STRUCTURE AND DYNAMICS OF THE SUN’S OPEN MAGNETIC FIELD

S. K. ANTIOCHOS, C. R. DeVORE, AND J. T. KARPEN

Naval Research Laboratory, Washington, DC 20375; antiochos@nrl.navy.mil

AND

Z. MIKić

Science Applications International Corporation, San Diego, CA 92121

Received 2007 May 25; accepted 2007 July 29

ABSTRACT

The solar magnetic field is the primary agent that drives solar activity and couples the Sun to the heliosphere. Although the details of this coupling depend on the quantitative properties of the field, many important aspects of the corona-solar wind connection can be understood by considering only the general topological properties of those regions on the Sun where the field extends from the photosphere out to interplanetary space, the so-called open field regions that are usually observed as coronal holes. From the simple assumptions that underlie the standard quasi-steady corona-wind theoretical models, and that are likely to hold for the Sun as well, we derive two conjectures as to the possible structure and dynamics of coronal holes: (1) coronal holes are unique in that every unipolar region on the photosphere can contain at most one coronal hole, and (2) coronal holes of nested polarity regions must themselves be nested. Magnetic reconnection plays the central role in enforcing these constraints on the field topology. From these conjectures we derive additional properties for the topology of open field regions, and propose several observational predictions for both the slowly varying and transient corona/solar wind.

Subject headings: Sun: corona — Sun: magnetic fields

1. INTRODUCTION

Decades of solar XUV/X-ray observations have shown that the Sun’s corona is divided into two distinct types of magnetic regions: those in which field lines are “closed,” i.e., rooted to the photosphere at two ends, and those in which field lines are “open,” rooted to the photosphere at only one end and extending out to the heliosphere. The fast solar wind is believed to emanate from these open field regions, which usually appear dark in X-rays, and therefore are referred to as coronal holes. Although a great deal of progress has been made on understanding the Sun’s open magnetic field since the seminal work of Parker (1958), many questions remain, especially concerning the dynamics of solar-heliospheric coupling. For example, we still do not understand the magnetic topology and evolution that give rise to the slow wind (e.g., Axford & McKenzie 1997; Zurbuchen 2007). At an even more basic level, the coronal and heliospheric observations themselves appear to be in conflict. Coronal images often show streamer evolution, and inflows/outflows that are interpreted as closed flux opening into the wind and open flux closing back down to the streamer (Hundhausen et al. 1984; Howard et al. 1985; Sheeley & Wang 2002). Furthermore, the long-standing observation that some coronal holes appear to rotate rigidly, impervious to differential rotation (Timothy et al. 1975; Zirker 1977), implies the continuous opening and closing of the coronal field (Wang et al. 1996; Lionello et al. 2006). However, the opening of closed flux implies the injection of field lines into the wind with both footpoints connected to the photosphere. Such field lines should exhibit a bidirectional heat flux (counterstreaming electrons) in heliospheric measurements, yet counterstreaming electrons are rarely seen outside of interplanetary coronal mass ejections (ICMEs; Gosling 1990). Conversely, the closing down of open flux would imply the formation of U-shaped field lines in the heliosphere with no connection to the Sun, resulting in a heat flux drop-out in the heliospheric measurements (McComas et al. 1989, 1991; Lin & Kahler 1992). Again, these are rarely seen outside of ICMEs (Pagel et al. 2005). Apparently, the heliosphere does not care what the corona is doing!

This apparent contradiction among the observations has led to conflicting approaches to modeling the topology and evolution of the open field. Motivated by their observations, some heliospheric researchers have proposed the interchange model, in which reconnection between open and closed flux is the dominant process by which the coronal open field evolves (Fisk et al. 1999; Fisk 2005; Fisk & Zurbuchen 2006). Note that reconnection between an open and a closed field line produces another pair of open and closed field lines, and no disconnected lines (Crooker et al. 2002). In this model the magnetic topology is presumed to be highly complex, with open fields mixing into closed field regions, as in a diffusion process.

On the other hand, motivated by their observations, solar researchers have primarily used the quasi-steady model in which the coronal magnetic field is calculated from the instantaneous normal component at the photosphere using some extrapolation procedure. The simplest and most widely used such procedure is the source-surface model, in which the field is assumed to be current free in the corona and purely radial at some fixed spherical surface (Altschuler & Newkirk 1969; Schatten et al. 1969; Hoeksema 1991). This model is calculated on a routine basis from photospheric flux data that is usually low resolution (or smoothed substantially). It appears to predict a topology of clearly separated open flux (coronal holes) and closed flux regions, at least for the observed photospheric flux distributions (Wang & Sheeley 1990; Arge & Pizzo 2000; Luhmann et al. 2002, 2003; Schrijver & DeRosa 2003). Although the dynamics are not explicitly calculated, field line opening and closing driven by the evolution of the photospheric flux are implicit in this model. The assumption is that the coronal field evolution can be approximated by a sequence of source-surface solutions. The quasi-steady
models have been extended in recent years to include solution of the full MHD equations, no longer requiring the assumption of a current-free magnetic field in the corona and an artificial source surface. Typically, the MHD models are used to compute a steady-state equilibrium (e.g., Linker et al. 1999; Riley et al. 2001; Odstrcil 2003; Roussey et al. 2003), but they can also be employed for fully dynamic simulations (Riley et al. 2003).

