Modelling nanoflares in active regions and implications for coronal heating mechanisms

P. J. Cargill1, 2, H. P. Warren3 and S. J. Bradshaw4

1 Space and Atmospheric Physics, The Blackett Laboratory, Imperial College, London SW7 2BW, UK
2 School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS, UK
3 Space Science Division, Naval Research Laboratory, Washington, DC 20375, USA
4 Department of Physics and Astronomy, Rice University, Houston, TX 77005, USA

Recent observations from the Hinode and Solar Dynamics Observatory spacecraft have provided major advances in understanding the heating of solar active regions (ARs). For ARs comprising many magnetic strands or sub-loops heated by small, impulsive events (nanoflares), it is suggested that (i) the time between individual nanoflares in a magnetic strand is 500–2000 s, (ii) a weak ‘hot’ component (more than $10^6.6$ K) is present, and (iii) nanoflare energies may be as low as a few $10^{23}$ ergs. These imply small heating events in a stressed coronal magnetic field, where the time between individual nanoflares on a strand is of order the cooling time. Modelling suggests that the observed properties are incompatible with nanoflare models that require long energy build-up (over 10 s of thousands of seconds) and with steady heating.

1. Introduction

The heating of the solar corona by small, impulsive heating events appears to date to a discussion by Gold [1], and the subsequent more quantitative proposal of Levine [2, 3] that small coronal current sheets were responsible for the heating. Subsequent analysis of Skylab data led to a quasi-consensus that the X-ray corona could be understood in terms of steady heating and associated scaling laws between the energy input and coronal plasma properties. It was not until Parker [4] proposed the idea of the ‘nanoflare’ as a unit of...
coronal energy release that impulsive heating began to be widely reconsidered. On the basis of transition region (TR) EUV brightenings, he suggested that the nanoflare energy was of order a few $10^{24}$ ergs and argued that this could be accounted for by a gradual ‘braiding’ of the coronal magnetic field in response to photospheric motions, followed by fast energy release associated with localized magnetic reconnection. An essential aspect is the existence of a critical state of the coronal magnetic field below which dissipation does not occur.

However, one difficulty was that, in the past, observations were unable to distinguish the signatures of impulsive from steady-state heating \[5,6\] and this led to efforts being made to identify distinctive properties of a nanoflare-heated corona \[7,8\] that might be observable by a future generation of instrumentation. This work has now been rewarded with results from the Hinode and Solar Dynamics Observatory (SDO) spacecraft that have inaugurated an era of quantitative studies of coronal heating. A key advance that is the subject of this paper is the ability to measure with some accuracy the distribution of coronal plasma as a function of temperature, especially in the brightest AR loops.

In this paper, we first review the response of coronal plasma to a nanoflare (§2) through one- and zero-dimensional hydrodynamic modelling. Definitions of nanoflare timings are presented in §3 and §4 then demonstrates how a combination of Hinode and SDO results and modelling can be used to constrain the properties of coronal nanoflares. In §5, we discuss the implications of these results.

2. Coronal plasma response to a nanoflare: generic behaviour

Much of the interpretation of modern coronal data depends on a good understanding of how a coronal ‘loop’ responds to a heating pulse. The loop can either be a monolithic structure as seen in observations, with characteristic transverse dimension of at least several megametres, or sub-structures within this large-scale magnetic field, referred to as a sub-loop or magnetic strand, with characteristic transverse dimension that may be as small as 100 km \[9,10\]. For a given rate and location of energy deposition, the problem reduces to one-dimensional hydrodynamics along a field line as the strong coronal magnetic field constrains mass motions and thermal conduction to be along the field due to the low plasma beta and small electron gyro-radius, respectively. Such an approach goes back to flare studies in the late 1970s \[11\] and was gradually extended to the non-flaring corona by, in particular, the NRL group \[12,13\]. A modern description of the relevant techniques can be found in \[14\].

