HEATING AND ACCELERATION OF THE SOLAR PLASMA

Thomas E. Holzer

NCAR/HAO, P.O. Box 3000, Boulder, CO 80307-3000, Email: holzer@ucar.edu

ABSTRACT
This tutorial focuses on basic physical concepts relevant to the heating of the corona and acceleration of the solar wind. We begin with a discussion of observations of the solar wind and solar corona and then develop a simple theoretical framework that allows us to explore some implications of the observations. Next we turn to observations of the lower solar atmosphere and their implications for the nature of the photosphere-to-corona energy flux that is required for coronal heating and solar wind acceleration. Finally, we take a brief look into the future, considering some observational and theoretical avenues that may prove productive.

1. INTRODUCTION
Active research on coronal heating and solar wind acceleration has been ongoing for more than five decades. Needless to say, a large number of researchers have contributed to our current understanding of these problems, and this tutorial will focus on a distillation of the basic physical concepts developed by these researchers. We will not, generally, try to connect names with concepts, except in the case of the most recent research. Those who are interested in such connections are referred to detailed reviews of these subjects. Two particularly useful recent reviews have been published by Cranmer (2002) and Marsch et al. (2003), and it is suggested that anyone who finds the material discussed below interesting would benefit from reading carefully both of these reviews.

In the following discussion, we first consider the coronal origin of the solar wind, beginning with a discussion of solar wind and coronal observations in §2 and developing a theoretical framework for understanding these observations in §3. Next we turn to the origin and nature of the energy flux that heats the corona and drives the solar wind, first discussing some basic observations of the solar chromosphere and photosphere in §4 and then outlining in §5 some of the ideas proposed for the transport of energy from the lower solar atmosphere to the corona. Finally, in §6 we consider briefly some of the observational and theoretical efforts that might be useful to pursue in the future.

2. OBSERVATIONS: WIND AND CORONA

Figure 1. The first and second orbits of Ulysses (adapted from McComas et al., 2003). The top panels show solar wind speed as a function of latitude, with the white circles denoting 400 and 800 km/s. The bottom panel shows the sunspot number through the two orbits. Note that each orbit begins in the east (to the left of the sun), in 1992 and 1998, and the orbit progresses counterclockwise.

2.1 The Solar Wind
The first and second orbits of the Ulysses spacecraft have provided us with a clear picture of the 3-D structure of the solar wind (cf. Fig. 1). In the years around solar minimum, there is a bimodal structure of the solar wind, with relatively quiet, high-speed wind flowing at middle to high solar latitudes, and noisier, low-speed wind flowing at low latitudes. In the years around solar maximum, noisy, low- to moderately-fast wind is found at all latitudes. The structure of the solar corona, which can also be seen in Fig. 1, exhibits a similar bimodal character. Near solar minimum coronal streamers are found at low latitudes and large coronal holes at high latitudes, while near solar maximum streamers and much smaller coronal holes are found at all latitudes.

Some more detailed characteristics of the bimodal wind that occurs in the years around solar minimum are summarized in Table 1, which is based on the reviews of Neugebauer (2001) and Feldman et al. (1976). The fast wind flows at nearly twice the speed...
of the slow wind, and, as noted above, shows much less variability. The proton flux density, however, is nearly twice as large in the slow wind as in the fast wind. Interestingly, the total energy flux density (including the energy needed to lift the wind out of the solar gravitational field and that needed to accelerate it to its asymptotic speed) that is required to drive the slow solar wind is very nearly the same as that required for the fast wind. The difference is that in the slow wind, nearly 70% of the energy goes into lifting the wind out of the gravitational field, while in the fast wind, some 60% of the energy is employed to bring the wind to its large asymptotic flow speed. Although only a small fraction of the solar wind energy resides in thermal energy, it is instructive to consider the temperatures of different particle species, particularly when we map the solar wind back to the solar corona (see below). In the slow wind, the electron temperature is much larger than the proton temperature, while the reverse is true in the fast wind. The minor ion temperatures are always greater than the proton temperature, but in the fast wind the ion/proton temperature ratio is even greater than the mass ratio, while it is less than the mass ratio in the slow wind. Finally, the radial component of the magnetic field is about the same in all types of wind, and it is not too variable. This result is particularly important for those using potential field source surface (PFSS) extrapolations of the photospheric field (see below), for the radial field measured at 1 AU provides a measure of the total magnetic flux integrated over the entire source surface. Such a constraint may provide some amelioration of the inaccuracies involved in measuring polar photospheric magnetic fluxes.

Before leaving solar wind observations, it is particularly important to consider measurements of the elemental abundance and ionization state of the wind, for these measurements provide some of the most important clues available in finding the coronal origin of the solar wind. We discuss here only two observational results, which are representative of the rich set of composition data available. The upper panel of Fig. 2 shows a scatter plot of the $O^{7+}/O^{6+}$ ratio as a function of solar wind speed. This ratio provides a measure of the coronal electron temperature in the region where the oxygen ionization state is frozen into the flow (i.e., where expansion dominates ionization and recombination), with larger ratios

$Table 1. Values of some relevant solar wind parameters ($\pm 3\sigma$) at 1 AU in the years around solar minimum$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol/Units</th>
<th>Slow Wind</th>
<th>Fast Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow speed</td>
<td>$u$ (km s$^{-1}$)</td>
<td>$430 \pm 100$</td>
<td>$750 \pm 50$</td>
</tr>
<tr>
<td>Proton flux density</td>
<td>$n u$ (cm$^{-2}$ s$^{-1}$)</td>
<td>$(3.5 \pm 2.5) \times 10^{9}$</td>
<td>$(2.0 \pm 0.5) \times 10^{9}$</td>
</tr>
<tr>
<td>Energy flux density</td>
<td>$\frac{1}{2} \rho u v_T^2 + \frac{1}{2} \rho u^2$ (erg cm$^{-2}$ s$^{-1}$)</td>
<td>$1.9 \pm 0.9$</td>
<td>$2.0 \pm 0.5$</td>
</tr>
<tr>
<td>Electron temperature</td>
<td>$T_e$ (10$^4$ K)</td>
<td>$1.3 \pm 0.5$</td>
<td>$1.0 \pm 0.2$</td>
</tr>
<tr>
<td>Proton temperature</td>
<td>$T_p$ (10$^5$ K)</td>
<td>$0.4 \pm 0.2$</td>
<td>$2.4 \pm 0.6$</td>
</tr>
<tr>
<td>Ion Temperature</td>
<td>$T_i/T_p$</td>
<td>$&lt; m_i/m_p$</td>
<td>$&gt; m_i/m_p$</td>
</tr>
<tr>
<td>Magnetic Field</td>
<td>$B_r$ (nT)</td>
<td>$3.1 \pm 1.1$</td>
<td>$3.1 \pm 1.1$</td>
</tr>
</tbody>
</table>

