Coronal Holes

Steven R. Cranmer
Harvard-Smithsonian Center for Astrophysics
60 Garden Street, Mail Stop 50
Cambridge, MA 02138, U.S.A.
email: scranmer@cfa.harvard.edu
http://www.cfa.harvard.edu/~scranmer/

Accepted on 15 September 2009
Published on 29 September 2009

Abstract
Coronal holes are the darkest and least active regions of the Sun, as observed both on the solar disk and above the solar limb. Coronal holes are associated with rapidly expanding open magnetic fields and the acceleration of the high-speed solar wind. This paper reviews measurements of the plasma properties in coronal holes and how these measurements are used to reveal details about the physical processes that heat the solar corona and accelerate the solar wind. It is still unknown to what extent the solar wind is fed by flux tubes that remain open (and are energized by footpoint-driven wave-like fluctuations), and to what extent much of the mass and energy is input intermittently from closed loops into the open-field regions. Evidence for both paradigms is summarized in this paper. Special emphasis is also given to spectroscopic and coronagraphic measurements that allow the highly dynamic non-equilibrium evolution of the plasma to be followed as the asymptotic conditions in interplanetary space are established in the extended corona. For example, the importance of kinetic plasma physics and turbulence in coronal holes has been affirmed by surprising measurements from the UVCS instrument on SOHO that heavy ions are heated to hundreds of times the temperatures of protons and electrons. These observations point to specific kinds of collisionless Alfvén wave damping (i.e., ion cyclotron resonance), but complete theoretical models do not yet exist. Despite our incomplete knowledge of the complex multi-scale plasma physics, however, much progress has been made toward the goal of understanding the mechanisms ultimately responsible for producing the observed properties of coronal holes.
**Imprint / Terms of Use**

*Living Reviews in Solar Physics* is a peer reviewed open access journal published by the Max Planck Institute for Solar System Research, Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany. ISSN 1614-4961.

This review is licensed under a Creative Commons Attribution-Non-Commercial-NoDerivs 3.0 Germany License: [http://creativecommons.org/licenses/by-nc-nd/3.0/de/](http://creativecommons.org/licenses/by-nc-nd/3.0/de/)

Because a *Living Reviews* article can evolve over time, we recommend to cite the article as follows:

Steven R. Cranmer,  
“Coronal Holes”,  
http://www.livingreviews.org/lrsp-2009-3

The date given as <date> then uniquely identifies the version of the article you are referring to.

**Article Revisions**

*Living Reviews* supports two ways of keeping its articles up-to-date:

**Fast-track revision** A fast-track revision provides the author with the opportunity to add short notices of current research results, trends and developments, or important publications to the article. A fast-track revision is refereed by the responsible subject editor. If an article has undergone a fast-track revision, a summary of changes will be listed here.

**Major update** A major update will include substantial changes and additions and is subject to full external refereeing. It is published with a new publication number.

For detailed documentation of an article’s evolution, please refer to the history document of the article’s online version at [http://www.livingreviews.org/lrsp-2009-3](http://www.livingreviews.org/lrsp-2009-3).
Contents

1 Introduction 5

2 Historical Overview 7

3 Properties of On-Disk Coronal Holes 10

4 Properties of Off-Limb Coronal Holes 14
   4.1 Magnetic connectivity with the solar wind 14
   4.2 Visible light observations 15
   4.3 Ultraviolet spectroscopy 17

5 Coronal Heating and Solar Wind Acceleration 23
   5.1 Sources of energy 23
   5.2 The Wave/Turbulence-Driven (WTD) solar wind idea 24
   5.3 The Reconnection/Loop-Opening (RLO) solar wind idea 28
   5.4 Kinetic microphysics 29

6 Summary and Conclusions 35

7 Acknowledgements 37

References 38

List of Tables

1 Tabular outline of suggested physical processes for preferentially heating and accelerating minor ions in coronal holes. 30
1 Introduction

Coronal holes are regions of low density plasma on the Sun that have magnetic fields opening freely into the heliosphere. Because of their low density, coronal holes tend to be the regions of the outer solar atmosphere that are most prone to behaving as a collisionless plasma. Ionized atoms and electrons flow along the open magnetic fields to form the highest-speed components of the solar wind.

The term “coronal hole” has come to denote several phenomena that may not always refer to the same regions. First, the darkest patches on the solar surface, as measured in ultraviolet (UV) and X-ray radiation, are called coronal holes. Second, the term also applies to the lowest-intensity regions measured above the solar limb, seen either during a total solar eclipse or with an occulting coronagraph. Third, there is a more theoretical usage that equates coronal holes with all open-field footpoints of time-steady solar wind flows. There are good reasons why these three ideas should be related to one another, but the overlap between them is not complete. To avoid possible confusion, this paper will mainly use the first two observational definitions, with the third one being only partly applicable.

During times of low solar activity, when the Sun’s magnetic field is dominated by a rotationally-aligned dipole component, there are large coronal holes that cover the north and south polar caps of the Sun. In more active periods of the solar cycle, coronal holes can exist at all solar latitudes, but they may only persist for several solar rotations before evolving into different magnetic configurations.

Despite not being as visually spectacular as active regions, solar flares, or coronal mass ejections (CMEs), coronal holes are of abiding interest for (at least) three main reasons.

1. The extended corona and solar wind connected with coronal holes tends to exist in an ambient time-steady state, at least in comparison with other regions. This makes coronal holes a natural starting point for theoretical modeling, since it often makes sense to begin with the simplest regions before attempting to understand more complex and variable structures.

2. Coronal hole plasma has the lowest density, which makes it an optimal testbed for studies of collisionless kinetic processes that are the ultimate dissipation mechanisms in many theories of coronal heating. Other regions tend to have higher densities and more rapid Coulomb collisions, and thus the unique signatures of the kinetic processes (in their particle velocity distributions) are not as straightforward to measure as in coronal holes.

3. Coronal holes and their associated high-speed wind streams are also responsible for a fraction of major geomagnetic storms at 1 AU. Corotating interaction regions (CIRs) form when fast and slow wind streams collide with one another, and the subsequent interaction between these structures and the Earth’s magnetosphere can give rise to long-lasting fluxes of energetic electrons.

This paper reviews measurements of the plasma properties of coronal holes and how these measurements have been used to put constraints on theoretical models of coronal heating and solar wind acceleration. There have been several earlier reviews that have focused mainly on the topic of coronal holes, including Zirker (1977), Suess (1979), Harvey and Sheeley Jr (1979), Parker (1991), Kohl and Cranmer (1999), Hudson (2002), Cranmer (2002a), Ofman (2005), de Toma and Arge (2005), Jones (2005), and Wang (2009). Interested readers are urged to survey these other reviews in order to fill in any gaps in topical coverage in the present paper.

The remainder of this paper is organized as follows. Section 2 gives a brief history of the discovery and early years of research on coronal holes. Section 3 summarizes the observations and derived plasma properties of “on-disk” coronal holes (i.e., primarily using the definition of holes as dark patches on the solar surface at UV and X-ray wavelengths). Section 4 reviews
the measurements of “off-limb” coronal holes and describes our current knowledge of how these structures are linked to various kinds of solar wind streams measured in situ. Section 5 discusses a broad range of possible theoretical explanations for how the plasma in coronal holes is heated and how the solar wind in these regions is accelerated. Section 6 concludes this paper with a few words about how the study of coronal holes helps to improve our wider understanding of heliophysics, astrophysics, and plasma physics.
2 Historical Overview

Although the term “coronal hole” was not first used until the middle of the 20th century, people have reported the existence of visible features associated with the Sun’s corona – seen during total eclipses – for centuries (see, e.g., Wang and Siscoe, 1980; Vaquero, 2003). A popular astronomy book from the first decade of the 20th century (Serviss, 1909) contained clear descriptions of coronal streamers, eruptive prominences, and polar plumes in coronal holes. The following description of the latter, from an eclipse in 1900, conveys that early speculation may sometimes be prescient:

“The sheaves of light emanating from the poles look precisely like the ‘lines of force’ surrounding the poles of a magnet. It will be noticed in this photograph that the corona appears to consist of two portions: one comprising the polar rays just spoken of, and the other consisting of the broader, longer, and less-defined masses of light extending out from the equatorial and middle-latitude zones. Yet even in this more diffuse part of the phenomenon one can detect the presence of submerged curves bearing more or less resemblance to those about the poles. Just what part electricity or electro-magnetism plays in the mechanism of solar radiation it is impossible to say, but on the assumption that it is a very important part is based the hypothesis that there exists a direct solar influence not only upon the magnetism, but upon the weather of the earth” (Serviss, 1909).

The first quantitative observations of coronal holes were made by Waldmeier (1956, 1957) at the Swiss Federal Observatory in Zürich. These features were identified as long-lived regions of negligible intensity in coronagraphic (off-limb) images of the 5303˚A green emission line (see also Waldmeier, 1975, 1981). Waldmeier called the features that appeared more-or-less circular when projected onto the solar disk Löcher (holes), and the more elongated features were called Kanal (channels) or Rinne (grooves).

In off-limb eclipse and coronagraph images, the darkest coronal hole regions are surrounded by brighter and more complex streamers. These wispy structures appear to be connected to closed magnetic field lines at the solar surface, but they are often stretched upwards to an elongated cusp-like point, with thin “stalks” of radial rays at the top. For this reason their appearance was likened to a pointed German helmet (or a brush-topped Greek or Roman helmet), and the common phrase helmet streamers is often seen. The earliest studies of coronal morphology tended to concentrate more on streamers than coronal holes because the former are significantly easier to see than the latter (see, e.g., Miller, 1908; Mitchell, 1932; Newkirk Jr, 1967; Pneuman, 1968). Piddington (1972) outlined some early ideas about the global structure of the “quiet” (i.e., solar minimum) corona. Figure 1 compares an adaptation of J. H. Piddington’s sketch of the quiet corona to a more recent photograph from another eclipse around solar minimum.

As the quality of the observations improved, coronal holes became objects of study in their own right. The largest coronal holes were observed to contain fine thread-like polar plumes that appear to follow the superradially expanding open magnetic field lines above the solar limb (Saito, 1958; Stoddard et al., 1966; Newkirk Jr and Harvey, 1968). These elongated structures were found to correlate with bright chromospheric faculae on the surface (e.g., Harvey, 1965) and with longer extensions for the small jet-like spicules that continually rise and fall above the limb (Lippincott, 1957; Beckers, 1968).

Coronal holes were essentially re-discovered in the late 1960s and early 1970s as discrete dark patches on the X-ray and ultraviolet solar disk. Newkirk Jr (1967) reviewed some of the earliest rocket-based measurements in the extreme UV, and Tousey et al. (1968) discussed how the UV emission was “usually weaker over the poles” in images from a series of rocket flights between 1963 and 1967 (around solar minimum). These regions on the solar disk came to be called coronal holes in parallel with the earlier off-limb usage. Munro and Withbroe (1972) analyzed OSO–4 observations to conclude that both the density and electron temperature were lower in these dark regions. In 1973 and 1974, solar instruments on the Apollo telescope mount (ATM) on Skylab confirmed many earlier ideas about coronal holes with data significantly better in quantity and
In addition to the large north and south polar holes, there were also found to be smaller coronal holes that exist at lower latitudes (often at times other than solar minimum). Sometimes the largest coronal holes can exhibit thin “peninsulas” that jut out from the main regions. Harvey and Recely (2002) called these regions “polar lobes.” Notable examples have been the so-called “Boot of Italy” seen by Skylab in 1974 (e.g., Zirker, 1977) and the “Elephant’s Trunk” seen by SOHO in 1996 (Del Zanna and Bromage, 1999). A third example, from December 2000, is shown in Figure 2.