(a) Methods

Equations of mass, momentum and energy conservation are solved for a defined magnetic field geometry, a heating pulse and appropriate boundary conditions at the chromosphere. In addition to thermal energy and mass motions, the energy equation must include thermal conduction parallel to the magnetic field, and optically thin radiative losses. Thus

$$\frac{\partial E}{\partial t} = -\frac{\partial}{\partial s} v(E + p) - \frac{\partial F_c}{\partial s} + Q - n^2 \Lambda(T), \tag{2.1}$$

where $v$ is the velocity, $E = p/(\gamma - 1) + 1/2 \rho v^2$, $F_c = -\kappa_0 T^{5/2} \partial T/\partial s$ is the heat flux, $Q(s,t)$ is a heating function that includes both steady and time-dependent components, $\Lambda(T) = \chi T^\alpha$ is the radiative loss function in an optically thin plasma \[15\] which is approximated by a piecewise-continuous function and $s$ is a spatial coordinate along the magnetic field. There is also an equation of state: $p = 2nk_BT$. For a given initial state and $Q(s,t)$, the plasma evolution can then be followed. Here equation (2.1) treats the plasma as a single fluid, but can readily be extended to a two-fluid approach \[14\].
While numerical solutions of the full hydrodynamic equations are desirable, approximate methods which solve for quantities averaged along the loop have now matured. These zero-dimensional models show good agreement with the one-dimensional numerical solutions, provide a transparent view of the relevant physics and are essential when modelling a corona comprising thousands of sub-loop elements. The enthalpy-based thermal evolution of loops (EBTEL) approach is the most recent example [15,16]. EBTEL splits the corona into two regions that interface at the point where thermal conduction changes from a loss to a gain. A formal definition of the TR is that it lies below this point [17] and the TR responds to what happens in the corona and vice-versa.

For subsonic flows, one can derive equations for the average coronal thermal energy, the TR energy balance and the coronal density [15]:

\[
\frac{1}{\gamma - 1} \frac{d\tilde{p}}{dt} = \tilde{Q} - \frac{1}{L} (R_c + R_{tr}),
\]
\[
\frac{\gamma}{\gamma - 1} (pv + F_c)_0 + R_{tr} = 0
\]

and

\[
\frac{d\tilde{n}}{dt} = \frac{(nv)_0}{L} = -\frac{\gamma - 1}{2kT_0L\gamma} (F_{c0} + R_{tr}),
\]

where all variables are now only functions of time. Here, overbar denotes a coronal average, \(Q(t)\) is the heating rate per unit volume, \(F_{c0} = -(2/7)\kappa_0T_0^{7/2}/L\) is the heat flux at the top of the TR, \(R_c \approx \tilde{n}^2 A(\tilde{T})L\) the integrated coronal radiation, \(R_{tr}\) the integrated TR radiation, subscript ‘0’ denotes a quantity at the top of the TR, \(L\) is the loop half-length and subscript ‘a’ a quantity at the loop apex. Solving this set of equations requires the specification of three (semi-)constants that are defined by
\[C_1 = R_{tr}/R_c, \quad C_2 = \tilde{T}/T_a \quad \text{and} \quad C_3 = T_0/T_a,\]
and the physics behind their determination is discussed fully in [15,16]. \(C_2\) and \(C_3\) can be taken as constant, with values of 0.9 and 0.6, respectively. \(C_1\) is, in the absence of gravity, 2 for equilibrium, static loops and 0.6 during radiative cooling. Cargill et al. [16] discusses further the full implementation of \(C_1 = C_1(T_a, L)\), in particular how it can model stratification due to solar gravity.

The first of these equations states that the energy balance of the whole loop is between heating and radiation, with thermal conduction and enthalpy redistributing energy between corona and TR. The second determines the direction of any flow into and out of the corona. When downward coronal conduction into the TR exceeds the radiation there, TR plasma is heated, moving into the corona so that the coronal density increases. In the opposite regime, when coronal conductive losses are small, there must be an enthalpy flux from the corona to power the TR losses, so the coronal density falls. This in turn leads to a characteristic evolution of the temperature and density.

(b) Difficulties

Despite the apparent simplicity, the zero- and one-dimensional hydro approaches have complications.