It should be emphasized that the quasi-steady models are nothing more than particular implementations of Parker’s basic theory of the solar wind. In his seminal work, Parker (1958) argued that if the gas pressure in the corona becomes larger than the magnetic pressure, the gas must expand outward as a wind, dragging the field lines with it. Hence, the open and closed regions are fundamentally determined by the magnitude of coronal heating and the magnitude of the field in the corona. Implicit in Parker’s theory is the assumption that field will open and close in response to changes in these quantities. In principle, the MHD models are a more physical representation of the theory; but since the coronal heating mechanism is not known, they are also ad hoc in practice.

The previous studies of the interchange and quasi-steady models imply that magnetic field topology is likely to be the key factor that can be used to distinguish between and test the two types of models. It may be, however, that for a sufficiently complex photospheric flux distribution the quasi-steady model will begin to approach the interchange model, in that open field regions will be mixed indiscriminately with closed field regions, and that interchange reconnection will dominate the evolution. Therefore, in this paper we examine the possible topologies predicted by the quasi-steady models, and in particular, the effect of a multipolar photospheric flux distribution. Using the two straightforward assumptions that underlie the quasi-steady models and some analytic arguments, we derive severe constraints on the open field topologies allowed by these models. We will conclude below that the quasi-steady model remains topologically simple for arbitrarily complex photospheric flux. However, it does begin to resemble the interchange model geometrically, in that the boundary between the open and closed field in the quasi-steady models develops extreme structure, with narrow corridors of open field extending deep into the closed field region. Both types of models agree that reconnection is the dominant physical process involved in the evolution of the open magnetic flux. Reconnection is also the process generally invoked for transient opening and closing of field lines with it. Hence, for a typical sound speed of \( \sim 10^7 \text{ cm s}^{-1} \) and scale of the order of the solar radius \( \sim 10^{10} \text{ cm} \), the restriction is that the photospheric field evolves on the timescale of hours or longer.

This restriction holds for fields of the scale of active regions, but not for the small magnetic-carpet bipoles. The effect of these small temporal and spatial scale bipoles is to add some high-frequency “noise” to the system. The interchange advocates would argue that it is exactly this “noise” that is responsible for the distinguishing feature of their model: the open field can diffuse freely into closed field regions. It seems unlikely, however, that the small-scale carpet fields can have such a dramatic effect on the field topology. The key point is that, even in the presence of the magnetic carpet, the structure of the coronal field is determined by the balance between the Lorentz force and gas pressure. This means that, in the low-beta corona where the field dominates, the time-averaged field line geometry must resemble that of a fully closed potential or, more generally, a force-free equilibrium field. Note that this constraint holds even in the open field regions. Field lines can deviate significantly from the potential/force-free state only at heights where the gas pressure becomes of order the magnetic pressure. Therefore, in those regions where the force-free field lines never reach large heights, and the field must be closed irrespective of any “noise” due to the presence of the carpet. The occurrence of open field lines in such a region would require a large deviation from the geometry of the force-free state, which would violate basic force balance.

The second assumption of the quasi-steady model is that there are no long-lived current sheets in the closed field corona. (In this paper we use the term “current sheet” to refer exclusively to a true discontinuity in the field, not to merely a large concentration of electric current.) This second assumption is also likely to be valid, because the corona is low beta and appears to be evolving slowly (except during CMEs and flares), so the field must be in a fairly robust equilibrium. It has been argued by several authors that current sheets are not likely to exist in the force-free corona (van Ballegooijen 1985; Antiochos 1987). Note that the small-scale, transient currents required by many models of coronal heating (e.g., Parker 1983) are allowed, but are not expected to influence the global corona. Furthermore, long-lived volumetric currents such as those induced by photospheric shearing and twisting motions are also allowed and can be arbitrarily large, as long as there are no true discontinuities in the magnetic field. Of course, in the source-surface models there are no currents whatsoever, the field is assumed to be potential, but volumetric currents are almost certain to be present in the MHD models.

3. CORONAL HOLE UNIQUENESS

Although the assumptions of smooth currents and a quasi-steady corona seem innocuous, they actually place severe restrictions on coronal hole topology. In particular, we claim that they imply the following “uniqueness conjecture”: Every unipolar region on the photosphere can contain at most one coronal hole.