Additional insights came from the fusion of spectroscopy and coronagraphic occultation. Inspired by rocket-borne UV observations of the extended corona during a solar eclipse in March 1970, Kohl et al. (1978) developed a UV coronagraph spectrometer to measure the profile shape of the bright H\textsc{i} Ly\textalpha emission line at 1216 Å. This line is sensitive to several key properties of the velocity distribution of coronal protons, and thus these observations could be used to begin distinguishing proton temperatures from electron temperatures in the collisionless outer regions of coronal holes (see Section 4.3). The rocket-borne UV coronagraph spectrometer was launched three times (in 1979, 1980, and 1982), and the results included the first direct evidence for proton heating and supersonic outflow in coronal holes (Kohl et al., 1980; Strachan et al., 1993; Kohl et al., 2006).

The fact that coronal holes coincide with regions of open magnetic field that expands out into interplanetary space was realized during the first decade of in situ solar wind observations (e.g., Wilcox, 1968; Altschuler et al., 1972; Hundhausen, 1972). Noci (1973) made a theoretical argument, on the basis of measured wave fluxes and heat conduction, that coronal holes should have the largest solar wind kinetic energy fluxes (i.e., the highest speeds). Pneuman (1973) argued that coronal holes need not have lower energy deposition than closed-field regions (as is suggested by the lower intensities of coronal holes) if the solar wind carries away much of that energy. Krieger et al. (1973) utilized X-ray sounding rocket images to identify a large coronal hole as the solar source of a strong high-speed stream as measured by the Pioneer 6 and Vela spacecraft. Around this time it was also realized that coronal holes and high-speed wind streams are also responsible for a fraction of the geomagnetic storms seen at 1 AU (Bell and Noci, 1973; Neupert and Pizzo, 1974; Bell and Noci, 1976; see also Tanskanen et al., 2005 and Zhang et al., 2007). Although there is still no complete understanding of which types of solar wind flow are connected with which types

Figure 1: Left: Adaptation of a sketch of the quiet solar corona made by Piddington (1972), based on prior drawings (Waldmeier, 1955) and photographs (Gold, 1955) of the 30 June 1954 eclipse. Right: Contrast-adjusted eclipse image taken with the POISE instrument on 26 February 1998, in Westpunt, Curacao. The original image was made available courtesy of the High Altitude Observatory (HAO), University Corporation for Atmospheric Research (UCAR), Boulder, Colorado. UCAR is sponsored by the National Science Foundation.
of coronal structures, the causal link between the largest coronal holes and high-speed streams remains firm (see also Section 4.1).
3 Properties of On-Disk Coronal Holes

The traditional observational distinction between coronal holes and their surroundings (i.e., active regions and "quiet Sun") is that coronal holes have the lowest emission in the UV and X-ray. This definition must be amended, however, to exclude filaments (which are often dark when projected against the solar disk) that are cool magnetic structures and not part of the corona (see, e.g., Scholl and Habbal, 2008; Krista and Gallagher, 2009). Coronal hole magnetic fields are known to be more uniform and unipolar than in other regions (see below). The boundaries between coronal holes and surrounding regions are sometimes sharp, sometimes diffuse, and sometimes filled with many small loops (Hudson, 2002).

Coronal holes are more or less indistinguishable from their surroundings in the photosphere and low chromosphere, and usually one cannot see any significant intensity contrast between hole and non-hole regions until the temperature exceeds about $10^5$ K. Spectra, however, can help make the distinction clearer. Teplitskaya et al. (2007) found that central self-reversals in the chromospheric Ca II H and K lines are noticeably different in coronal holes as opposed to surrounding quiet-Sun regions. A frequently used observational diagnostic of coronal hole boundaries is the He I 10830 Å near-infrared absorption line triplet (Harvey et al., 1975; Harvey and Recely, 2002). At these wavelengths, the absorption is weakest in coronal holes (i.e., the intensity is highest) and spectroheliogram images show the coronal holes quite clearly. It is somewhat counterintuitive that a spectral line of a neutral species (He I) should be sensitive to the properties of the hot corona. However, the atomic level populations that determine the He I 10830 Å source function are unusually responsive to photoionization from UV wavelengths shortward of 504 Å. The overlying solar corona emits these wavelengths in abundance. In coronal hole regions, though, the corona generally has a lower density and temperature, and thus there is less intense UV and X-ray emission to populate the upper levels of the He I triplet state (see, e.g., Zirin, 1975; Andretta and Jones, 1997; Centeno et al., 2008). This gives rise to reduced absorption. The He I 10830 Å lines are also good probes of supersonic outflow velocities in distant stellar winds (see, e.g., Dupree et al., 2005; Kwan et al., 2007; Sanz-Forcada and Dupree, 2008).

Another observational diagnostic of coronal holes is their elemental abundance signature (e.g., Feldman, 1998; Feldman and Widing, 2003). In the upper chromosphere, transition region, and low corona, holes exhibit abundances very close to those measured in the photosphere. This stands in contrast to other regions (quiet Sun and active regions), which show significant enhancements in the abundances of elements with low first ionization potential (FIP). This selective fractionation is believed to occur in the upper chromosphere, where low-FIP elements become ionized and high-FIP elements remain more neutral. These patterns extend into the heliosphere, where high-speed flows associated with coronal holes are often identifiable from their near-photospheric FIP abundances (von Steiger et al., 1995; Zurbuchen et al., 2002). However, there is still no widespread agreement about the exact physical processes that give rise to this preferential ionization.

The number, sizes, and heliographic locations of coronal holes vary as a function of the solar activity cycle. Large polar holes exist for about 7 years around solar minimum, and are not present for about 3 or 4 years around solar maximum. However, in the declining phase of activity soon after the maximum, it is possible to see the gradual growth of the new-polarity polar coronal holes. This occurs as a number of smaller high-latitude holes eventually collect together at the poles (Timothy et al., 1975; Harvey and Recely, 2002). Taking this growth phase into account, there are only about 1 or 2 years at solar maximum without any distinct high-latitude coronal hole presence. Figure 3 illustrates the growth phase around the peak of solar cycle 23 in 2001. The growth and full "reappearance" of polar coronal holes at this time was described by Miralles et al. (2001a) and McComas et al. (2002). The post-maximum growth of new polar coronal holes lasts about twice as long as their disappearance in the rising phase of the next maximum (Waldmeier, 1981; Fisher and Sime, 1984).
Many low-latitude coronal holes tend to be situated near the edges of magnetically complex active regions. Sometimes active regions emerge within the coronal holes themselves; these have been called “sea anemone” type regions from their spiky, flower-like appearance (Shibata et al., 1994; Asai et al., 2008). Evolving magnetic interactions between active regions and coronal holes have been studied both as a means of enhancing the mass flux of the associated solar wind on nearby open flux tubes (e.g., Habbal et al., 2008; Wang et al., 2009) and as a possible explanation for the nearly rigid rotation of coronal holes (Wang et al., 1996; Antiochos et al., 2007). A burst of emerging magnetic flux in one of these active regions may give rise to new systems of loop connections in the area bordering the coronal hole, and thus cause the coronal hole to decrease in size. This kind of rapid topological evolution of the magnetic field may be relevant in explaining extreme space weather events such as “the day the solar wind disappeared” (i.e., the dramatic drop in the in situ density seen on 11 May 1999; see Janardhan et al., 2008).

Photospheric magnetograms show that coronal holes are more unipolar than other regions; i.e., they have a larger degree of magnetic flux imbalance between the two polarities (Levine, 1982; Wang, 2009). For the large polar coronal holes, this appears to be the long-term outcome of the decay of active-region magnetic fields and their eventual diffusion up to the poles. The unipolar nature of coronal holes is likely to be related to their connection with open-field solar wind streams. As described in Section 2, one reason why coronal holes are dark is that the solar wind carries away both mass and energy, leaving a lower density and pressure. In addition, Abramenko et al. (2006a) and Hagenaar et al. (2008) found that the local rate of emergence of small-scale magnetic flux (mostly in the form of “ephemeral” small-scale bipoles) is substantially lower in unipolar regions than in more mixed or balanced regions of positive and negative magnetic polarity. In most theories of coronal heating of closed loops, the total flux and the overall complexity of the field both drive the total heating rate. Thus, the lower emergence rate of new flux elements in coronal holes may be another factor determining why they have lower densities and pressures (i.e.,
less coronal heating; see, e.g., Rosner et al., 1978) and thus why they are dark.

Over the past few decades, the magnetic connection between coronal holes and the fast solar wind has been traced down to ever smaller spatial scales. We now can say with some certainty that much of the plasma that eventually becomes the time-steady solar wind seems to originate in thin magnetic flux tubes (with observed sizes of order 50 – 200 km) observed mainly in the dark lanes between the ∼ 1000 km size photospheric granulation cells. These strong-field (1 – 2 kG) flux tubes have been called G-band bright points and network bright points, and groups of them have been sometimes termed “solar filigree” (e.g., Dunn and Zirker, 1973; Spruit, 1984; Berger and Title, 2001; Tsuneta et al., 2008). These flux tubes are concentrated most densely in the supergranular network (i.e., the bright lanes between the larger ∼ 30,000 km size supergranulation cells). Somewhere in the low chromosphere, the thin flux tubes expand laterally to the point where they merge with one another into a more-or-less homogeneous network field distribution of order ∼ 100 G. This merged (mainly vertically oriented) field is associated mainly with the lanes and vertices between supergranular cells. Because this field does not fill the entire coronal volume, it is still susceptible to an additional stage of lateral expansion and weakening. Thus, at a larger height in the chromosphere, these network flux bundles are believed to undergo further broadening into so-called “funnels” (Gabriel, 1976; Dowdy et al., 1986). However, it is still unclear to what extent the closed fields in the supergranular cell centers affect the canopy-like regions between funnels (e.g., Schrijver and Title, 2003). Figure 4 illustrates the successive merging of unipolar field in coronal holes.

Observations of blueward Doppler shifts in supergranular network lanes and vertices, especially in large coronal holes, may be evidence for either the “launching” of the solar wind itself or for upward-going waves that are linked to wind acceleration processes (e.g., Hassler et al., 1999; Peter and Judge, 1999; Aionaz et al., 2005; Tu et al., 2005). These measurements are consistent with several models of the dynamic, multi-species solar wind in superradially expanding funnels (Byhring et al., 2008; Marsch et al., 2008). However, this interpretation of the data is still not definitive, because there are other observational diagnostics that have shown more of a blueshift in the supergranular cell-centers between funnels (e.g., HeI 10830 Å; Dupree et al., 1996; Malanushenko and Jones, 2004). There may be subtle radiative transfer effects that preferentially brighten regions of upflow or downflow (see, e.g., Chae et al., 1997; Avrett, 1999), and these may need to be taken into account in order to understand the meaning of the measured Doppler shifts in these regions.

Lastly, it is important to mention the phenomenon of transient coronal holes (sometimes known...
as “coronal dimmings”). These are rapid reductions in the UV and X-ray intensity in discrete patches that appear to be spatially and temporally correlated with the liftoff of CME plasma (e.g., Rust, 1983; Thompson et al., 2000; Yang et al., 2008). Spectroscopically, these regions are seen to exhibit similar characteristics as normal coronal holes, including Doppler blueshifts (Harra et al., 2007) and large amplitudes of nonthermal wave broadening (McIntosh, 2009). UV coronal dimmings are beginning to be used as diagnostics for the amount of plasma released – i.e., the total mass – in the CME event (e.g., Aschwanden et al., 2009). It should be noted that transient coronal holes represent just one kind of observed dimming that is associated with time-dependent flare/CME activity; there are other types of dimmings that do not resemble coronal holes (Hudson, 2002).
4 Properties of Off-Limb Coronal Holes

Coronal holes observed above the solar limb usually trace out the same regions that are identified as dark coronal-hole “patches” directly on the solar disk. Thus, the lower plasma density that more or less defines the off-limb coronal hole is directly related to the lower density measured on the disk. Section 4.1 briefly discusses how these regions are believed to be connected magnetically with the broader heliosphere. The observations of off-limb coronal holes made with visible-light imaging and polarimetry (Section 4.2) and ultraviolet spectroscopy (Section 4.3) are also summarized below.