1. The corona and chromosphere are linked by the thin TR. In a static loop, this is a domain where the heat flux from the corona is balanced by radiative losses: indeed in that case the radiation from the TR is twice that from the corona [16,18]. The strong dependence (at constant pressure) of conduction and radiation on temperature leads to the TR having a very steep temperature gradient, with the total TR thickness being less than 10% of the loop length, and the temperature scale heights are of order 1 km or less in places. Correct numerical modelling of the heat flux requires a concentration of grid points in the TR. If the TR is under-resolved, artificially low coronal densities are the result [14]. The zero-dimensional models seem to handle this problem remarkably well.
(2) The chromosphere is a very complex region and a careful treatment of radiation and atomic physics is required to reproduce the observed emission. However, from the viewpoint of the corona, the chromosphere is best treated as an extended mass reservoir at a fixed temperature. Radiative losses below 20 000 K are assumed to be negligible. It is advisable in one-dimensional models to have a much deeper chromosphere than one thinks is needed.

(3) Below a few MK, coronal radiation is largely due to atomic transitions associated with heavier elements, especially Fe. The losses peak at roughly $10^5$ K and fall off at higher temperatures. Continued improvement in the understanding of the relevant atomic physics has led to an increase in the energy estimated to be radiated to space at coronal temperatures. The change is a factor of four between the 1970s and now [19,20]. This can have significant consequences for the rate of plasma cooling, especially below 1 MK [20,21]. There are also several problems specifically related to loop magnetic geometry that further complicate the comparison of hydrodynamic simulations to observations:

(4) While the large-scale corona is often successfully described by a potential extrapolation of the photospheric field, such a model does not adequately describe the topology of the magnetic field in an active region (AR). The difficulty is that the plasma beta in the photosphere is not small and magnetic field extrapolations are poorly constrained. Considerable effort has been put into developing techniques to approximate the magnetic field at the top of the chromosphere where the plasma beta becomes small, but these efforts have not been entirely successful [22]. The inability to accurately describe the coronal field leads to uncertainties in the loop length, which has important implications for many loop properties, such as the predicted EUV intensity and total loop cooling time. However, preliminary modelling efforts that incorporate magnetic field variation as a function of height in a controlled manner within the three-dimensional MHD equations are starting to quantify these issues [23].

(5) The internal structure of coronal loops presents another complication. Magnetic flux tubes are ultimately rooted in very narrow inter-granular lanes of the photosphere and expand through the chromosphere, TR and the corona. This change in flux tube cross section with height is not well understood. Observations suggest that the flux tube expansion in the corona is much less than anticipated [24,25] and a constant loop cross section is often assumed in models without any physical explanation. Additionally, most observations of the corona have been carried out with a spatial resolution of 1000 km or more. Only very recently has coronal plasma been imaged at a resolution below 200 km [26]. It is possible that what appear to be isolated, individual loops are actually composed of much smaller scale threads. The decomposition of loops into smaller sub-elements that evolve independently may explain why coronal loops persist for longer than a radiative cooling time [27].

(6) The complex, three-dimensional structuring of the solar upper atmosphere further complicates the comparison of hydrodynamic simulations and observations. Magnetic flux closes at different spatial scales and this gives rise to many unrelated structures along the line of sight. In the quiet Sun, for example, about 50% of the flux closes at heights below 2.5 Mm [28] and the footpoint of a coronal loop, which should provide important boundary conditions for simulations, is likely to be buried in a forest of low lying structures at TR and chromospheric temperatures. AR plage appears to be essentially unipolar, but spicular emission may still obscure loop footpoints and make quantitative comparisons more uncertain [29].

(c) Sample results

Figure 1 shows the response of a loop of half-length 33.75 Mm to a nanoflare train of square-wave heating pulses each lasting 200 s. The upper plot shows the heating per unit volume, the lower the temperature and density, the EBTEL and Hydrad [30] being the dashed and solid lines,
Figure 1. The response of the plasma to a nanoflare train in a short coronal loop ($2L = 67.5$ Mm) over an interval of roughly 7 h. The nanoflare duration is 200 s. Panel (a) shows the heating per unit volume and panel (b) the temperature (uppermost pair of curves) and the density (lower pair). The dashed (solid) lines are the zero-dimensional EBTEL (one-dimensional Hydrad) models. (Online version in colour.)

