The way forward for coronal heating

Ineke De Moortel, Philippa K Browning, Stephen J Bradshaw, Balázs Pintér and Eduard P Kontar consider approaches to the longstanding and enigmatic problem of coronal heating, as presented at the RAS discussion meeting on 11 January 2008.

In the 1940s it was demonstrated that the temperature of the plasma in the Sun’s outer atmosphere (the solar corona) is over a million degrees Kelvin (Edlén 1942). This is rather surprising, given that the Sun’s surface temperature is only about 6000 K. Although this discovery removed the difficulty of an otherwise unknown element (“coronium”), it presented a new puzzle, explaining how the coronal plasma is heated to such high temperatures. The coronal heating problem requires us to find a mechanism (or mechanisms) to supply the energy losses of the hot coronal plasma. These energy losses are due to conduction and radiation — estimated to be up to $10^4 \text{W m}^{-2}$ in active regions. The key to this problem is accepted to be the strong coronal magnetic field, which plays a crucial role in the solar corona (the plasma β—the ratio of thermal energy density to magnetic energy density — is around $10^{-4}$).

There is a very strong correlation between the brightness of coronal emission and the strength of the magnetic field. Indeed, active regions, which are bright in X-ray and EUV images and hence hotter, have a much greater heating requirement than the quiet Sun, and also have the strongest magnetic field. There is now a substantial body of evidence from other stars that X-ray coronae are associated with magnetic fields. Indeed, there is a good correlation of X-ray luminosity with magnetic flux over many orders of magnitude, ranging from solar quiet regions through active regions, to dwarf and T Tauri stars (Pevtsov et al. 2003).

Energy transfer

The basic paradigm of coronal heating is that there is an energy transfer from kinetic energy of flows below the solar surface into free magnetic energy in the corona, through motion of the footpoints of the coronal magnetic fields. Existing theories can be classified according to the ratio of the timescale of this photospheric driving and the Alfvén wave transit time along a coronal field line. When the driving is slow compared to the Alfvén travel time, we have DC heating, with quasi-static currents in the corona; in the opposite case, we have AC heating, with coronal magnetohydrodynamic (MHD) waves.

In both cases, the main challenge is to explain how energy is dissipated in the solar corona: because the conductivity is extremely high, the ratio of the ohmic dissipation time to the Alfvén time (the Lundquist number) is very large, around $10^{13}$. Thus, in order to explain coronal heating, the energy dissipation must take place on scales smaller than typical MHD scales, where kinetic effects are likely to be significant. Moreover, the theories must include the coupling between global scales, i.e. on which the photospheric driving occurs, and local scales, where the dissipation must take place. Further difficulties for theory are presented by the coupling between the dense interior and tenuous outer atmosphere, as well as the complex geometry and topology of coronal magnetic fields, and the dynamic nature of the corona.

Theories of wave heating have to account both for the transmission of waves into the corona and for dissipating the wave energy with adequate efficiency. Fast waves are totally reflected before reaching the corona, and the energy flux in acoustic waves is too low, so attention mainly focuses on Alfvén waves. It has been suggested (Hollweg 1984) that in long loops (>10000 km), enough energy for coronal heating can be transmitted by loop resonances. Efficient dissipation requires the generation of small length scales, with favoured mechanisms, both relying on the inhomogeneity of the corona, being phase mixing and resonant absorption. Spatial gradients of Alfvén speed lead to phase mixing of shear Alfvén waves (Heyvaerts and Priest 1983). In a closed field configuration, with the wavelength fixed by the fieldline length, each fieldline oscillates with a different frequency; thus, neighbouring fieldlines become increasingly out of phase and strong spatial gradients build up, leading to enhanced dissipation. The damping time scales with the cube root of the Lundquist number. The theory has been extended to account for effects such as the generation of Kelvin–Helmholtz instability due to the velocity shear (Browning and Priest 1984), nonlinearities and mode coupling (e.g. Nakariakov et al. 1997), gravitational stratification (De Moortel et al. 1999) and a diverging field geometry (Ruderman et al. 1998, De Moortel et al. 2000). More recently, phase mixing in collisionless plasmas was modelled by Tsihkarlu et al. (2005).