4.1 Magnetic connectivity with the solar wind

Although the magnetic field in the solar corona is generally too weak to be measured directly, the overall morphology of the field lines can be extrapolated from magnetograms taken at the level of the photosphere. One popular technique is the “potential field source surface” (PFSS) method, which assumes the corona is current-free out to a spherical surface (set typically at a radius between 2.5 and 3.5 solar radii, or $R_\odot$), above which the field is radial (e.g., Schatten et al., 1969; Altschuler and Newkirk Jr, 1969; Hoeksema and Scherrer, 1986). The PFSS method has been shown to create a relatively good mapping between the Sun and the heliosphere (Arge and Pizzo, 2000; Luhmann et al., 2002; Wang and Sheeley Jr, 2006; Wang, 2009), although the results can be problematic for regions dominated by stream-stream interactions (Poduval and Zhao, 2004).

By far, the strongest causal link between a specific type of coronal structure (measured via remote sensing) and a particular type of quasi-steady solar wind flow (measured in situ) is the connection between large coronal holes and high-speed streams (Wilcox, 1968; Krieger et al., 1973). The general interpretation of this correlation – together with the results of magnetic extrapolation models such as PFSS – is that coronal holes represent a bundle of open flux tubes that flare out horizontally as distance increases. In other words, the coronal hole flux tubes expand superradially. Although there are some observations that appear to support other interpretations (Woo et al., 1999; Habbal et al., 2001; Woo, 2005; Woo and Druckmüllerová, 2008), the preponderance of evidence seems clearly to support the idea that fast solar wind streams emerge mainly from superradially expanding coronal holes (e.g., Guhathakurta et al., 1999b; Cranmer et al., 1999b; Jones, 2005; Wang and Sheeley Jr, 2006; Wang et al., 2007).

In contrast to the rather definitive correlation between large coronal holes and the fast solar wind, the coronal sources of the more chaotic slow-speed solar wind are not as well understood (see Schwenn, 2006). Two regions that are frequently cited as sources of slow wind are: (1) boundaries between coronal holes and streamers, and (2) narrow plasma sheets that extend out from the tops of streamer cusps (Wang et al., 2000; Strachan et al., 2002; Susino et al., 2008). These regions tend to dominate around solar minimum. Note that the former type of boundary region tends to contain flux tubes that may be classified as coronal holes when using the theoretical definition (i.e., footpoints of field lines that are open; see Section 1) but would not be defined as such when using the observational definitions (i.e., low emission or low density).

During more active phases of the solar cycle, there is evidence that slow solar wind streams also emanate from small coronal holes (e.g., Nolte et al., 1976; Neugebauer et al., 1998; Zhang et al., 2003) and active regions (Hick et al., 1995; Liewer et al., 2004; Sakao et al., 2007). During the rising phase of solar activity, there seems to be a relatively abrupt ($< 6$ month) change in the locations of slow wind footpoints: from the high-latitude hole/streamer boundaries to the low-latitude active region and small coronal hole regions (Luhmann et al., 2002). The ability of many of these kinds of regions to produce slow wind was modeled by Cranmer et al. (2007) and Wang et al. (2009); see also Section 5.

Living Reviews in Solar Physics
http://www.livingreviews.org/lrsp-2009-3
4.2 Visible light observations

Measurements of the plasma properties in the extended corona (i.e., \( r \approx 1.5 \) to \( 10 R_{\odot} \), where the main solar wind acceleration occurs) require the bright solar disk to be occulted. The coronal emission is many orders of magnitude less bright than the emission from the solar disk, so even the vast majority of “stray light” that diffracts around the occulting edge must be eliminated. Visible-light coronagraphs that combine stray-light rejection with linear polarimetry have the ability to measure the Thomson-scattered polarization brightness \( (pB) \) in the corona. The use of \( pB \) rather than the total coronal brightness eliminates the contribution from the dust-scattered F-corona, which is believed to be unpolarized up to distances of about \( 5 R_{\odot} \). Because the coronal plasma is optically thin to the Thomson-scattered photons, \( pB \) is proportional to the line-of-sight integral of the electron density \( n_e \), multiplied by a known scattering function. Methods for inverting this integral to derive \( n_e \) as a function of position in various coronal structures have been developed and improved over the years (e.g., van de Hulst, 1950; Altschuler and Perry, 1972; Munro and Jackson, 1977; Guhathakurta and Holzer, 1994; Frazin et al., 2007). For coronal holes, the LASCO (Large Angle and Spectrometric Coronagraph) instrument on SOHO has also been used to probe the superradial expansion of open magnetic flux tubes (DeForest et al., 1997, 2001) and the evolution of transient polar jets (Wang et al., 1998; Wood et al., 1999). The White Light Coronagraphs on Spartan 201 (Fisher and Guhathakurta, 1995) and on the UVCS (Ultraviolet Coronagraph Spectrometer) instrument aboard SOHO (e.g., Kohl et al., 1995; Romoli et al., 2002) have provided electron densities between \( 1.5 \) and \( 5 R_{\odot} \) in coronal holes.

Figure 5 illustrates a selection of visible-light measurements of the electron density in coronal holes and compares them to similar measurements of streamers and to a semi-empirical model of the chromosphere, transition region, and low corona (Avrett and Loeser, 2008). The blue coronal hole curves were adapted from the results of Fisher and Guhathakurta (1995) (dotted), Cranmer et al. (1999b) (solid), Doyle et al. (1999) (dashed), and Guhathakurta et al. (1999a) (dot-dashed). The red curves for equatorial helmet streamers were adapted from the results of Sittler Jr and Guhathakurta (1999) (solid) and Gibson et al. (1999) (dashed).

Note that streamers are denser than coronal holes by about a factor of 10, but the hole measurements themselves can often exhibit variations in the electron density by factors of the order of \( 2 \) – \( 3 \). Much of this spread is likely to be the result of different lines of sight passing through regions that contain varying numbers of polar plumes (see, e.g., Cranmer et al., 1999b, 2008). Some of the cited \( pB \) observations were optimized to avoid bright concentrations of plumes, and others have been purposefully averaged over the full range of coronal hole substructure. It is also possible that absolute calibration uncertainties may still exist between the different instruments used to determine \( pB \) and \( n_e \), and this could compound the reported range of variation in coronal hole electron densities.

For a steady-state solar wind outflow, the conservation of mass demands that the product of density, flow speed, and cross-sectional area of the flux tube remain constant. Thus, if the magnetic geometry and the electron density are known, mass conservation allows the solar wind outflow speed to be computed. Kohl et al. (2006) used the representative values of \( n_e \) shown in Figure 5 together with a range of estimates for the superradial flux-tube expansion of coronal holes to determine outflow speeds in coronal holes. Figure 41a of Kohl et al. (2006) illustrates the result of this process, which shows a large range of values reflecting the uncertainties in both \( n_e \) and the flux-tube area factor. Despite these uncertainties, though, the electron densities that became available in the 1990s demonstrated that the fast solar wind accelerates rapidly in coronal holes – probably reaching half of its asymptotic terminal speed \( (u_\infty \approx 700 – 800 \text{ km s}^{-1}) \) by heights no larger than \( 2 – 4 R_{\odot} \).

The increased sensitivity of the LASCO instrument over earlier coronagraphs revealed a nearly continual release of low-contrast density inhomogeneities, or “blobs,” from the cusps of helmet

---

Living Reviews in Solar Physics

http://www.livingreviews.org/lrsp-2009-3
Figure 5: Comparison of empirically determined densities in the upper solar atmosphere. Avrett and Loeser (2008) values of electron number density (solid black curve) and total hydrogen number density (dot-dashed black curve) are compared with various visible-light $pB$ electron number densities for coronal holes (blue curves) and streamers (red curves); see text for details.

streamers (Sheeley Jr et al., 1997; Tappin et al., 1999; Wang et al., 2000; Chen et al., 2009). These features were seen to accelerate up to speeds of order 300–400 km s$^{-1}$ by the time they reached the outer edge of the LASCO field of view ($r \approx 30 R_\odot$); see also Figure 8 below. The blobs are typically only about 10% to 15% brighter or dimmer than the surrounding streamer material. Because of this low contrast, these features do not appear to comprise a large fraction of the mass flux of the slow solar wind. However, it is still unclear whether blobs are passive “tracers” that flow with the solar wind speed, or whether they are wavelike fluctuations that propagate relative to the background solar wind reference frame. This diagnostic tool has been much more difficult to apply in coronal holes than it has in the bright streamers, so no firm measurements of the fast wind acceleration yet exist from this technique.

Visible light measurements have also revealed evidence for compressive magnetohydrodynamic (MHD) waves that propagate along open field lines in coronal holes. Intensity oscillations measured by the UVCS and EIT instruments on SOHO were found to have periodicities between about 10 and 30 minutes and are consistent with being upwardly propagating slow-mode magnetosonic waves (DeForest and Gurman, 1998; Ofman et al., 1999, 2000). The relative amplitude of the density fluctuations ($\delta n/n_0$) for these waves was found to range between about 0.03 and 0.15 (see Cranmer, 2004a). This is consistent with measurements of the density fluctuation amplitudes made at larger distances via radio scintillations (Spangler, 2002) and in situ instruments (Tu and Marsch, 1994). There have also been claims that low-frequency oscillations have been measured in H I Ly$\alpha$ emission (Morgan et al., 2004; Bemporad et al., 2008; Telloni et al., 2009). In these cases, however, it is extremely important to take into account all of the relevant instrumental effects. These measurements still appear to be provisional.

As seen in Section 2 above, coronal holes have long been observed as the sites of thin, ray-like polar plumes. The earliest measurements of polar plume properties were made in broad-band visible light, and these dense strands are often seen to stand out distinctly from the ambient
Coronal Holes

interplume corona. It is not clear, though, to what extent off-limb observations (which integrate over long optically thin lines of sight) ever capture only the “pure” plume or interplume plasmas. Space-based observations from, e.g., Spartan 201 and SOHO improved our ability to measure the physical properties in and between plumes (e.g., Fisher and Guhathakurta, 1995; DeForest et al., 1997; Cranmer et al., 1999b; DeForest et al., 2001). Although the brightest plumes are still discernible at the uppermost heights observed by LASCO (i.e., 30 – 40 $R_\odot$), the plume/interplume density contrast becomes too low to measure clearly in interplanetary space ($r > 60 R_\odot$). However, indirect and possibly plume-related signatures in the in situ data have been reported by Thieme et al. (1990), Reisenfeld et al. (1999), and Yamauchi et al. (2002). The disappearance of plumes probably is the result of some combination of transverse pressure balance (i.e., dense plumes expanding to fill more of the available volume; see Del Zanna et al., 1998) and MHD instabilities that can mix the two components (Parhi et al., 1999; Andries and Goossens, 2001).

4.3 Ultraviolet spectroscopy

Ultraviolet spectroscopy of the corona is a powerful tool for obtaining detailed empirical descriptions of solar plasma conditions (see, e.g., Kohl and Withbroe, 1982; Withbroe et al., 1982). Coronal holes, being the lowest density regions of the outer solar atmosphere, exhibit a complex array of plasma parameters due to their nearly collisionless nature. As a result, every ion species tends to have its own unique temperature, its own type of departure from a Maxwellian velocity distribution, and its own outflow speed. After a brief summary of near-limb measurements made with the SUMER instrument on SOHO, this section mainly describes results at larger heights (in the more clearly collisionless extended corona) from UVCS.

The un-occulted SUMER (Solar Ultraviolet Measurements of Emitted Radiation) spectrometer has observed off-limb regions up to heights of approximately 1.3 $R_\odot$ in coronal holes (Wilhelm et al., 1995, 2000, 2004, 2007). In polar regions at solar minimum, ion temperatures exceed electron temperatures even at $r \sim 1.05 R_\odot$, where densities were presumed to be so high as to ensure rapid collisional coupling and thus equal temperatures for all species (Seely et al., 1997; Tu et al., 1998; Landi and Cranmer, 2009). Spectroscopic evidence is also mounting for the presence of transverse Alfvén waves propagating into the corona (Banerjee et al., 1998, 2009; Dolla and Solomon, 2008; Landi and Cranmer, 2009).