$Figure 2. Ratios of O$^{7+}/O^{6+}$ and of Mg to O as functions of solar wind speed (from Richardson and Kane, 2004). The photospheric value of the Mg/O ratio is 0.04.$
corresponding to higher coronal electron temperatures. Although the scatter is large, the trend is clear: higher speed wind originates in coronal regions with lower electron temperatures. The lower panel of Fig. 2 shows a scatter plot of the Mg/O ratio as a function of solar wind speed. This ratio provides a measure of the first-ionization-potential (FIP) effect, which is characterized by an elevated ratio (relative to the photospheric value) of the abundance of a low-FIP element to that of a high-FIP element (the dividing energy between low FIP and high FIP is about 10 eV). In this case Mg (FIP of 7.6 eV) is the low-FIP element, and O (FIP of 13.6 eV) is the high-FIP element. Since the photospheric value of Mg/O is about 0.04, we see from the linear fit to the scatter plot that the high-speed wind exhibits little, if any, FIP bias, while the low speed wind shows a FIP bias of about 3. Whereas the ionization state measurements reflect processes (viz., electron collisional ionization and coronal expansion) operating in the corona, the FIP effect is most likely associated with processes occurring in the upper chromosphere, where low-FIP elements are largely ionized, and high-FIP elements are largely neutral.

### 2.2 Mapping from the Wind to the Corona

With these solar wind observations in hand, let us turn to the mapping of measurements at 1 AU back to the point of origin of the wind in the corona. Although many researchers have considered this mapping problem over the past three decades, we base the following discussion on the comprehensive work of Neugebauer and her colleagues (Neugebauer et al., 1998, 2002; Liewer et al., 2004). These authors use a two-step mapping process involving radial, constant-speed mapping from the point of observation in the wind to a spherical surface (the so-called source surface) at 2.5 R⊙. From this surface they use a potential field source surface (PFSS) model to map along potential magnetic field lines to any height between the photosphere (at 1.0 R⊙) and the source surface (at 2.5 R⊙). Although there are uncertainties involved in both steps of the mapping (particularly in the PFSS mapping in the vicinity of active regions), the results shown in Table 2 are not likely to be strongly dependent on these uncertainties. As has been known from the Skylab era (e.g., Hundhausen, 1977), in the years around solar minimum, fast wind comes from polar coronal holes, while slow wind comes from near the boundaries between these holes and the low-latitude coronal streamer belt. During this period of the solar cycle, there is also a slow to moderately fast component of the wind (belying our oversimplified bimodal description) that originates in mid- to low-latitude coronal holes (the faster wind) and their boundary regions (the slower wind). Near solar maximum, the very fast wind disappears and there is only the slow to moderately fast wind observed. As near solar minimum, the solar maximum wind can originate in mid- to low-latitude coronal holes and their boundary regions, but it is also found to originate in the immediate vicinity of active regions (or, perhaps more appropriately, from strong field regions, since there remains controversy over exactly how to define an active region).

### 2.3 The Solar Corona

Now that we have mapped the solar wind back to the corona, let us ask what are the properties of the coronal regions from which the solar wind appears to originate. Table 3 provides a summary of some relevant observations obtained by the UVCS, SUMER, and CDS instruments on SOHO (Raymond, 1999; Parenti et al., 2000; Ko et al., 2002; Frazin et al., 2003; Mancuso et al., 2003; Uzzo et al., 2003; Cranmer, 2005; Vasquez and Raymond, 2005). Polar coronal holes exhibit a very rapid acceleration of the flow in the inner corona, which is consistent with the very high-speed wind thought to arise from these holes (cf. §3). The coronal holes also show little or no FIP bias, which again is consistent with the high-speed wind. Finally, the electron, proton, and ion temperatures in

<table>
<thead>
<tr>
<th>Coronal Region</th>
<th>Region Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polar Coronal</td>
<td>Rapid Acceleration</td>
</tr>
<tr>
<td>Hole (PCH)</td>
<td>Little or No FIP Bias</td>
</tr>
<tr>
<td>Equatorial</td>
<td>PCH Characteristics</td>
</tr>
<tr>
<td>Coronal Hole</td>
<td>Moderated</td>
</tr>
<tr>
<td>Coronal Streamer</td>
<td>FIP Bias = 3 (legs)</td>
</tr>
<tr>
<td></td>
<td>( T_i \approx T_p \approx 1.4 T_e ) (legs)</td>
</tr>
<tr>
<td></td>
<td>( T_i \approx T_p \approx T_e ) (core)</td>
</tr>
<tr>
<td></td>
<td>( \frac{[\text{Ion}^{\text{(core)}}]}{[\text{Ion}^{\text{(core)}}]} \approx 0.1 - 0.2 )</td>
</tr>
<tr>
<td>Active Region</td>
<td>Similar to Streamers</td>
</tr>
<tr>
<td>High ( T_e ) Component</td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 2. Mapping the Solar Wind back to the Sun**

<table>
<thead>
<tr>
<th>Period of Cycle</th>
<th>Type of Wind</th>
<th>Solar Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>Fast</td>
<td>Polar Coronal Holes (PCH)</td>
</tr>
<tr>
<td></td>
<td>Slow</td>
<td>Near PCH Boundaries</td>
</tr>
<tr>
<td>Slow to Moderately Fast</td>
<td>Mid- to Low-Latitude Coronal Holes and Their Boundaries</td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>Slow to Moderately Fast</td>
<td>Mid- to Low-Latitude Coronal Holes &amp; Boundaries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active Regions</td>
</tr>
</tbody>
</table>
holes exhibit the same trends as observed in the high-speed wind (cf. Tables 1 and 3), so the mapping of the high-speed solar wind to polar coronal holes appears to be accurate. Equatorial coronal holes have properties very similar to those of polar coronal holes, although slightly moderated, and this is consistent with the moderately high-speed wind that has been mapped back to these holes.

The properties of coronal streamers compared with those of coronal holes are of particular interest, since these properties offer the opportunity to determine whether the low-speed wind mapped to coronal hole boundary regions originates in the hole itself (i.e., a quasi-steady magnetically open region) or from the leg of the bounding streamer. The streamer leg characteristics shown in Table 3 are clearly much more consistent with slow solar wind characteristics than are those of polar coronal holes. It is thus tempting to infer that magnetic field lines defining streamer legs, which need to be closed in the corona to produce the observed streamer properties (cf. §3), are transiently opened to the heliosphere in order to supply streamer-like plasma to the slow solar wind. Similarly, active regions, whose plasma properties are not too different from streamers (except for the presence of a component that exhibits very high electron temperatures), are not an unreasonable source for slow solar wind, as has been inferred from the solar wind mapping process. The difference is that there may be an adequate heat source in these strong magnetic field regions to provide the plasma with streamer-like properties even if active region field lines are open to the solar wind on a quasi-steady basis. Of course, in this case, it would be necessary to explain the development of a strong FIP bias on quasi-steady open magnetic field lines, and in the absence of such an explanation it would seem more likely that the active region source of slow solar wind involves the transient opening of field lines.