respectively. The agreement between zero and one dimension is very acceptable (Cargill et al. [16] discuss the errors fully). As an example of the plasma evolution, consider the nanoflare at roughly 3 h. The temperature peaks first during the heating phase, then falls quite rapidly. The density increases over a few hundred seconds and then both temperature and density fall (the enthalpy/radiation phase). There are three phases: (i) heating is balanced by conductive losses up to 200 s, (ii) conductive losses driving an upflow from the TR between 200 and approximately 1000 s at which stage peak coronal density is reached, and (iii) after 1000 s, the loop cools by radiation to space and an enthalpy flux to the TR with an undershoot below the equilibrium temperature at around 2500 s. Thus the loop evolves through a cycle when first conductive cooling dominates, then radiative cooling. The time to cool from the peak temperature to 1 MK is roughly 1700 s.

3. Steady and impulsive heating

There is sometimes considered to be a bimodal approach to coronal heating: steady versus impulsive. From the viewpoint of commonly advocated heating mechanisms, there is no process that is time-independent, although plasma microinstabilities that operate at marginal stability come close. Magnetic reconnection, commonly assumed to be the process going on in nanoflares, is intrinsically unsteady [31,32]. Wave heating is also unsteady [33] due to the feedback of the continually changing coronal density on the dissipation process. So the question is not one of ‘steady versus impulsive’, but the degree of unsteadiness, which can be defined as the ratio of the delay between heating events on a sub-loop or strand ($T_N$) compared to a characteristic plasma cooling time ($\tau_{\text{cool}}$).

$\tau_{\text{cool}}$ can be defined in various ways, depending on the time one chooses to start. From the initial temperature due to a nanoflare, $\tau_{\text{cool}}$ involves conductive and radiative/enthalpy-driven phases. Simple expressions can be found in [34,35]. It is approximately proportional to the loop length and roughly independent of the nanoflare energy. Alternatively, one is often interested in the time taken to cool from the temperature at which the emission measure is a maximum ($T_m$).
In this case, the cooling time is approximately given by radiative and enthalpy-driven cooling. In this phase, one has $T \sim n^l$, with $l = 2$ being used frequently [35–37], while $l \sim 1$ for very long loops [38] and $l \sim 2.5$ for short ones [37,38]. One can derive analytical expressions for such cooling (Appendix of [35]) and find that the cooling time is of order $\tau_{\text{cool}} \sim 3Zk_B T^{1-\alpha}/n$, where $Z$ is a numerical factor involving $l$ and $\alpha$. In general, the cooling is catastrophic (i.e. reaches a zero temperature in finite time) and $\tau_{\text{cool}}$ is of order 1000–2000 s for both scenarios.

One can then define two nanoflare regimes: low frequency (LF) when $TN \gg \tau_{\text{cool}}$ and high frequency (HF) when $TN \ll \tau_{\text{cool}}$. In the former case, a loop cools through all temperatures from $T_m$ down to some minimum (typically assumed as 20 000 K), and in the latter the cooling terminates at some coronal temperature less than $T_m$ at which time the loop is reheated. This definition is perhaps overly simple since as we shall see the important results occur when $TN < \tau_{\text{cool}}$, rather than when $TN \ll \tau_{\text{cool}}$.

4. Emission measure results

Extensive studies of AR core loops have been carried out in the last few years. These are the brightest loops at the heart of ARs and have intense emission over a range of wavelengths. We now try and understand those results in terms of nanoflare heating.

(a) Results below $T_m$

Defining the emission measure $EM(T) = \int n_e^2 dh$, where $dh$ is the distance along a line of sight, the broad temperature coverage of the EUV Imaging Spectrometer (EIS) and X-ray Telescope (XRT) instruments on Hinode has led to determinations of $EM(T)$ from multiple ARs [39–42]. $EM(T)$ has a peak at around $T = T_m = 10^{6.5–6.6}$ and falls away sharply to higher and lower temperatures. Below $T_m$, $EM(T) \sim T^a$ to 1 MK, with $a$ ranging from 2 to 5. Figure 2a shows a typical $EM(T)$ distribution from an AR [39] and for this case $a = 3.1$. It should however be noted that there still remain uncertainties in the determination of the slopes both below and above $T_m$ [43,44].