We do not have a rigorous mathematical proof of this conjecture; however, we present below compelling arguments that it should hold for observed solar conditions, as long as the assumptions above are valid. In particular, it should hold for the quasi-steady, source-surface, and MHD models. We believe, however, that the conjecture is valid in general.

3.1. Bipolar Topology

To clarify our arguments for uniqueness, consider first the simplest possible coronal topology: a bipolar field with one coronal hole in each unipolar region. An analytic source-surface
solution for such a field can be obtained by using the method of images (Jackson 1962). For a dipole $d$ located at a point $r_d$ inside the Sun, and a source surface at radius $R_S$, the magnetic field potential, $\Phi = -\nabla \Phi$ is given by

$$\Phi = \frac{d \cdot (r - r_d)}{|r - r_d|^3} = \frac{R_S^3 d \cdot (R_S^2 r - r_d)}{|r_d|^3 r - R_S^2 r_d|^3}.$$  \hspace{1cm} (1)$$

It is straightforward to verify that $\Phi = 0$ at the source surface, $r = R_S$, so the field is purely radial there. Equation (1) is highly useful, because one can build up a field of arbitrary complexity simply by adding more dipoles. For the simplest possible case of a single dipole at Sun center pointed along the vertical axis, equation (1) reduces to

$$\Phi = \frac{(d \cdot r)(R_S^3 - r^3)}{R_S^3 r^3}$$  \hspace{1cm} (2)$$

and is shown in Figure 1a for $R_S = 2.5$ with the solar radius normalized to unity.

Topologically, the magnetic field can be considered to define a mapping that connects the points along any field line in the volume. This mapping is given by integrating the equations for a field line:

$$\frac{dx}{ds} = \frac{B_x}{B}, \quad \frac{dy}{ds} = \frac{B_y}{B}, \quad \text{and} \quad \frac{dz}{ds} = \frac{B_z}{B}.$$  \hspace{1cm} (3)$$

where $s$ is distance along a field line, and $x, y,$ and $z$ are the coordinates of the field line at the point $s$. It is apparent from these formulas that, in order for the mapping to be discontinuous, i.e., for field lines to “split,” either one of the components $B_x, B_y,$ or $B_z,$ must be discontinuous, which implies the presence of a singularity in the current (a current sheet), or the magnitude $B$ must vanish. All the arguments below follow from this straightforward, but very powerful result: In the absence of singular currents, magnetic field lines can split only at locations where the field vanishes, such as true null points. Although we will use this result only in the context of solar coronal magnetic fields, we emphasize that it holds in general. For example, the classical model for the magnetosphere, a dipole embedded in a background field, exhibits the topological feature that the field lines split only at the magnetopause and distant null points (Cowley 1973; Stern 1973; Lau & Finn 1990).

Note that the field of Figure 1a does vanish at the null line located on the equator of the source surface, and in fact, the field lines do split there. (For a source surface solution the null is of the X-type, whereas for a solution with a solar wind it would be of the Y-type; but this difference is irrelevant to our argument.) Equations (1)–(3) confirm our expectation that the mapping must be discontinuous across a coronal-hole boundary. This boundary is a true separatrix surface, just like the well-known fan surface described below. However, the mapping defined by equation (2) is continuous everywhere else, in particular in the closed field region. Since there are no currents in the closed field corona of Figure 1a, the field line mapping is especially simple; but even if photospheric flows were applied that greatly deformed the polarity distribution and that generated strong current in the corona, the field line mapping would remain continuous as long as the flows were smooth (Antiochos 1987). Note that we allow for the possibility of quasi-separatrix layers (QSL), where the mapping exhibits large gradients (Titov et al. 2002) as long as it is not truly discontinuous.

The key point in the argument for uniqueness is that the field lines in the closed field region of Figure 1a cannot split, because neither of the necessary conditions is met: a simple bipolar region contains no nulls and, by assumption, no current sheets are present there. If the field lines cannot split, it is straightforward to demonstrate that the positive and negative polarity regions of Figure 1a can each contain only one coronal hole. Assume that a second disconnected hole exists, for example, in the northern hemisphere (Fig. 1b). This second hole must be fully inside the northern polarity region, because as a direct consequence of the absence of current sheets: A coronal hole boundary cannot intersect a polarity inversion line (PIL). This condition must be generally valid; otherwise the heliospheric current sheet would extend all the way down to the photosphere, violating our fundamental assumptions. (Note that to avoid confusion with magnetic null lines, we will use the more precise term PIL rather than the term “photospheric neutral line” that is also commonly used.) Since the coronal holes do not intersect each other or the PIL, there must exist an annulus of flux in the closed region that completely encircles one of the holes but not the other, as sketched in Figure 1b. This annulus flux must map to a negative region across the equatorial PIL. Because the open flux of the coronal holes constitutes a barrier that extends to infinity, it is evident from Figure 1b that the flux of any such annulus must go around at least one of the
coronal holes in order to connect across the PIL; i.e., the field lines must split in the closed corona region. But this is not possible in the absence of current sheets; hence, a second disconnected coronal hole is forbidden.

