Determining the heating mechanism (or mechanisms) that causes the outer atmosphere of the Sun, and many other stars, to reach temperatures orders of magnitude higher than their surface temperatures has long been a key problem. For decades, the problem has been known as the coronal heating problem, but it is now clear that ‘coronal heating’ cannot be treated or explained in isolation and that the heating of the whole solar atmosphere must be studied as a highly coupled system. The magnetic field of the star is known to play a key role, but, despite significant advancements in solar telescopes, computing power and much greater understanding of theoretical mechanisms, the question of which mechanism or mechanisms are the dominant supplier of energy to the chromosphere and corona is still open. Following substantial recent progress, we consider the most likely contenders and discuss the key factors that have made, and still make, determining the actual (coronal) heating mechanism (or mechanisms) so difficult.

**Keywords:** corona; chromosphere; magnetic fields; coronal heating

### 1. Introduction

The coronal heating problem is one of the longest running solar physics puzzles and is still a highly controversial topic. Considerable progress has been made in modelling possible heating mechanisms, but discriminating among these to discover if any one process dominates is extremely difficult to do. This is because the corona is not energetically isolated from the other regions of the atmosphere, such as the chromosphere, but instead the whole of the solar atmosphere forms a highly coupled system, with both energy and mass transferred in both directions between the chromosphere and corona through the transition region. The chromosphere has very different plasma properties to the corona and so explaining the heating of the solar atmosphere is now recognized to be considerably more complicated than had been appreciated during much of the twentieth century.

The coronal heating problem first arose following the work of Grotrian [1] (erroneously the author on the paper is given as W. Grotian) and Edlen [2], who discovered that emission lines observed during a total solar eclipse in 1869 were not owing to a new element called coronium, but were the result of highly ionized...
iron. This work demonstrated that the temperature of the corona is in excess of a million kelvin. In comparison, the temperature near the Sun’s surface, the photosphere, is just 6000 K. At the same time, the density falls off by six orders of magnitude from the photosphere to the corona.

The Sun is not unique in having such a hot outer atmosphere. Indeed, most stars in the cool half of the Hertzsprung–Russell (H-R) diagram have a corona and can be observed in X-rays. Their average temperatures range from 1 to 45 million kelvin with the hottest associated with the most rapidly rotating stars. Using observations from the Einstein Observatory, Vaiana [3] was the first to show that most stars, as opposed to just the odd exotic star, have soft X-ray emission and, hence, a corona. Linsky [4] reviewed stars across the H-R diagram with emissions in ultraviolet and X-ray emission lines that are produced by plasma hotter than $10^4$ K. He found that dwarf stars of spectral type G-M and rapidly rotating subgiants and giants of spectral type F-K in spectroscopic binary systems are most likely to be solar-like in nature, i.e. they are likely to have ‘a turbulent magnetic field sufficiently strong to control the dynamics and energetics of their outer atmosphere’. Furthermore, he found that T Tauri stars, other pre-Main Sequence stars and single giants of spectral type F to early K are also probably solar-like.

X-ray, EUV and UV images of the Sun (figure 1a–e), taken by the Atmospheric Imaging Assembly (AIA) and Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics Observatory (SDO), show beautiful, finely detailed loop structures, as well as more diffuse emission which is likely to be formed from...
a mass of unresolved loops. The resolved loops either connect regions of opposite-polarity magnetic field seen in the photosphere, or they are extended open structures that are anchored in a photospheric magnetic feature. As a general rule, the hotter the loops, the stronger the regions of magnetic field they are associated with [5].

The structure of the magnetic field is highlighted in coronal images because heat is spread efficiently along field lines, but transport across them is greatly inhibited (thermal conduction in the corona is 13 orders of magnitude stronger along field lines than across them [6]). All of these factors signify that the magnetic field plays a major role in heating the solar corona. Moreover, there is strong evidence from other stars that X-ray coronae are associated with magnetic field because a power-law relation exists between X-ray luminosity and magnetic flux over many orders of magnitude, ranging from solar quiet regions, through active regions, to dwarf and T Tauri stars [5,7].

The loops observed, for example in figure 1a–e, are visible because they contain plasma of the appropriate density and temperature to emit at the given wavelength. The surrounding space contains loops that are either at a different temperature (hotter or cooler) or insufficiently dense to be observed. In equilibrium, and without any heating, a loop will be cool and rarefied; so any heating mechanism must explain both how the material fills the loop and heats it to coronal temperatures.