Resonant absorption

Resonant absorption (Ionson 1978) considers resonances of incoming waves where the wave frequency matches the local Alfvén frequency, creating narrow layers where the wave amplitude builds up and energy is dissipated. Interestingly, resonant absorption may lead to sporadic heating, more akin to the nanoflare scenario described below; since the localized heating at the resonant layer will create a rise in density due to chromospheric evaporation, altering the Alfvén speed profile and creating new resonant layers (Ofman et al. 1998).

At the RAS meeting, T Van Doorsselaere (University of Warwick) compared resistive and
viscous dissipative effects on footpoint heating by waves. According to analytical and numerical results, ohmic heating has to dominate any wave heating mechanisms in order to accommodate the constraint of footpoint heating (Van Doorsselaere et al. 2007).

Both ground- and space-based observations have recently found evidence that waves are omnipresent in the solar atmosphere. For example, Tomczyk et al. (2007) report observations of upwardly propagating coronal waves from intensity, line-of-sight velocity and linear polarization measurements obtained with the Coronal Multi-Channel Polarimeter (CoMP) instrument. However, the detected waves have an energy flux much too low for coronal heating requirements, although unresolved waves may have more energy.

Early results from Hinode show similar evidence for the ubiquitous nature of waves in the solar corona. Using the Hinode Solar Optical Telescope (SOT), De Pontieu et al. (2007) find strong transverse displacements of spicules in the upper chromosphere, with an energy flux (~100 W m$^{-2}$) that does appear sufficient to heat the quiet Sun corona or drive the solar wind. Similarly, Okamoto et al. (2007) report transverse oscillations of filamentary structures in prominences observed by SOT. All the above authors have interpreted their respective observed oscillations as Alfvén waves. However, this interpretation was recently challenged by Van Doorsselaere et al. (2008), who claim that an explanation in terms of guided kink magneto-acoustic waves is more appropriate.

Wave heating is a strong candidate for open field regions but is probably less viable for closed field regions, partly due to the short Alfvén time in active region loops (1–30 s). Even if not the prime source of coronal heating, coronal waves are important as a diagnostic of coronal parameters, through the burgeoning science of coronal seismology (see, for example, reviews by Chaplin and Ballai 2005, De Moortel 2005, Nakariakov and Verwichte 2005).

Consider now the alternative scenario, DC heating: the Poynting flux of energy into the corona can be expressed as:

$$F = \frac{1}{2} B_v B_h v_{ph}$$

where $B_v$ and $B_h$ are the vertical and horizontal magnetic field components, respectively, and $v_{ph}$ is the photospheric footpoint velocity. Taking typical coronal values, $B_v = 0.01$ T and $v_{ph} = 1$ km s$^{-1}$, gives a Poynting flux of about $10^4$ W m$^{-2}$: sufficient for active region heating, as long as the horizontal field component is around 10% or more of the vertical field in magnitude. The latter condition imposes an important constraint on theories, since it is required to explain why dissipation only “switches on” when this value is reached. While energy flux is sufficient, global ohmic dissipation of the DC currents in the corona is far too inefficient for coronal heating purposes. However, there is strong evidence that current sheets or filaments are omnipresent in the corona, and these form sites for localized efficient dissipation of energy by magnetic reconnection (see, for example, Priest and Forbes 2000 and Biskamp 2000 for an overview of this important physical process).

Reconnection models

Models of magnetic reconnection may be classified as: spontaneous linear instability (such as tearing mode), steady driven reconnection (such as Sweet-Parker or Petschek) and forced reconnection (triggered by an external disturbance; Hahm and Kulsrud 1983). The latter could be most relevant to coronal heating, with disturbances arising from photospheric footpoint motions, for example. An analytical theory of coronal heating by forced reconnection has been presented by Vekstein and Jain (1998), with numerical simulations of the nonlinear effects recently performed by Jain et al. (2006). An important challenge for reconnection theory is to incorporate kinetic effects, which are significant at small length scales. For example, the Hall effect, which allows separation of electron and ion fluids, can significantly affect the reconnection rate (Vekstein and Bian 2006). Magnetic reconnection during collisionless X-point collapse is also an efficient source of heat, according to the computational simulations by D Tsiklauri (Salzburg), presented at the meeting.