This part of $EM(T)$ is the observed consequence of radiation/enthalpy-dominated loop cooling and one can derive a simple expression for $EM(T)$ there under the assumption that the nanoflare-heated loop cools from $T_m$ to below 1 MK [7,45]. The emission around any given temperature will be proportional to the time the plasma spends there, so $EM(T) \sim n_e^2 \tau_{\text{rad}}$, where $\tau_{\text{rad}} \sim T^{1-\alpha}/n$ is the instantaneous radiative cooling time. With $T \sim n^2$ in the radiative phase (§3), the result $EM \sim T^{3/2-\alpha} \sim T^2$ arises from a simple fit to the radiative losses. For $T \sim n^l$, $EM \sim T^{l/\alpha+1-\alpha}$ [45]. For the relatively short hot loops seen in AR cores, $l = 2$ is reasonable, so that $a = 3.1$ suggests the loop may have inhibited cooling below $T_m$. Warren et al. [39] interpreted this as being evidence for ‘steady’ heating through a one-dimensional hydro simulation that had $TN \sim 150$ s. Figure 2b shows that HF nanoflares produce a very sharp emission peak around $T_m$, suggesting that this value of $TN$ may be too small. On the other hand, a LF nanoflare model gives too broad a distribution with slope $a = 2.17$.

In fact, the values of $a$ found in [39–42] suggests that there are a range of AR properties some with $TN > \tau_{\text{cool}}$ and some with the opposite property. The EBTEL approach was used to investigate many hundreds of possible heating models involving lengthy nanoflare ‘trains’ such as shown in figure 1 with different nanoflare energy distributions [34]. It was concluded that only nanoflare trains where (i) the time between each event was proportional to the energy of the second event and (ii) the nanoflares had a power law distribution were satisfactory. Nanoflares that were equally spaced in time could not reproduce the range of values of $a$ even though they had random or power law energy distributions. The difficulty lay in obtaining values of $a > 2$ because loops either cooled to below 1 MK or had sharply peaked EM. The importance of the energy-dependent waiting time was that a subset of the loops (those with the largest energy release) were able to cool to near 1 MK, while many others did not. Figure 3 demonstrates this further.
Figure 2. Panel (a) shows \( EM(T) \) derived from Hinode EIS and XRT observations of an AR core. The parabolas are EM loci curves (line intensity over emissivity) for each line. The black curves represent 250 Monte Carlo simulations of the DEM. Panel (b) shows comparisons of the observed DEM with the DEMs derived from a single loop heated at LF (\( T_N = 1800 \) s: LF) and HF (\( T_N = 150 \) s: HF). Neither model provides satisfactory agreement with data (central curve) suggesting that more complicated heating scenarios are required. (Adapted from [39].) (Online version in colour.)

Figure 3. Panel (a,b) shows the parameter \( a \) where \( EM(T) \sim T^a \) as a function of the waiting time (\( T_N \)) between nanoflares. The left and right columns show power law distributions of nanoflares without and with an energy-dependent waiting time, respectively. Panels c and d (e and f) show a typical profile of temperature (density) as a function of time for these two cases.
Panels (a,b) show the outcome of a sequence of 20 loop models where we calculate the parameter $a$ as a function of the average waiting time between nanoflares, $(T_N)$ where a power law distribution of nanoflare energy with slope $-2.5$ is assumed. The left (right) columns show cases where the nanoflares are equally spaced (have a waiting time proportional to the energy of the second nanoflare). In all cases, $2L = 80$ Mm and the energy input is at a level commensurate with the AR presented in [39] leading to $T_m = 10^{6.6}$ K. It can be seen that in the absence of an energy-dependent waiting time, there is a very abrupt transition between the expected LF value of $a$ (approx. 2), and cases where there is no plasma below approximately $10^{6.3}$ K. On the other hand, when such waiting time is included, there is a continual range of $a$ between 2 and 5. The other panels show temperature (c,d) and density (e,f) as a function of time without (left) and with (right) a waiting time. The important point is that when an energy-dependent waiting time is included, the loop temperature cools to below 1 MK before large events frequently enough to extend $EM(T)$ over the entire observed temperature domain. Equal spacing between nanoflares implies that the loop either always or never cools to 1 MK, hence giving the abrupt transition in $a$.