In the corona, the combined radiative and conductive losses in active regions are $10^7$ erg cm$^{-2}$ s$^{-1}$ and in the quiet Sun are $3 \times 10^5$ erg cm$^{-2}$ s$^{-1}$, whereas in the chromosphere they are $2 \times 10^7$ and $4 \times 10^6$ erg cm$^{-2}$ s$^{-1}$ in active regions and the quiet Sun, respectively [8]. The source of this energy is convective churning of the plasma at and below the photosphere. These motions continuously move the footpoints of the ubiquitous magnetic fields that thread through the surface of the Sun with the time scale of these motions, relative to the end-to-end Alfvén travel time, leading to one of two fundamental types of heating mechanism. Slow (long time scale) motions result in a quasi-static stressing of the field, while fast (short time scale) motions generate waves. The dissipation of magnetic stresses, which typically manifest themselves in the form of current sheets, leads to magnetic reconnection and is known as DC heating, while the dissipation of waves is referred to as AC heating. In both cases, the actual dissipation will occur at kinetic scales, because the Lundquist number (the ratio of the ohmic dissipation time to the Alfvén crossing time) is very large (approx. $10^{10}$–$10^{13}$). On the other hand, the photospheric driving and the complex geometry/topology of the magnetic field are determined globally. Not surprisingly, this massive range of scales which need to be addressed by coronal heating models poses a problem for both theorists and observers. Current parallel computers have neither the memory nor the power to permit well-resolved models on all scales, and current imaging telescopes have neither the required spatial and temporal resolution nor the sensitivity to see anything at the kinetic scale. Additional difficulties arise in the form of the coupling between the dense interior and tenuous outer atmosphere (which require very different physics) and, for open magnetic field regions, mechanisms must also account for the mass flux loss owing to the solar wind. With all these difficulties, it is not surprising that, to date, researchers have been unable to properly explain the solar heating problem. Nonetheless, our understanding has advanced significantly.
Since the identification of the hot solar atmosphere, there have been a huge number of heating mechanisms proposed that involve reconnection or magnetic waves. A full review of the analytical and numerical studies carried out is far beyond the scope of this review; so we refer the interested reader to the reviews by Aschwanden [9], Goedbloed & Poedts [10], Klimchuk [11], Aschwanden et al. [12], Goossens et al. [13] and Reale [14]. Instead, we focus on those mechanisms currently most favoured. First, though, in §2, we consider the source of the corona’s energy, before giving a brief selective discussion on the backgrounds to wave heating (§3) and magnetic reconnection (§4). Nanoflare observations are reviewed in §5, and the heating of coronal loops, including numerical modelling and the recent observational results on spicules and Alfvénic waves, is the focus of §6. Finally, in §7, we present our conclusions.

2. Poynting flux

The energy transported into or out of the solar atmosphere is provided by a Poynting flux through the surface of the Sun,

\[ S = \frac{1}{\mu_0} \int \int_{S_s} E \times B \cdot dS, \]

where \( E \) is the electric field, \( B \) the magnetic field, \( \mu_0 \) the magnetic permeability and \( S_s \) the solar surface. By assuming that motions are ideal and by writing the magnetic field and velocity in terms of components that are parallel or perpendicular to the solar surface (denoted by subscripts \( p \) or \( n \), respectively), the Poynting flux per unit area through the solar surface can be written as

\[ \frac{1}{\mu_0} (E \times B) \cdot \hat{n} = -\frac{1}{\mu_0} (B_p \cdot v_p) B_n + \frac{1}{\mu_0} (B_p \cdot B_p) v_n, \]

where \( \hat{n} \) is the normal to the solar surface directed radially outwards. The first term corresponds to the movement of existing magnetic footpoints by horizontal convective motions, whereas the latter represents the emergence or submergence of magnetic flux.

Typically, the magnetic field is tilted back away from the direction of motion and so \((B_p \cdot v_p) B_n/\mu_0 \approx B_p v_p B_n/\mu_0\) provides an inflow of energy into the solar atmosphere (irrespective of the sign of \( B_n \)). If instead the field was tilted forward towards the direction of motion, then the Poynting flux would actually remove energy from the solar atmosphere. Magnetic fields are thought to be held in intense kilogauss flux tubes and are moved about with a typical local speed of 1 km s\(^{-1}\). Assuming that the magnetic field is inclined at an angle of 20°, and that the filling factor is 1 per cent in the quiet Sun (giving a mean field strength of 10 G), then the Poynting flux per unit area is about 2.5 \times 10^7 \text{erg cm}^{-2} \text{s}^{-1}. Active regions have much higher filling factors of around 50 per cent (implying a mean field strength of 500 G), which gives a Poynting flux per unit area of 1.2 \times 10^9 \text{erg cm}^{-2} \text{s}^{-1}. In both cases, simply considering the energy injected by moving the footpoints seems to give more than enough energy to heat both the chromosphere and corona, even if we assume that half the motions result in a loss rather than a gain in energy. Of course, though, energy may also be required to raise plasma from the chromosphere into the

\[ \text{Phil. Trans. R. Soc. A (2012)} \]
corona, accelerate particles and drive the solar wind in open-field regions. With this in mind, we also estimate the energy contribution from the emergence of flux.

To obtain an inflow of energy to the solar atmosphere from the second term, we require an upflow, and hence the emergence of flux. Newly emerging bipolar regions of flux become highly shredded as they rise through the top layers of the convection zone into the photosphere. They appear as a host of small-scale loops that are aligned so that after emergence they form a dipole with anything up to $\approx 5 \times 10^{19} - 10^{20}$ Mx of flux. Larger bipoles are found, but these are associated with active regions. The Poynting flux associated with the emergence of a single kilogauss intense flux tube is $2 \times 10^{12}$ erg cm$^{-2}$ s$^{-1}$, assuming an upflow of 2 km s$^{-1}$. However, the total emergence rate of new flux is just $450$ Mx cm$^{-2}$ d$^{-1}$ [15] suggesting a filling factor of $6 \times 10^{-7}$ and so, in the quiet Sun, the Poynting flux owing to the emergence of new magnetic flux is around $10^6$ erg cm$^{-2}$ s$^{-1}$. This is a substantial fraction of the $4.3 \times 10^6$ erg cm$^{-2}$ s$^{-1}$ that is required, but is a factor 25 less than that from the moving of footpoints. In active regions, the Poynting flux from moving flux is similarly dominant.

Despite their simplicity, the order of magnitude estimates presented here are qualitatively in agreement with the results from the sophisticated magnetohydrodynamic (MHD) modelling of the upper convection zone to the lower corona using radiative transfer by Abbett & Fisher [16]. It is clear that the energy that heats the solar atmosphere comes from the convection zone and this energy manifests itself both as waves and, through slow stressing, current sheets leading to reconnection. Which of these is most important for chromospheric and coronal heating is still very much an open question.