Electron temperatures derived from line ratios (Habbal et al., 1993; David et al., 1998; Doschek et al., 2001; Wilhelm, 2006; Landi, 2008) exhibit relatively low values in the off-limb coronal holes ($T_e \sim 800,000$ K) that are not in agreement with higher temperatures derived from “frozen-in” in situ charge states (Ko et al., 1997). It is difficult to reconcile these observations with one another in the absence of either non-Maxwellian electron velocity distributions or strong differential flows between different ion species near the Sun (Esser and Edgar, 2000). However, some models consistent with both the SUMER temperatures and the in situ charge states are being produced (e.g., Laming and Lepri, 2007).

Figure 6 displays a range of temperatures measured in coronal holes and high-speed wind streams, and it shows how the SUMER electron temperatures ($T_e$) are noticeably lower than the heavy ion temperatures (e.g., shown for the O$\,^\mathrm{+5}$ ions that correspond to the bright O vi 1032, 1037 Å spectral lines) even in the low corona. The degree of agreement between the spectroscopic measurements and one-fluid models of the low corona (i.e., the semi-empirical model of Avrett and Loeser (2008) and the theoretical model of Cranmer et al. (2007)) depends on the height of the sharp transition region between chromospheric ($10^4$ K) and coronal ($10^6$ K) temperatures.

The UVCS instrument on SOHO is a combination of an ultraviolet spectrometer and a linearly occulted coronagraph that observes a 2.5 $R_\odot$ long swath of the extended corona, oriented tangentially to the radial direction at heliocentric distances ranging between about 1.5 and 10 $R_\odot$ (Kohl et al., 1995, 1997, 1998, 2006). In coronal holes, UVCS measurements have allowed many
key details about the velocity distributions of H\(^0\), O\(^{+5}\), and Mg\(^{+9}\) to be derived. For the resonantly scattered emission lines seen at large heights with UVCS, the most straightforward plasma diagnostic is to use the Doppler-broadened line width as a sensitive probe of the overall variance of random particle motions along the line of sight. In other words, measuring the line width provides a constraint on the so-called “kinetic temperature” (i.e., a combination of microscopic stochastic motions and macroscopic [but unresolved] motions due to waves or turbulence) along the direction perpendicular to the (nearly radial) magnetic field lines.

In the ionized solar corona, a given hydrogen nucleus spends most of its time as a free proton, and only a small fraction of time as a bound H\(^0\) atom. Thus, the measured plasma properties of neutral hydrogen are considered to be valid proxies of the proton properties below about 3 \(R_\odot\) (Allen et al., 2000). Spartan 201 and UVCS observations of the H\(^i\) Ly\(\alpha\) emission line in coronal holes indicated rather large proton kinetic temperatures in the direction perpendicular to the magnetic field (\(T_{p\perp} \sim 3\) MK) and also the possibility of a mild temperature anisotropy (with \(T_{p\perp} > T_{p||}\)) above heights of 2 – 3 \(R_\odot\) (Kohl et al., 1997; Cranmer et al., 1999b; Antonucci et al., 2004; Kohl et al., 2006).

UVCS observations indicated that the O\(^{+5}\) ions are much more strongly heated than protons in coronal holes, with perpendicular temperatures in excess of 200 MK (see Figure 7). This exceeds the temperature at the central core of the Sun by an order of magnitude! The UVCS measurements also provided signatures of temperature anisotropies possibly greater than \(T_{i\perp}/T_{i||} \approx 10\) (e.g., Cranmer et al., 1999b, 2008). The measured kinetic temperatures of O\(^{+5}\) and Mg\(^{+9}\) are significantly greater than mass-proportional when compared with protons, with \(T_i/T_p > m_i/m_p\) (see also Kohl et al., 1999, 2006). The surprisingly “extreme” properties of heavy ions in coronal holes have led theorists...
Coronal Holes

On-disk profiles: $T = 1–3$ MK

Off-limb profiles: $T \gtrsim 200$ MK

**Figure 7:** Combined image of the solar corona from 17 August 1996, showing the solar disk in Fe XII 195 Å intensity from EIT (yellow inner image) and the extended corona in O vi 1032 Å intensity from UVCS (red outer image). Axisymmetric field lines are from the solar-minimum model of Banaszkiewicz et al. (1998), and O vi emission line profiles (bottom) are from SUMER (Warren et al., 1997, left) and UVCS (Kohl et al., 1997, right).

to develop a number of new ideas regarding the heating and acceleration of the solar wind; these are discussed further in Section 5.4.

Figure 6 shows UVCS perpendicular temperatures for protons and O$^{+5}$ ions in coronal holes. The O$^{+5}$ data points were taken from the recent re-analysis of solar minimum data from 1996–1997 by Cranmer et al. (2008). The proton temperature data were assembled by Cranmer (2004b) from a number of individual measurements of the H I Lyα profile at solar minimum. The sources of these measurements are: Cranmer et al. (1999b) (squares), Esser et al. (1999) (diamonds), Zangrilli et al. (1999) (asterisks), and Antonucci et al. (2000) (triangles). The kinetic proton temperatures are of order 2–3.5 MK, but in Figure 6 we attempted to remove the contribution of nonthermal wave broadening. The semi-empirical model of Cranmer and van Ballegooijen (2005) was used to specify the amplitude of transverse Alfvén waves as a function of height, and their contribution to the line widths was subtracted. The remaining “microscopic” $T_{\perp p}$ does not show as clear a signal of “preferential” proton heating as would be apparent from the larger kinetic temperature. Although one can still marginally see that $T_p > T_e$, the existing measurements of $T_p$ and $T_e$ do not fully overlap with one another in radius. Improved measurements are needed in order to better constrain the proton and electron heating rates in the corona.

The UVCS emission line data contain information about the Doppler motions of atoms and ions along the magnetic field (i.e., transverse to the line of sight). The so-called “Doppler dimming” diagnostic technique provides constraints on both the bulk outflow speed along the field and the
Figure 8: Radial dependence of solar wind outflow speeds. UVCS Doppler dimming determinations for protons (red; Kohl et al., 2006) and O$^{+5}$ ions (green; Cranmer et al., 2008) are shown for polar coronal holes, and are compared with theoretical models of the polar and equatorial solar wind at solar minimum (black curves; Cranmer et al., 2007) and the speeds of “blobs” measured by LASCO above equatorial streamers (open circles; Sheeley Jr et al., 1997).

parallel kinetic temperature (for more details, see Kohl and Withbroe, 1982; Noci et al., 1987; Kohl et al., 2006). In coronal holes, Doppler-dimmed line intensities from UVCS are consistent with the outflow velocity for O$^{+5}$ being larger than the outflow velocity for protons by as much as a factor of two at large heights (Kohl et al., 1998; Li et al., 1998; Cranmer et al., 1999b). Figure 8 illustrates the outflow speeds measured by UVCS in coronal holes, and compares with the theoretical model of the fast solar wind presented by Cranmer et al. (2007). Also shown for comparison are observational and theoretical data for the slow solar wind associated with equatorial helmet streamers at solar minimum.

In contrast to many prior analyses of UVCS data, which concluded that there must be both intense preferential heating of the O$^{+5}$ ions and a strong field-aligned anisotropy, Raouafi and Solanki (2004), Raouafi and Solanki (2006), and Raouafi et al. (2007) reported that there may not be a compelling need for O$^{+5}$ anisotropy depending on the assumptions made about the other plasma properties of the coronal hole (e.g., electron density). However, Cranmer et al. (2008) performed a detailed re-analysis of these observations and concluded that there remains strong evidence in favor of both preferential O$^{+5}$ heating and acceleration and significant O$^{+5}$ ion anisotropy (in the sense $T_{\perp} > T_{\parallel}$) above $r \approx 2.1 R_\odot$ in coronal holes. In determining these properties, it was found to be important to search the full range of possible ion temperatures and flow speeds, and not to make arbitrary assumptions about any given subset of the parameters.

The UVCS results discussed above are similar in character to in situ measurements made in the fast solar wind, but they imply more extreme departures from thermodynamic equilibrium in the extended corona. For example, proton velocity distributions measured in the fast solar wind between 0.3 and 1 AU have anisotropic cores with $T_{p\perp} > T_{p\parallel}$, and their magnetic moments increase with increasing distance; this implies net input of perpendicular energy on kinetic scales
Coronal Holes

(e.g., Marsch et al., 1982; Marsch, 2006). Many heavy ions appear to flow faster than the protons by about the local Alfvén speed (Hefti et al., 1998; Reisenfeld et al., 2001), and in the fastest solar wind streams they are also heated preferentially in the same sense as in the corona (Collier et al., 1996).

In the years since the solar minimum of 1996–1997, UVCS observed a large number of other coronal holes that appeared throughout the maximum of solar cycle 23 and the new-millennium solar minimum of 2007–2009. UVCS tends to observe only the largest coronal holes, since when the smallest holes rotate to the solar limb their UV line profiles tend to be contaminated by emission from streamers in the foreground and background. This selection effect naturally screens out small coronal holes that have been correlated with slow solar wind streams at 1 AU (Nolte et al., 1976; Neugebauer et al., 1998). In the cases where UVCS and in situ measurements were made for the same regions associated with large coronal holes, high speeds in excess of 600 km s\(^{-1}\) were inevitably seen in interplanetary space (Miralles et al., 2004, 2006). However, the O\(^{+5}\) outflow speeds measured in the extended corona via Doppler dimming have showed a substantial range of variation. For example, Miralles et al. (2001b) found that the outflow speeds at 2–3 \(R_\odot\) in an equatorial coronal hole were approximately three times lower than those measured in the polar coronal holes from 1996–1997 at similar heights (see also Poletto et al., 2002). This implies that the range of coronal heights over which the acceleration of the solar wind occurs can vary greatly, even when the wind at 1 AU ends up similarly fast.

UVCS also has measured the plasma properties in bright polar plumes. The densest concentrations of polar plumes along the line of sight are seen to exhibit narrower line widths – i.e., lower kinetic temperatures – than the lower-density interplume regions (Kohl et al., 1997; Noci et al., 1997; Kohl et al., 2006). Similarly, plumes are seen to have lower outflow speeds than the interplume regions (Giordano et al., 2000; Teriaca et al., 2003; Raouafi et al., 2007), although at low heights the data are not as clear-cut (e.g., Gabriel et al., 2003). UVCS also has put constraints on the plume/interplume density contrast and the filling factor of polar plumes in coronal holes. Cranmer et al. (1999b) used a large number of UVCS synoptic measurements to determine statistically that, at heights around \(r \approx 2 R_\odot\), the plume/interplume density ratio is approximately 2, and polar coronal holes are comprised of about 25% plume and 75% interplume plasma (corresponding to \(\sim 40\) individual plumes distributed throughout the coronal hole). Earlier measurements made closer to the limb showed a higher density contrast and a smaller filling factor, so the UVCS data are generally consistent with lateral expansion of polar plumes with increasing distance.

It is still relatively unknown how much of the mass, momentum, and energy flux of the fast solar wind comes from polar plumes. Despite that uncertainty, though, there have been several reasonably successful models of polar plume formation. Wang (1994, 1998) presented models of polar plumes as the extensions of flux tubes with concentrated bursts of added coronal heating at the base – presumably via nanoflare-like reconnection events in X-ray bright points (see also DeForest et al., 2001). In these models, the extra basal heat input is balanced by conductive losses to produce the larger plume density. The heating rate in the extended corona is not affected by the basal burst, but the larger density in the flux tube gives rise to less heating per particle at all heights, which leads to lower temperatures in the extended corona and a smaller gas pressure force for solar wind acceleration. This model is consistent with the smaller temperatures and outflow speeds measured in plumes with UV spectroscopy.