(b) High-temperature non-flaring coronal plasmas

The ‘hot’ coronal component is defined as the plasma above $T_m$. The LF nanoflare scenario predicts plasma well in excess of $T_m$ [7,45] and this is seen in the LF $EM(T)$ plot on the right of figure 2. Conduction dominates the cooling when $T > T_m$ which leads to $EM(T) \sim T^{-11/2}$ [45]. However, there are theoretical difficulties in characterizing this component due to, inter alia, departures from ionization equilibrium [46] and the inapplicability of the standard Spitzer–Härm treatment of thermal conduction [47].

It is also non-trivial to detect such plasmas with contemporary instruments. The difficulty is primarily due to the large separation in wavelength between the strong emission lines formed at coronal temperatures ($T \sim 10^{5.7-10.7}$ K) and those formed at temperatures generally observed only in flares ($T \sim 10^{7.0}$ K). This difference in wavelength necessitates the use of multiple instruments. The spectral regions observed with Hinode/EIS (170–211 Å and 245–290 Å), for example, include emission lines from Fe VIII–Fe XVI and Ca XIV–XVII, which provide good coverage of coronal temperatures. To observe higher temperature emission with good signal-to-noise requires observations at soft X-ray wavelengths with telescopes such as XRT on Hinode. As temperature discrimination in these instruments is achieved through the use of thick focal plane filters, they cannot easily observe high-temperature emission at low-emission measure [48].

Despite this, evidence for the presence of such plasmas is becoming overwhelming. Hinode measurements [49–51] were made using XRT and EIS, and [50] removed any ambiguity in the temperatures measured by XRT channels through the confirmation of hot plasma by EIS Ca XVII. Ugarte-Urra & Warren [52] studied the Fe XVIII line and constructed a distribution of the duration and number of ‘brightenings’ in this line. They concluded that these brightenings occurred at least two to three times an hour along the line of sight through any pixel. Further evidence for such hot plasmas has been found in a recent rocket flight [53].

Spectrally resolved observations at soft X-ray wavelengths were in fact obtained with the Flat Crystal Spectrometer and the Bent Crystal Spectrometer on the Solar Maximum Mission (SMM) in the 1980s. A re-analysis of these observations of quiescent ARs suggests that the slope of the differential emission measure (DEM) at temperatures above $10^{6.7}$ K can be extremely steep [54]. Parametrizing the DEM as $EM(T) \sim T^{-b}$, they obtained values for $b$ between 14 and 24. (This result may be affected by the limited number of emission lines formed near the peak in the DEM and the relatively small sample size (only six of the 40 available SMM observations were considered to be quiescent).) Warren et al. [40] analysed 15 Hinode and SDO AR observations and obtained somewhat smaller values for the slope above the maximum ($b \sim 6–10$), but this study only included a single emission line formed near $T \sim 10^{7.0}$ K. Both values are larger than those predicted in [45].
Figure 4. Time-dependent simulations of AR 11089. Panel (a) shows the pre-processed magnetic field and randomly selected field lines. The nonlinear force-free extrapolations were computed by Wiegelmann et al. [60]. Panels (b,c) show the observed and simulated Fe XVIII emission from the AIA 94 Å channel respectively. The scaling for both images is the same. Panel (d) shows the AIA 171 Å image. The simulated DEM from a $10 \times 10$ region in the AR core is also shown for comparison with the observed DEM shown in figure 2. (Online version in colour.)

(c) Global active region results

The computational efficiency of the EBTEL model makes it possible to consider time-dependent heating scenarios for global AR simulations for the first time. Previous work on full AR modelling focused on steady heating, which captured some of the morphological properties of high-temperature AR emission but did not account for the temporal variability in the observations [55–58]. Recently, Warren et al. [59] combined nonlinear force-free extrapolations, volumetric
heating rates for each field line derived from the magnetic field strength and loop length, and the heating event waiting time distribution suggested in [34] to simulate the high-temperature emission in 15 solar ARs as a function of time. These simulations are able to reproduce the variation of the integrated Fe XVIII emission with total unsigned magnetic flux, the general shape of the observed DEM above about 10^6.2K, and many aspects of the observed intensity fluctuations. An example simulation for AR 11089, the same AR from which the DEM shown in figure 2 is derived, is illustrated in figure 4.