3. Acoustic and magnetohydrodynamic waves

Almost immediately after the identification in the early 1940s of highly ionized spectral lines, such as FeIX and CaXIV, mechanisms based on the dissipation of acoustic waves were proposed [17–19]. However, a heating theory based on waves must overcome several obstacles, namely: (i) waves must be generated in (or below) the solar surface layers, (ii) sufficient energy flux has to be transported as only a fraction of the wave energy will be transmitted into the corona, and (iii) the waves have to dissipate (very) efficiently to convert the energy flux into heat in the right place, at the right time. The first of these aspects (i) is probably the easiest to address. Indeed, solar convection generates a mixture of upward propagating waves with an energy flux of several times $10^7$ erg cm$^{-2}$ s$^{-1}$ [20], which would be more than adequate to heat the solar corona (and accelerate the solar wind). However, (ii) is not so easy to satisfy since most of the (magneto)acoustic waves do not propagate into the corona owing to strong reflection and refraction off the rapid density and temperature increase in the Transition Region (in the traditional, plane-parallel view of the solar atmosphere). Additionally, mode coupling is expected to take place in regions where the sound and Alfvén speed are approximately equal (see the numerical simulations by Rosenthal et al. [21] and Bogdan et al. [22]) reducing the flux reaching the corona still further. Alfvén waves, on the other hand, are generally seen as very efficient carriers of energy over large distances [23–25]. With regard to issue (iii), although Wentzel [26,27]...
pointed out that Alfvén waves could heat the corona, their weak damping makes them problematic as a coronal heating mechanism. This led to the development of a variety of mechanisms that are likely to enhance the dissipation of Alfvén waves such as resonant absorption [28–31] and phase mixing [32]. Both resonant absorption and phase mixing are based on the property that in an inhomogeneous plasma, individual surfaces can oscillate with their own Alfvén frequency. In resonant absorption, on a specific magnetic surface, an incoming wave motion can be in resonance with local oscillations, allowing the transfer of energy from large-scale motions to small length scales where dissipation can become effective. Similarly, phase mixing describes how Alfvén waves on neighbouring field lines (with different Alfvén speeds owing to a transverse inhomogeneity in the density and/or the magnetic field strength) gradually become out of phase, again generating increasingly large transverse gradients until dissipative length scales are reached. Although mostly treated separately in the literature, examples of the close interplay between resonant absorption and phase mixing can be found in Ruderman et al. [33,34]. Apart from resonant absorption and phase mixing, other mechanisms such as (nonlinear) mode conversion (to modes which might more readily dissipate) and ion-cyclotron resonance have also been proposed (see for example, reviews by Aschwanden [9], Goedbloed & Poedts [10], Klimchuk [11], Erdélyi & Ballai [35] and references therein).

Although many wave-based heating theories were developed, especially in the 1980s and early 1990s (see [9,10,13] for more detailed reviews), reconnection-based mechanisms were generally favoured to address the coronal heating problem. Porter et al. [36] point out that observed non-thermal line broadenings imply sufficient wave fluxes to account for the energy requirements of both quiet Sun and active regions, but that this does not necessarily mean that these fluxes are readily converted into heat. Indeed, even if the line broadenings are actually produced by waves, the efficiency of many of the theoretically studied wave-based heating mechanisms relies on the presence of waves of the right type and the right frequency, which, at the time, could not be verified because direct observations of waves in the solar atmosphere were largely absent. This situation has changed with the advent of space-based EUV imagers (Solar and Heliospheric Observatory (SOHO)/Extreme Ultraviolet Imaging Telescope (EIT) and Transition Region and Coronal Explorer (TRACE)), which now have sufficiently high spatial and temporal resolution to resolve oscillatory motions. A large variety of waves and oscillations have since been observed in the solar atmosphere. However, the observed energy budget contained within the observed waves was generally several orders of magnitude too small to account for the heating of coronal loops. Additionally, given the rapid damping of many of the observed oscillations, it is not at all clear that the time scales involved even allow for heating to occur during the actual oscillations. Hence, wave-based coronal heating mechanisms initially gained little ground. The discovery of several types of MHD waves in the solar corona did however lead to the rapid development of coronal seismology (see the review by De Moortel & Nakariakov [37]).

More recently observations of flows and Alfvén waves have led to a renewed interest in wave heating mechanisms (see §6c,d), both observationally and theoretically. For example, Van Doorsselaere et al. [38] look at coronal loop heating profiles and find that, based on observational evidence, a resistive wave heating mechanism is more likely than a viscous one. Similarly, Taroyan et al. [39]
show that power spectra of Doppler shift oscillations might allow a distinction between models based on footpoint heating and uniform heating. A recent study by Antolin & Shibata [40] investigates heating based on torsional Alfvén waves and suggests that, within the constraints of their 1.5 D model, this mechanism is viable only in a rather narrow regime, requiring long and thick loops (see Jess et al. [41] for a possible observational identification of torsional Alfvén waves in the solar chromosphere). We refer the interested reader to Taroyan & Erdélyi [42] for a review of heating diagnostics with MHD waves.

4. Magnetic reconnection

One consequence of new flux (greater than or equal to $10^{19}$ Mx) emerging into the photosphere is the formation of new coronal loops connecting the new flux features to pre-existing ones. The only way these new loops could have formed is through magnetic reconnection (see article by Pontin [43] or books by Priest & Forbes [44] and Biskamp [45] for more details), allowing a global restructuring of the magnetic field. The removal of flux from the photosphere is also associated with magnetic reconnection because previously unconnected pairs of magnetic features must connect before they can cancel. In both cases, the reconnection driven by emergence or cancellation of flux may be associated with not only the formation and heating of new loops, but also the bulk acceleration of plasma, particle acceleration and the lifting of mass into the chromosphere/corona (e.g. [46–48]). Depending on the location and rate of reconnection and also the amount of plasma lifted, such processes may result in loops that can range from hot coronal loops to cool dense chromospheric loops.

