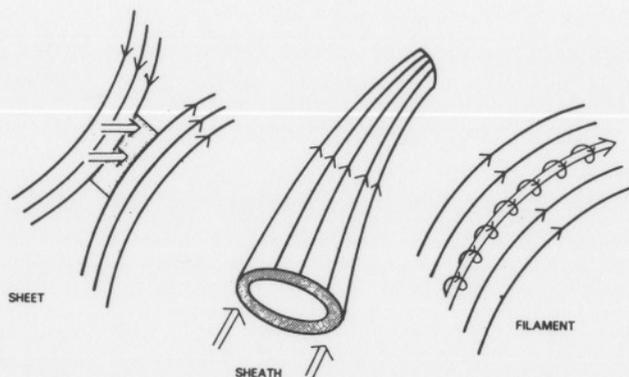
Fig. 6.4.   Three geometries for a current concentration in which enhanced magnetic field dissipation may occur. Light arrows label magnetic field lines, whereas large arrows indicate electric current directions.

extremely intense *current sheets*, *current sheaths* (around flux tubes) or *current filaments* (Figure 6.4). If the current density is so strong that the width of such a current concentration is less than (typically) a few metres, the dissipation may be considerably enhanced by the presence of plasma turbulence (Section 2.1.5).

Provided current sheets, sheaths or filaments can be formed, they produce a rapid conversion of magnetic energy into heat (by ohmic dissipation), bulk kinetic energy and fast-particle energy, in a manner that has been studied extensively in connection with the more violent heating of a solar flare (Section 10.1). This suggests that, especially in the strong magnetic field of an active region, the corona is in a state of *ceaseless activity* and is being heated by many tiny micro-flarings that are continually generated by the photospheric motion below. The coronal loops that stand out in soft X-ray pictures are those in which most heat is being released and then conducted efficiently along the magnetic field.

The features of heating by *magnetic* (or *current*) *dissipation* that need to be understood concern the *way in which current sheets, sheaths or filaments are formed, are maintained* (if necessary) *and decay*. The order-of-magnitude estimates of Tucker (1973) and others are described below, to determine how thin the resistive regions need to be to provide the necessary heating. A discussion of the formation of current sheets or filaments is also given, but no convincing explanation for sheaths has yet been put forward. Of the three proposed alternatives, *current sheets* have received by far the most attention; they may be formed either by pushing topologically distinct regions against one another or by magnetic non-equilibrium. In the former case, they are maintained for as long as the external footpoint motion continues. Current filaments may be created as a result of *tearing-mode instability* (Section 7.5.5) or *thermal instability* (Section 6.4.4C).

6.4.4A. Order of Magnitude

Tucker (1973) and Levine (1974) were among the first to suggest coronal heating by the dissipation of non-potential magnetic fields. They considered neutral current sheets dispersed throughout active regions, and they established qualitatively that

current dissipation could provide enough heat for the corona. Tucker supposes that magnetic energy is being stored at a rate

$$\frac{dW_m}{dt} \approx \frac{vB^2}{2\mu} L \qquad (6.19)$$

by photospheric motions (v) that twist a magnetic field of strength B over an area L^2. The energy is at the same time being dissipated ohmically, at a rate

$$D \approx \frac{j^2}{\sigma} L^3 \qquad (6.20)$$

for currents (j) distributed uniformly through the volume (L^3) of the active region. If the magnetic field is being twisted up faster than it is relaxing ohmically, the excess energy will be stored until it is released as, for instance, a *solar flare*. But, if the two rates (6.19) and (6.20) are equal, the active region will maintain a steady state. The effective twisting speed (v) that is needed to provide a heat input of, say, 3000 W m^{-2} to the corona can be found from Equation (6.19) as $v \approx 100$ m s^{-1} for a photospheric field strength of 100 G. Furthermore, uniform dissipation throughout the active region with a classical Coulomb electrical conductivity (σ) requires a current density that can be estimated by equating DL^{-3} from Equation (6.20) to 3000 W m^{-2}. With $L \approx 10\ 000$ km, Tucker finds $j \approx 30$ A m^{-2}. Since this corresponds to the rather large magnetic field gradient of 0.4 G m^{-1}, he suggests that the dissipation is concentrated at thin current sheets rather than distributed uniformly. The ohmic dissipation inside sheets may be greatly enhanced above normal because of the much larger electric currents and the possibility of plasma turbulence, but the sheets occupy only a small fraction of the active-region volume. For each sheet of thickness l^* and area L^{*2} with an electric current $j^* \approx B/(\mu l^*)$ and a turbulent electrical conductivity (σ^*), the rate of heat generation is

$$D^* \approx \frac{j^{*2}}{\sigma^*} L^{*2} l^* \quad \text{or} \quad D^* \approx \frac{B^2}{\mu^2 \sigma^*} \frac{L^{*2}}{l^*}.$$

Tucker adopts a turbulent conductivity that is about a million times smaller than the classical value and assumes a sheet width of 10 m, consistent with the critical current for turbulence onset. He finds that only a few current sheets of length $L^* = 1000$ km are necessary to generate the heat that is required for an active region. Levine (1974) suggested that the tangled nature of coronal magnetic fields produces many small current sheets that are collapsing. During the collapse, particles are accelerated and then thermalised by Coulomb collisions in the surrounding region.

Rosner *et al.* (1978) support Tucker's ideas for direct coronal heating by magnetic dissipation. They point out that the observed intensity of active regions in X-rays appears to be directly related to the level of photospheric magnetic activity. Early in the life of an active region, the magnetic field is complex and the corona bright, whereas one rotation later the region often possesses a more dispersed field with a coronal plasma whose pressure is an order of magnitude lower. Instead of heating in current sheets, Rosner *et al.* suggest that the heating is concentrated in current *sheaths*, which are thin, annular regions near the edges of coronal flux tubes or loops. The loops are stressed by twisting motions at the footpoints, and, in a steady state, the work done

by photospheric motions balances the coronal dissipation. Once the heat has been released, it is conducted *along* the magnetic field very efficiently (and so produces loop-like structures), but the distance that it is transported *across* the field depends on the nature of the small-scale instabilities that may be present. Golub *et al.* (1980) discussed the problem further and suggested that heating due to twisting motions (v_ϕ) in a loop is $E_H \sim B_\phi B_z v_\phi / L$ per unit volume. Equating this to Equation (6.40) gives a scaling law $p \sim B^{12/7}$, which relates the loop pressure and field strength and agrees reasonably with observations.

The main uncertainty with the model concerns the way in which the necessary strong currents are created and maintained. In order to provide sufficient heating, the current sheath in the loop needs to be typically 50 m thick and the sheath plasma must be turbulent, with an electrical conductivity smaller than normal by a factor of 10^4. However, (apart from flaring situations (Chapter 10)) it seems unlikely that large enough current densities can be produced in a loop to create and maintain a strongly turbulent plasma. A more likely alternative is magnetic dissipation in the loop by means of *resistive instabilities* (Section 7.5.5), such as the collisionless (or collisional) tearing mode, which occurs for much smaller current densities.