UVCS made the first spectroscopic measurements of polar jets in coronal holes (Dobrzycka et al., 2000, 2006). The events observed by EIT, LASCO, and UVCS during the 1996–1997 solar minimum tended to be “cool jets” with higher densities, lower temperatures, and faster outflows than the surrounding coronal holes (see also Wang et al., 1998). More recently, Hinode has observed a new population of “hot” X-ray jets in coronal holes (Culhane et al., 2007; Cirtain et al., 2007; Shimojo et al., 2007; Filippov et al., 2009). UVCS found that some of these events persist up to heights of at least 1.7 \(R_\odot\) and that the jet protons remain hotter than the surrounding coronal hole.
(Miralles et al., 2007). Thus, there appear to exist two distinct kinds of polar jets (cool and hot), with differences possibly related to the relative degrees of heating and adiabatic expansion of the jet parcels. Jets and plumes have roughly similar angular sizes and intensity contrasts near the solar limb, and there is growing evidence that they share a common origin (e.g., Raouafi et al., 2008). It is possible that the only substantial difference between the two phenomena is the duration of the bursts of basal heating; i.e., jets seem to be the result of short-lived bursts of heating, whereas plumes may be the product of base-heating events that last longer than several hours.
5 Coronal Heating and Solar Wind Acceleration

Despite more than a half-century of study, the basic physical processes that are responsible for heating the million-degree corona and accelerating the supersonic solar wind are not known. This section broadens the topic of this paper a bit beyond just coronal holes, since an understanding of solar wind acceleration naturally encompasses not only the question of why fast solar wind streams are fast, but also why (various kinds of) slow solar wind streams are slow. Section 5.1 summarizes some of the major issues regarding coronal energy deposition. The next two subsections describe two alternate views of solar wind acceleration via waves and turbulence in open flux tubes (Section 5.2) and reconnection between open and closed flux tubes (Section 5.3). Lastly, Section 5.4 reviews how collisionless kinetic effects in coronal holes (i.e., preferential ion heating and temperature anisotropies) can be used to more conclusively identify the detailed physical processes that produce the solar wind.

5.1 Sources of energy

Different physical mechanisms for heating the corona probably govern active regions, closed loops in the quiet corona, and the open field lines that give rise to the solar wind (see other reviews by Marsch, 1999; Hollweg and Isenberg, 2002; Longcope, 2004; Gudiksen, 2005; Aschwanden, 2006; Klimchuk, 2006). The ultimate source of the energy is the solar convection zone (e.g., Abramenko et al., 2006b; McIntosh et al., 2007). A key aspect of solving the “coronal heating problem” is thus to determine how a small fraction of that mechanical energy is transformed into magnetic free energy and thermal energy above the photosphere. It seems increasingly clear that loops in the low corona are heated by small-scale, intermittent magnetic reconnection that is driven by the continual stressing of their magnetic footpoints. However, the extent to which this kind of impulsive energy addition influences the acceleration of the solar wind is not yet known.

Intertwined with the coronal heating problem is the heliophysical goal of being able to make accurate predictions of how both fast and slow solar wind streams are accelerated. Empirical correlation techniques have become more sophisticated and predictively powerful (e.g., Wang and Sheeley Jr, 1990, 2006; Arge and Pizzo, 2000; Leamon and McIntosh, 2007; Cohen et al., 2007; Vršnak et al., 2007) but they are limited because they do not identify or utilize the physical processes actually responsible for solar wind acceleration. There seem to be two broad classes of physics-based models that attempt to self-consistently answer the question: “How are fast and slow wind streams heated and accelerated?”

1. In wave/turbulence-driven (WTD) models, it is generally assumed that the convection-driven jostling of magnetic flux tubes in the photosphere drives wave-like fluctuations that propagate up into the extended corona. These waves (usually Alfvén waves) are often proposed to partially reflect back down toward the Sun, develop into strong MHD turbulence, and dissipate over a range of heights. These models also tend to explain the differences between fast and slow solar wind not by any major differences in the lower boundary conditions, but instead as an outcome of different rates of lateral flux-tube expansion over several solar radii as the wind accelerates (see, e.g., Hollweg, 1986; Wang and Sheeley Jr, 1991; Matthaeus et al., 1999; Cranmer, 2005; Suzuki, 2006; Suzuki and Inutsuka, 2006; Cranmer et al., 2007; Verdini and Velli, 2007; Verdini et al., 2009).

2. In reconnection/loop-opening (RLO) models, the flux tubes feeding the solar wind are assumed to be influenced by impulsive bursts of mass, momentum, and energy addition in the lower atmosphere. This energy is usually assumed to come from magnetic reconnection between closed, loop-like magnetic flux systems (that are in the process of emerging, fragmenting, and being otherwise jostled by convection) and the open flux tubes that connect
to the solar wind. These models tend to explain the differences between fast and slow solar wind as a result of qualitatively different rates of flux emergence, reconnection, and coronal heating at the basal footpoints of different regions on the Sun (see, e.g., Axford and McKenzie, 1992, 1997; Fisk et al., 1999; Ryutova et al., 2001; Markovskii and Hollweg, 2002, 2004; Fisk, 2003; Schwadron and McComas, 2003; Woo et al., 2004; Fisk and Zurbuchen, 2006).

It is notable that both the WTD and RLO models have recently passed some basic “tests” of comparison with observations. Both kinds of model have been shown to be able to produce fast ($v > 600 \text{ km/s}$), low-density wind from coronal holes and slow ($v < 400 \text{ km/s}$), high-density wind from streamers rooted in quiet regions. Both kinds of model also seem able to reproduce the observed in situ trends of how frozen-in charge states and the FIP effect vary between fast and slow wind streams.

The fact that both sets of ideas described above seem to mutually succeed at explaining the fast/slow solar wind could imply that a combination of both ideas would work best. However, it may also imply that the existing models do not yet contain the full range of physical processes – and that once these are included, one or the other may perform noticeably better than the other. It also may imply that the comparisons with observations have not yet been comprehensive enough to allow the true differences between the WTD and RLO ideas to be revealed.

Several recent observations have pointed to the importance of understanding the relationships and distinctions between the WTD and RLO models. The impulsive polar jets discussed in Section 4 may be evidence that that magnetic reconnection drives some fraction of the fast solar wind (see also Fisk, 2005; Moreno-Insertis et al., 2008; Pariat et al., 2009). Also, direct observations of Alfvén waves above the solar limb indicate the highly intermittent nature of how kinetic energy is distributed in spicules, loops, and the open-field corona (De Pontieu et al., 2007; Tomczyk et al., 2007; Tomczyk and McIntosh, 2009). Spectroscopic observations of blueshifts in the chromospheric network have long been interpreted as the launching points of solar wind streams, but it remains unclear how nanoflare-like events or loop-openings contribute to the interpretation of these diagnostics (He et al., 2007; Aschwanden et al., 2007; McIntosh et al., 2007). Even out in the in situ solar wind – far above the roiling “furnace” of flux emergence at the Sun – there remains evidence for ongoing reconnection (Gosling et al., 2005; Gosling and Szabo, 2008). There is also evidence that the dominant range of turbulence timescales measured in interplanetary space (i.e., tens of minutes to hours) is related to the timescale of flux cancellation in the low corona (Hollweg, 1990, 2006).

Determining whether the WTD or RLO paradigm – or some combination of the two – is the dominant cause of global solar wind variability is a key prerequisite to building physically realistic predictive models of the heliosphere. Many of the widely-applied global modeling codes (e.g., Riley et al., 2001; Roussev et al., 2003; Tóth et al., 2005; Usmanov and Goldstein, 2006; Feng et al., 2007) continue to utilize relatively simple empirical prescriptions for coronal heating in the energy conservation equation. Improving the identification and characterization of the key physical processes will provide a clear pathway for inserting more physically realistic coronal heating “modules” into three-dimensional MHD codes.

5.2 The Wave/Turbulence-Driven (WTD) solar wind idea

There has been substantial work over the past few decades devoted to exploring the idea that the plasma heating and wind acceleration along open flux tubes may be explained as a result of wave damping and turbulent cascade. No matter the relative importance of reconnections and loop-openings in the low corona, we do know that waves and turbulent motions are present everywhere from the photosphere to the heliosphere, and it is important to determine how they affect the mean state of the plasma. A review of the observational evidence for waves and turbulence in the solar wind is beyond the scope of this paper, but several recent reviews of the remote-sensing and
Coronal Holes

in situ data include Tu and Marsch (1995), Mullan and Yakovlev (1995), Goldstein et al. (1997), Roberts (2000), Bastian (2001), and Cranmer (2002a, 2004a, 2007). Although this subsection mainly describes recent work by the author, these results would not have been possible without earlier work on wave/turbulent heating by, e.g., Coleman (1968), Hollweg (1986), Hollweg and Johnson (1988), Isenberg (1990), Li et al. (1999), Matthaeus et al. (1999), Dmitruk et al. (2001, 2002), and many others.

Cranmer et al. (2007) described a set of models in which the time-steady plasma properties along a one-dimensional magnetic flux tube are determined. These model flux tubes are rooted in the solar photosphere and are extended into interplanetary space. The numerical code developed in that work, called ZEPHYR, solves the one-fluid equations of mass, momentum, and energy conservation simultaneously with transport equations for Alfvénic and acoustic wave energy. ZEPHYR is the first code capable of producing self-consistent solutions for the photosphere, chromosphere, corona, and solar wind that combine: (1) shock heating driven by an empirically guided acoustic wave spectrum, (2) extended heating from Alfvén waves that have been partially reflected, then damped by anisotropic turbulent cascade, and (3) wind acceleration from gradients of gas pressure, acoustic wave pressure, and Alfvén wave pressure.

The only input “free parameters” to ZEPHYR are the photospheric lower boundary conditions for the waves and the radial dependence of the background magnetic field along the flux tube. The majority of heating in these models comes from the turbulent dissipation of partially reflected Alfvén waves (see also Matthaeus et al., 1999; Dmitruk et al., 2002; Verdini and Velli, 2007; Chandran et al., 2009a). Photospheric measurements of the horizontal motions of strong-field intergranular flux concentrations (i.e., G-band bright points) were used to constrain the Alfvén wave power spectrum at the lower boundary. This empirically determined power spectrum is dominated by wave periods of order 5–10 minutes. It is important to note, however, that radio and in situ measurements find that most of the fluctuation power in the solar wind is at lower frequencies (i.e., periods of hours). We still do not yet know (1) if the shape of the power spectrum evolves significantly between the lower solar atmosphere and interplanetary space, or (2) if some low-frequency power is missed by the existing measurements of G-band bright point motions. In any case, as seen below, the resulting wave reflection and turbulent dissipation that comes from just the 5–10 minute periods appear to be sufficient to explain the observed levels of coronal heating and solar wind acceleration.

Non-WKB wave transport equations were solved to determine the degree of linear reflection at heights above the photospheric base (see Cranmer and van Ballegooijen, 2005). The resulting values of the Elsasser amplitudes $Z_{\pm}$, which denote the energy contained in upward ($Z_+$) and downward ($Z_-$) propagating waves, were then used to constrain the energy flux in the cascade. Cranmer et al. (2007) used a phenomenological form for the damping rate that has evolved from studies of Reduced MHD and comparisons with numerical simulations. The resulting heating rate (in units of erg s$^{-1}$ cm$^{-3}$) is given by

$$Q = \rho \left( \frac{1}{1 + [t_{\text{eddy}}/t_{\text{ref}}]^n} \right) \frac{Z_+^2 Z_+ + Z_-^2 Z_-}{4L_{\perp}}$$

(1)

where $\rho$ is the mass density and $L_{\perp}$ is an effective perpendicular correlation length of the turbulence (see, e.g., Hossain et al., 1995; Zhou and Matthaeus, 1990; Breech et al., 2008; Podesta and Bhatnagar, 2009; Beresnyak and Lazarian, 2009). Cranmer et al. (2007) used a standard assumption that $L_{\perp}$ scales with the cross-sectional width of the flux tube (Hollweg, 1986). The term in parentheses above is an efficiency factor that accounts for situations in which the cascade does not have time to develop before the waves or the wind carry away the energy (Dmitruk and Matthaeus, 2003). In open field regions, the cascade is “quenched” when the nonlinear eddy time scale $t_{\text{eddy}}$ becomes much longer than the macroscopic wave reflection time scale $t_{\text{ref}}$. In closed field regions, the correction factor may behave in an opposite sense as it does for open field regions.
Figure 9: Summary of Cranmer et al. (2007) models: (a) The adopted solar-minimum field geometry of Banaszkiewicz et al. (1998), with radii of wave-modified critical points marked by symbols. (b) Latitudinal dependence of wind speed at ∼ 2 AU for models with $n = 1$ (multi-color curve) and $n = 2$ (brown curve), compared with data from the first Ulysses polar pass in 1994–1995 (black curve; Goldstein et al., 1996). (c) $T(r)$ for polar coronal hole (red solid curve), streamer edge (blue dashed curve), and strong-field active region (black dotted curve) models.