Emerging and/or cancelling flux models have been used to explain X-ray bright points [49–52] and X-ray jets [53–56], although the basic principles of these models also work on larger scales in active regions (e.g. [46,48,57,58]), indicating that many of the observed coronal loops (large and small) may well result from reconnection.

The processes of emergence and cancellation, however, are not essential for reconnection to occur. The mere movement of magnetic footpoints can also lead to reconnection [52,59–63]. Close et al. [64] used a magnetic-feature-tracking algorithm to identify and track magnetic features observed in quiet-Sun SOHO/Michelson Doppler Imager (MDI) high-resolution magnetograms. They then extrapolated the magnetic field into the solar atmosphere and quantified the changes in magnetic connections between features as the photospheric field evolved. Their work suggested that all the flux in the corona is recycled (all the connections are broken and reconnected) in around 1.4 hours, a rate at least 10 times faster than the recycling time of the photospheric flux as observed by MDI [65]. Furthermore, because the flux in the photosphere is contained predominantly within small-scale magnetic features, the complexity in the structure of the magnetic field falls off rapidly, with more than 50 per cent of the flux closing down below 2.5 Mm and just 10 per cent reaching above 25 Mm [66]. These results indicate that the majority of the magnetic interactions (reconnections) between loops will occur below these heights, quite probably in the chromosphere.

The models of Close et al. [64,66] are oversimplified, and more sophisticated models are required to quantify these effects more accurately. For instance,
in the solar atmosphere, reconnection occurs more readily in the cool dense chromospheric region than in the rarefied hot corona because diffusivity is inversely proportional to temperature. Furthermore, reconnection will convert magnetic energy into both thermal and kinetic energy (generating waves and shocks), as well as particle acceleration. In particular, there will be direct heating from ohmic dissipation of the plasma in field lines threading the reconnection site, as well as indirect heating of the surrounding plasma owing to the viscous damping of waves, outflow jets and shocks created by the reconnection. In some reconnection models (e.g. [67]), it is believed that the majority of the heating comes from the shocks, although recent numerical experiments suggest that ohmic heating dominates, in both the chromosphere and corona, in other reconnection scenarios [68,69]. Thus, it is still an open question as to whether viscous or ohmic heating mechanisms dominate and what the location of the energy deposition is relative to the reconnection site. However, it is clear that the precise nature will depend on the particular reconnection regime operating and on the properties of surrounding plasma.

Gudiksen & Nordlund [70,71] were the first to run three-dimensional MHD models solving an energy equation that included field aligned thermal conduction, as well as radiative losses. Their results qualitatively match coronal observations in many ways but, in such complex numerical experiments, it was not possible, at the time, to identify exactly which physical processes were responsible for the heating. For instance, some key questions that need to be addressed are, in their model is reconnection the key heating mechanism and, if so, is the reconnection occurring principally in the corona or the chromosphere? Also, the cadence of their experiments was such that analysing the role of waves was not possible. Yet more sophisticated modelling [48,72], including additional physics such as optically thick radiation in the chromosphere, has also been performed. These models find episodic heating events in the chromosphere and it is claimed that these events are a result of ‘stressing of the magnetic field’ by photospheric motions, as well as flux emergence, but a detailed analysis on the true nature of these events has yet to be undertaken.

5. Nanoflares: observations

One important heating mechanism, the basics of which were proposed almost 40 years ago, is the dissipation of a multitude of small-scale current sheets [73,74]. In particular, Parker [73] hypothesized that slow convective motions of magnetic footpoints could lead to a braiding of the magnetic field, creating small current sheets that would be continually dissipated and reformed, providing direct heating via ohmic dissipation. This concept may be thought of as a braiding between distinct magnetic loops or a self braiding of the magnetic field within a single loop. In either case, magnetic loops need not necessarily wrap around each other, but simple shearing motions suffice to produce current sheets.

Following the discovery by Lin et al. [75] from rocket-borne observations of many small, hard X-ray spikes with energies between $10^{24}$ and $10^{27}$ erg, Parker [76] speculated that a sea of small flares (‘nanoflares’), with energies nine orders of magnitude less than large flares (i.e. $10^{24}$ erg), could heat the corona if they occurred in sufficient numbers. This is known as Parker’s braiding or nanoflare theory.
In reality, a wide range of events (sizes and energies) have been observed in the solar atmosphere following power-law distributions of frequency with varying indices. Hudson [77] considered the implications of a power-law distribution of energies. His important paper pointed out that the index of the power-law distribution of energies is critical in establishing whether small-scale or large-scale flares provide more heat. If the power-law index $\alpha > -2$, large flares dominate the heating of the corona, whereas if $\alpha < -2$, small flares dominate. Note, though, even if the slope is steep enough, nanoflares can only heat the corona provided there are enough of them to produce sufficient energy.

Using Solar Maximum Mission/Hard X-ray Burst Spectrometer, Yohkoh/Soft X-ray Telescope, SOHO/EIT and TRACE, the distributions of flares, microflares and/or nanoflares were determined by many [78–84]. Power-law indices for the distribution of event energies were reported ranging from $-2.7 < \alpha < -1.5$, leaving the validity of nanoflares as a heating mechanism completely open. Similarly, power-law indices in the range $-2.5 < \alpha < -1.85$ have been reported for stellar flares observed on a range of late-type cool stars (spectral type F-M) [85–87].