The possibility of magnetic dissipation was later considered in more detail. Sakurai and Levine (1981) analysed the generation of force-free fields and the storage of magnetic energy in the corona due to small photospheric motions at the footpoints of both a uniform and a bipolar configuration. Sturrock and Uchida (1981) estimated the rate of increase of magnetic energy due to a *stochastic* motion of the photospheric footpoints of a coronal loop. They obtained scaling laws for a loop's temperature in terms of its length and magnetic field strength. Also, Ionson (1981, preprint) set up an interesting LRC circuit analogue for coronal loops.

6.4.4B. Current Sheets

Current sheets may be formed in several ways. One is by the *interaction of topologically separate parts* of the magnetic configuration of, say, an active region. High-resolution observations of the *photospheric* magnetic field (e.g., Figure 1.10) exhibit a highly complex magnetic pattern with frequent changes of polarity. The *coronal* field is also complex, with many distinct magnetic flux tubes shown up by X-ray and EUV pictures (e.g., Figure 1.15). As the photospheric footpoints of coronal loops move, so the neighbouring coronal flux-tubes will respond and interact with one another, either moving further apart or coming closer together. At the interface between the two tubes, a *current sheet* is formed, the magnetic field reconnects, and magnetic energy is released in the process. Such magnetic dissipation takes place not only when neighbouring magnetic field lines are oppositely directed, as in Figure 6.5, but also when the field lines are inclined at a non-zero angle (Priest and Sonnerup, 1975).

The formation of current sheets when *new magnetic flux is emerging* from below the photosphere has been studied in connection with solar flares by several authors (see Section 10.2.1), but the same calculations are applicable when *magnetic flux is evolving* rather than emerging. In particular, it must be stressed that the current sheet is a response to the applied photospheric motions. If the neighbouring footpoints move relative to one another at a certain speed, then the *corona will just respond*
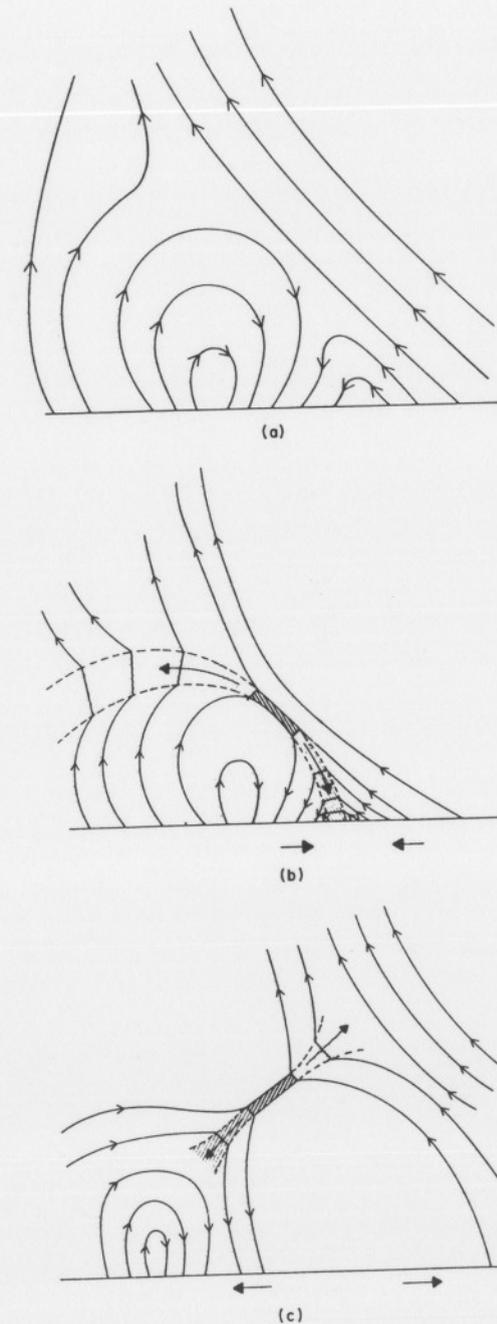
Fig. 6.5. Magnetic dissipation due to the relative motion of: (a) two neighbouring flux tubes when they (b) approach one another or (c) move further apart.

by creating a current sheet and allowing *magnetic reconnection* at that speed (provided the speed is less than a maximum value (Section 10.1)). Furthermore, the reconnection and associated dissipation is maintained as long as the relative footpoint-motion continues, with the dimensions of the current sheet depending on the magnetic field strength and photospheric speed. Conditions inside the sheet will only be turbulent if the resulting sheet width is small enough. Indeed, it is when the width becomes less than a critical value that a subflare or flare is triggered, according to the *emerging flux mechanism* (Section 10.2.1). It should also be noted that slow magneto-acoustic shock waves radiate from the ends of the current sheet and that fine jets of plasma are emitted between pairs of shocks. As plasma comes in slowly from the sides, the *bulk of the heat is released at these shock waves* rather than in the central current sheet itself.

Current sheets may also develop when magnetostatic equilibrium becomes *unstable* or even ceases to exist, a situation known as *non-equilibrium*. In a simple bipolar magnetic field when the photospheric footpoints move slowly, the low-β (Section 2.5) corona responds by establishing a series of *force-free* configurations (Section 3.5). In general, however, the coronal magnetic field is much more complex than this, and it contains topologically distinct flux systems. Parker (1972) and Syrovatsky (1978) have demonstrated that, as the footpoints of such a magnetic field move, the *corona cannot adjust to a new force-free equilibrium* and current sheets are formed instead. These current sheets are themselves not in equilibrium, since they allow a rapid reconnection at some fraction of the Alfvén speed (Section 10.1), and the magnetic configuration reduces to the state of lowest potential energy. Parker referred to such a process as *topological dissipation*. Continual footpoint motion means that the coronal field is all the time responding by reconnecting and so converting magnetic energy into heat.

Parker (1972) establishes that, *if the pattern of small-scale variations is not uniform* along a large-scale field, then the field *cannot be in magnetostatic equilibrium*. In other words, equilibrium exists only if the field variations consist of a simple twist extending from one footpoint to another. More complex topologies (such as braided flux tubes with several field lines wrapped around each other) are not in equilibrium. To obtain his result, Parker considers a uniform configuration, having a plasma pressure p_0 and magnetic field $B_0\hat{z}$, with the footpoints anchored at the planes $z = \pm L$ (Figure 6.6). Suppose that a displacement of the footpoints by at most $\lambda (\ll L)$ leads to small deviations ($\mathbf{B}_1$ and p_1) from the uniform field and pressure, respectively, so that $\mathbf{B} = B_0\hat{z} + \mathbf{B}_1(x, y, z)$, $p = p_0 + p_1(x, y, z)$, where $B_1/B_0 \approx p_1/p_0 \ll 1$. Then the equation $-\nabla p + (\nabla \times \mathbf{B}) \times \mathbf{B}/\mu = 0$ for magnetostatic equilibrium (Section 3.1) becomes, to first order,

$$-\nabla\left(p_1 + \frac{B_0 B_{1z}}{\mu}\right) + \frac{B_0}{\mu}\frac{\partial \mathbf{B}_1}{\partial z} = 0. \tag{6.21}$$