While these simulations capture several important properties of observed ARs, they also expose a number of limitations in our understanding of the coronal heating process. For example, to bring the models into better agreement with the observations, the heating on field lines connected to the strong fields in sunspots and on the very short field lines across the neutral line had to be turned off. The strong magnetic fields in sunspots appear to produce only weak heating, perhaps because convective motions in the photosphere are inhibited there. Similarly, because the total cooling time for an impulsively heated loop scales linearly with the loop length [35], any discontinuities in the heating lead to rapid cooling for the very short (approx. 20 Mm) field lines that span the AR neutral line. This is inconsistent with the observed 171 Å images, which, as is illustrated in figure 4 do not show short, bright loops in the core of the AR. The behaviour of the million-degree emission in an AR is generally not well reproduced by these models. It is also clear that the NLFF extrapolations for these regions [60], while clearly more realistic than simple potential calculations, only partially capture the morphology of the AR.

5. Discussion

(a) Magnetic environment for heating

We noted in the Introduction that the nanoflare scenario involved the stressing (or braiding) of the coronal magnetic field by photospheric motions until a critical shear was reached. It is commonly assumed that this dissipation leads to the relaxation of the field to a nearly potential state, whereupon the energy build up recommences (the dissipation timescale is much more rapid than the build-up time). Defining the coronal magnetic field in terms of axial and transverse components (B_a and B_t, where B_a lies in the direction between the photospheric footpoints and B_t is perpendicular to B_a), then for a photospheric velocity of v, B_t ~ vB_a/2L provided B_t(t = 0) = 0. The energy injected into the corona (Poynting flux) is vB_aB_t/4π and equating this to the nanoflare energy Q = 2LA_h(B_t^2/8π) gives Q ~ T_N^2. Here, A_h is the cross-sectional area of the magnetic strand and B_tN = B_t(t = T_N). To account for AR losses, the time taken to build up the required energy is of order tens of thousands of seconds for typical parameters. If the nanoflare occurred sooner, with a lower level of stress in the field, then the observed coronal radiation cannot be accounted for in the case of the AR loops discussed here. In turn, this requires that B_tN must be in the range 0.25–0.5B_a which in turn gives nanoflare energies of a few 10^{24} ergs. This has several important consequences.

1. Models relying on the reduced MHD approximation [61–63] cannot account for AR heating as their fundamental assumption of B_t ≪ B_a will be violated. In fact when Dahlburg et al. [62] included equation (2.1), temperatures of order 1 MK were obtained. Cargill [34] and López Fuentes & Klimchuk [64] have presented comparable but slightly different arguments as to how this can happen. If a stressed field relaxes by only a small amount, the Poynting flux into the loop is large, due to its proportionality to B_t. This leads to an energy input proportional to T_N^2 rather than T_N but the larger Poynting flux ‘wins’ over the slower build up to give the required values of T_N. This leads to a very different scenario for AR heating which requires a highly stressed magnetic field undergoing small relaxations. This in turn reduces the nanoflare energy by roughly an order of magnitude to a few 10^{23} ergs while increasing
the number of nanoflares [34]. The following questions then arise. How does the loop initially reach the required stressed state? Does it by-and-large emerge highly stressed? What MHD process gives these small relaxations? The answers are not known yet, but the final question may be resolved by ‘sandpile-like’ models where a stressed field relaxes through a range of event sizes [64,65].

(3) Complete models of wave heating in closed magnetic structures have been developed recently [66–68]. Alfvén waves from both loop footpoints interact, leading to a turbulent cascade and dissipation on small scales. For reasonable footpoint parameters, an adequate energy supply reaches the corona. However, van Ballegooijen et al. [66] noted that dissipation of these waves leads to a corona with only small temperature fluctuations, in other words, something close to steady heating which does not seem to be consistent with the AR emission measure results. It should be noted that obtaining good resolution in the plane perpendicular to the loop axis is difficult in this class of models, so future developments with more computing power could change this conclusion.