Observational determinations of $\alpha$ come from the analysis of images taken by different instruments (e.g. SOHO/EIT, TRACE), with different detection routines (e.g. local peaks in emission or intensity in particular lines in a specific order of some specified size in area and peak), different energies (e.g. thermal energy or radiative/conductive losses) determined and different assumptions made. Parnell [88] examined the effects of making different, but reasonable, assumptions to detect flares, determine their energies and calculate the resulting slope of the power law. The approach involved using just two datasets, one of the quiet Sun and one of an active Sun taken by TRACE.

In the end, Parnell [88] determined 1200 different estimates of the power-law index by considering different sets of the possible parameters as used by others in their analysis. For both datasets (quiet Sun and active region), $\alpha$ lay in the range $-2.5$ to $-1.6$ with a mean of $-1.9$. The most reasonable conclusion that one can draw from these results is that it is not possible to accurately determine the distribution of individual nanoflare/flare energies using the method of direct detection. Instead, alternative methods that attempt to detect ensembles of nanoflares have been sought [89,90], providing support, but not conclusive observational proof, that nanoflares heat the corona. Thus, nanoflare heating is still regarded as one of the most promising coronal heating mechanisms for both coronal loops and the background quiet Sun.

6. Heating of coronal loops

Figure 1 clearly shows that the solar corona is made up of many loops that range from just a few megametres in length to hundreds of megametres. Even the diffuse emission seen in many of the images in figure 1 is highly likely to be made up of unresolved mini loops. The footpoints of coronal loops are not made up of single monolithic magnetic features of one polarity, but are made up of many tiny intense flux tubes, or strands, which are continuously being sheared, rotated and compressed as a result of the underlying granular and supergranular motions. Modelling has shown that any one of these motions can create tiny current sheets within a coronal loop, i.e. as in Parker’s braiding/nanoflare [76], the kink
instability model of Hood et al. [91] or the tectonics idea of Priest et al. [92]. The idea is that, if there are enough current sheets distributed throughout the loop cross section, and these are dissipated and regenerated at an appropriate rate, then enough energy may be provided to maintain the loop at more than a million K for several hours, as observed. However, it is also possible that small-scale heating events may result from a range of different wave dissipation mechanisms such as phase mixing [32] and resonant absorption [28,93].

(a) Three-dimensional magnetohydrodynamic loop modelling

Models in which simple magnetic fields have been shown to produce multiple current sheets through random driving motions have been conducted by many (e.g. [94–97]). For instance, Galsgaard & Nordlund [95] conducted three-dimensional MHD numerical experiments in which they sheared an initially uniform vertical magnetic field first in one direction then at right angles. The driving pattern was repeated continuously with random driving periods in each direction forming a quasi-steady state, involving a chaotic distribution of multiple small-scale current sheets throughout the domain, where the current sheets continuously dissipate and reform elsewhere. Each small-scale reconnection event may be thought of as a single nanoflare that locally heats a few strands of the coronal loop. However, by averaging over the depth of the domain and over a short time period, the loop appeared to be approximately uniformly heated (figure 2a). If the same braiding experiment is carried out, but this time with a stratified loop, thermal conduction and optically thin radiation, then within
the chromospheric regions of the loop considerably more current sheets are both formed and dissipated in comparison with the higher temperature coronal regions (figure 2b) [98]. This imbalance in the distribution of heating events is a result of the significant (a factor of 10) increase in Alfvén speed from the photosphere to the corona. The energy density from propagating perturbations is therefore much higher, per unit length, below the transition region than above. Furthermore, the transition region acts as a reflector preventing much of the Poynting flux injected by the boundary motions from reaching the corona.

Instead of a complex driving pattern on a simple magnetic field creating multiple forced reconnection events, a multitude of small-scale reconnection events can also be created from a complex magnetic field which undergoes an instability [91,99,100]. Wilmot-Smith et al. [100] and Pontin et al. [99] considered the resistive relaxation of a coronal loop consisting of a complex-braided magnetic field in which an instability leads to spontaneous reconnection producing a multitude of tiny nanoflares (figure 2c–e). They conducted three-dimensional MHD experiments in which this initially braided magnetic field (containing just six crossings), formed by slow footpoint motions, is allowed to relax via resistive MHD. Initially, the magnetic field has large-scale diffuse currents, but, as the field evolves, intense current layers are naturally generated until a resistive instability spontaneously triggers reconnection and the start of a cascade in energy release to small scales. Most importantly, they show that in a magnetic loop with what appears to be just six reconnection sites (braidings), many hundreds of little energy release events occur, distributed throughout the coronal loop (figure 2c). Moreover, there is much more reconnection than one would first predict. This is because the magnetic field is not clever: it does not follow the route to the quickest and simplest untangling of itself, but rather reconnects wherever the conditions for the onset of reconnection are met [101,102]. This results in a small reduction in the magnetic stresses in the field, but not a complete untangling, and a continuous chain of reconnection events must occur before the field can completely untangle itself. Indeed, Pontin et al. [99] find that the flux in their braided flux tubes can reconnect up to 4.4 times, confirming that the flux is not untangled via the quickest route. Such a process is known as multiple (or recursive) reconnection and was first found by Parnell et al. [103]. With so many small-scale energy releases throughout the coronal loop, the entire loop can easily be heated and not just a few strands from within it.