But the basic equation $\nabla \cdot \mathbf{B} = 0$ gives $\nabla \cdot \mathbf{B}_1 = 0$, and so the divergence of Equation (6.21) implies

$$\nabla^2\left(p_1 + \frac{B_0 B_{1z}}{\mu}\right) = 0. \tag{6.22}$$

Now, variations of $(p_1 + B_0 B_{1z}/\mu)$ extend at most a distance of order $\lambda (\ll L)$ into
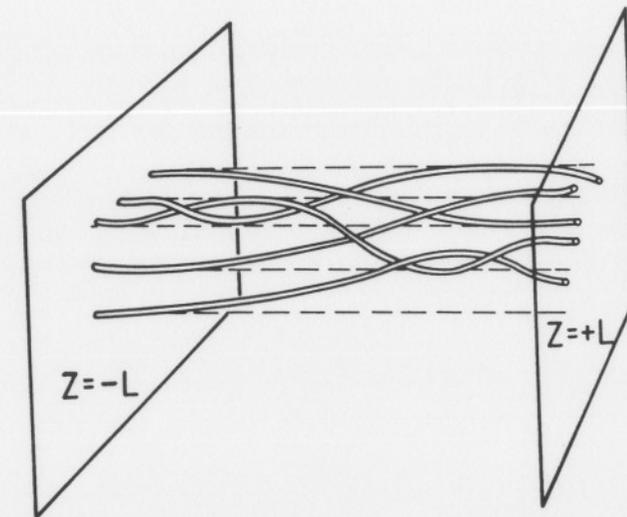
Fig. 6.6. Schematic drawing of the topology of magnetic tubes of force following a displacement of the ends of the tubes where they intersect $z = +L$ (from Parker, 1972).

the region $-L < z < L$ from the end-planes ($z = \pm L$), so the solution to Laplace's equation (6.22) is

$$p_1 + \frac{B_0 B_{1z}}{\mu} = \text{constant}, \tag{6.23}$$

save in boundary layers of width λ near $z = \pm L$. However, the z-component of Equation (6.21) is $\partial p_1/\partial z = 0$, and hence the z-derivative of Equation (6.23) implies $\partial B_{1z}/\partial z = 0$. Thus, if a field is in magnetostatic equilibrium, its pattern does not vary along the general direction of the field, which establishes Parker's result. Sakurai and Levine (1981) have since established that the perturbed field *is* determined uniquely by a small motion of the footpoints*. Nevertheless, the detailed consequences of Parker's result remain to be seen when finite-amplitude displacements at the boundary are taken into account. Later, Parker (1981) suggested that large displacements of the feet of a flux tube could cause the tube to become dislocated from its initially neighbouring field. In its new quasi-equilibrium position the tube would become flattened and eventually dissipated.

Green (1965) and Syrovatsky (1971) demonstrated that the slow, continuous deformation of a two-dimensional potential magnetic field containing *neutral points* leads to the production of neutral current sheets in the *perfectly conducting* limit. Consider the initial magnetic field $B_x = y$, $B_y = x$, *which may be written in* terms of the complex variable $z = x + iy$ as

$$B_y + iB_x = z, \tag{6.24}$$

and contains a neutral point at the origin (Figure 6.7). Next, suppose that this configuration is deformed by the imposition of a uniform electric field (perpendicular

* All the variations occur in the boundary layers near $\pm L$, and so Figure 6.6 is misleading.

to the z-plane) that drives a motion normal to the magnetic field. If the plasma is regarded as perfectly conducting, the field lines are carried with the motion and Ohm's Law may be written

$$E + vB = 0. \tag{6.25}$$

It is clear that the instantaneous plasma speed is determined by Equation (6.25) everywhere except at the neutral point where B vanishes. Furthermore, suppose that the flow speed and plasma pressure are so small that the configuration passes through a series of equilibria with vanishing Lorentz force, so that in this two-dimensional situation

$$0 = jB. \tag{6.26}$$

The peculiar role of the neutral point is evident in both Equations (6.25) and (6.26), which imply that a continuous deformation of the original field (6.24) through a series of potential configurations with $j = 0$ is possible everywhere *save in the vicinity of the neutral point*. At the neutral point itself, Equation (6.26) allows non-zero currents. Indeed, a solution to the problem is that a current sheet develops, represented by a cut from $z = -iL$ to $z = +iL$, say, in the complex plane. The resulting magnetic field components are given by $B_y + iB_x = (z^2 + L^2)^{1/2}$, and the field is sketched in Figure 6.7. This complex variable technique for obtaining the position of the current sheet that forms as oppositely directed fields approach one another has been extended to the case of approaching bipolar fields that are equal (Priest and Raadu, 1975) or unequal (Tur and Priest, 1976). In the latter case the sheet is curved. All these calculations use the assumption that the plasma remains frozen to the field everywhere, but in practice this approximation fails inside the current sheet. During the approach of the two flux systems, the current sheet bifurcates into two slow magnetoacoustic shocks and reconnection occurs in the manner described in Section 10.1. When the footpoint motion ceases, the magnetic configuration rapidly reduces (over an Alfvén travel-time) to its lowest energy state, namely a potential field.

Syrovatsky (1978) generalised his previous result for potential fields to show that a
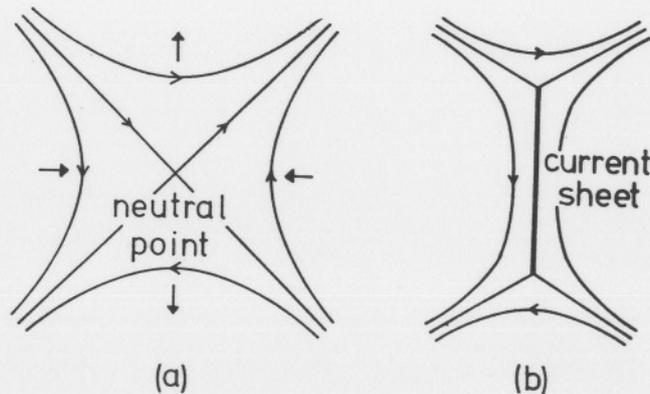
Fig. 6.7. (a) A potential magnetic field near an X-type neutral point. (b) The field produced by the slow motion indicated in (a) by solid-headed arrows. The plasma is assumed perfectly conducting.

continuous deformation of a *force-free* field in general leads to the production of *current sheets*. It is only for rather simple fields and simple footpoint motions that no current sheets are generated (see Section 3.5). Suppose a magnetic configuration evolves through a series of force-free states such that

$$(\nabla \times \mathbf{B}) \times \mathbf{B} = \mathbf{0}, \tag{6.27}$$

with $\nabla \cdot \mathbf{B} = 0$. Then, in general, this evolution cannot maintain the condition of frozen-in flux, since, given a time-sequence of magnetic fields satisfying Equation (6.27), the solution of the induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) \tag{6.28}$$

for $\mathbf{v} \times \mathbf{B}$ requires it to have (in general) a component parallel to $\mathbf{B}$. This is unacceptable since $\mathbf{v} \times \mathbf{B}$ must be normal to $\mathbf{B}$ by definition. The difficulty in solving Equation (6.28) for a given magnetic field shows up particularly clearly when a closed magnetic field line (C) exists. Simply integrate over a surface (S) bounded by such a C to give

$$\frac{\partial}{\partial t} \int_S \int \mathbf{B} \cdot \mathbf{dS} = \int_C (\mathbf{v} \times \mathbf{B}) \cdot \mathbf{ds}. \tag{6.29}$$

Then, for a given sequence of solutions to Equation (6.27), there is in general no reason to suppose that the flux through S should remain constant. This implies that the left-hand side of Equation (6.29) is non-zero, whereas the right-hand side must vanish (since $\mathbf{v} \times \mathbf{B}$ is normal to the field-line curve (C)).