Parker (1988) proposed that the corona is heated by the combined effect of many small (and currently unobservable) nanoflares – tiny heating events with the same energy release mechanism as large-scale flares, namely magnetic reconnection, but much weaker (of the order of $10^{24}$ ergs). Typical solar flares, releasing $10^{23}–10^{24}$ ergs, are not frequent and therefore they have inadequate total energy output for coronal heating purposes. However, if small events occur sufficiently frequently, their combined effect may be more significant; the issue depends on whether the spectrum of the frequency of occurrence vs energy is steep enough (Hudson 1991). Hugh Hudson (University of California, Berkeley) in his invited review talk proposed the interesting and rather controversial idea that flares and microflares actually cool the corona. The argument is that flares reduce the mean magnetic energy and hence, because heating is known to correlate with magnetic field strength, reduce the coronal heating.

A theoretical model by Browning et al. (2003) predicts sporadic heating of a single coronal loop by a series of discrete events of varying magnitude, which may be regarded as nanoflares. Each heating event is triggered by kink instability, while the energy release is predicted by assuming a helicity conserving relaxation to a minimum energy state. The theory has been confirmed recently by 3-D numerical simulations of a kink unstable loop (Browning et al. 2008). This approach allows for the distribution of nanoflares to be predicted ab initio.

New results from RHESSI have put significant
limits on heating by microflares with typical energies of $10^{28} - 10^{29}$ ergs. A study of 25000 microflares by Hannah et al. (2008) shows that these events, like ordinary flares, exhibit high temperatures and a non-thermal component correlated with the thermal component. All occur within active regions and are thus concentrated in bands of latitude. The frequency distribution has a similar exponent to flares, with the spectrum being insufficiently steep for coronal heating requirements. However, this still leaves open the possibility that smaller nanoflares, with somewhat different physics, may be responsible for heating the corona. Investigating the observational consequences of nanoflare heating is thus very important. SXT/Yohkoh and TRACE observations have recently been analysed by G Vekstein (Manchester) and co-workers to find evidence for significant nanoflare heating, and to determine the properties of nanoflares, extending the methodology outlined in Vekstein and Katsukawa (2000).

A promising approach is to discriminate observationally between different heating models by comparing the predicted scaling of the heating rate of various models (particularly in terms of the field strength $B$ and loop length $L$, since other parameters can be expressed in terms of these) and to see which fits best with observations (Mandrini et al. 2000). Unfortunately, the initial results are inconclusive, in that most models are consistent with the data. A more advanced version is to create synthetic images of extrapolated coronal loops from measured photospheric magnetograms, using different scalings for the heating rate ($\dot{E} \propto B^\beta L^\alpha$), as shown in figure 1. By comparing with actual images, the best fit exponents ($\alpha$ and $\beta$) can be found (Warren and Winebarger 2006). Using the same empirical scaling laws for the heating function, Stéphane Régnier (St Andrews) presented recent work at the meeting on a comparison between potential and nonlinear force-free magnetic field extrapolations, showing the importance of both currents and the complexity of the field in studying the heating mechanism.

The approach seems very promising but, as pointed out by Hudson in his review, there are some difficulties: the active region filling factor is small (i.e. how much of the coronal volume is actually contributing to the emission observed in a particular wavelength?) and adjacent flux tubes have almost identical $B$ and $L$, while having very different emission (see for example, the image from XRT/Hinode in figure 2). Simple scaling models cannot account for filamentary structure!