Parnell & Haynes [101] showed that multiple reconnection has a series of important consequences for coronal heating. Not only does it result in a multitude of small-scale reconnection events, and therefore widespread heating throughout the coronal loop, but considerable energy may be released over a prolonged period. Parnell & Haynes [101] considered an experiment where two opposite-polarity magnetic sources, in a background uniform overlying field, are driven past each other at a constant slow speed of 5 per cent of the Alfvén crossing time. They ran a series of numerical three-dimensional MHD experiments that were exactly the same except for the value of the resistivity. They found that as the resistivity decreased and became more realistic, the magnetic reconnection between the opposite-polarity sources was more complex in nature and thus took longer. This resulted in an increase in the horizontal component of the magnetic field at the base and thus an increase in the Poynting flux injected into the system. With more energy in the system, more energy had to be released to get back to the
same relaxed state in each case. In their braiding experiment, Pontin et al. [99] note similar findings. Thus, the nature, location and timing of magnetic energy release events are important to understand, not only because of their feedback on the system, which may affect the amount of Poynting flux subsequently injected, but also because the plasma/observational consequences will depend greatly on these factors too.

(b) One-dimensional loop modelling

Numerical three-dimensional MHD experiments, such as those described above, are very useful in that they can give a good idea as to the distribution, frequency and size of energy release events and can provide an indication of the mechanism(s) heating the corona. However, owing to the large global scales they cover they rarely include all the appropriate physics. In particular, a very simple energy equation is often used, neglecting thermal conduction, radiation, radiative transfer and ionization non-equilibrium effects. Thus, the plasma response seen in MHD models is not correct and so comparing with observations is extremely difficult. However, the very limited energy transport across fieldlines means that one-dimensional hydro models that can incorporate many of these important physical processes can be used to model the plasma response in a single strand, and a whole coronal loop, by combining the results from many single strand runs (see for a review [11,14]).

It is highly unlikely that cool dense loops thread the corona waiting to be heated. Instead any coronal heating mechanism must explain both how the plasma is transported into the corona and how it is heated. One crucial result of one-dimensional hydrodynamic models has been to create a clear picture of the behaviour of the plasma within a strand during an impulsive coronal heating event (note that, although often referred to as a ‘nanoflare’, this event may be either AC or DC in origin) from the starting point of an initially cool, low density, equilibrium strand, i.e. one that will not be emitting in EUV or X-rays [11,104,105]. An impulsive coronal event will cause the strand to evolve through four phases (figure 3). During the first phase (I), heat from the event is conducted along the field rapidly raising the temperature of the whole strand up to a maximum. The density and emission show very little change during this very short phase, creating a strand that is underdense relative to static equilibrium theory. A second longer phase (II), during which the impulsive event continues to heat, follows in which the temperature remains fairly constant, but the chromosphere is strongly heated, causing it to expand, leading to the evaporation of plasma that rises and fills the strand. During this so-called evaporation phase, the emission rises rapidly and the strand becomes denser, but it still remains underdense. The third phase (III) starts when the impulsive event stops and conduction starts to cool. The emission first peaks, then starts to decrease, but the density continues to increase. As the strand cools and becomes more dense, radiative cooling effects become more important until, at the end of this phase, radiative and conductive cooling finally balance. In phase IV, the loop appears overdense, radiative cooling dominates and the density starts to decrease, slowly at first, until the temperature decreases to a point at which catastrophic cooling occurs and the density plummets until the temperature, density and emission measure return to their pre-event values.
Figure 3. (a) Graph depicting the temporal evolution of the temperature (dot-dashed), light curve (solid) and density (dashed) in a loop (strand) owing to pulsed heating from a single coronal heating event. (b) A temperature versus density graph of a loop during the same event (phases I–IV). A long-duration heating pulse may lead to an evolution through phases I, II* and IV*. The dotted line indicates the equilibrium state in which conductive and radiative cooling balance. Graphs based on figs 14 and 15 from Reale [14].

Clearly, this scenario has a number of key observables that can be tested. Namely, at the start, extremely hot plasma, which, it has been calculated, must be in excess of 4 MK, should be present [106]. However, because the loop plasma density is low during this phase, any emission will be very weak and so such hot plasma will be difficult to detect. Schmelz et al. [107] have observed hot plasma outside solar flares, which is widespread throughout active regions, but weak. A recent paper by Reale et al. [108] claims to have now found thin, high-temperature strands of around 10 MK within coronal active regions using SDO/AIA. However, such hot emission is likely to be rare because, Bradshaw & Klimchuk [109] argue, non-equilibrium ionization effects will suppress the hot emission lines and in some cases render them invisible.

Secondly, evaporation upflows should be present. There appears to be a straightforward split with hot chromospheric UV emission lines and coronal EUV lines, in the quiet Sun, with temperatures of formation above 0.5 MK having blue Doppler shifts (e.g. Ne VIII, Mg X, Fe XII), while the cooler (less than 0.5 MK) emission lines (e.g. C II, C IV, Si IV, S VI, N V, O IV, O V, O VI) are red shifted [110]. In active regions, the split is normally higher at around 1 MK. These results may seem consistent with a hot evaporative upflow and a cooler downflow; however, there are other interpretations. In particular, the hot upflows may very well indicate that the actual heating mechanism is not located in the corona, but that plasma in the chromosphere and transition region is heated in situ, resulting in its expansion, creating a flow that fills the loop with hot plasma. There is currently a hot debate over which interpretation is correct, especially because a direct signature of evaporation such as the impact of accelerated particles on the chromosphere is likely to be weak and, hence, very difficult to identify. Most likely though both mechanisms are in operation, but which is most dominant is still an open question.