This simple but compelling argument implies several important points. If the open field is intermixed with the closed field, as in the interchange model, the topology is equivalent to many small coronal holes embedded in the larger scale closed field. Therefore, by reversing our argument, we conclude that the interchange model requires the presence of many current sheets in the corona. Furthermore, the current sheets must be inherently transient, because unlike the quasi-steady heliospheric current sheet, gas pressure cannot maintain current sheets in the low-beta corona.

3.2. Open Field Corridors

Another important point is that the coronal hole topology of Figure 1b requires only minor modification to make it agree with the uniqueness hypothesis: the addition of a very thin corridor of open field connecting the two holes, as illustrated in Figure 1c. The geometry of the corridor can be arbitrary as long as it connects the holes, in which case it becomes impossible to find an annulus of closed field that passes between the two holes, and field line splitting is no longer an issue. If the holes are connected, however, only one continuous coronal hole exists in the northern hemisphere, and uniqueness still holds. Note, however, that the corridor may be below observable resolution limits, so that the corona would appear to contain two disconnected holes, reconciling observations with our hypothesis. We contend that such narrow open field corridors are likely to be present in the real corona, providing a natural explanation for the well-observed phenomenon of seemingly disconnected coronal holes (Kahler & Hudson 2002).

It is easy to find open field corridors in the source-surface models of observed solar fields. Figure 2a shows the flux distribution for Carrington rotation 1922 (1997 April 24—May 22), and Figure 2b shows the regions of open and closed fields as calculated by the standard source-surface model (with a source-surface radius of 2.5 \( R_\odot \)) for this flux distribution. Although the polarity regions and the PILs are much more complex than the single dipole of Figure 1, both systems exhibit one dominant polarity in the north and one in the south, with a PIL separating them. There are also numerous small opposite-polarity regions with their PILs separating them from the main polarities, but none of these appear to contain a coronal hole. It should be noted that the actual flux distribution at the photosphere used to calculate the source-surface model of Figure 2b contains many more opposite polarity regions than can be seen in Figure 2a, which, for ease of viewing, shows the flux slightly above the photosphere (at \( r = 1.04 R_\odot \)). These opposite polarity regions are responsible for the numerous, small closed field circular regions near the boundary of the coronal hole in Figure 2b. In fact, there are undoubtedly many more such regions throughout the polar coronal holes on the Sun than can be observed with present instrumentation, so a true coronal hole map must actually resemble a "swiss cheese" pattern. We will consider the effect of such opposite polarity regions later in this paper.

In the northern hemisphere near the center of Figure 2b, the arrow points to a coronal hole that appears to be well separated from the main polar hole, but still in the same polarity region. If so, this would clearly violate uniqueness. In order to investigate this "disconnected" hole in detail, we calculate an analytic approximation to the field of Figure 2a. Note that for Figure 2a the solution was calculated numerically on a fixed grid using a finite-difference scheme to solve Laplace’s equation

\[ \nabla^2 \Phi = 0, \tag{4} \]

with \( \Phi = 0 \) at the source surface. A finite-difference solution of equation (4) is convenient for deriving initial conditions to a time-dependent code, but for examining the detailed topological properties of the source-surface model, a numerical solution is not as effective as an analytic one. Therefore, we have taken the magnetic flux distribution of Figure 2a and calculated its spherical harmonic expansion out to large order \( l = 51 \). Of course, a finite-order expansion does not return the identical flux distribution on the boundary as in Figure 2a, but this is not significant. Our only requirement is that the flux distribution be approximated sufficiently accurately that the "disconnected" hole is preserved.

Given the coefficients for the boundary-flux expansion, \( C_{lm} \), we can then write the exact source-surface solution in the domain (e.g., Wang & Sheeley 1992)

\[ \Phi = \sum_{l=0}^{L} \sum_{m=-l}^{l} C_{lm} Y_{lm}(\theta, \phi) \left( \frac{r^{2l+1} - R_x^{2l+1}}{r^{l+1}} \right) \frac{1}{\left[ I + (l+1)R_x^{2l+1} \right]}, \tag{5} \]