(see, e.g., Gómez et al., 2000; Rappazzo et al., 2008). In most of the solar wind models, though, Cranmer et al. (2007) used $n = 1$ in Equation (1) based on analytic and numerical results (Dobrowolny et al., 1980; Oughton et al., 2006), but they also tried $n = 2$ to explore a stronger form of the quenching effect.

Figure 9 summarizes the results of varying the magnetic field properties while keeping the lower boundary conditions fixed. For a single choice for the photospheric wave properties, the models produced a realistic range of slow and fast solar wind conditions. A two-dimensional model of coronal holes and streamers at solar minimum reproduces the latitudinal bifurcation of slow and fast streams seen by Ulysses. An active-region-like enhancement of the magnetic field strength in the low corona generates a high mass flux and a slow wind speed, in agreement with observations of open field lines connected with active regions (see also Wang et al., 2009). As predicted by earlier studies, a larger coronal “expansion factor” naturally gives rise to a slower and denser wind, higher temperature at the coronal base, and lower-amplitude Alfvén waves at 1 AU.

In these models, the radial gradient of the Alfvén speed affects where the waves are reflected and damped, and thus whether energy is deposited below or above the Parker (1958a) critical point. Early studies of solar wind energetics (e.g., Leer and Holzer, 1980; Pneuman, 1980; Leer et al., 1982) showed that if there is substantial heating below the critical point, its primary impact is to “puff up” the hydrostatic scale height, drawing more particles into the accelerating wind and thus producing a slower and more massive wind. If most of the heating occurs at or above the critical point, the subsonic atmosphere is relatively unaffected, and the local increase in energy flux has nowhere else to go but into the kinetic energy of the wind (leading to a faster and less...
dense outflow). The ZEPHYR results shown in Figure 9 display this kind of dichotomy because the superradial expansion creates a much higher critical point over the equatorial regions than over the poles. Additional studies of how and where the mass flux and wind speed are determined include Withbroe (1988), Hansteen and Leer (1995), Hansteen et al. (1997), Janse et al. (2007), and Wang et al. (2009).

Perhaps more surprisingly, varying the coronal expansion factor in the models shown in Figure 9 also produces correlative trends that are in good agreement with in situ measurements of commonly measured ion charge state ratios (e.g., O\textsuperscript{7+}/O\textsuperscript{6+}) and FIP-sensitive abundance ratios (e.g., Fe/O). Cranmer et al. (2007) showed that the slowest solar wind streams – associated with active-region fields at the base – can produce a factor of \(\sim 30\) larger frozen-in ionization-state ratio of O\textsuperscript{7+}/O\textsuperscript{6+} than high-speed streams from polar coronal holes, despite the fact that the temperature at 1 AU is lower in slow streams than in fast streams. Furthermore, when elemental fractionation is modeled using a theory based on preferential wave-pressure acceleration (Laming, 2004, 2009; Bryans et al., 2009), the slow wind streams exhibit a substantial relative buildup of elements with low FIP with respect to the high-speed streams. Although the WTD models utilize identical photospheric lower boundary conditions for all of the flux tubes, the self-consistent solutions for the upper chromosphere, transition region, and low corona are qualitatively different. Feedback from larger heights (i.e., from variations in the flux tube expansion rate and the resulting heating rate) extends downward to create these differences.

Another empirical “marker” of heliospheric stream structure is the proton specific entropy, or entropy per proton, which is often approximated as being proportional to \(\ln\left(T_p/n_p^{-1}\right)\), where \(\gamma \approx 1.5\) is an empirical adiabatic index for solar wind protons (e.g., Burlaga et al., 1990; Pagel et al., 2004). When measured in regions of the (non-CME) heliosphere where corotating interaction regions have not yet formed shocks, this quantity is seen to clearly distinguish slow wind streams from fast wind streams. Figure 10 shows how the specific entropy is positively correlated with wind speed, both in measurements made by the Solar Wind Electron Proton Alpha Monitor (SWEPAM) instrument on ACE (McComas et al., 1998) and in the Cranmer et al. (2007) ZEPHYR models discussed above. Each model data point was computed independently of the others. The models had identical lower boundary conditions at the photosphere, and they differed from one another only by having a different radial dependence of the magnetic field. Because entropy should be conserved in the absence of significant small-scale dissipation, the quantity that is measured at 1 AU may be a long-distance proxy for the near-Sun locations of strong coronal heating. In other words, the comparison of measured and modeled solar wind entropy variations may be a key way to discriminate between competing explanations of solar wind acceleration.

Although Equation (1) describes the plasma heating rate in terms of the local properties of MHD turbulence, it is also possible to see that this expression gives a heating rate proportional to the mean magnetic flux density at the coronal base. As illustrated above in Figure 4, the mean field strength in the low corona is determined by both the photospheric field strength in the intergranular bright points and the total number of bright points that eventually merge their fields together in the low corona. The field strength at this merging height can thus be estimated as \(B \approx f_\perp B_\perp\), where \(B_\perp \approx 1500\) G is the (nearly universal) photospheric bright-point field strength and \(f_\perp\) is the area filling factor of bright points in the photosphere. The latter quantity appears to vary by more than an order of magnitude in different regions on the Sun, from about 0.002 (at low latitudes at solar minimum) to \(\sim 0.1\) (in active regions). If the regions below the merging height can be treated using approximations from “thin flux tube theory” (e.g., Spruit, 1981; Cranmer and van Ballegooijen, 2005), then it is possible to express each term in Equation (1) as a function of \(f_\perp\) and the photospheric properties. For example, \(B \propto \rho^{1/2}\) applies to thin flux tubes in pressure equilibrium, and thus \(\rho\) at the merging height can be estimated as \(f_\perp^2 \rho_\perp\) (where \(\rho_\perp\) is the photospheric density). For Alfvén waves at low heights, \(Z_\perp \propto \rho^{-1/4}\), and so \(Z_\perp\) at the merging height scales like \(Z_\perp/f_\perp^{1/2}\). Also, we assumed that \(L_\perp \propto B^{-1/2}\). If the quenching factor in
Figure 10: Solar wind specific entropy plotted as a function of solar wind speed, computed for both the ZEPHYR models at 1 AU (black symbols, curve) and from ACE/SWEPAM data (blue points).

parentheses in Equation (1) is neglected, then

\[ Q \approx \rho Z^3 L_\perp \approx \left( \frac{\rho_* Z_*^3}{L_*} \right) \frac{f_*^{3/2} f_*^{-3/2}}{f_*^{-1/2}} \approx Q_* f_* . \]  

(2)

Equivalently, Equation (2) implies that \( Q/Q_* \approx B/B_* \), and thus that the heating in the low corona scales directly with the mean magnetic field strength there. In a more highly structured field, the latter is equivalent to the magnetic “flux density” averaged over a given region. Observational evidence for such a linear scaling has been found for both a variety of solar regions and other stars as well (see, e.g., Pevtsov et al., 2003; Schwadron et al., 2006; Suzuki, 2006; Kojima et al., 2007; Pinto et al., 2009).

5.3 The Reconnection/Loop-Opening (RLO) solar wind idea

It is clear from observations of the Sun’s highly dynamical “magnetic carpet” (Schrijver et al., 1997; Title and Schrijver, 1998; Hagea et al., 1999) that much of coronal heating is driven by the continuous interplay between the emergence, separation, merging, and cancellation of small-scale magnetic elements. Reconnection seems to be the most likely channel for the injected magnetic energy to be converted to heat (e.g., Priest and Forbes, 2000). Only a small fraction of the photospheric magnetic flux is in the form of open flux tubes connected to the heliosphere (Close et al., 2003). Thus, the idea has arisen that the dominant source of energy for open flux tubes is a series of stochastic reconnection events between the open and closed fields (e.g., Fisk et al., 1999; Ryutova et al., 2001; Fisk, 2003; Schwadron and McComas, 2003; Feldman et al., 2005; Schwadron et al., 2006; Schwadron and McComas, 2008; Fisk and Zhao, 2009).

The natural appeal of the RLO idea is evident from the fact that open flux tubes are always rooted in the vicinity of closed loops (e.g., Dowdy et al., 1986) and that all layers of the solar atmosphere seem to be in continual motion with a wide range of timescales. In fact, observed correlations between the lengths of closed loops in various regions, the electron temperature in
the low corona, and the wind speed at 1 AU (Feldman et al., 1999; Gloeckler et al., 2003) are highly suggestive of a net transfer of Poynting flux from the loops to the open-field regions that may be key to understanding the macroscopic structure of the solar wind. The proposed RLO reconnection events may also be useful in generating energetic particles and cross-field diffusive transport throughout the heliosphere (e.g., Fisk and Schwadron, 2001).

Testing the RLO idea using theoretical models seems to be more difficult than testing the WTD idea because of the complex multi-scale nature of magnetic reconnection. It can be argued that one needs to create fully three-dimensional models of the coronal magnetic field (arising from multiple magnetic elements on the surface) to truly assess the full range of closed/open flux interactions. The idea of modeling the coronal field via a collection of discrete magnetic sources (referred to in various contexts as “magneto-chemistry,” “tectonics,” or “magnetic charge topology”) has been used extensively to study the evolution of the closed-field corona (e.g., Longcope, 1996; Schrijver et al., 1997; Longcope and Kankelborg, 1999; Sturrock et al., 1999; Priest et al., 2002; Beveridge et al., 2003; Barnes et al., 2005; Parnell, 2007; Ng and Bhattacharjee, 2008), but applications to open fields and the solar wind have been rarer (see, however, Fisk, 2005; Tu et al., 2005).

In order to develop the RLO paradigm to the point where it can be tested more quantitatively, several key questions remain to be answered. For example, how much magnetic flux actually opens up in the magnetic carpet? Also, what is the time and space distribution of reconnection-driven energy addition into the (transiently) open flux tubes? Lastly, how is the reconnection energy distributed into various forms (e.g., bulk kinetic energy in “jets,” thermal energy, waves, turbulence, and energetic particles) that each affect the accelerating solar wind in different ways? Combinations of simulations, analytic scaling relations, and observations are needed to make further progress.

5.4 Kinetic microphysics

The theoretical models discussed in the previous two subsections mainly involved a “one-fluid” or MHD approach to the coronal heating and solar wind acceleration. However, at large heights in coronal holes, the collisionless divergence of plasma parameters for protons, electrons, and heavy ions allows the multi-fluid kinetic processes to be distinguished in a more definitive way. The UVCS measurements of strong O$^{+5}$ preferential heating, preferential acceleration, and temperature anisotropy have spurred a great deal of theoretical work in this direction (see reviews by Hollweg and Isenberg, 2002; Cranmer, 2002a; Marsch, 2005, 2006; Kohl et al., 2006). Specifically, the observed ordering of $T_i > T_p > T_e$ and the existence of anisotropies of the form $T_{\perp} > T_{\parallel}$ in coronal holes led to a resurgence of interest in models of ion cyclotron resonance.

The ion cyclotron heating mechanism is a classical resonance between left-hand polarized Alfvén waves and the Larmor gyrations of positive ions around the background magnetic field. If the wave frequency and the natural ion gyrofrequency are equal, then in the rest frame of the ion the oscillating electric and magnetic fields of the wave are no longer felt by the ion to be oscillating. The ion in such a frame senses a constant DC electric field, and it can secularly gain or lose energy depending on the relative phase between the ion’s velocity vector and the electric field direction. In a wave field with random phases, an ion will undergo a random walk in energy. Thus, on average the ions can be considered to “diffuse” into faster (i.e., wider) Larmor orbits with larger perpendicular energy (see Rowlands et al., 1966; Galinsky and Shevchenko, 2000; Isenberg, 2001; Cranmer, 2001; Isenberg and Vasquez, 2009).