More details of the formation and properties of curent sheets can be found in the reviews by Priest (1976, 1981b).

6.4.4C. Current Filaments

A magnetic configuration that is non-potential, such as a sheared force-free structure, may become unstable in several ways with the electric current concentrating into filaments. One mechanism is the tearing-mode instability, which is described in Section 7.5.5 (see also Galeev *et al.* (1981), who suggest nonlinear kinetic tearing as an effective heating mechanism for coronal loops), and another is thermal instability. Heyvaerts (1974) has described two such types of instability, namely the *Joule mode* and the *antidiffusion mode*. They both cause a uniform electric current to concentrate into many small *current threads* (or filaments) parallel to the magnetic field. The main conditions for the validity of his analysis are that the temperature be about 10^5 K, so that the radiative loss function ($Q(T)$) be approximately constant (Section 2.3.3), and that the perturbation wave-number (k) be both small enough that thermal conduction is swamped by Joule heating and also large enough that a local stability analysis is applicable.

For a disturbance propagating at an angle (θ) to the ambient magnetic field with $\gamma = \frac{5}{3}$, and an electrical conductivity (σ) proportional to $T^{3/2}$, the *Joule mode* has a growth-rate

$$\omega = (\sin^2 \theta - \cos^2 \theta) \frac{j_0^2}{\sigma p_0}, \tag{6.30}$$

where j_0 is the equilibrium current density and p_0 the equilibrium pressure. It occurs most effectively for $\theta = \frac{1}{2}\pi$, and so it forms fine structures aligned along the magnetic field and consisting of hotter current concentrations separated by cooler regions. The mode is, however, restricted to wavenumbers $k \gg (\omega\mu\sigma)^{1/2}$, such that the plasma is not frozen to the field, and another limitation in applicability is that the current density needs to be rather large for the growth-rate to be reasonably high. For example, suppose we adopt $T_0 = 10^5$ K, $N_0 = 10^{15}$ m^{-3}, together with a turbulent conductivity that is a factor of 100 smaller than the Coulomb value (2.17) and a current density $j_0 = 1$ A m^{-2} (corresponding to a magnetic field change of 100 G over 1000 km). Then (6.30) gives a growth-rate of only 4×10^{-4} s^{-1}.

The *antidiffusion mode* causes the magnetic flux to concentrate rather than diffuse, and it occurs only for θ close to $\frac{1}{2}\pi$. Its growth-rate is

$$\omega = \frac{k^2}{\mu\sigma(\theta - \frac{1}{2}\pi)^2}, \tag{6.31}$$

so that $\omega \gg k^2/(\mu\sigma)$ and the plasma is almost frozen to the field. A local increase in temperature enhances the electrical conductivity and current density, which in turn produces more heating and drives the instability.

6.5. Coronal Loops

The solar atmosphere, which has a vertical stratification produced by the force of gravity, is by no means uniform in the horizontal direction and possesses a complex
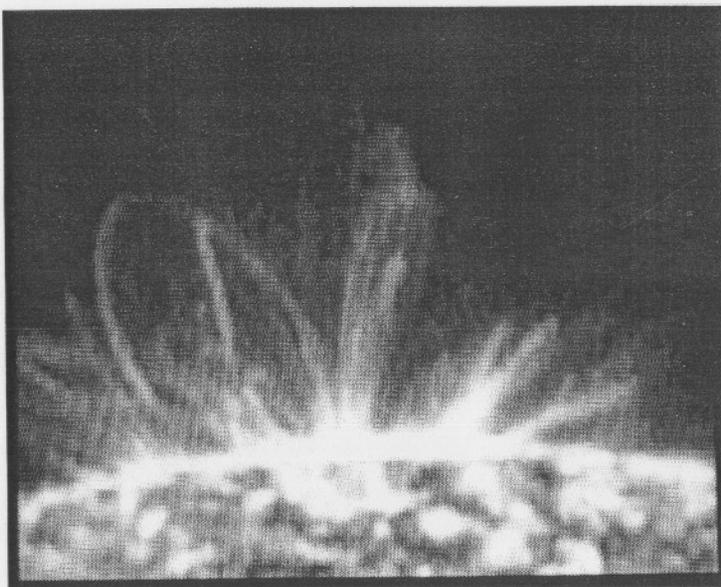
Fig. 6.8.   An active-region loop system in the EUV line of O VI (3×10^5 K) (courtesy R. Levine, Centre for Astrophysics, Harvard).

TABLE 6.2.
Typical length $2L(\times 1000$ km$)$, temperature $T(K)$ and density n(m^{-3}) for the different kinds of coronal loop

	Interconnecting	Quiet-region	Active-region	Post-flare	Simple-flare
$2L$	20–700	20–700	10–100	10–100	5–50
T	$2 - 3 \times 10^6$	1.8×10^6	$10^4 - 2.5 \times 10^6$	$10^4 - 4 \times 10^6$	$\lesssim 4 \times 10^7$
n	7×10^{14}	$0.2 - 1.0 \times 10^{15}$	$0.5 - 5.0 \times 10^{15}$	10^{17}	$\lesssim 10^{18}$

structure dominated by the magnetic field. X-ray and EUV observations such as those from Skylab have indicated that the corona (outside *coronal holes*) consists largely of loop structures that presumably outline the magnetic field. Five morphological types of loop are found, namely, *interconnecting loops*, *quiet-region loops*, *active-region loops* (an example being shown in Figure 6.8), '*post*'-*flare loops* and *simple (compact) flare loops*. General properties of their structure have been discussed, for instance, by Priest (1978) and Chiuderi *et al.* (1981), and their observational characteristics are summarised in Sections 1.3.4B and 1.4.1C and Table 6.2. We describe below a model for the temperature-density structure of coronal loops, together with some general comments about the types of flow that may be expected in them.