2.4 Three-Component Magnetic Field Model

In summarizing the observations over the last decade of the solar wind and solar corona, it is interesting to step back some three decades to a prediction of the three-dimensional solar wind that the Out-of-the-Ecliptic Mission (which became Ulysses) would observe. Fig. 3 shows Axford’s (1977) three-component model of the coronal magnetic field. He suggests that high-speed wind comes from coronal holes which are threaded by magnetic field lines that are open for long periods, while the lower-speed wind comes from transiently open magnetic field lines near the boundaries of holes or from strong magnetic field regions (i.e., active regions), where magnetic field lines can be either transiently open or open for long periods. His definition of transient (versus long period) is an occurrence on a time scale less than several hours, which is the characteristic time required to establish a quasi-steady flow on a newly opened field line. Upon comparing Axford’s 1977 predictions with the observations from Ulysses and SOHO, his prescience seems rather remarkable.

3. THEORETICAL FRAMEWORK

3.1 Mass Balance

Let us take as the foundation of our theoretical framework the equations for conservation of mass, momentum, and energy for the $j^{th}$ particle species of a multicomponent plasma. The general form of a conservation equation is well illustrated by the mass balance equation:

$$\frac{\partial \rho_j}{\partial t} + \nabla \cdot (\rho_j \mathbf{u}_j) = m_j (Q_j - L_j) \quad (1)$$

Here, $\rho_j$ is the mass density, $\mathbf{u}_j$ the flow velocity, $m_j$ the mass, $Q_j$ the volumetric production rate, and $L_j$ the volumetric loss rate for $j^{th}$ species particles. The first term in Eq. 1 can be thought of as the time rate of change of mass in a unit volume element, the second term as the rate at which mass is flowing out of the volume element, and the term on the right side as the net rate at which mass is produced in the volume element. All our conservation equations will look the same: the time derivative of a density, plus the divergence of a flux density, equals a net volumetric production rate. Eq. 1 can be divided by the particle mass to obtain what is usually called the continuity
equation. Both forms of this equation are often written without the production and loss terms, but if we are to describe the upper chromosphere and transition region, which are crucial to coronal energy balance, these terms must be retained, since the ionization state of all ion species changes dramatically from the upper chromosphere to the corona. Even in a description starting at the coronal base, production and loss must be included for ions heavier than He, since their ionization state is not fixed until well above the coronal base.

3.2 Momentum Balance

By analogy with Eq. 1, momentum balance is described by

\[
\frac{\partial \rho_j}{\partial t} + \nabla \cdot (\rho_j \mathbf{u}_j) = -\nabla \cdot \mathbf{p}_j - \nabla \cdot \mathbf{\Pi}_j + n_j e Z_j (\mathbf{E} + \mathbf{u}_j \times \mathbf{B}) - \rho_j (G M / r^2) \mathbf{\hat{e}}_r + \mathbf{F}_{fj} + \mathbf{F}_{j} + m_j \left( \sum_l Q_{jk} \mathbf{u}_k - L_j \mathbf{u}_j \right)
\]

Here, \( p_j \) is the pressure, \( \mathbf{\Pi}_j \) the stress tensor, \( e \) the electron charge, \( Z_j \) the charge number, \( \mathbf{E} \) and \( \mathbf{B} \) the electric and magnetic fields, \( G \) the gravitational constant, \( M \) the solar mass, \( r \) the heliocentric radial distance, \( \mathbf{\hat{e}}_r \), the radial unit vector, \( \mathbf{F}_{fj} \) the frictional force density, \( \mathbf{F}_j \) the thermal force density, \( F_{fj} \) any additional force density, and \( m_j \left( \sum_l Q_{jk} \mathbf{u}_k - L_j \mathbf{u}_j \right) \) the force density arising from production and loss of \( j \)-type particles. The momentum density \( (\rho_j \mathbf{u}_j) \) and momentum flux density \( (\rho_j \mathbf{u}_j \mathbf{u}_j) \) appearing on the left side of Eq. 2 are associated with the bulk flow of the \( j^\text{th} \) particle species, but we could have included in these terms the effects of the internal energy or pressure, which we have instead dealt with as force terms.

All the terms on the right side of Eq. 2 are force densities, which serve as the sources (and sinks) of the momentum density. On the right side of Eq. 2, the first term is the pressure gradient force, which is a major contributor to the driving of all types of solar wind. The second term is associated with the viscous stress, which has two components: the shear stress (associated with the off-diagonal terms in \( \mathbf{\Pi}_j \)) and the dilatation stress (associated with the diagonal terms in \( \mathbf{\Pi}_j \)). Both the shear stress and the dilatation stress can be important in the damping of waves, but the dilatation stress also often plays a role in shock maintenance, and it is particularly important for the background flow whenever there is a significant pressure anisotropy. In this case, the pressure gradient and viscous stress terms are normally combined and rewritten in terms of the components of the pressure parallel and perpendicular to the magnetic field (cf. Marsch et al., 2003, eq. 19.6), and separate energy equations for the parallel and perpendicular pressures are then used. The third and fourth terms represent the familiar electromagnetic and gravitational forces. The fifth and sixth terms are both interspecies collisional forces (Braginskii, 1965): the frictional force is associated with relative flow between species, while the thermal force arises from the presence of a temperature gradient. The frictional force is generally important from the upper chromosphere to the lower corona, and the thermal force is particularly important in the transition region between the upper chromosphere and corona, where the temperature rises steeply from 10^6 K to 10^7 K.

Indeed, in the transition region, the momentum balance for minor ions is essentially a balance between the upward thermal force and the downward frictional force, both of which arise from the collisional interaction of these ions with protons. The additional force, \( F_j \), is normally associated with waves and involves a wave pressure gradient component and sometimes another component as well. The association of \( F_j \) with waves implies that Eq. 2 describes the background flow, while a separate equation describes the waves; of course, as the waves become nonlinear, this type of separation is not necessarily possible. Finally, the force term arising from production and loss of \( j \)-type particles is not important for our discussion here, although it is important in the outer heliosphere, where the charge transfer between interstellar hydrogen and the solar wind slows the wind through this force.