Comparing theory and observations is obviously crucial to “proving” any coronal heating theory. However, this task is not always as straightforward as it may sound. Theoretical models do not often yield parameters that are easily compared to observed quantities and, at the same time, the interpretation of observational data is subject to many possible sources of error and confusion, such as the optically thin radiation in the solar corona or the unknown filling factor. To compare theory and observations, one could try to deduce basic parameters such as temperature or density from the observational data (e.g. Prato et al. 2006). However, it has long been known that the inversion of data is often an “ill-posed” problem, i.e. that the solutions are non-unique and additional constraints are needed to resolve this non-uniqueness. Various inversion algorithms for model-independent reconstruction of electron spectra have proved to work well for continuum X-ray emission (Piana et al. 2003, Kontar et al. 2005), although the situation is more complicated for line emission. For example, Wikstøl et al. (1998) argued that line emission observations alone are not sufficient to explain the nature of the solar transition region and additional information is needed. The other approach, often termed “forward modelling”, is to calculate observable parameters (for example, intensities or Doppler shifts) for instruments from theoretical models. These “synthesized” observations can then be directly compared to real observational data.

**Forward modelling**

Forward modelling provides a robust link between theory and observations, allowing us to start from a certain physical model and subsequently to examine the observable consequences of this model. This then provides a way to test the sensitivity of the observables on the input parameters of the model or to compare the observational consequences of different theoretical models put forward to explain the same solar phenomenon. At best, this will allow us to distinguish between different theoretical suggestions or even to eliminate those models that cannot “reproduce” the observations. On the other hand, it is of course possible that substantially different theoretical models yield observable parameters which are too similar to distinguish the different models confidently. However, the strength of forward modelling lies in its predictive character: predicting observational consequences of theoretical models at its most basic level, providing guidance for the design of specific observing programmes, but also predicting the potential achievement and hence guiding the development of future instrument design. Crucial to forward modelling of line emission are atomic data packages, such as CHIANTI (Dere et al. 1997, Young et al. 2003), which provide essential parameters such as element abundances, ion population fractions and ion emissivities.

Generating observable quantities from theoretical models through forward modelling has been successfully demonstrated in the past by numerous authors, for various solar phenomena.
but the greatest effort has focused on the effects of coronal nanoflares. With a few selected examples, we will try to demonstrate the potential of forward modelling. During the 1990s, a series of papers by Wikstøl and Hansteen determined a range of observables associated with wave propagation in the solar atmosphere. Using a basic, 1-D model (including the effect of non-equilibrium ionization), Hansteen (1993) shows that the redshifts observed in transition-region spectral lines can be explained in terms of downward propagating, compressive (acoustic) waves, generated at the apex of coronal loops by nanoflares (or some other form of episodic heating). More recently, modelling of the response of transition region lines to nanoflares by Taroyan et al. (2006) also suggests that nanoflares might be responsible for the observed transition-region redshifts. Demonstrating the predictive power of forward modelling, Wikstøl et al. (1997) look for observable signatures of wave energy, propagating in different directions through the solar atmosphere, which survive spatial and temporal averaging. As a basic model, the flow of wave energy in different directions can be associated with two different coronal heating models, namely upward propagating acoustic waves.
and nanoflare heating (generating downward propagating waves). In particular, these authors predict that the associated spectral signatures should be observable by the SUMER instrument onboard SOHO, both in line profiles and line ratios of density-sensitive line pairs. Such an approach could not only be applied to the solar corona, but also to (unresolved) stellar sources.

**Nanoflare heating**

Other authors have also considered forward modelling of nanoflare heating in the solar corona, in particular Cargill, Klimchuk and co-authors (Cargill 1993, 1994, Cargill and Klimchuk 1997, Klimchuk and Cargill 2001, Cargill and Klimchuk 2004, Petsourakos and Klimchuk 2006, Sarkar and Walsh 2006, Buchlin et al. 2007, 2008). Often using basic 0-D or 1-D hydro-models, the effect of nanoflare heating on observable parameters such as emission measure, line profiles and intensities is investigated in detail and again, the predictive power of forward modelling is exploited. For example, observations of the filling factor (the ratio of volume radiating in EUV or X-ray to total volume) could be used to deduce the average size of the energy-release site. Modelling a coronal active region as a collection of several hundred thin loops, heated randomly by thousands of nanoflares, predictions are made in terms of the emission measure (which is shown to exhibit a sharp change in gradient above temperatures of about 1 MK) and filling factors for both broadband instruments such as SXT/Yohkoh and narrowband (spectroscopic) instruments.