6.5.1. STATIC ENERGY-BALANCE MODELS

For a loop in *hydrostatic equilibrium* and in *thermal equilibrium* between conduction, radiation and heating, the temperature (T) and electron density (n_e) satisfy (see Equations (2.32) and (3.5))

$$\frac{1}{A}\frac{d}{ds}\left(\kappa_0 T^{5/2}\frac{dT}{ds}A\right) = \chi n_e^2 T^\alpha - H \tag{6.32}$$

and (for fully-ionised hydrogen)

$$\frac{1}{\cos\theta}\frac{dp}{ds} = -m_p n_e g, \tag{6.33}$$

where the pressure

$$p = 2n_e k_B T; \tag{6.34}$$

$A(s)$ is the cross-sectional area of the loop at a distance s along it from the base, and $\theta(s)$ is the inclination of the loop to the vertical (Figure 6.9). For given forms of $A(s)$ and $\theta(s)$, the Equations (6.32) and (6.33) are to be solved subject to: (i)

$$p = p_0, \qquad T = T_0, \tag{6.35}$$

at the base ($s = 0$) of the loop, and (ii) the symmetry condition

$$\frac{dT}{ds} = 0 \tag{6.36}$$

at the summit ($s = L$).
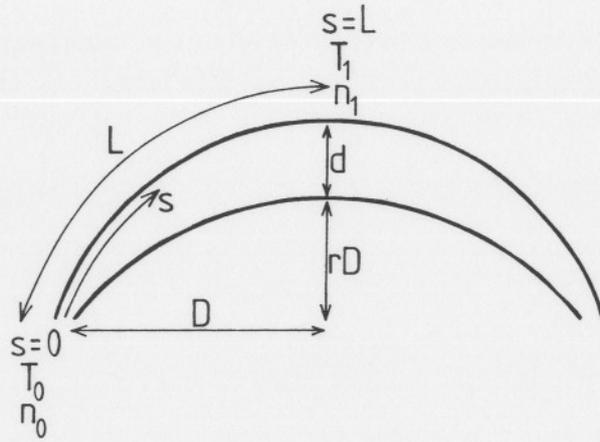
Fig. 6.9.   The notation for a symmetric coronal loop of length $2L$ with temperature T_0 and density n_0 at the footpoint ($s = 0$), and T_1 and n_1 at the summit ($s = L$). r is the ratio of loop height to half the base length (D), and d is the ratio of the diameter of the loop cross-section at the top to that at the footpoint.

With T_0 fixed, the temperature profiles and, in particular, the *summit temperatures* (T_1) are determined by three parameters, namely the loop length ($2L$), the base pressure (p_0) and the heating rate (H), so that $T_1 = T_1(L, p_0, H)$. The value of T_0 is fairly arbitrary, but, when a value near the temperature minimum or a chromospheric plateau is taken, we have a so-called *thermally isolated loop*, for which the above three boundary conditions must be supplemented by

$$\frac{dT}{ds} = 0 \quad \text{at } s = 0. \tag{6.37}$$

In this case, the summit temperature is a function of only two parameters: $T_1 = = T_1(L, H)$, since the base pressure cannot be freely prescribed as in Equation (6.35) but must be adjusted to a value $p_0 = p_0(L, H)$ determined by the remaining boundary conditions. Of course, in practice there may be some complicated feedback if H itself depends on p_0.

6.5.1A. Uniform Pressure Loops

For very *low-lying loops*, whose summits are much below a coronal scale-height of roughly 80 000 km (Equation 3.10), the loop pressure is uniform, and so one needs to solve just Equation (6.32), with n_e given by Equation (6.34) and p constant.

Consider first the *thermally isolated loops* with uniform cross-sectional areas (Rosner *et al.*, 1978). Their summit temperatures may be estimated in order of magnitude as follows, by using the fact that, whereas the relative sizes of C, R, H vary locally, their global (or integral) values are similar. Since the heating and radiation in Equation (6.32) are globally of the same order, the heating is roughly

$$H \approx \frac{p^2 \chi}{4 k_B^2} T_1^{-5/2}, \tag{6.38}$$

in terms of the summit temperature. (Here α is approximated by $-\frac{1}{2}$.) Furthermore, the thermal conduction term has a similar size to the radiation term, and so Equation (6.32) gives, in order of magnitude,

$$\frac{\kappa_0 T_1^{7/2}}{L^2} \approx \frac{p^2 \chi}{4 k_B^2} T_1^{-5/2},$$

or, after rearranging,

$$T_1 \sim (pL)^{1/3}. \tag{6.39}$$

The constant of proportionality, namely $(4 k_B^2 \kappa_0 / \chi)^{-1/6}$, is about 10 000 in m.k.s. units (when $\kappa_0 = 10^{-11}, \chi = 10^{-32}, k_B = 1.4 \times 10^{-23}$). Substitution of T_1 from Equation (6.39) in (6.38) then gives

$$H \sim p^{7/6} L^{-5/6} \tag{6.40}$$

for the mechanical heating. If one regards the loop heating as being prescribed, so that its pressure just responds to preserve equilibrium, Equations (6.39) and (6.40) may be rearranged to give

$$T_1 \sim H^{2/7} L^{4/7}, \qquad p \sim H^{6/7} L^{5/7}. \tag{6.41}$$

This implies that both *the temperature and pressure are increased by either stretching a thermally isolated loop or enhancing its heating.*

When $T_1 \gg T_0$, the above order-of-magnitude expressions (6.39), (6.40) for heating and summit temperature may be derived rigorously from Equation (6.32), subject to Equations (6.35) to (6.37), as follows. Multiply Equation (6.32) by $T^{5/2} \, dT/ds$ and integrate from the loop summit, so that

$$\frac{1}{2} \kappa_0 \left(T^{5/2} \frac{dT}{ds} \right)^2 = \int_{T_1}^{T} T^{5/2} \left(\frac{p^2 \chi}{4 k_B^2} T^{-5/2} - H \right) dT,$$

or

$$\frac{1}{2} \kappa_0 T^5 \left(\frac{dT}{ds} \right)^2 = \frac{p^2 \chi}{4 k_B^2} (T - T_1) - \frac{2H}{7} (T^{7/2} - T_1^{7/2}). \tag{6.42}$$

Now, since $T_1 \gg T_0$ and the conductive flux ($\kappa_\parallel \, dT/ds$) vanishes at the base, Equation (6.42) gives

$$H = 3.5 \frac{p^2 \chi}{4 k_B^2} T_1^{-5/2},$$

which is just Equation (6.38), apart from the factor 3.5. After substituting for H and taking the square-root, Equation (6.42) then reduces to

$$T^2 \frac{dT}{ds} = \left(\frac{p^2 \chi}{2 k_B^2 \kappa_0} \right)^{1/2} \cdot \left(1 - \frac{T^{5/2}}{T_1^{5/2}} \right)^{1/2},$$

which in turn integrates to give the loop length as

$$Lp = \left(\frac{2 k_B^2 \kappa_0}{\chi} \right)^{1/2} I T_1^3,$$

where $I = \int_0^1 t^2(1 - t^{5/2})^{-1/2}\,dt$ is approximately 0.72. Finally, taking the cube-root gives the same scaling law as Equation (6.39), but with a constant of proportionality larger than before by a factor of only $2^{1/6}I^{-1/3}$. Full numerical solutions of the equations of energy balance and hydrostatic equilibrium by Wragg and Priest (1982) show that the scaling law is accurate for short loops. However, for sufficiently long (or rare) loops, the temperature may be lower by a factor of two or so than the scaling law predicts, and too large an increase in loop length may actually cause the temperature to fall.