3.3 Energy Balance

Turning now to energy balance, we can write

\[
\frac{\partial}{\partial t} \left( \frac{1}{2} \rho_j \mathbf{u}_j^2 + \frac{1}{2} p_j \right) + \nabla \cdot \left( \frac{1}{2} \rho_j \mathbf{u}_j \mathbf{u}_j + \frac{1}{2} p_j \mathbf{u}_j + \mathbf{u}_j \cdot \mathbf{\Pi}_j + \mathbf{q}_j \right) = n_j e Z_j \mathbf{u}_j \cdot \mathbf{E} - \rho_j (G M / r^2) \mathbf{\hat{e}}_r + \mathbf{F}_{j} + S_j + \sum_l \left\{ \frac{m_j}{m_j + m_l} \left[ 3 n_l \nu_{jk} k_b \left( T_l - T_j \right) \right] \right\} + m_j \left( \sum_l Q_{jk} \frac{3 k_b T_k + u_k^2}{m_j} - L_j \frac{3 k_b T_j + u_j^2}{m_j} \right)
\]

Here, \( \mathbf{q}_j \) is the heat flux density, \( S_j \) is an additional energy source/sink term, \( \nu_{jk} \) is the momentum transfer collision frequency for collisions of species \( j \) and \( k \).
particles, \( k_b \) is the Boltzmann constant, and \( T_f \) is the temperature. As in the structure of Eqs. 1 and 2, the left side of Eq. 3 contains the time rate of change of an energy density and the divergence of an energy flux density. The energy density has a flow energy component and an internal energy component. The first two terms in the energy flux density include the advection of the energy density, but the second term includes the effect of the work term, \( p_j \mathbf{u}_j \). The third term is also a work term associated with the viscous stress, so the second and third terms together can be considered a generalized enthalpy flux density. The final component of the energy flux density is the heat flux density, \( q_j \).

The source/sink terms on the right side of Eq. 3 involve work done on or by the \( \theta \)th species fluid and the heating or cooling of the fluid. The first term includes only the electric work, since the magnetic field cannot do work (i.e., \( \mathbf{u}_j \cdot (\mathbf{u}_j \times \mathbf{B}) = 0 \)). The second term reflects the work required to lift the wind out of the gravitational field, which, as noted earlier, represents a substantial fraction of the energy required to drive the solar wind. The additional source/sink term, \( S_j \), includes, for example, the effects of wave heating and radiative cooling, as well as work done on the plasma by the wave pressure gradient force. The fourth term on the right side of Eq. 3 involves the collisional transfer of thermal energy between different species, and this term tends to drive the various species toward equal temperatures. Finally, the last term reflects the energy addition/loss associated with the production or loss of particles, and as noted in the context of Eq. 2, this term is not important in the corona and inner heliosphere, but is quite important in the outer heliosphere.

Before leaving the subject of energy balance, it is useful to consider two one-fluid descriptions: a conservation equation and a thermal energy equation. The one-fluid energy conservation equation is derived by summing Eq. 3 over all particle species:

\[
\frac{\partial}{\partial t} \left( \frac{1}{2} \rho u^2 + \frac{1}{2} p \right) + \mathbf{V} \cdot \left( \frac{1}{2} \rho u^2 \mathbf{u} + \frac{1}{2} p \mathbf{u} \mathbf{u} + \mathbf{u} \mathbf{u} + q \right) = \mathbf{J} \cdot \mathbf{E} - \rho \left( G M / r^2 \right) \mathbf{u} \cdot \mathbf{e}_r + S
\]  

(4)

where \( \mathbf{J} \) is the current density. Evidently, Eq. 4 is very similar to Eq. 3, although the interspecies terms have disappeared in the summation. Let us focus here on the \( \mathbf{J} \cdot \mathbf{E} \) term, which describes the transfer of energy between the plasma and the electromagnetic field. Using Maxwell’s equations (in particular, Faraday’s law and Ampère’s law) one can show that

\[
-\mathbf{J} \cdot \mathbf{E} = \frac{e}{\sigma} \left( \frac{1}{2} \mathbf{B}^2 / \mu_0 + \frac{1}{2} \mathbf{E}^2 \right) + \mathbf{V} \cdot \left( \mathbf{E} \times \mathbf{B} / \mu_0 \right)
\]  

(5)

We notice that Eq. (5) has the form of a conservation equation—this time for the electromagnetic energy density. Here \( -\mathbf{J} \cdot \mathbf{E} \) is the source term, which is consistent with \( \mathbf{J} \cdot \mathbf{E} \) being the source term in the plasma energy equation. We can delve more deeply into the meaning of \( \mathbf{J} \cdot \mathbf{E} \) by considering the simplest form of Ohm’s law, which can be derived by summing the momentum equation (Eq. 2) over all ion species and subtracting the electron momentum equation:

\[
\mathbf{E} + \mathbf{u} \times \mathbf{B} = \hat{\mathbf{\eta}} \cdot \mathbf{J}
\]  

(6)

Here, \( \hat{\mathbf{\eta}} \) is the tensor resistivity. (The tensor form is required for all magnetized plasmas, even if they are fully ionized.) Now, if we take the scalar product of \( \mathbf{J} \) with Eq. 6, we obtain

\[
\mathbf{J} \cdot \mathbf{E} = \mathbf{u} \cdot \left( \mathbf{J} \times \mathbf{B} \right) + \mathbf{J} \cdot \hat{\mathbf{\eta}} \cdot \mathbf{J}
\]  

(7)

The first term on the right side of Eq. 7 represents the work done on the plasma by the electromagnetic field, and the second term describes resistive heating, which includes a frictional heating component and another component associated with wave-particle interactions (i.e. heating associated with “anomalous” resistivity). When we write a one-fluid equation for the thermal energy, only the resistive heating term survives. Such an equation can be derived by combining one-fluid mass and momentum conservation equations with Eq. 4:

\[
\frac{1}{\rho} \frac{\partial}{\partial t} \left( \rho T \right) + \mathbf{V} \cdot \left( \frac{1}{2} \rho u T + p \mathbf{u} \right) = -\mathbf{V} \cdot \mathbf{q} + \mathbf{J} \cdot \mathbf{E} + S_T
\]  

(8)

where \( S_T \) is heating/cooling portion of \( S \). The term in parentheses on the left side of Eq. 8 is the Lagrangian derivative (i.e., the time derivative following the flow) of the temperature. On the right side, the first term is associated with compressive heating and expansive cooling, the second with heating through dissipation of the heat flux, the third with viscous heating, and the fourth with resistive heating.