Finally, because the density behaviour lags that of the temperature during the heating, evaporation and conductive cooling phases (I, II and III) the strand appears underdense relative to static equilibrium predictions. Observations
suggest that hot loops \((T > 2\text{MK})\) are underdense \([111]\), and so they are consistent with loops made up of multiple impulsively heated strands, where the majority of strands are in phases I, II and III. However, the so-called warm EUV loops \((T \approx 1–2\text{MK})\) are observed to be overdense relative to equilibrium models and are found to have a narrow temperature range, appearing practically isothermal along their length \([112,113]\). More recently, observations from the Hinode/EUV Imaging Spectrometer indicate that the plasma in these loops is not actually isothermal, but rather multi-thermal with a narrow temperature range \((10^6 \pm 3 \times 10^5 \text{K})\) and a volumetric filling factor of just 10 per cent \([114]\). Thus, it is thought that these warm loops are highly filamentary and cooling from a hotter state \([115,116]\). A possible explanation is that the loop is made up of multiple strands each of which is impulsively heated by a single, short-lived (in comparison to the characteristic cooling time) event. Hence, the plasma in the strand evolves as it is first heated and then cools, such that overall the loop appears to be overdense (in phase IV) for an extended period \([11]\).

For a heating event that is longer than a conductive cooling time, the plasma in a strand may follow the phases I, II* and IV* (\textit{figure 3b}) instead. In fact, strands that are heated by either a long slow heating event or a series of repeated short impulsive events could evaporate sufficient density such that the strand reaches a hydrostatic equilibrium and may remain in equilibrium (between phases II* and IV*), if the heating balances the conductive and radiative cooling. When the heating event ceases, radiation is likely to dominate and the temperature and density decrease (phase IV*) \([105,117]\). Some types of loops have a wide range of densities and lengths and, in general, follow a series of scaling laws derived from hydrodynamic equilibrium modelling (such as Rosner \textit{et al.}\[118]\)). They appear to be consistent with steady heating, i.e. a deposition of heating at a rate that balances the conductive and radiative cooling of the loop \([111,119,120]\). Indeed, models of entire active regions or the whole global corona have been made by applying these scaling laws to potential and nonlinear force-free magnetic field extrapolations \([121–124]\).

The coronally heated, multi-strand scenarios discussed earlier fit quite a few, but not all observations. In particular, their main problem is that they do not explain how the chromosphere itself is heated, although, if the heating events are due to braiding, then, as the three-dimensional braiding experiments of Galsgaard \([98]\) seem to suggest, this should not be a problem as an even greater number of heating events should occur in the chromosphere than in the corona. Furthermore, as discussed in §1, waves excited by footpoint motions are likely to be abundant in the chromosphere. Nevertheless, a recent series of observations and numerical experiments now suggest that the corona may actually be heated by strong upflows that are driven from the photosphere/chromosphere, as discussed in the next section.

\textit{(c) Spicules and propagating disturbances}

Following the launch of SOHO and TRACE, several authors such as Berghmans \& Clette \([125]\) and De Moortel \textit{et al.}\ [126] reported the observation of propagating disturbances in large, quiescent (fan) loops at the edges of active regions, which were interpreted as slow magnetoacoustic waves propagating along coronal loops. However, the energy budget contained within these observed
propagating disturbances was generally several orders of magnitude too small to account for the heating of coronal loops. Even though the coronal energy budget of the propagating slow waves was thought to be very small, Fontenla et al. [127] and Jefferies et al. [128] point out that in the lower atmosphere, the energy flux contained in the low-frequency waves (less than 5 mHz) is much higher than the flux in their high-frequency counterparts. These high-frequency waves had long been considered as prime candidates to address the heating of the chromosphere, an idea that was rejected by Fossum & Carlsson [129]. However, their low-frequency counterparts seem to contain much more energy than previously thought.

The potential of the low-frequency propagating disturbances has very much come to the forefront again with recent observational results from Hinode and SDO/AIA, as they have been linked with the supply of mass and energy to the solar corona and solar wind. De Pontieu et al. [130,131] suggest that a significant fraction of the coronal mass is supplied by chromospheric jets (‘type II spicules’). These authors analyse Hinode and SDO data, in which they find faint, but ubiquitous, quasi-periodic disturbances, interpreted by these authors as flows rather than slow magnetoacoustic waves (see De Pontieu & McIntosh [132] and Verwichte et al. [133] for more details on this debate), and estimate that the mass contained within these spicules could supply a substantial part of the coronal hot plasma. Recently, McIntosh et al. [134] reported on observations of slow, counterstreaming downflows (approx. 10 km s$^{-1}$) in cooler spectroscopic lines, which the authors suggest could be the return-flow of coronal material (i.e. the end of the mass cycle). Although this mechanism is under considerable debate, it certainly highlights that the corona cannot be treated in isolation.