where, as in equation (2), the solar radius is normalized to unity. The advantage of this formulation is that, in principle, we can use the analytic solution given by equation (5) to determine the field line mapping with arbitrary accuracy. (In practice, however, the computational time required may be prohibitive.) The other important advantage of the expansion above is that simply by using different values for the order of the expansion, \( L \), we can investigate the effect of applying different levels of smoothing to the photospheric flux distribution.
We find that, for spherical harmonic solutions with orders ranging up to $L = 51$, all of the detailed open field structure evident in Figure 2b disappears except for the “disconnected” hole in the north. Evidently, this hole is a robust feature of the photospheric polarity distribution. Figure 3 presents results from the $L = 31$ solution. The first panel (Fig. 3a) shows the domain used for plotting, along with the PIL on the photosphere (thin blue line) and the PIL on the source surface (thick blue line). The photospheric polarity is predominantly positive in the region shown, the north pole is all positive, but near the center of the region there is an extended tongue of negative polarity oriented east-west. The “disconnected” hole of Figure 2b lies south of this negative tongue. The next panel (Fig. 3b) shows two sets of open field lines (red and green lines) traced from the source surface down to the photosphere. The starting footpoints for the two sets are two lines of constant longitude separated by approximately 20° in longitude on the source surface. Each set begins with a field line whose starting point is very close to the source-surface PIL, and whose end point on the photosphere lies in the “disconnected” hole, south of the negative polarity tongue. It is evident from Figure 3b that the field maps a line of constant longitude on the source surface to a line on the photosphere that passes from the small “disconnected” hole, around the tongue, and into the main northern polar hole.

If the small hole were truly disconnected, then the field line mapping would be discontinuous, and the line on the photosphere defined by each field line set would contain a break. Figures 3c and 3d show closer views, (and from different perspectives), of the field line mapping at the photosphere. We note that the mapping is continuous, but as it passes around the tongue-like PIL, the mapping develops very large gradients there. We had to increase the density of starting footpoints on the source surface by 2 orders of magnitude in order to find field lines that map to locations near the tip of the tongue. For expansions of significantly higher order, such as $L = 41$, the negative polarity tongue becomes more extended, and the mapping develops such large gradients that it becomes impossible with our numerical integration routine to find open field lines near the tip of the tongue by tracing downward from the source surface. To find such field lines one needs to begin the trace near the photosphere. This region of the field line mapping can be considered to be an extreme example of a QSL.

The key point, however, is that the field line mapping is indeed continuous for all values of $L$, implying that a narrow corridor of open field must connect the southern hole to the main polar hole. Consequently, there is only one hole per unipolar region, in agreement with uniqueness.

As calculated from the spherical harmonic solution with $L = 31$, the corridor has an extremely narrow photospheric width, of the order of only 10 km at some locations. Note that the photospheric footpoints of the two field line sets are indistinguishable in Figure 3d, whereas the source-surface footpoints are separated by scales of order the solar radius. Such a small scale for the corridor width calls into question the quasi-steady assumption. A corridor of this scale on the Sun would be highly dynamic, with field constantly opening and closing in response to small changes in the photospheric flux. But as long as

Figure 3. — (a) Plotting domain for a spherical harmonic expansion solution with $L = 31$. The blue contours correspond to the PILs on the photosphere and on the source surface. (b) Sets of field lines (red and green lines) traced from lines of constant longitude on the source surface down to the photosphere. (c) View of the photospheric footpoint positions of the field lines of the previous panel. (d) A close-up of the region near the tip of the negative polarity tongue showing that the footpoints of the red and green field lines become unresolvably close in this region.
the seemingly disconnected hole is present, then on average, a continuous corridor must exist. Note that additional, wider corridors can clearly be seen in the upper left of Figure 2b. If enough corridors are present, then on a large scale the open and closed fields will appear to be intermixed, so that near the coronal hole boundary the structure of the quasi-steady models may begin to resemble that of the interchange model. Furthermore, since they are likely to be continually dynamic, the corridors could become an important source of the slow wind. This issue clearly needs further study.

3.3. Multipolar Topology

The argument above was developed for a simple bipolar magnetic topology. The real corona, as seen in Figure 2, usually contains other large-scale structures, such as active regions. These add topological complexity to the coronal field, especially null points where field lines do split. Therefore, the next step in the proof for uniqueness is to consider the effect on the arguments above of adding an active-region dipole to the topology of Figure 1. If the active region merely distorts the equatorial PIL, as in the well-studied case of 1997 May 12 (Arge et al. 2004), the topology of the closed field is still bipolar, and the annulus argument above holds. Similarly, if the active region emerges inside one of the open field regions, then the only effect is to produce a small closed field region inside one of the polar holes, which again has no effect on the topology of the main closed region or on our argument. A significant change in topology occurs only if the active region produces a new PIL in the closed field photosphere, as in the field of Figure 4a, where we have added a new low-latitude dipole source to equation (1). The potential is now given by

$$\Phi = \Phi_0 + \frac{d_1 \cdot r}{|r - r_1|} - \frac{R^2_s r^2 d_1 \cdot r}{|r|^2 |r - R^2_s r_1^2|},$$

where $$\Phi_0$$ is the potential of equation (2), $$d_1 = (0, A, 0)$$, and $$r_1 = (0.9^\circ, 60^\circ, 0^\circ)$$. In other words, a dipole pointing due north is placed at a latitude of 30° and a depth of 0.1 $$R_s$$.