In the actual solar corona, however, it is not likely that the situation is as straightforward as summarized above. Instead of a population of pre-existing, linear cyclotron waves that are dissipated, there may be a rich variety of nonlinear plasma mechanisms at play. The observed ion heating is likely to be just the final stage of a multi-step process of energy conversion between waves, turbulent motions, reconnection structures, and various kinds of distortions in the particle velocity distributions. Table 1 surveys the field of suggested possibilities, and the remainder of
Table 1: Tabular outline of suggested physical processes for preferentially heating and accelerating minor ions in coronal holes.

<table>
<thead>
<tr>
<th>Reconnection events in the low corona generate:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ion cyclotron resonant Alfvén waves .................</td>
<td>Tu and Marsch (1997); Cranmer (2000, 2001)</td>
</tr>
<tr>
<td>Fast collisionless shocks ................................</td>
<td>Lee and Wu (2000)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MHD turbulence in the extended corona generates:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ion cyclotron waves (“parallel cascade?”), with:</td>
<td></td>
</tr>
<tr>
<td>&gt;&gt; Alfven and fast-mode nonlinear coupling ............</td>
<td>Chandran (2005)</td>
</tr>
<tr>
<td>&gt;&gt; Three-wave (ion-sound/parametric) coupling ...........</td>
<td>Yoon and Fang (2008)</td>
</tr>
<tr>
<td>&gt;&gt; Fermi-like diffusion between inward/outward wave resonances ....</td>
<td>Isenberg (2001)</td>
</tr>
<tr>
<td>Kinetic Alfvén waves (“perpendicular cascade”), with:</td>
<td></td>
</tr>
<tr>
<td>&gt;&gt; Shear instabilities → ion cyclotron waves ............</td>
<td>Markovskii et al. (2006)</td>
</tr>
<tr>
<td>&gt;&gt; Debye-scale electron holes ........</td>
<td>Mattheeus et al. (2003); Cranmer and van Ballegooijen (2003)</td>
</tr>
<tr>
<td>Oblique MHD waves (high $k_\parallel$, high $k_\perp$) ..</td>
<td>Li and Habbal (2001)</td>
</tr>
<tr>
<td>Current sheets → coherent Fermi acceleration .............</td>
<td>Dmitruk et al. (2004)</td>
</tr>
<tr>
<td>Transverse density gradients → drift currents .............</td>
<td>Markovskii (2001); Zhang (2003)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low-frequency Alfvén waves in the corona directly undergo:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization drift → lower-hybrid waves ..........</td>
<td>Singh and Khazanov (2004); Khazanov and Singh (2007)</td>
</tr>
<tr>
<td>Nonresonant stochastic heating ......................</td>
<td>Lu and Li (2007); Wu and Yoon (2007)</td>
</tr>
<tr>
<td>Stochastic heating at fractional cyc. resonance ......</td>
<td>Chen et al. (2001); Guo et al. (2008)</td>
</tr>
</tbody>
</table>

| Heavy ion velocity filtration ..................... | Pierrard and Lamy (2003); Pierrard et al. (2004) |

this section discusses these ideas in more detail.

One potential obstacle to the idea of ion cyclotron heating is that the required gyroresonant wave frequencies in the corona are of order $10^2$ to $10^4$ Hz, whereas the dominant frequencies of Alfvén waves believed to be emitted by the Sun are thought to be much lower (i.e., less than 0.01 Hz, corresponding to periods of minutes to hours). Axford and McKenzie (1992) suggested that the right kinds of high-frequency waves may be generated in small-scale reconnection events in the chaotic “furnace” of the supergranular network. These waves could propagate upwards in height – and downwards in magnetic field strength – until they reached a location where they became cyclotron resonant with the local ions, and thus would damp rapidly to provide the ion heating (see also Schwartz et al., 1981; Tu and Marsch, 1997).

The above scenario of “basal generation” of ion cyclotron waves has been called into question for several reasons. Cranmer (2000, 2001) argued that the passive sweeping of a pre-existing fluctuation spectrum would involve ions with low gyrofrequencies (i.e., small ratios of charge to mass; $q_i/m_i \approx 0.1$ to 0.2 in proton units) encountering waves of a given frequency at lower heights than the ions that have been observed to exhibit preferential heating, like O$^{+5}$ ($q_i/m_i = 0.31$) and Mg$^{+9}$ ($q_i/m_i = 0.37$). Thus, the resonances of many minor ion species may be strong enough to damp out a base-generated spectrum of waves before they can become resonant with the observed species. Furthermore, Hollweg (2000) found that a base-generated spectrum of ion cyclotron waves would exhibit a very different appearance in interplanetary radio scintillations than the observed radio data. There remains some uncertainty about these criticisms of a basal spectrum of ion cyclotron waves, and definitive conclusions cannot yet be made (see discussions in Tu and Marsch, 2001; Hollweg and Isenberg, 2002).

There are several other interesting consequences of the Axford and McKenzie (1992) idea of rapid reconnection events at the coronal base. It is possible, for example, that such microflaring...
activity would give rise to intermittent bursts of parallel electron beams that propagate up into the extended corona. Sufficiently strong beams may be unstable to the growth of wave power at the ion cyclotron frequencies (Markovskii and Hollweg, 2002, 2004; Voitenko and Goossens, 2002), and these waves would then go on to heat the ions. Also, Lee and Wu (2000) suggested that small-scale reconnection events could produce fast collisionless shocks in the extended corona. For shocks sufficiently thin and strong (i.e., with a bulk velocity jump of at least $\sim 0.3$ times the Alfvén speed), ions that cross from one side of the shock to the other remain “nondeflected” by the rapid change in direction of the magnetic field. Thus, they can convert some of their parallel motion into perpendicular gyration. Mancuso et al. (2002) suggested this mechanism may be applied to understanding UVCS measurements of ion heating in shocks associated with CMEs. However, it is unclear to what extent the open magnetic regions in coronal holes are filled with sufficiently strong shocks to enable this process to occur (see also Hollweg and Isenberg, 2002).

In contrast to the ideas of base-generation of ion cyclotron waves, there have been several proposed mechanisms for “gradual generation” of these waves over a range of distances in the corona and solar wind. A natural way to produce such an extended source of fluctuations is MHD turbulent cascade, which continually transports power at large scales to small scales via the stochastic shedding of transient eddies. A strong turbulent cascade is certainly present in interplanetary space (see reviews by Tu and Marsch, 1995; Goldstein et al., 1997). It is well known, though, that in both numerical simulations and analytic descriptions of Alfvén-wave turbulence (with a strong background “guide field” like in the corona) the cascade from large to small length scales (i.e., from small to large wavenumbers) occurs most efficiently for modes that do not increase in frequency. In other words, the cascade acts most rapidly to increase the perpendicular wavenumber $k_\perp$ while leaving the parallel wavenumber $k_\parallel$ largely unchanged (e.g., Strauss, 1976; Shebalin et al., 1983; Goldreich and Sridhar, 1995; Cho et al., 2002; Oughton et al., 2004). This type of cascade is expected to generate so-called kinetic Alfvén waves (KAWs) with $k_\perp \gg k_\parallel$, but not ion cyclotron waves.

Under typical “low plasma beta” conditions in the corona and fast solar wind, the linear dissipation of KAWs would lead to the preferential parallel heating of electrons (Leamon et al., 1999; Cranmer and van Ballegooijen, 2003; Gary and Borovsky, 2008). This is essentially the opposite of what has been observed with UVCS. However, there have been several suggestions for more complex (nonlinear or multi-step) processes that may be responsible for ions to receive perpendicular heating from KAW-type fluctuations.

1. Markovskii et al. (2006) discussed how a perpendicular turbulent cascade produces increasingly strong shear motions transverse to the magnetic field, and that this shear may eventually be unstable to the generation of cyclotron resonant waves that can in turn heat protons and ions (see also Mikhailenko et al., 2008). This effect may also produce a steepening in the power spectrum of the magnetic field fluctuations that agrees with the observed “dissipation range” (Smith et al., 2006).

2. Voitenko and Goossens (2003, 2004) suggested that high-$k_\perp$ KAWs with sufficiently large amplitudes could begin to exhibit nonlinear resonance effects (“demagnetized wave phases”) leading to rapid ion perpendicular heating. There are definite thresholds in the KAW amplitude that must be exceeded for these mechanisms to be initiated, and it is unclear whether the actual coronal turbulent spectrum has enough power in the relevant regions of wavenumber space (see also Wu and Yang, 2006, 2007).

3. The damping of low-frequency KAWs may give rise to substantial parallel electron acceleration. If the resulting electron velocity distributions were sufficiently beamed, they could become unstable to the generation of parallel Langmuir waves. In turn, the evolved Langmuir wave trains may exhibit a periodic electric potential-well structure in which some of
the beam electrons can become trapped. Adjacent potential wells can then merge with one another to form isolated “electron phase space holes” of saturated potential. Ergun et al. (1999), Matthaeus et al. (2003), and Cranmer and van Ballegooijen (2003) described how these tiny (Debye-scale) electrostatic structures can heat ions perpendicularly via Coulomb-like quasi-collisions.

4. Obliquely propagating MHD waves with large perpendicular \textit{and} parallel wavenumbers – including KAWs and fast-mode waves – can interact resonantly with positive ions via channels that are not available when either $k_\parallel$ or $k_\perp$ are small. Li and Habbal (2001) found that oblique fast-mode waves with large wavenumbers may be even more efficient than Alfvén waves at heating ions under coronal conditions. Hollweg and Markovskii (2002) discussed how the higher-order cyclotron resonances become available to obliquely propagating waves with large wavenumbers, and how these can lead to stochastic velocity-space diffusion for ions.

5. On the smallest spatial scales, the plasma in numerical simulations of MHD turbulence is seen to develop into a collection of narrow current sheets undergoing oblique magnetic reconnection (i.e., with the strong “guide field” remaining relatively unchanged). Dmitruk et al. (2004) performed test-particle simulations in a turbulent plasma and found that protons can become perpendicularly accelerated around the guide field because of coherent forcing from the perturbed fields associated with the current sheets (see also Parashar et al., 2009). It remains to be seen whether this process could lead to more than mass-proportional energization for minor ions.

6. If the plasma contains sufficiently small-scale density gradients transverse to the magnetic field ($\nabla_\perp \rho$), then drift currents can be excited that are unstable to the generation of high-frequency waves (Markovskii, 2001; Zhang, 2003; Vranjes and Poedts, 2008; Mecheri and Marsch, 2008). These instabilities depend on both the amplitudes and scale lengths of $\nabla_\perp \rho$. To measure the latter, it is important to take into account both remote-sensing measurements of coronal density inhomogeneities (e.g., Woo, 2006; Pasachoff et al., 2007) and constraints from radio scintillation power spectra at larger distances (Bastian, 2001; Spangler, 2002; Harmon and Coles, 2005).

Despite the fact that theory predicts a predominantly perpendicular cascade, there is some evidence that the turbulent fluctuations in the solar wind have some energy that extends up to large $k_\parallel$ values in a power-law tail (see, e.g., Bieber et al., 1996; Dasso et al., 2005; MacBride et al., 2008). Whether or not this means that true “parallel cascade” occurs in the corona and solar wind is still not known. However, some progress has been made using a phenomenological approach to modeling the cascade as a combination of advection and diffusion in wavenumber space. In the model of Cranmer and van Ballegooijen (2003), the relative strengths of perpendicular advection and diffusion determine the slope of the power-law spectrum in $k_\parallel$, and thus they specify the amount of wave energy that is available at the ion cyclotron frequencies (see also Cranmer et al., 1999a; Landi and Cranmer, 2009; Jiang et al., 2009).