Understanding the (differential) emission measure is an important step towards identifying the distribution of coronal densities and temperatures. Klimchuk and Cargill (2001) actually provide tabulated values for line intensity and emission measure for a range of spectral lines, which can directly be compared with spectroscopic observations from, e.g. CDS onboard SOHO, or EIS on Hinode. Such a comparison would be a profound test for the nanoflare heating model! One of the predictions of the studies by Cargill and Klimchuk is the possibility of very low filling factors in the solar corona. Numerical simulations by Bradshaw and Cargill (2006) investigate how nanoflares might heat such a tenuous plasma to very high temperatures and whether any specific observational signatures exist. Departures from ionization equilibrium of up to an order of magnitude are found for iron and the authors suggest that an associated blue-shift at the loop footpoints might be observable by EIS/Hinode.

The study of Petsourakos and Klimchuk (2006) calculates nanoflare line profiles for Ne VIII, Mg X and Fe XVII (see figure 3). The Ne VIII and Mg X profiles agree well with existing observations. The predicted profile for the hot Fe XVII line shows distinct enhancements in the line wings, providing another powerful test for the nanoflare model, as this line is observable with EIS/Hinode. Going one step further still, recent work of Buchlin and co-authors (Buchlin et al. 2007, 2008) couples physical models of nanoflare heating (anisotropic turbulence driven by Alfvén waves) and cooling processes (convection, conduction, and radiation based on atomic physics) in a coronal loop. The heating is found to be intermittent and sufficient to heat the loop to temperatures of more than a million degrees, with realistic values of the amplitude of the forcing (corresponding to motions of the photospheric footpoints of the loop). Spectral line profiles of the UV emission and their time evolution are obtained by forward modelling, and are used to identify signatures of heating processes in observations by, for example, Hinode and STEREO.

**Beyond simple models**

The advantage of using relatively simple 0-D or 1-D hydro-models is that it allows one to vary many of the model input parameters such as the number of loop strands and the location and size (or distribution of sizes) of the heating events. As already mentioned before, the frequency of occurrence of nanoflares vs their energy is thought to follow a power-law, based on the extension of the observed power-law distribution of larger flares down to nanoflare energies (see e.g. Hudson 1991). Hence, of particular interest is the distribution of the (input) nanoflare energy and whether this will be reflected in the observations. This question was addressed by Parenti et al. (2006), who looked at the statistical properties of coronal loops, subject to turbulent heating, to test whether the plasma response simply transmits the statistical distribution of events, without modifying it (in other words, whether the injected and radiated energies follow the same power-law). Two different cooling mechanisms are taken into account, namely conduction and radiation, which both turn out to play an important role. So how does coronal energy transport affect the event distribution, for a known distribution of energy input? It appears that the resulting distribution of intensities strongly depends on the chosen spectral lines (see figure 4); for cool (EUV) lines, the power-law index of the output distribution is strongly modified. However, for hotter lines ($T \sim 10^7$ K, i.e. those lines where thermal conduction is the dominant cooling mechanism), the distribution is well preserved, highlighting the importance of choosing the “right” emission lines for the analysis of the statistical distribution of observed nanoflares.

Although many studies have focused on nanoflares, one of the strengths of forward modelling is that it can be applied to almost any theoretical model. Substantial progress can be made by forward modelling of large-scale, 3-D MHD simulations such as those by Gudiksen and Nordlund (2005), from which a range of
observables can be predicted. These particular simulations model the solar corona ab initio, where the necessary energy to heat the model corona is provided through slow braiding of magnetic field lines by photospheric footpoint motions. Peter et al. (2004, 2006) calculate both intensity and Doppler shifts, providing a range of different diagnostics which can be compared to both imaging and spectroscopic observations. The synthesized emission shows a nice qualitative agreement with the general appearance of the solar corona, whereas the Doppler maps show an intriguing amount of detail. In fact, there seems to be substantially more spatial variability in the (coronal) Doppler shifts than in the associated intensity, emphasizing again the need for combined imaging and spectroscopic observations. For a range of spectral lines, a direct comparison is made between the observed Doppler shifts and emission measures and the corresponding values derived from their synthesized spectra (see figure 3). Taking into account the limitations of the underlying 3-D MHD model, these graphs show a remarkable agreement for both parameters, in almost the entire range of calculated spectral lines.