The addition of the active-region dipole produces a new PIL separating the negative active-region spot from its positive surroundings. We will use the term “nested” polarity region to refer to a configuration like the negative spot, which is wholly surrounded by a larger opposite-polarity region. The fact that some of the surroundings are in the form of a strong positive spot just south of the strong negative spot is not important to the topology. Associated with the active-region PIL is a null point in the corona, along with the usual dome-shaped fan separatrix surface and pair of spine lines (e.g., Greene 1988; Lau & Finn 1990; Antiochos 1990; Priest & Titov 1996). The intersection of the fan with the photosphere forms a closed separatrix curve defining the boundary between the positive flux connecting across the active-region PIL and that connecting across the equatorial PIL. Note that the field lines of the fan and spines all connect to the null and split there; consequently, the mapping is discontinuous at the fan and spines. The magnetic field of Figure 4a is simply the well-known embedded bipole, the most likely topology if there are two PILs (three polarity regions) on the photosphere (Antiochos 1998).

The only other possible topology for a two-PIL photosphere is that of a “bald patch” in which the null point occurs below the surface and the magnetic field over part of the nested PIL is concave up (e.g., Titov et al. 1993). Bald patch topologies can occur only if the nested polarity region is small; therefore, they are unlikely to play a significant role in determining the large-scale solar field, such as the coronal hole topology. On the other hand, they have interesting implications for coronal dynamics, because we do not expect bald patch topologies to survive in open field regions (Mueller & Antiochos 2007, in preparation). Determining the evolution of bald patch topologies is problematic, however, because simple line-tied boundary conditions cannot be used wherever the coronal field is concave up at the photosphere (Antiochos 1990; Karpen et al. 1990). Consequently, we will not consider bald patch topologies further in this paper.

It should be emphasized that the null-point topology of Figure 4a is observed to be a generic feature of coronal magnetic fields. It was immediately seen by Skylab, where it was referred to as a “fountain” region (Tousey et al. 1973; Sheeley et al. 1975), and by Yohkoh, where it was referred to as an “anemone” region (Shibata et al. 1994; Vourlidas et al. 1996). As demonstrated by numerous extrapolations of observed photospheric fields, it is ubiquitous throughout the Sun on a broad range of scales (e.g., Aulanier et al. 2000; Fletcher et al. 2001; Luhmann et al. 2003; Ugarte-Urra et al. 2007). Of course, the true solar field is almost always more complex than that of a single active-region bipole. However, if the photospheric flux consisted of clearly separated embedded bipoles, each with its own PIL, then the topology would simply be that of a collection of nonintersecting fans and spines. Even for complex active regions, we expect that such regions also appear mainly bipolar, on the large scale that is
important for determining the global field. Therefore, if uniqueness holds for the topology of Figure 4a, it is likely to hold in general for all observed solar fields.

We now add a disconnected coronal hole to the system shown in Figure 4a and consider the implications of the null-point topology for our uniqueness conjecture. If the separatrix curve of the nested polarity does not intersect a coronal hole boundary as in Figure 4b, then it is always possible to find an annulus of closed flux surrounding either hole that maps across the equatorial PIL. In this case, the presence of the nested polarity is irrelevant, and the field of the annulus must split around one of the holes, which is disallowed. Thus, our earlier argument for uniqueness applies without change. But what happens if the active region moves or expands, such that its associated separatrix curve intersects the coronal holes as sketched in Figure 4c? In this case there is no annulus of flux closing across the equator that encircles only one of the holes—all such annuli encircle both holes, so that no splitting of the field is required. Furthermore, any annulus of closed flux that encircles only one of the holes must cross the nested-polarity separatrix curve, in which case the annulus flux is allowed to split at the null point.

Although a rigorous treatment of the topology implied by Figure 4c requires a fully dynamic calculation, we can gain useful insight by considering the topology predicted by a sequence of static models in which a nested-polarity separatrix curve approaches a coronal hole boundary. Figure 5 shows the source-surface model topology for two slightly different positions of the embedded dipole. The dipole in Figure 5a is located only $1^\circ$ south of that in Figure 5b. The changes in location and shape of the PIL and the separatrix curve between Figures 5a and 5b are imperceptible, but the open field topologies are dramatically different. The coronal hole boundary passes completely north of the fan separatrix curve in Figure 5a and completely south in Figure 5b, so the separatrix and the coronal hole never actually intersect. A point of clarification is that, when the nested polarity is inside the coronal hole as in Figure 5b, the separatrix itself can be considered to define a coronal hole boundary, because it separates the flux that closes across the nested polarity PIL from the surrounding open flux. Note, however, that this type of coronal hole boundary is confined to low heights, well below the source surface, because the fan field lines all connect to an X-type null low in the corona. For the issue of uniqueness, the only coronal hole boundaries that are important, and that we will refer to, are those that connect to the source surface.