Now, in general, the conductive flux may not vanish at the temperature (T_0) that is chosen as the base of the loop. This means that the loop is not thermally isolated, and it leads not only to *more general expressions* for the summit temperature than Equation (6.39), with the heating rate as an extra parameter, but also to the possibility of *thermal non-equilibrium* and *instability*. These more general loop models are especially useful if the loop base is taken at, say, 5×10^5 K, in view of the difficulties and uncertainties in modelling the transition region. Starting at a value of, say, 2×10^4 K for T_0 (rather than, say, 10^6 K) may give significant errors if Equation (6.32) describes the transition-region physics inadequately. It would be invaluable to extend the models down to the temperature-minimum region where the conductive flux vanishes, but, at such low temperatures, radiative transfer effects would have to be included and these introduce a coupling to the region below.

Numerical solutions for uniform-pressure loops have been found by many authors, including Hood and Priest (1979a), who treat the case when T_0 is 10^6 K and the heating is proportional to density: $H = hn_e$. These solutions depend on the three parameters
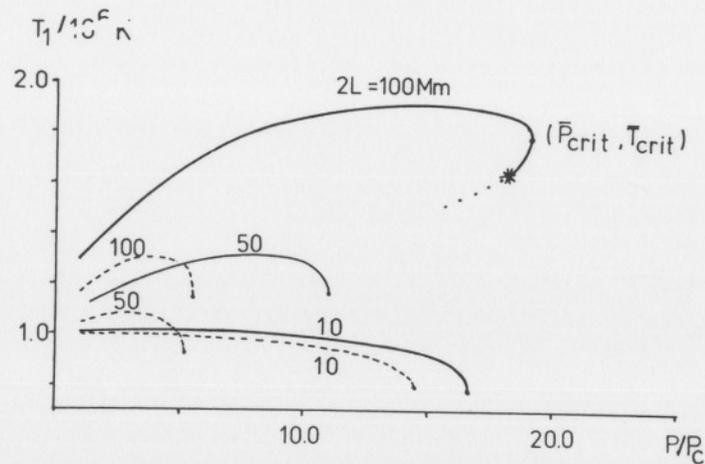
Fig. 6.10. The loop summit temperature (T_1) as a function of the pressure (p) and half-length (L) for a low-lying static coronal loop. p_c is the pressure for a standard plasma of density 5×10^{14} m^{-3} and temperature 10^6 K. The solid (or dashed) curves are for mechanical heating ten (or five) times larger than the radiation from the standard plasma. The curves end at critical conditions (p_{crit}, T_{crit}), indicated here for hot loops of length 100 000 km. The lower unstable solutions are also included for this loop; the star indicates a thermally isolated loop and the dots show oscillatory solutions (from Hood and Priest, 1979a).

L, p, h through just the two combinations Lp and h/p. (The base conductive flux can be regarded as an alternative parameter to p, and in practice there may be some limitation on the range of values of this flux which the atmosphere below T_0 can cope with statically.) Some typical numerical results are shown in Figure 6.10. They suggest explanations for many observed loops properties. For example, the fact that *shorter loops often appear brighter* may be because they possess consistently higher heating rates, or it may be caused by the higher pressures (p_{crit}) that are allowed for shorter loops. Furthermore, the relatively *small variation in observed X-ray temperature compared with pressure* is clearly present in the results. In order to produce a temperature range of 2.2 to 2.8×10^6 K typical of active-region loops, one requires a heating rate that is between 10 and 15 times bigger than the standard radiation. Quiet-region loops need a heating rate about half as big as this. For the above numerical modelling, the energy balance Equation (6.32) was solved along a single magnetic field line, but it has also been solved more generally in the two cases of a cylindrically symmetric structure (Hood and Priest, 1979a) and a force-free arcade of loops (Priest and Smith 1979), as described in Sections 11.1.1 and 11.1.2.

Chiuderi *et al.* (1981) suggest that the observational errors in L, p_0 and T_1 are so great that no meaningful information can yet be inferred about the heating processes. Since the temperature profile is flat over most of a loop's length, a good approximate scaling law can be obtained by neglecting the radiative loss, in agreement with Roberts and Frankenthal (1980).

6.5.1B. Cool Cores

A striking feature of the results in Figure 6.10 is that, as the loop pressure slowly increases (with its heating and length held fixed), so the summit temperature (T_1) rises to a maximum and then decreases to a critical value (T_{crit}) at which dT_1/dp becomes infinite and a *catastrophe* occurs, as indicated schematically in Figure 6.11. If the *pressure exceeds the value* p_{crit}, the loop is therefore in a state of *thermal non-equilibrium*. There is no neighbouring equilibrium, and so the plasma cools along the dotted line seeking a new equilibrium below 10^5 K. As it cools, much of the plasma is likely to drain out of the loop, since it cannot all maintain hydrostatic equilibrium at lower temperatures. The existence of a lack of equilibrium and the consequent cooling is also present when a *loop is (slowly) stretched* far enough at constant p and H or when its *heating is slowly decreased* at constant p and L. It provides an explanation for the existence of extremely cool cores (Section 1.4.1C) that Foukal (1975, 1976) and Jordan (1975) have observed in some coronal loops. (In addition it may be the cause of the sudden loop evacuation that Levine and Withbroe (1977) described.) In extreme cases, Hood and Priest (1979a) propose this as the mechanism for creating active-region filaments or prominences (Section 11.1.1). The idea that active-region filaments are just stretched magnetic flux-ropes containing cool plasma is consistent with the frequently reported observation of motions along filaments, presumably guided by the magnetic field. The cool cores contain too much plasma to be in hydrostatic equilibrium at such large heights, and so they must be dynamic. The cool plasma may have been injected up as spicules, or its temperature may have decreased by thermal instability.
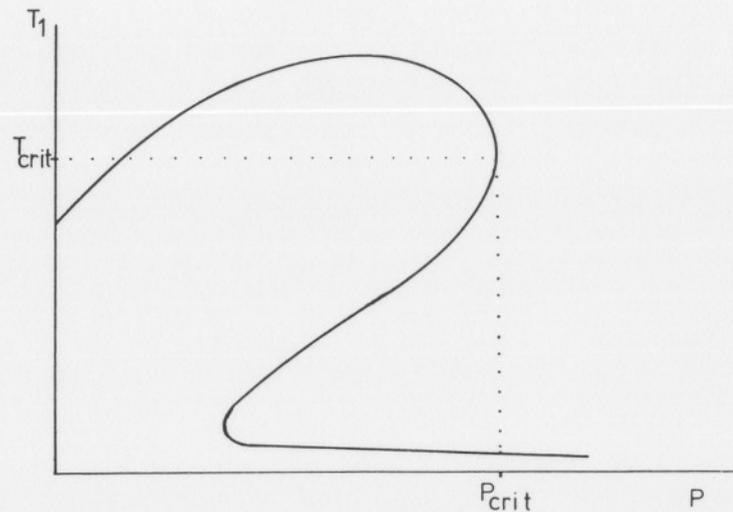
Fig. 6.11. The summit temperature (T_1) for a static coronal loop shown schematically as a function of its pressure (p). When p_{crit} is reached, the plasma cools along the dotted line to a new equilibrium well below T_{crit}.