### 3.4 Kinetic Descriptions

One difficulty with the preceding discussion is that not all of the important physical processes involved in
coronal heating and solar wind acceleration can be adequately described using fluid equations, even if all particle species are treated separately (e.g., Shoub, 1983; Marsch, 2004). Thus we need an equation describing the evolution of the $j^\text{th}$ species velocity distribution function, which is also called the (six-dimensional) phase space density. Such an equation can be written in conservation form:

$$\frac{\partial f_j}{\partial t} + \nabla \cdot (\mathbf{V}_j f_j) + \mathbf{V}_j \cdot (\partial f_j/\partial t) = \left( \frac{\delta f_j}{\delta t} \right),$$

(9)

Here, $f_j(r, \mathbf{v}, t)$ is the phase space density, which is defined as the number of particles residing in both an infinitesimal configuration space volume element bounded by $x, x + dx, y, y + dy, z, z + dz$ and an infinitesimal velocity space volume element bounded by $v_x, v_x + dv_x, v_y, v_y + dv_y, v_z, v_z + dv_z$. The configuration space and velocity space flux densities are then $\mathbf{V}_j$ and $\mathbf{V}_j f_j$, where $\mathbf{V}$ is the particle velocity and $\mathbf{V}$ is the particle acceleration—the divergence of the phase space flux density is written in Eq. 9 as the sum of a configuration space divergence ($\nabla \cdot (\mathbf{V} f_j)$) and a velocity space divergence ($\mathbf{V} \cdot (\partial f_j/\partial t)$). The two terms on the right side of Eq. 9 are source/sink terms: the first includes the effects of particle sources and sinks, while the second includes the effects of particle collisions. The naming of kinetic equations usually reflects the nature of the source/sink terms on the right side of Eq. 9. For example, the Boltzmann equation employs a specific form of the collision term called the Boltzmann collision integral, and it contains no particle source/sink term. Similarly, the Fokker-Planck equation simply has a different form of the collision term. Perhaps of most interest to us in considering kinetic effects in the coronal heating problem is the Vlasov equation, in which the right side of Eq. 9 is set to zero, and $\mathbf{V}_j$ is associated only with the electromagnetic force. This is the form that is used to describe coronal heating processes such as Landau damping and cyclotron damping of plasma waves.

It has long been argued by some authors (e.g., Lemaire and Scherer, 1973) that a kinetic treatment is required not only for microscopic processes (like those just mentioned), but also for an adequate description of the background wind, as Coulomb collisions decrease in importance with distance from the sun. This contention is certainly correct if one is interested in the details of the velocity distribution function (phase space density) for any of the solar wind particle species, although it is also true that the basic fluid parameters (flow speed, density, and to a large extent temperature) are well described by fluid treatments. As Lemaire and Scherrer (1973) have noted, the kinetic and fluid treatments are complementary, so it is necessary to determine which is more useful for describing a particular aspect of the solar wind. An example where a kinetic treatment is required is in the description of the electron velocity distribution function in high-speed solar wind. Lie-Svendsen and Leer (2000) have shown that this distribution function is well described in the inner heliosphere by a test-particle Fokker-Planck approach using Coulomb collisions only in proton-electron solar wind.

### 3.5 Mass Flux and Asymptotic Flow Speed

Having established a basic theoretical framework, let us consider the determination of the solar wind mass flux and asymptotic flow speed. For this purpose, we use a one-fluid description: the summation of Eq. 1 over all particle species leads to

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

(10)

For simplicity, let us consider a steady, one-dimensional flow along a radial flow tube of cross-sectional area $A$. Then the first term in Eq. 10 vanishes, and we obtain

$$\frac{1}{A} \frac{d}{dr} (\rho u A) = 0$$

(11)

If we deal only with the subsonic region of the flow, and we are only interested in the pressure and density, then we can treat our atmosphere as essentially hydrostatic from the coronal base to the critical point, where the flow speed equals the sound speed (e.g., Parker, 1963). Momentum balance (described by the summation of Eq. 2 over all particle species) involves a balance between the upward pressure gradient force and the downward gravitational force:

$$\frac{d p}{d r} = -\rho \frac{GM}{r^2}$$

(12)

For the simplest case of an isothermal atmosphere (where the constant temperature can be thought of as an average temperature, $\langle T \rangle$, through the subsonic region), Eq. 12 can be integrated from the coronal base ($r_i$) to the critical point ($r_c$), yielding

$$\rho = \rho_i \exp \left[ -\frac{GMm}{2k\langle T \rangle r_i} \left( \frac{d \ln A}{d \ln r} \right) \right]$$

(13)
From Eq. 11, we see that the mass flux, $\rho u A$, is a constant of the flow, and we know that the flow speed at the critical point is given (in the isothermal approximation) by

$$u_c = \sqrt{\frac{2k_b \langle T \rangle}{m}}$$  \hspace{1cm} (14)

Thus we can evaluate the mass flux at the critical point, and from Eqs. 13 and 14 we have

$$\rho u A = \rho_0 A_0 \left( \frac{A_c}{A_0} \right) \sqrt{\frac{2k_b \langle T \rangle}{m}} \cdot \exp \left[ -\frac{GMm}{2k_b \langle T \rangle_0} + \left( \frac{d \ln A}{d \ln r} \right) \right]$$  \hspace{1cm} (15)

One last point before discussing Eq. 15 is that the density at the base of the corona is determined by the location of the transition region, which in turn is determined by the downward heat flux and upward enthalpy flux through the transition region (Rosner et al., 1978; Hansteen and Leer, 1995):

$$\rho_0 \propto \left( |u_c| - \frac{1}{2} p_0 u_c \right)$$  \hspace{1cm} (16)

Since the heat flux is carried primarily by electrons, the coronal electron temperature determines the downward heat flux and thus the pressure at the coronal base. Hence, in coronal holes, where protons are much hotter than electrons, the overall (i.e., one-fluid) plasma temperature can be relatively high while the coronal base pressure stays relatively low. Returning to our consideration of Eq. 15, we see that a number of factors control the solar wind mass flux in a quasi-steady flow. Specifically, the mass flux is increased by (1) increasing the coronal base density, (2) increasing the mean one-fluid temperature in the subsonic region, (3) increasing the expansion of a flow tube between the coronal base and the critical point, and (4) increasing the flow tube expansion rate at the critical point.