Figure 5 shows the change in topology for a $1^\circ$ shift in the embedded dipole’s position, but we find the same result no matter how small the shift: either the separatrix curve is fully inside the closed field region, as in Figure 5a, or it is fully inside the coronal hole, as in Figure 5b. The implication of the quasi-steady model, therefore, is that the open field topology undergoes a discontinuous jump as a nested polarity region approaches a coronal hole boundary. This result may seem unphysical, but it follows inevitably from the fan-spine topology of Figure 4a, in which the spine lines split at the null to form the fan. The actual amount of flux in the spines and fan is a set of measure zero, so this picture is to be taken in the sense of a limit. The relevant point, however, is that any arbitrarily small but finite flux bundle enclosing the outer spine maps to an arbitrarily narrow but finite-width annulus on the photosphere surrounding the separatrix curve. Therefore, if the outer spine is closed (connects to the photosphere), then the fan is surrounded by closed flux, and the nested polarity must be considered to be in the closed field region. For this case the null is inside the closed field region as in Figure 5a. Conversely, if the outer spine is open (connects to the heliosphere), the fan must be surrounded by open flux, and the nested polarity along with the null is in the coronal hole, as in Figure 5b. The case where the outer spine is exactly on the coronal hole boundary corresponds to the fan being surrounded by an open region of vanishing width—a singular case that can be neglected.

We conclude that the configuration shown by Figure 4c is impossible, because: A nested polarity region must be surrounded by either all open or all closed field. In other words, a separatrix curve cannot intersect a coronal hole boundary, and consequently, the correct topology for the system of Figure 4c must actually be that shown in Figure 5c: a seemingly disconnected coronal hole connected by an open field corridor. But if separatrix curves and coronal hole boundaries cannot intersect, then the annulus argument can always be applied and uniqueness holds even in a multi-polar topology like that of Figure 4.

Since Figure 5 shows only a sequence of potential field states, a critical question that immediately arises is whether a true dynamical evolution will be compatible with this sequence. When a dipole is convected by photospheric flows toward a coronal hole, we expect that the null point will deform into a current sheet similar to the classic Syrovatskii (1981) theory, and that magnetic reconnection will occur between the spine and external flux. Such reconnection has been observed in many numerical experiments (e.g., Parnell & Galsgaard 2004; Pontin & Galsgaard 2007). This spine reconnection will act so as to exchange the outer spine with external flux, effectively moving the spine flux toward the coronal hole boundary (and also destroying the current sheet). When the spine reaches the coronal hole boundary, reconnection between open and closed flux will move the outer spine into the open field region. Once inside the coronal hole, any subsequent motion will result in further interchange reconnection, as has been proposed in models for heating the solar wind and coronal hole plumes.

![Figure 5](image-url)
(Parker 1992; Axford & McKenzie 1992; DeForest & Gurman 1998). Although this scenario awaits verification with fully time-dependent calculations, which are still in progress (Antiochos 2006), it seems clear that magnetic reconnection can readily produce the evolution implied by the source-surface solutions of Figure 5.

4. CORONAL HOLE NESTING

An important feature of the topology of Figure 5b is the very thin (below the resolution of the figure) open field corridor, attached to the polar coronal hole at both ends, that passes around the south side of the fan, and right over the center of the strong positive spot there. Although the presence of the corridor implies that the coronal hole is no longer simply connected, it is still a single, unique coronal hole. Such corridors should form naturally on the Sun any time a bipolar region moves into a coronal hole. In fact, as noted in §3.2, such corridors can be seen at the edges of the polar coronal holes in Figure 2b.

Open field corridors play a critical role in reconciling multipolar topologies with our uniqueness conjecture. Let us return to the topology of Figure 4a, in which the active-region separatrix curve is well removed from the polar coronal-hole boundary, and consider the effect of opening a coronal hole inside the nested polarity; as sketched in Figures 6a and 6b. This should be allowed by the uniqueness conjecture, because there would still be only one coronal hole per unipolar region on the Sun. But the annulus argument forbids the topology of Figures 6a and 6b on two counts. First, consider any annulus of closed flux that surrounds the inner spine. The inner spine maps to the whole fan surface, which surrounds the nested polarity, therefore, any annulus surrounding the spine must map to an annulus surrounding the nested polarity. But in order to pass around the embedded polarity coronal hole, any such annulus of closed flux would have to split, which is not allowed. This problem is easily taken care by insisting that: *Any coronal hole that opens inside a nested polarity must encompass the spine.* This conclusion has important implications for models of CME initiation, such as the breakout model (Antiochos et al. 1999). It predicts that prior to eruption the inner spine should appear to move toward that part of the sheared PIL that eventually erupts. Furthermore, the amount of energy available for eruption will depend on how much the spine can move (Antiochos et al. 1999; DeVore & Antiochos 2005).