Another feature of the numerical solutions (Figure 6.10) is that thermally isolated loops, in particular, are *thermally unstable* to perturbations that preserve the base temperature (Antiochos, 1979; Hood and Priest, 1980a Chiuderi *et al.*, 1981). This may provide an explanation for the *ceaseless activity* of the solar atmosphere and, in particular, for *spicules*. However, the loops are close to marginally stable conditions with an extremely small growth-rate (Craig *et al.*, 1982), although an adequate coupling to the chromosphere has not yet been incorporated.

6.5.1C. *Hydrostatic Equilibrium*

For loops that are about a coronal scale-height or greater in vertical extent, the pressure decreases substantially from the loop base to its summit, and so the full equations (6.32) to (6.34) need to be solved. Wragg and Priest (1981a) have done so for a loop that is an arc of a circle. The ratio of height to foot-point separation is denoted by $\frac{1}{2}r$, and the increase in cross-sectional diameter from base to summit is d. Figure 6.12 gives the results for the summit temperature of thermally isolated loops and shows the effect on the summit temperature of increasing the loop height or divergence. See also Vesecky *et al.* (1979) and Serio *et al.* (1981). The latter have modelled thermally isolated loops in hydrostatic equilibrium with a heating that declines away from the base like $\exp(-s/s_H)$; they derive scaling laws and find that loops longer than $2s_H$ develop a temperature minimum at the summit, which may be relevant to prominence formation (Section 11.1).
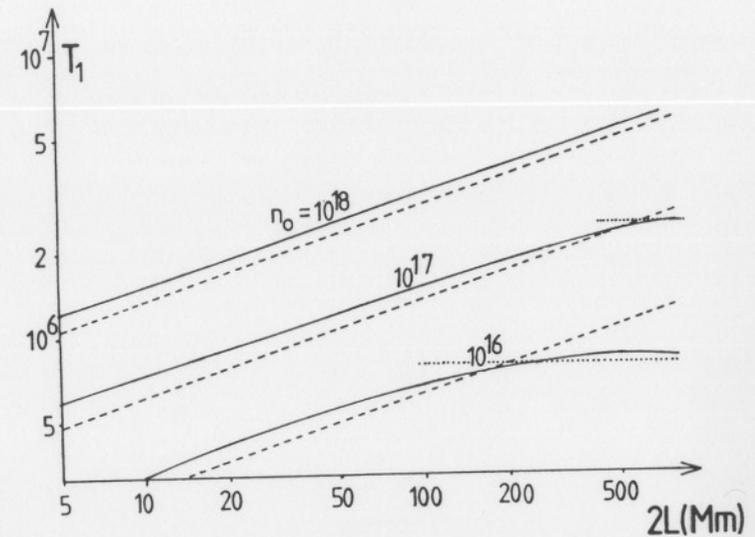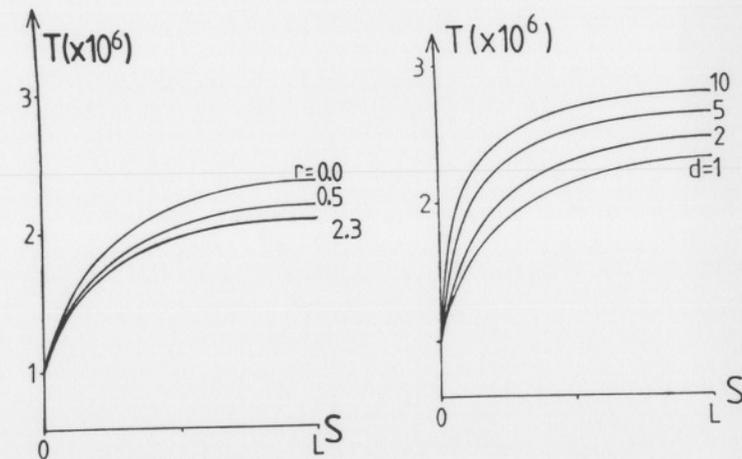
Fig. 6.12. Coronal loops in hydrostatic equilibrium. (a) The summit temperature (T_1) as a function of loop length ($2L$) in Mm ($= 10^6$ m) and base density (n_0) in m^{-3} for thermally isolated loops. The scaling laws $T_1 \sim (p_0 L)^{1/3}$ for short loops and $T_1 \sim p_0^{1/2}$ for long loops are shown dashed and dotted, respectively. (b) The effect of loop geometry on the temperature profile of an interconnecting loop of length 225 Mm and heating $\bar{h} = 7$. The ratio of loop height to footpoint separation is denoted by $\frac{1}{2}r$: $r = 0$ gives the uniform pressure case, while $r = 2.3$ gives the loop of maximum height for a given L. (c) The effect of loop divergence on the temperature profile $T(s)$ for a semicircular active-region loop of length 80 Mm and heating $\bar{h} = 20$. The ratio of summit diameter to base diameter is denoted by d. (From Wragg and Priest, 1981a.)

6.5.2. FLOWS IN CORONAL LOOPS

For the most part the effect of flows in models of the solar atmosphere has been neglected, and yet both steady and unsteady flows are universally present (Figure 1.23). The present rudimentary state of the theory of such flows is summarised by Priest (1981c) and observations of the many different types have been described in Section 1.4.1D. Ground-based observations reveal Evershed outflow (6 to 7 km s^{-1}), Evershed inflow (20 km s^{-1}), network downflow (0.1 to 2 km s^{-1}), surges (20 to 30 km s^{-1}), spicules (20 to 30 km s^{-1}) and coronal rain (50 to 100 km s^{-1}). Space observations show both transient, small-scale, fast flows (0 to 150 km s^{-1}), lasting for minutes or less, and persistent, large-scale slow flows (2 to 10 km s^{-1}), lasting for an hour or more.

From a theoretical viewpoint there are several ways of generating a flow in a coronal loop, which are briefly outlined in the remainder of this section (Figure 6.13). *Siphon flow* would be driven by a pressure difference between the two footpoints. It has been invoked by Meyer and Schmidt (1968) to explain Evershed motions along low-lying photospheric and chromospheric loops, but it may also occur along coronal loops. If one starts with a static loop and switches on a pressure difference, an *accelerated flow* will be driven from the high-pressure footpoint. But, if one starts with a loop containing a flow and then a small pressure difference is imposed in opposition to the flow, it is possible for a *decelerated* flow to be set up towards the higher pressure. There are several ways in which different footpoint pressures may be maintained. For example, the constancy of total base pressure (plasma plus magnetic) would imply that regions of high magnetic field strength possess a low plasma pressure. Also, a converging photospheric flow could compress both magnetic field and plasma, and so enhance the pressure locally. Again, a supergranular flow could drive a downflow by viscous coupling in the intense tubes (Section 8.7) that make up the boundary of a supergranule cell. Finally, the pressure at a loop footpoint may be increased by enhancing the heating there.