There have also been proposals for additional mechanisms that could allow a parallel cascade to occur in the corona and solar wind. Nonlinear couplings between the dominant Alfvén waves and other modes such as fast magnetosonic waves (Chandran, 2005; Luo and Melrose, 2006) and ion-acoustic waves (Yoon and Fang, 2008) have the potential to enhance the wave power at high frequencies. It has also been known for some time that nonlinear coupling between Alfvén and fast-mode waves may help explain why the measured \textit{in situ} magnetic field magnitude $|\mathbf{B}|$ remains roughly constant while its direction varies strongly (Barnes and Hollweg, 1974; Vasquez and Hollweg, 1996, 1998).
If MHD waves have sufficiently large amplitudes, they may undergo nonlinear wave steepening, which leads to density variations as well as oscillations in the parallel components of the velocity and magnetic field. These may generate progressively smaller scales along the magnetic field (e.g., Medvedev, 2000; Suzuki et al., 2007). There is also a “bootstrap” kind of effect for ion cyclotron wave generation that was discussed by Isenberg et al. (2001). If some outward-propagating cyclotron waves exist, the resonant diffusion may act to produce proton velocity distributions that are unstable to the generation of inward-propagating cyclotron waves. In response, the proton distributions would become further deformed and thus could become unstable to the growth of both inward and outward waves. It is not yet known if this process could reach the point of becoming self-sustaining, but if so, it may also serve as an extended generation mechanism for high-$k_{\parallel}$ waves.

In addition to the above ideas that involve large wavenumbers and kinetic effects, there have been other suggested physical processes that do not require high-$k_{\parallel}$ resonances to be initially present in order to heat the ions.

1. Particles in large-amplitude Alfvén waves exhibit both $E \times B$ drift motions (i.e., their standard velocity amplitude) and a polarization drift velocity $V_{\text{pol}}$ that is smaller than the former by the ratio $\omega/\Omega_i$, where $\omega$ is the wave frequency and $\Omega_i$ is the ion cyclotron frequency. A sufficiently large $V_{\text{pol}}$ can lead to cross-field currents unstable to the generation of high-frequency waves, and to eventual equipartition between $V_{\text{pol}}$ and the ion thermal speed (Singh and Khazanov, 2004; Singh et al., 2007; Khazanov and Singh, 2007). It is not yet known whether the effective $V_{\text{pol}}$ for the coronal fluctuation spectrum is large enough to provide a significant fraction of the ion thermal speeds.

2. Recently there have been suggested some completely nonresonant mechanisms that depend on the stochasticity of MHD turbulence to produce an effective increase in random ion motions (Lu and Li, 2007; Wu and Yoon, 2007; Bourouaine et al., 2008). Questions still remain, though, concerning the spatial scales over which one should refer to particle motions as “heating” versus “wave sloshing.” This energization mechanism may be just a more chaotic form of the standard velocity amplitude that an ion feels when in the presence of a spectrum of Alfvén waves (see, e.g., Wang and Wu, 2009). In this case, the maximum amount of heating from this process would provide mass-proportional heating for minor ions and protons (i.e., $T_i/T_p = m_i/m_p$), and it is clear that the UVCS measurements for O$^{+\delta}$ show heating in excess of this amount (see Section 4.3).

3. Both numerical and analytic studies of Alfvén waves show that, at sufficiently large amplitudes, there can be gyroresonance-like ion energization for sets of frequencies at specific fractions of the local ion cyclotron resonance frequency (e.g., Chen et al., 2001; Guo et al., 2008). Like several other processes listed above, this effect becomes active only above certain thresholds of wave amplitude. Also, Markovskii et al. (2009) showed that mildly nonlinear Alfvén waves – with frequencies slightly below the local proton gyrofrequency and power in both the upward and downward directions along the field – can also undergo additional modes of dissipation and proton heating that are not anticipated in linear ion cyclotron resonance theories.

Finally, there has been some development of the so-called velocity filtration theory, which requires neither direct heating nor wave damping in order to energize coronal ions. Spacecraft measurements of plasma velocity distributions, both in the solar wind and in planetary magnetospheres and magnetosheaths, have revealed that “suprathermal” power-law tails are quite common. These observations led to the suggestion by Scudder (1992a, 1994) of an an alternative to theories that demand explicit energy deposition in the low corona (see also Parker, 1958b; Levine, 1974). A velocity distribution having a suprathermal tail will become increasingly dominated by its high-energy particles at larger distances from the solar gravity well. Thus an effective “heating” occurs
as a result of particle-by-particle conservation of energy. The major unresolved issue is whether suprathermal tails of the required strength can be produced and maintained in the upper chromosphere and transition region – where Coulomb collisions are traditionally believed to be strong enough to rapidly drive velocity distributions toward Maxwellians. Whether the solar atmosphere actually plays host to strong nonthermal tails is still under debate, with some evidence existing in favor of their presence (e.g., Esser and Edgar, 2000; Ralchenko et al., 2007) and other evidence against them (Anderson et al., 1996; Ko et al., 1996; Feldman et al., 2007).

The original (Scudder, 1992a,b) ideas about suprathermal velocity filtration were applied only to the primary (proton and electron) coronal plasma. More recently, Pierrard and Lamy (2003) and Pierrard et al. (2004) have shown that this mechanism can produce extremely high temperatures for heavy ions in the corona – providing they had suprathermal tails in the chromosphere. The primary quantities presented in these papers, however, were integrated isotropic temperatures $T$. No information was given about the predicted sense of the temperature anisotropy for the minor ions. For a collisionless exospheric model, there is a suspicion that a combination of several effects (e.g., the initial velocity filtration and subsequent magnetic moment conservation) would result in velocity distributions with $T_\parallel \gg T_\perp$, which is not what is observed.

Since it is obvious that not all of the proposed mechanisms described above (and shown in Table 1) can be the dominant cause of the collisionless ion energization in coronal holes, there is a great need to “cut through the jungle” and assess the validity of each of these processes. For many of these suggested ideas, further theoretical development is required so that specific observational predictions can be made. However, there are also several types of measurement that have not been widely recognized or utilized as constraints on theoretical models. A prime example is the use of radio sounding (i.e., interplanetary scintillations and Faraday rotation) to measure the fine structure of the corona and solar wind in density, velocity, and magnetic field strength (see, however, Hollweg and Isenberg, 2002; Spangler, 2002; Harmon and Coles, 2005; Chandran et al., 2009b). Another example is the use of high-resolution UV spectral line profiles to probe departures from Maxwellian or bi-Maxwellian ion velocity distributions (e.g., Cranmer, 2001).
6 Summary and Conclusions

The last decade has seen significant progress toward identifying and characterizing the processes that produce coronal holes. As remote-sensing plasma measurements have become possible in the extended solar corona (i.e., the region of primary acceleration of the solar wind), the traditional gap between solar physics and in situ space physics has become narrower. However, there are still many unanswered questions: How and where in the solar atmosphere are the relevant waves and turbulent motions generated? Which kinds of fluctuation modes (i.e., linear or nonlinear; Alfvén, fast, or slow; high $k_\parallel$ or high $k_\perp$) are most important? What frequencies dominate the radially evolving power spectrum? What fraction of the interplanetary solar wind comes from filamentary structures such as polar plumes and polar jets? Are there relatively simple “scaling laws” that will allow us to use only the measured properties at the solar surface to predict the resulting amount of coronal heating and solar wind acceleration?

Answering the above questions involves moving forward in both the theoretical and observational directions. Section 5 described the large number of suggested physical processes for energizing the plasma in coronal holes. The validity of many of these processes still needs to be assessed, and their relative contributions to the heating and acceleration of the actual solar corona need to be quantified. If, at the end of this process, there are still a number of mutually inconsistent theories that are still viable, the only way forward is to determine what future measurements would best put the remaining models to the test. These activities are ongoing with the planning of inner heliospheric missions such as Solar Probe (McComas et al., 2007) and Solar Orbiter (Marsden and Fleck, 2007), as well as next-generation ultraviolet coronagraph spectroscopy missions that would follow up on the successes of UVCS/SOHO (see, e.g., Cranmer, 2002b; Gardner et al., 2003; Kohl et al., 2006).

The development of more physically sound models of the solar wind feeds back in many ways to a wider understanding of stellar outflows and star/planet evolution. Figure 11 shows some of the early stages of evolution for a representative solar-type star. At all ages, cool stars are inferred to exhibit some kind of wind or jet-like outflow (Lamers and Cassinelli, 1999; Wood, 2004; Güdel, 2007; Cranmer, 2008a). Young stars first become visible as dust-obscured cloud cores and protostars (e.g., Lada, 1985; Hartmann, 2000), and these objects are often associated with bipolar, collimated jets. These outflows indicate some kind of transfer of energy from the accretion disk’s orbital motion to torqued magnetic fields (rooted on the stellar surface) that relieve the buildup of angular momentum and eject plasma out the poles (e.g., Blandford and Payne, 1982; van Ballegooijen, 1994). As the accretion rates decrease over time, protostars become visible as classical T Tauri stars (CTTS), and there remains ample evidence for polar outflows in the form of both “disk winds” and true stellar winds (Hartigan et al., 1995; Ferreira et al., 2006; Cranmer, 2008b). The primordial accretion disk is dissipated gradually as the star enters the weak-lined T Tauri star (WTTS) phase, with a likely transition to a protoplanetary dust/debris disk. Strong stellar magnetic activity remains evident during these stages from, e.g., X-rays (Feigelson and Montmerle, 1999). Many “post T Tauri” stars, once they reach the zero-age main sequence (ZAMS), remain rapidly rotating, and for young ZAMS stars such as AB Dor there is evidence for a range of X-ray emitting plasma from dark polar spots (which probably do not correspond to open magnetic field regions like coronal holes) to huge “slingshot prominences” extending over several stellar radii (e.g., Güdel et al., 2001, 2003; Jardine and van Ballegooijen, 2005).

Learning about the fundamental physics responsible for solar coronal holes has relevance that reaches into other areas of study besides astrophysics, including plasma physics, space physics, and astronautical engineering. The practical benefits of improving long-term predictions for the conditions of the Earth’s local space environment are manifold (see, e.g., Feynman and Gabriel, 2000; Eastwood, 2008). In addition, parallel research into the expansion of the solar wind from the Earth’s ionosphere has led to an improved understanding of kinetic processes in plasmas on
the boundary between collisional and collisionless conditions (Lemaire and Pierrard, 2001; Barakat and Schunk, 2006). Finally, a growing realization that strict topical compartmentalization is often a hindrance to making progress has given rise to greater interest in interdisciplinary studies of “universal processes in heliophysics” (Crooker, 2004; Davila et al., 2009).
7 Acknowledgements

I would like to thank the editors of *Living Reviews in Solar Physics* for the invitation to prepare this review. I gratefully acknowledge John Kohl, Adriaan van Ballegooijen, Leonard Strachan, John Raymond, William Matthaeus, Philip Isenberg, Joseph Hollweg, S. Peter Gary, Eckart Marsch, Marco Velli, Mari Paz Miralles, Larry Gardner, and Enrico Landi for many valuable discussions and collaborations. I would especially like to thank the colleagues with whom I have engaged in more adversarial debates over the years, such as Ian Axford, Chmamyi Tu, Nathan Schwadron, and Nour-Eddine Raouafi. “As brothers fight ye!”

This work was supported by the National Aeronautics and Space Administration (NASA) under grants NNG04GE77G, NNX06AG95G, NNX07AL72G, and NNX09AB27G to the Smithsonian Astrophysical Observatory. *SOHO* is a project of international cooperation between ESA and NASA. This research made extensive use of NASA’s Astrophysics Data System (ADS) and the John G. Wolbach Library at the Harvard-Smithsonian Center for Astrophysics.
References


*Living Reviews in Solar Physics*

http://www.livingreviews.org/lrsp-2009-3
Coronal Holes


*Living Reviews in Solar Physics*

http://www.livingreviews.org/lrsp-2009-3
Coronal Holes


*Living Reviews in Solar Physics*

Coronal Holes


---

Living Reviews in Solar Physics

http://www.livingreviews.org/lrsp-2009-3


Living Reviews in Solar Physics
http://www.livingreviews.org/lrsp-2009-3


Coronal Holes


http://www.livingreviews.org/lrsp-2009-3