A lot of forward modelling so far has in some way or another been related to the long-standing coronal heating problem. However, a few recent studies have focused on forward modelling of MHD wave behaviour in the solar atmosphere, which is of particular relevance to the rapidly developing field of coronal seismology (De Moortel 2005, Nakariakov and Verwichte 2005). For example, the interpretation of loop oscillations observed by SUMER/SoHO and SXT/Yohkoh in terms of slow standing modes, initiated by a microlare at the loop footpoint, is investigated by Taroyan et al. (2007). Synthetic spectra are produced for the FeXIX, CaXV and CaXIII lines, which are directly comparable with SUMER observations. The time profiles of the synthesized intensity and Doppler shifts show a good qualitative agreement with the observed oscillations and confirm that these particular hot loop oscillations will show up much stronger in Doppler shift than in intensity. Although not modelling a particular physical process, De Moortel and Bradshaw (2008) investigate intensity perturbations associated with coronal density oscillations and show that care has to be taken with the interpretation of some parameters derived from observed intensity oscillations. In particular, it is found that the damping rate of the “input” density perturbations and the resulting, synthesized intensity oscillations can be substantially different.

Many other examples of forward modelling can be found in the literature. To mention but a few, Innes and Töth (1999) and Sarro et al. (1999) study the dynamical response of several different spectral lines to Petschek-like reconnection and apply their results to explosive events in the chromospheric network. The latter authors pay particular attention to the effects of non-equilibrium ionization. Observables associated with catastrophic cooling and downflows in coronal loops are investigated by Müller et al. (2004, 2005), Lundquist et al. (2004) demonstrate how forward modelling can lead to observational constraints for coronal heating mechanisms and Aioouaz et al. (2005) show that different heating profiles in coronal funnels lead to different Doppler shifts in the Ne VIII line. Martínez-Sykora et al. (2008) describe the response of the solar atmosphere to magnetic flux emergence with magnetograms, synthesized continuum and CaII H-line images.

For the future

From the theoretical point of view, a number of areas of future work are required in order to make progress on coronal heating: further studies of reconnection (including 3-D models, kinetic effects and relation to external driving) and waves (including nonlinear studies, kinetic effects, more realistic geometries and coronal seismology). Observational challenges include a better understanding of the fine structure of coronal plasma: what are the minimum temperature and density in an active region loop? Also, further consideration must be given to the time series analysis of X-ray brightness, and to the electrodynamics of the chromosphere. Clearly, furthering our understanding of the elusive coronal heating mechanism depends on progress in both observations and theoretical modelling but most importantly, by successfully comparing and linking theory and observations. This necessary link can be provided by forward modelling as its predictive power is an excellent tool to study a wide variety of physical processes. Forward modelling can both aid the (correct) interpretation of observational data to allow comparison with theoretical models and provide guidance and constraints derived from theoretical modelling on observational campaigns and the design of future instrumentation.●

Ineke De Moortel, University of St Andrews (ineke@msc.st-and.ac.uk); Philippa Browning, University of Manchester; Stephen J Bradshaw, Imperial College, London; Balázs Pintér, University of Aberystwyth; Edward F Kontar, University of Glasgow.

Acknowledgments. The authors would like to thank all the participants and speakers, in particular the invited speakers, P Cargill (Imperial College, London), V Hansteen (University of Oslo) and H Hudson (University of California, Berkeley), IDM and SB acknowledge support of a Royal Society University Research Fellowship and STFC Postdoctoral Fellowship, respectively. EPK acknowledges support of an STFC Advanced Fellowship. BP acknowledges the financial support from STFC.

References

Browning P K et al. 2003 A&A 400 355.
Chaplin B and Baliou I 2005 A&G 44 47.
De Pontieu B et al. 2007 Science 318 1574.
Lundquist L L et al. 2004 ESA-SP 575 306.
Okamoto S et al. 2007 Science 318 1577.
Sarkar A and Walsh R W 2006 ESA-SP 617 118.
Tomczyk S et al. 2007 Science 317 1192.
Vekstein G and Jain R 1996 Science 277 1138.