Now we are prepared to see how the asymptotic flow speed of the wind is determined. First, in Eq. 4, let us write the energy addition terms, $\mathbf{J} \cdot \mathbf{E}$ and $S$, in terms of the divergence of an energy flux, $\Phi$:  

$$-\nabla \cdot \left( \Phi / A \right) = \mathbf{J} \cdot \mathbf{E} + S$$  \hspace{1cm} (17)

We can also rewrite the gravitational work term as

$$-\rho \left( GM/r^2 \right) \mathbf{u} \cdot \mathbf{e}_r = \nabla \left[ \rho \left( \frac{1}{2} v^2 \right) \right]$$  \hspace{1cm} (18)

where we recall that we are assuming steady, radial flow, and where we have defined the gravitational escape speed as

$$v_e = \sqrt{\frac{2GM}{r}}$$  \hspace{1cm} (19)

Neglecting the viscous term, we can now write Eq. 4 for a steady flow in the following form

$$\rho u A \left( \frac{1}{2} u^2 + \frac{1}{2} p/\rho - \frac{1}{2} v^2 \right) + \Phi = \text{const}$$  \hspace{1cm} (20)

where $\Phi \left( = \Phi_e + qA \right)$ is the total non-advective energy flux, and we have dropped the vector notation, since both $u$ and $\Phi$ are the magnitudes of radial vectors. At the coronal base, the gravitational potential energy per unit mass dominates the other two advected energy terms in Eq. 20, and at large distances from the sun, the entire solar wind energy flux resides in the flow energy. It follows from Eq. 20 that

$$\Phi_0 - \frac{1}{2} \rho u A v^2_\infty = \frac{1}{2} \rho u A u^2_\infty$$  \hspace{1cm} (21)

There are two other useful ways to combine terms in Eq. 21:

$$\Phi_0 = \frac{1}{2} \rho u A v^2_\infty + \frac{1}{2} \rho u A u^2_\infty$$  \hspace{1cm} (22)

$$u_\infty = \frac{2\Phi}{\sqrt{\rho u A} - v^2_0}$$  \hspace{1cm} (23)

Eq. 22 tells us that we have already assumed in §2: that the energy flux entering the corona from below goes into lifting the wind out of the solar gravitational field and accelerating it to its asymptotic flow speed. Eq. 23 tells us that the asymptotic flow speed increases with increasing energy flux and decreases with increasing mass flux. Evidently, that portion of the energy flux dissipated in the region of subsonic flow increases the mass flux, since it will tend to increase the coronal temperature in that region. In contrast, the energy deposited in the region of supersonic flow has no effect on the mass flux. Consequently, the recipe for a high speed wind is to deposit less energy in the subsonic region and more in the supersonic region. It is also clear from Eq. 23, along with Eq. 15, that for a given energy flux deposited in the region of subsonic flow, the asymptotic flow speed will be lowered by a more rapid flow tube divergence, since this will increase the solar wind mass flux. This result is, of course, qualitatively consistent with the predictive scheme of Wang and Sheeley (1990; Arge and Pizzo, 2000). This consistency does not necessarily mean that the degree of flow-tube expansion is the dominant factor in determining solar wind flow speed. Yet the impressive success of this predictive scheme must
indicate something significant—perhaps, for example, the proximity of rapidly diverging magnetic geometries to regions of transiently open magnetic fields.

3.6 Modeling High- and Low-Speed Wind

Modelers of the high-speed wind must keep in mind the observation (cf. §2) that the coronal hole proton temperature is much greater than the electron temperature. In fact, the proton temperature is large enough that the proton pressure gradient force must be a very significant (if not the dominant) driver of the high-speed wind. The large proton temperature does not, however, imply a large heat flux downward into the chromosphere (and a resultant large coronal base pressure; Rosner et al., 1978; Hansteen and Leer, 1995), since this heat flux is carried primarily by electrons. The significant difference between the proton and electron temperatures in coronal holes, together with the need to include the upper chromosphere and transition region in high-speed wind models, indicates that such models must be either two-fluid or multi-fluid in nature. As noted in the preceding paragraph, driving the high-speed wind generally requires significant energy deposition in the region of supersonic flow. Such deposition is certainly aided in coronal holes by the observed rapid acceleration of the wind (cf. §2), since this places the sonic point (beyond which the deposition must occur) much closer to the coronal base. Evidently, we will not really understand the acceleration of the high-speed solar wind until we have a good handle on the nature of the energy flux that enters the corona from the lower solar atmosphere (cf. §4 and §5). It is clear from observations that a significant fraction of this energy flux must damp in such a way as to produce a large perpendicular-to-parallel proton and ion temperature anisotropy in coronal holes, leading to the idea of cyclotron resonance as an important damping mechanism. But there may be some significant portion of the energy flux that has a different nature, leading to a different damping mechanism operating in a different location from the cyclotron resonant damping.

It appears that the slow solar wind requires a somewhat different modeling approach. Although it is possible that at least some of the slow wind may arise from magnetic geometries that are open to the heliosphere for long periods, it is likely that most of the slow wind comes from transiently open magnetic geometries. Consider an active region, where the electron temperature and the coronal base pressure can be large and thus can produce slow wind with a high mass flux on field lines that are open for long periods. The question is whether such wind will exhibit a significant FIP bias, since it would seem that the high-speed wind from coronal holes shows little or no FIP bias because of its origin on long-term open field lines, where the settling of high-FIP atoms does not occur. Certainly, any model of slow wind on long-term open field lines must address the question of FIP bias. A more natural approach to slow wind modeling would seem to involve the transient opening and closing of magnetic field lines. On closed field lines, the coronal electron temperature will naturally be higher than on open field lines (for similar energy input rates) because the energy deposited on these field lines must be conducted down to the chromosphere, where it is radiated away, while on open field lines some 80% to 90% of the deposited energy is advected outwards as the solar wind (Hansteen and Leer, 1995). Furthermore, the lack of a quasi-steady outflow on closed field lines presumably permits the settling of high-FIP atoms in the upper chromosphere and the development of a significant FIP bias. If a field line is opened for a period comparable to or less than that required to establish a quasi-steady outflow (several hours), then the plasma characteristics of a closed field line will determine the nature of the outflow, leading to a slow solar wind. Two types of mechanisms for transiently opening magnetic field lines have been suggested: the reconnection of open magnetic field lines with closed magnetic field lines (e.g., Axford, 1977; Fisk, 2005), and the instability of coronal streamers (e.g., Dahlburg and Einaudi, 1999; Endeve et al., 2004). The latter type of mechanism has been discussed in considerably greater quantitative detail than the former, but both types of mechanism seem likely to be operating in the solar corona.

4. OBSERVATIONS: LOWER ATMOSPHERE

4.1 Chromospheric Network and Supergranules

Let us take a brief tour of the photosphere, chromosphere, and lower transition region, focusing on features that are closely connected with the structure of the solar magnetic field (cf. Lites and Berger, 2002) and that seem likely to be particularly relevant to the non-radiative energy flux flowing from the lower atmosphere to the corona. We begin in Fig. 4a with an image of the sun taken in the Ca II K resonance line at the Big Bear Solar Observatory on May 25, 2003 at 16:26:02 UT. The line cores of both the Ca II H and K lines are formed largely in the middle chromosphere and are well-correlated with the presence of relatively strong magnetic fields. The brightest regions in the image are active regions, but there is a somewhat fainter pattern of brightness appearing as dark cells with bright boundaries. An entire cell, including the cell interior and its boundary, is called a supergranule, while the bright (but highly non-uniform) cell boundary is called the chromospheric network.
interior of a supergranular cell is commonly called the internetwork). The supergranular pattern is more clearly visible in the 4:1 zoom shown in Fig. 4b (cf. white box in Fig. 4a). Although not visible in this observation, the supergranule is characterized by an upflow in the cell interior, a horizontal flow outward toward the cell boundary (i.e., the network), and a strong downflow in the network, where the strong magnetic fields are concentrated.