Coronal siphon flow has been analysed by Cargill and Priest (1980) and Noci (1981). A simple case is that of steady flow along a loop of uniform cross-section, satisfying conservation of mass, momentum and energy in the form

$$\frac{d}{ds}(\rho v) = 0; \qquad \rho v \frac{dv}{ds} = -\frac{dp}{ds} - \rho g \cos\theta, \qquad \frac{d}{ds}\left(\frac{p}{\rho^{\gamma}}\right) = 0,$$

where s is the distance measured along a loop of length $2L$ and $\theta(s)$ is the local inclination of a section of the loop to the vertical, so that for a semicircular loop $\theta = \pi s/(2L)$. The adiabatic law has been assumed for simplicity. Eliminating p and ρ between the three equations and writing $c_s(s)$ as the sound speed yields

$$\left(v - \frac{c_s^2}{v}\right)\frac{dv}{ds} = -g\cos\frac{\pi s}{2L},$$

which is similar in form to the solar wind equations (Section 12.2) and possesses a critical point ($v = c_s$) at the loop summit ($s = L$). More general solutions for a varying cross-sectional area and a full energy equation have also been produced (Cargill and Priest 1982b). The main feature is that for small pressure differences the flow is *subsonic*, but for larger pressure differences the flow becomes *supersonic* near the loop summit and is then slowed down by a shock wave in the downflowing leg (Figure 6.14).

The cause of *spicular motions* (Figure 6.13(b)), in which cool plasma is propelled up a loop leg and then falls back down, has not yet been adequately explained. One possibility is that they are driven by granular buffeting (Section 8.7.3), with a resonance between the forcing granular motion at the edge of an intense magnetic flux tube and the vertical plasma motion within the tube (Roberts, 1979). Alternatively, a similar resonance may occur when wave motions within a supergranule cell impinge on its boundaries, or spicules may be a result of a lack of thermal stability in the corona.

Fig. 6.13.   The main types of flow in coronal loops are: (a) siphon flow, (b) spicule flow, (c) loop draining, (d) loop filling.
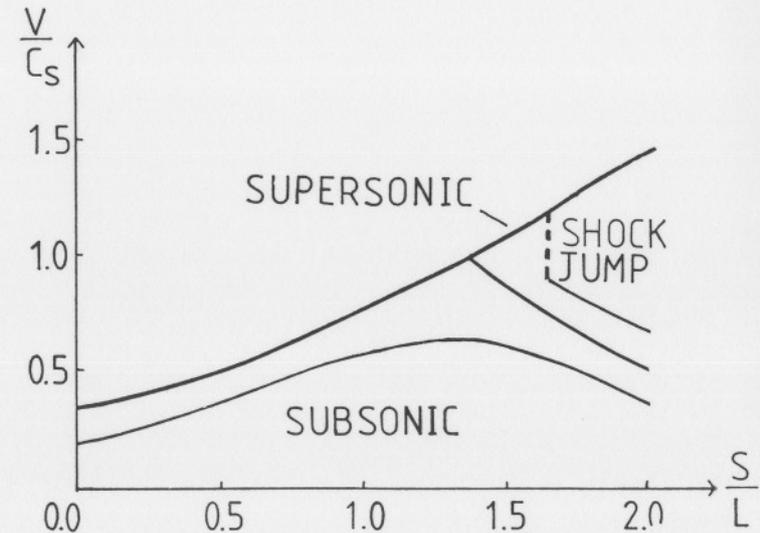
Fig. 6.14.   The siphon flow speed (v) at a distance s along a converging coronal loop of length $2L = 100\,000$ km (from Cargill and Priest, 1980).

the loop plasma (Section 6.5.1B, Hood and Priest, 1980a).

Surges may be caused by the reconnection between *newly emerging* or *evolving* satellite flux and the ambient sunspot field (Heyvaerts *et al.*, 1977). Instead, they may be the result of *non-equilibrium*, when the pressure within a closed magnetic structure exceed a critical value.

Evershed flow may be evidence of siphon flow, or it may be a result of the interaction between the convection rolls and the magnetic field within the penumbra of a sunspot.

Downflow in both legs of a loop (Figure 6.13(c)) may occur if there is a *condensation* or *prominence* at the summit of the loop. It may, alternatively, be simply plasma that is returning to the lower atmosphere from spicules and surges and it may be heated adiabatically in the process (Poletto (1980) *Ap.J.* **240**, L 69). Downflow may also take place during the formation of *cool cores* as follows. Suppose a hot loop is stretched (or its pressure increased or heating reduced) until critical conditions for thermal non-equilibrium are reached. Plasma in the core of the loop near the summit will then cool down. Since it is no longer in hydrostatic equilibrium, most of the cool plasma will drain out of the core until a new hydrostatic equilibrium is reached. Once a cool core is produced, small-scale magnetohydrodynamic instabilities may drive a circulation of plasma from the ambient corona across the interface and into the core where it falls down. The generation mechanism for this downflow may instead operate in the photosphere rather than the corona. For example, super-granular flow may drag plasma downwards in the network, or a downflow may be associated with the transient formation of an intense tube by the *intense magnetic field instability* (Section 8.7.2).

A final category of flow is upflow in both legs (Figure 6.13(d)), which may be part of an overall circulation within a coronal arcade containing a *quiescent prominence* (Section 11.1.2). It is also driven during the rise phase of a *simple loop* (or *compact*) *flare* (Section 10.2), when the presence of an extra source of heating in the loop means there is insufficient plasma for hydrostatic equilibrium, and so extra material is sucked up from below. Another way of driving an upflow, for example above a sun-spot, is by Alfvén or magnetosonic waves that are propagating upwards and dumping their momentum as well as their energy in the plasma.
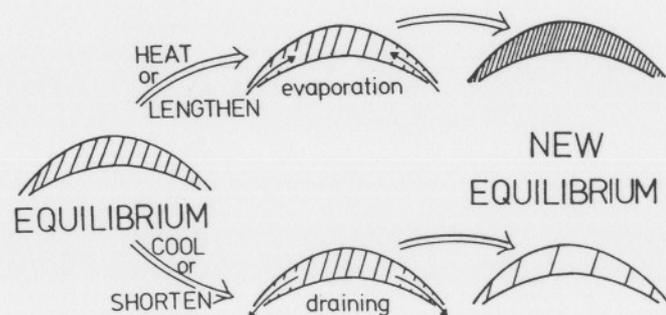
Draining or filling (i.e., *evaporation*) may also take place if a loop passes through a series of equilibria with a different heating (H) or loop length (L). Suppose H or L increases in value; then Equation (6.41) implies that the new equilibrium possesses a higher density, and so extra material must be brought up (or *evaporated*) from below along the loop (Figure 6.15). Similarly, if the heating or length are reduced in value, there is too much plasma in the loop for equilibrium; some of it must drain down until the pressure gradient balances gravity and all the energy terms balance.

There are many other ways in which flows may arise. As the large-scale magnetic field evolves through a series of largely force-free states in response to the motion of the photospheric footpoints, it may occasionally find that the threshold for the onset of a *magnetohydrodynamic instability* is passed (Section 7.5); instead, there may be no magnetic equilibrium at all and a state of non-equilibrium arises (Section 6.4.4).

Fig. 6.15. The evolution of a coronal loop from one equilibrium to another by means of an evaporation or draining, depending on whether the heating rate (or loop length) increases or decreases in value.