(d) **Transverse, Alfvénic waves**

As well as quasi-periodic propagating disturbances (flows/waves) propagating along coronal loops, recent observations have revealed transverse, propagating
oscillations along a variety of structures such as spicules [135–137], X-ray jets [138], prominence threads [139] and coronal loops [140,141,142]. An example of the active region analysed by McIntosh et al. [140] is shown in figure 4, who report on small-amplitude transverse displacements in coronal holes, quiet-Sun and active-region loops. An example of the loop (thread) displacements can be seen in the unsharp masked image inset in figure 4a. As these footpoint-driven, oscillatory displacements are interpreted as ‘Alfvénic’ (figure 4b), they are of particular interest to the coronal heating question. Indeed, De Pontieu [135] estimate that the energy flux carried by chromospheric Alfvénic waves (observed by Hinode/Solar Optical Telescope) is sufficient to accelerate the solar wind and possibly heat the quiet-Sun corona. Although Tomczyk et al. [141] report that the energy budget estimated from the Coronal Multi-channel Polarimeter (CoMP) observations of propagating disturbances in off-limb coronal loops is too small to account for coronal loop heating requirements, McIntosh et al. [140] not only demonstrate that small-amplitude, transverse, Alfvénic perturbations are ubiquitous in the solar corona but that the energy contained in these oscillations appears sufficient to address the requirements of the quiet Sun and coronal holes (approx. $10^5 \times 2 \times 10^5$ erg cm$^{-2}$ s$^{-1}$). In active regions, the reported energy (approx. $10^5$ erg cm$^{-2}$ s$^{-1}$) is too small to account for the heating of coronal loops (which according to Withbroe & Noyes [8] would require about $2 \times 10^6$ erg cm$^{-2}$ s$^{-1}$). However, in active regions, the superposition of a large number of randomly directed oscillating loop (strands) along the line of sight implies that the estimate reported by McIntosh et al. [140] is likely to be a lower limit. Similarly, a large number of loops, with transverse displacement in random directions, is likely to be present off-limb along the line of sight, which could explain why the energy budget estimated by Tomczyk et al. [141] seemed very small. Using three-dimensional MHD numerical simulations, De Moortel & Pascoe [143] show that the energy budgets estimated from observed, line-of-sight integrated Doppler velocities (such as those observed by Tomczyk et al. [141]) will be at least an order of magnitude smaller than the energy present in the three-dimensional coronal volume.

Although there is currently still considerable discussion about the interpretation of the propagating disturbances and the transverse displacements, what is clear is that the role of waves in the coronal heating problem has to be reassessed in light of the recent observations by SDO and Hinode. It now appears that Alfvénic waves are ubiquitous in the solar atmosphere, potentially containing a substantial amount of energy, and that chromospheric jets form an integral part of the mass supply cycle for the corona and solar wind. However, this does not necessarily imply a solution to the coronal heating problem. Although observations have unambiguously shown that waves are present, theoretical modelling now needs to address how these lead to heating, in the right locations, at the right time. Indeed, one has to keep in mind that although there appears to be a considerable amount of energy in the observed Alfvénic waves (see McIntosh et al. [140]), this does not translate automatically into heating of the plasma. In particular, several authors have recently interpreted the observed damping of, for example, the transverse oscillations observed by CoMP [141,142] as mode coupling, whereby the amplitude of the footpoint-driven, observed, transverse oscillations decreases owing to mode coupling to an azimuthal Alfvén wave in the tube boundary (see De Moortel & Nakariakov [37] for a detailed discussion).
Potentially, the dissipation of the azimuthal Alfvén waves, enhanced by phase mixing in the inhomogeneous tube boundaries could lead to local plasma heating. However, the viability of this remains to be demonstrated.

7. Discussion and conclusions

For decades after the discovery of the very high temperatures in the solar atmosphere, solving the coronal heating problem was mostly framed as an ‘AC or DC question’, or ‘waves versus reconnection’. Within the last few years, observational results have shown that the coronal heating problem has to be thought of in a different way. The highly dynamical nature of the solar atmosphere will clearly lead to the occurrence of both types of processes and the true question is to determine the relative contribution of the different mechanisms in different coronal structures. Wave-based heating mechanisms very much seem to have come full circle, from being the first mechanism suggested to provide the necessary energy input into the solar atmosphere, to being mostly discarded and recently back to the forefront again, driven by the spectacular observations by Hinode and SDO. Additionally, the discovery of the abundance of type II spicules demands that the impact of the equally dynamic chromosphere has to be assessed much more carefully and that it is essential that the ‘coronal’ heating question must be studied as part of the highly coupled solar atmosphere. Indeed, it is well known (but often ignored) that the energy requirements of the chromosphere far exceed those of the corona.

Heating by multiple small-scale reconnection events has long been considered a viable mechanism and much support has been lent to this mechanism both observationally and theoretically, but definitive observations of the smoking-gun are still missing. Indeed for all coronal heating mechanisms, there are still issues that need to be addressed before a definitive conclusion can be reached. For instance, the inability to directly ‘observe’ coronal heating has meant that proxies have had to be found. However, it is very difficult using these proxies to distinguish absolutely between wave heating or reconnection mechanisms. For example, the drifting of heating layers in phase mixing/resonant absorption means that such a wave-based mechanism will actually be impulsive in nature and so appear very much like a nanoflare heating event \([11,31,144,145]\). Antolin \textit{et al.} \([146,147]\) have recently investigated the different observational signatures of coronal heating by either nanoflares or the dissipation of Alfvén waves. These authors suggest that coronal rain could be a marker for coronal heating mechanisms, shedding light not only on the spatial distribution of heating events in coronal loops, but potentially also on the actual heating mechanism itself.

Additionally, on the theoretical front, it is clear that simplified models are not sufficient for comparison with observations. The response of the plasma in the solar atmosphere is very complex and interpreting responses of emission lines is not simple either. Forward modelling of theoretical experiments that have physically realistic effects included is essential, but the analysis of these results must be done with considerable care to determine which are important effects and will take time to complete. In the meantime, the argument over which mechanism(s) heat(s) the corona will continue.
I.D.M. acknowledges support from a Royal Society University Research Fellowship, and C.E.P. acknowledges the support of STFC through the solar group rolling grant. They thank Dr Brian Welsch for helpful discussions and anonymous referees for their constructive comments, which led to significant improvements in the paper.

References

Review. Coronal heating


Phil. Trans. R. Soc. A (2012)


*Phil. Trans. R. Soc. A* (2012)


Review. Coronal heating


135 De Pontieu, B. et al. 2007 Chromospheric Alfvénic waves strong enough to power the solar wind. Science 318, 1574–1577. (doi:10.1126/science.1151747)