4.2 Network, Supergranules, and Granules

Fig. 5a shows an image taken with the Swedish Solar Telescope (Berger et al., 2004) on the same day (at 11:22:32 UT) in the H line of Ca II. This image of the remnant plage network of AR 10365 corresponds to the white box in Fig. 4b (a 6:1 zoom from that image). Fig. 5b shows an image of the same region in the G-
band, which is a photospheric absorption feature arising primarily from a spectral band of the CH molecule, and to a lesser extent from Fe I lines near 430.5 nm. In both images of Fig. 5, another structure, some 30 times smaller than the supergranule becomes clearly visible. This is the granule, which in the G-band appears as a bright feature bounded by a dark lane and in the Ca II H line as a dark feature bounded by a bright lane. As in the case of the supergranule, the granule has a flow characterized by an upflow in the cell interior, horizontal flow towards the intergranular lane, and a downflow in these lanes. These characteristics are shown more clearly in Fig. 6, which is an 12:1 zoom from Fig. 5 (cf. the white boxes in Figs. 5a and 5b). Comparing the magnetogram (c) with the Ca II (a) and G-band (b) images, we see that the strong magnetic fields (dark in the magnetogram) correspond well with the bright features in Ca II and G-band and appear primarily in intergranular lanes. In these lanes, the Dopplergram (d) shows strong downflow, which is consistent with the picture of granules just mentioned.

4.3 Lower Transition Region Emission

Finally, let us consider the appearance of emission at temperatures corresponding to the lower transition region (viz., \(2 \times 10^5 \text{ K} < T < 2 \times 10^6 \text{ K}\)). Fig. 7 shows a C III 97.7 nm image taken by the SOHO/SUMER instrument on January 28, 1996 (Feldman et al., 2000). The C III 97.7 nm line is formed between about \(5 \times 10^4 \text{ K}\) and \(1.1 \times 10^5 \text{ K}\), so it falls under our definition of a lower transition region line. The bright regions in Fig. 7 are centered over the chromospheric network, but are clearly much broader than the network and extend a considerable distance into the internetwork region. (Note that the 30” distance indicated in the Fig. 7 corresponds to about 22 Mm.) We must also assume that the bright regions are highly time variable (e.g., Habbal and Withbroe, 1981), although that fact cannot be determined from this single image. One might interpret this image (along with the preceding discussion) as indicating that lower transition region emission comes predominantly from transient low-lying loops.

4.4 Magnetic Structure of the Quiet Atmosphere

We can summarize and interpret §§4.1-4.3 with the cartoon shown in Fig. 8, which is based on the work of several authors (e.g., Gabriel, 1976; Dowdy et al., 1986; Axford and McKenzie, 1995; Feldman, 1983; Aiouaz et al., 2005). This cartoon shows a schematic,
two-dimensional representation of three and one-half supergranular cells. In this picture, the coronal magnetic field (whether closed in the corona or open to the heliosphere) originates in the network while the magnetic field in the internetwork region is shown as a collection of low-lying bipoles on the granular scale. The intergranular magnetic field is not well observed, but the unsigned intergranular magnetic flux is thought to be not too much smaller than the network flux (e.g., Lites, 2002), which, of course, means that the magnetic field intensity is considerably smaller in the internetwork, but by no means negligible. One other magnetic structure shown in Fig. 8 is composed of field lines connecting the network field with the internetwork field—field lines that cannot be directly observationally inferred, but which are likely to exist.

Axford and McKenzie (1995) have suggested that the advection of small-scale internetwork magnetic fields into the network by the horizontal supergranular flow leads to continual magnetic-field-line reconnection that converts magnetic energy into plasma energy and in this way supplies the energy to the corona that is required to heat it and drive the solar wind (cf. §5). This conversion can lead to many different forms of plasma energy (e.g., Axford, 2002), including hydromagnetic waves, energetic particles, large-scale plasma flows, and high plasma temperatures. Evidently, if this process is occurring and if the basic magnetic structure of Fig. 8 is correct, then the plasma along the network-internetwork field lines could be transiently heated, leading to transient emission at lower transition region temperatures. This process is clearly similar to that commonly postulated for post-eruption loops, where reconnection at the top of a newly formed loop leads to a burst of heating along a loop, followed by radiative cooling that takes the loop through temperatures ranging from more than 1,000,000 K to less than 5,000 K. Presumably, the much higher density on low-lying network-internetwork field lines could lead to somewhat lower maximum temperatures (viz., < 200,000 K) and thus to emission at lower transition region temperatures. If this is the case, then these low-lying, transiently heated magnetic loops could give rise to the majority of emission at lower transition region temperatures, as is implied in Fig. 8. Clearly, the entire transition region on network field lines in Fig. 8 can be described by the normal picture of a thermal interface between the hot corona and the cool chromosphere, and this picture would lead to the majority of emission at upper transition region temperatures arising from these network field lines. This model of transition region emission, which is consistent with the ideas of Feldman (1983), is certainly not the only viable transition region model, and the ultimate determination of the origin of transition region emission awaits observations with higher spatial and temporal resolution, as well as improved forward modeling studies (e.g., Wikstøl et al., 1998).

5. ENERGY FLUX ENTERING THE CORONA FROM BELOW

We briefly summarize here the principal ideas proposed for the transport of energy from the lower atmosphere to the corona. As we will see, several interesting proposals have been put forth, but unfortunately these proposals have not been placed on a quantitative footing. There have, however, been recent modeling efforts that show the potential for the exploration of each of the following mechanisms in a
quantitative way. This possibility will be discussed further in §6.2.

5.1 Low-Frequency Waves from the Photosphere

The photosphere sits atop the solar convection zone and exhibits fluctuations over a broad range of frequencies, but most of the power in the photospheric fluctuation spectrum lies at relatively low frequencies ($f \ll 1$ Hz). One possibility for transferring energy from the photosphere to the corona is by way of a hydromagnetic wave flux generated by these photospheric fluctuations and propagating through the chromosphere and transition region into the corona (e.g., Osterbrock, 1961). In the low-$\beta$ plasma of the upper chromosphere, transition, region and corona, slow-mode and intermediate-mode (Alfvén) waves propagate primarily along the magnetic field, while fast-mode waves can propagate in any direction. Because of the rapidly increasing wave speed through the transition region (where the density rapidly declines, while the temperature rapidly rises), fast-mode waves are strongly refracted toward the regions of lower wave speed and never reach the corona. Slow-mode waves rapidly steepen into shocks and largely dissipate before arriving in the corona, while Alfvén waves suffer substantial (non-WKB) reflection in the transition region, where their wavelengths are long in comparison with the Alfvén speed scale height. Only if Alfvén waves were to have a sufficiently high frequency (and short wavelength) would they be able to pass through the transition region without significant reflection (cf. §5.2). Thus the low-frequency waves that one would expect to pervade the lower solar atmosphere are not good candidates for heating and acceleration of the coronal plasma.