These comments should not be construed as saying that the various processes such as braiding with $B_l \ll B_a$ or Alfvén wave cascades do not occur in the closed corona. Indeed in other locations such as the coronal background, they could well be important processes. But it is clear that they do not, at this time, seem to be able to account for the properties in AR cores.

(b) Energetic particles

The recent results of Testa et al. [69] from the Interface Region Imaging Spectrograph (IRIS) mission are a major step forward in understanding energy release in the corona. They detected unambiguous signatures of mass motions due to precipitation of energetic particles (tens of keV) in the Si IV line as measured by the IRIS spacecraft. The data appear to be fitted by events that release of order a few $10^{25}$ ergs in the particles, so the whole nanoflare is of order $10^{26}$ ergs. While it has long been believed that the reconnection process is an efficient particle accelerator [31], this is the first identification of particle acceleration at such event energies.

We can make an estimate of what happens to coronal plasma properties when a fraction of a nanoflare energy goes into energetic particles by a simple modification of the EBTEL approach. By defining a mean particle energy ($E$) and an electron flux ($F$), the injection of a particle beam can be incorporated simply into the EBTEL formalism (equation (2.2)) when $E \gg kT$ by adding a term $|F|/L$ to the first equation and $|F|$ to the fluxes in the second and third equations (see [15, §2.1] for a more general implementation). If the nanoflare energy and the partition between heating and accelerated particles are the two input parameters, in general terms the coronal density (temperature) will rise (fall) as the fraction of accelerated particles increases. Klimchuk et al. [15] showed an example where 50% of the nanoflare was in the form of energetic particles (similar to what is argued in [69]), and the hot plasma component above $T_m$ was largely suppressed. It was also pointed out that such an electron beam is the equivalent of footpoint heating, from the viewpoint of the coronal temperature and density.

6. Conclusion

The advances in understanding coronal heating discussed here and elsewhere in this issue have been made possible by a combination of a new generation of coronal observations and increasingly capable modelling techniques. Looking to the future, there would seem to be several desirable lines of inquiry.

(1) High-resolution observations of the corona have been carried out on the pioneering HiC rocket flight which has provided much preliminary information of how the TR and corona looks on scales of 0.2 [26]. For example, Testa et al. [70] noted that brightenings in the Fe XIII moss suggested nanoflare energies perhaps as low as a few $10^{23}$ ergs.
(2) Efforts to characterize the ‘hot’ coronal component need to be continued in conjunction with modelling efforts to understand how coronal plasmas at such temperatures behave. Key to this are non-equilibrium ionization and the treatment of non-classical thermal conduction. It seems likely that the emission at these temperatures is a very sensitive diagnostic of nanoflare energies.

(3) Despite the fact that the magnetic reconnection process has long been believed to be an effective particle accelerator, the results of Testa et al. [69] represent an important breakthrough. However, the inference of these particles is through a proxy, namely characteristic mass motions. Radiation signatures associated directly with the particles would be desirable. All of these goals require major new projects to observe the Sun from space either from existing instruments, such as Nuclear Spectroscopic Telescope Array for energetic particles, or future projects with high resolution and advanced spectroscopy such as the Marshall Grazing Incidence X-ray Spectrograph and the JAXA-led Solar-C.

(4) Theoretical modelling of plasma processes needs to be continued and, as computational abilities are enhanced, remove many of the existing commonly used assumptions. Of particular interest would be an assessment of the reality of sandpile models within the formal MHD framework. It is perhaps too much to expect small-scale models of magnetic field dissipation to be included in global models, but the former do give details of dissipation processes.

Acknowledgements. We thank many colleagues for their advice and input, in particular Jim Klimchuk. We also thank ISSI Bern for hosting the International Team led by S.J.B. and Dr Helen Mason, during whose meetings many of the issues presented here were discussed in detail.

Funding statement. H.P.W. was supported by NASA’s Hinode programme. S.J.B. thanks NASA for support under grant no. NNX11AF13G.

References