5.2 High-Frequency Waves from Reconnection

As noted in §4.4, Axford and McKenzie (1995) have proposed that the principal source of the energy required to heat the corona and accelerate the solar wind is produced through the advection of internetwork magnetic fields into the network and their continual reconnection there with network magnetic fields. These authors suggest that the main consequence of this reconnection for the corona is the generation of high-frequency Alfvén waves that can readily propagate through the transition region without reflection (cf. §5.1) and into the corona, where they are dissipated primarily through cyclotron resonance with the various ion species composing the coronal plasma. It is proposed that such a mechanism can explain coronal energy balance in both magnetically open and closed regions, and that there is no need for any other mechanism. As is the case with the other proposals discussed here, quantitative support is not yet available, although it certainly may be in the next few years.

5.3 Heat Flux from Reconnection

Markovskii and Hollweg (2004) have also employed the idea that reconnection in the network can produce a significant upward energy flux, but they have proposed that the continual reconnection occurs near the base of the corona and drives an intermittent heat flux into the corona. An electron heat flux is associated with a skewing of the velocity distribution function that is associated with a high-energy tail in the direction of heat transport and a drift of the core of the distribution function in the opposite direction (in the electron gas rest frame), and this structure leads to a current-driven instability that can produce both ion-acoustic and ion-cyclotron waves. Markovskii and Hollweg (2004) suggest that the ion-cyclotron waves produced in this manner dissipate in coronal holes, producing high proton temperatures (with strong anisotropy) and the high-speed solar wind.

5.4 Flux-Tube Twisting

Parker (1981) has noted that the inherent vorticity in photospheric motions will generate twists in magnetic field lines closed in the corona and thus lead to the formation of current sheets where magnetic energy can be transformed into plasma energy through the reconnection process mentioned above. Indeed, an optimal place for such twisting occurs in the network, where the strong downflows associated with concentrations of magnetic flux also are characterized by a significant amplification of the twisting motions that lead to energy storage and release in the corona, through what are commonly called nanoflares. Although, as in the case of the mechanisms discussed above, it is difficult to give a quantitative estimate of the rate of energy transfer from the photosphere to the corona, it seems clear that this mechanism must be operative in all magnetically closed coronal regions, whether or not it is the dominant mechanism in these regions.

5.5 Velocity Filtration

Scudder (1992) has argued that if the electron velocity distribution at the top of the chromosphere exhibits a strong skewing associated with an upward heat flux, then the electrons carrying the heat flux will be able to penetrate substantial distances before depositing their energy in the coronal plasma, thus serving to provide the principal source of heat for the corona. This suggestion is based on the fact that the Coulomb collision cross section varies inversely as the fourth power of the interparticle speed, so that while the bulk
of the electron distribution function may be collision-dominated, the high-energy tail carrying the heat flux does not necessarily suffer collisions (depending on the structure of the tail in velocity space). The main criticism of Scudder’s hypothesis has been that no mechanism has been suggested to produce the required highly skewed electron distribution function at the top of the chromosphere. It is interesting in this context to consider the proposal of Markovskii and Hollweg (2004), but to postulate that the reconnection producing their intermittent heat flux occurs in the upper chromosphere rather than the corona. In this case, the Markovskii and Hollweg mechanism could become a mechanism for producing the highly skewed distribution function that Scudder needs.

6. A LOOK INTO THE FUTURE

In closing, let us take a brief look into the future, first considering some useful future observations, and then looking to some fruitful directions for theoretical research. Since the subject of coronal heating and solar wind acceleration is in what one might call the detailed study phase, we will focus on observational efforts consistent with this phase of study. It is, of course, possible that an exploratory mission like the long-proposed Solar Probe will provide us with crucial insights into energy balance, but the exploratory nature of the mission precludes the prediction of such insights.

6.1 Future Observations

Perhaps the single most important instrument for improving our understanding of heating and acceleration of the coronal plasma would be a coronagraphic imaging instrument with the full spectral capabilities of the SOHO/UVCS instrument and a 360° field of view extending from as near the sun as possible (say 0.1 solar radii above the photosphere), to some 6 solar radii from sun center. This instrument can be thought of as a melding of the SOHO/UVCS and SOHO/LASCO instruments. All we have to do to understand the value of such an experiment is to consider the remarkable new understanding that has come from UVCS, with its narrow slit field of view.

Another experiment that would be very useful is a high spatial and temporal resolution instrument that provides images in spectral lines formed at lower transition region temperatures. In a sense, this could be considered an advanced SOHO/SUMER instrument. The observations obtained would likely begin to determine what extent the pictures of Feldman (1983), Wikstøl et al. (1998), and others describe the formation of the lower transition region emission. To the extent that these spectral lines are formed on the network-internetwork field lines described in Fig. 8, such observations could provide significant constraints on theoretical models of the network reconnection mechanisms that have been proposed.

Finally, and again with an eye to the reconnection of internetwork fields with network fields as the internetwork plasma is swept into the network, high resolution vector magnetic field measurements providing a clearer view of the internetwork field structure should prove extremely valuable. Fortunately, such observations will become available with the launch of Solar B.

We have focused here on solar observations, primarily because of the need for significant improvement in the instruments making these observations. Yet it will be crucial to have cotemporal solar wind observations of the sort described in §2.1, if we are to be able to place the new solar observations in a useful context.

6.2 Future Theory

There is a critical need for the quantitative modeling of energy transport from the lower solar atmosphere to the corona. The several interesting ideas described in §5 are actually of little use to us until they have been tested by detailed quantitative models appropriately constrained by observations (cf. §6.1). Two recent studies by Cranmer and van Ballegooijen (2005) and by Gudiksen and Nordlund (2005) serve as excellent examples of the way in which such models need to be developed and applied. Although the coronal mechanisms for dissipation of this energy flux generally stand on a much more quantitative basis than do those for the origin of the energy flux, there remains a good deal of more detailed work to do, particularly in the area of kinetic processes (e.g., Marsch, 2004). Finally, to make use of the improved understanding of the transport and dissipation of energy in the corona, it is necessary to develop and apply multifluid MHD models (e.g., Ofman, 2004) with realistic energy balance and extending from the upper chromosphere to the heliosphere, in order to describe adequately the solar wind produced by the transport and dissipation of the photosphere-to-corona energy flux.

REFERENCES


Neugebauer, M., The Solar Wind and Heliospheric Magnetic Field in Three Dimensions, in *The...*