The second application of the annulus argument leads to a more surprising conclusion. Any annulus of closed flux that surrounds the polar coronal hole of Figure 6a would clearly have to split even if the embedded coronal hole had a permissible, open-spine topology. It would appear, therefore, that the annulus argument implies not just one hole per unipolar region, but only one hole per polarity (i.e., only two on the whole Sun), which cannot be right. The resolution to this conundrum is illustrated by Figure 6c—the formation of an open field corridor that passes completely around the nested polarity region. In Figure 6c, field line splitting is no longer a problem, because any annulus of closed flux that surrounds the main coronal hole must also surround the active-region hole. Also, any annulus of closed flux inside the nested region that surrounds the nested hole closes completely within the closed flux region of the nested polarity. These arguments imply that multiple coronal holes are, indeed, possible, but they must obey the following nesting conjecture: *Coronal holes of nested polarity regions must themselves be nested.*

Although, the nesting conjecture imposes a powerful constraint on coronal hole topology, its importance for observed solar fields is uncertain. Nested coronal holes imply the presence of small conical current sheets in addition to the main heliospheric current sheet. In situ measurements usually indicate a single current sheet in the heliosphere, implying that there are only two coronal holes on the Sun. If so, then the issue of nested coronal holes becomes moot. Note that the nested polarity flux must be large in order to obtain a nested coronal hole within the context of the source-surface model, in other words, a large active region far from the equator. This combination is rarely observed on the Sun, but still, it would be intriguing to search for any such nested holes in the published source-surface maps, and then to search for their current sheets in the heliosphere.

Furthermore, the source-surface model is likely to underestimate the occurrence of nested coronal holes, because this model does not include force balance between field and plasma. The topology of Figure 5b contains a simple X-type null on the fan surface separating open and closed fields. If plasma is added to the configuration of Figure 5b, a difference in gas pressure will develop between the confined plasma inside the fan and the unconfined, solar wind plasma outside, just as there is a well-observed difference in the gas pressure between solar closed field regions and coronal holes. For low beta, a gas pressure gradient across the fan or any other magnetic surface can readily be balanced by a small magnetic pressure gradient there; but, this is not possible near the null where the beta becomes infinite. It seems, therefore, that the plasma pressure would strongly deform the field near the null, and in some cases, may force open a finite region of flux around the inner spine, even though the source-surface model predicts no nested coronal hole. Of course, the nested polarity regions on the Sun are likely to be evolving via flux emergence or cancellation and photospheric motions. In fact, several models for accelerating the wind (Parker 1992; Axford & McKenzie 1992) and for
forming polar plumes (DeForest & Gurman 1998) invoke this process of interchange reconnection between the closed flux of an embedded bipole and surrounding open field. Therefore, it may be that any small nested holes are masked by the reconnection and dynamics. There are bound to be cases, however, where large quasi-static bipoles appear inside coronal holes. We predict that, at least for these cases, there would be a significant and possibly observable difference between the topology predicted by the source surface and the MHD model.

One situation in which the nesting conjecture is quite likely to play an important role is in long-lived dimming regions formed by CMEs. The dimming regions are believed to be transient coronal holes where the magnetic field has been forced open by a CME (Thompson et al. 2000). Since they are transient, it is not clear that our arguments above apply. If the holes are sufficiently long-lived (timescales of tens of hours), however, the quasi-steady assumption may still be valid. If so, then we can make two predictions on such “not-too-transient” holes. First, any dimming region that forms inside an active-region PIL must encompass the inner spine. This prediction probably is difficult to test, because the inner spine is not easily observed (Aulanier et al. 2000). Second, the formation of such a dimming region must be accompanied by the formation of a transient coronal hole (possibly a very narrow open field corridor) surrounding the PIL. This latter prediction may well be testable in some well-observed CMEs.

5. DISCUSSION

Let us summarize our main findings and predictions on coronal hole topology. First, we list two supporting “lemmas”: this statement is rigorously valid for the quasi-steady models, because the heliospheric current sheet cannot extend down to the photosphere. In the interchange models, however, open flux is presumably free to diffuse across PILs, which emphasizes the striking difference between the two models.

A nested polarity region must be surrounded by either all open or all closed field. The prediction from this result is that coronal hole boundaries undergo discontinuous jumps in response to bipolar regions entering or exiting the holes.

Next, application of the annulus of closed flux argument leads to our two main results, the uniqueness and nesting conjectures: Every unipolar region on the photosphere can contain at most one coronal hole. We predict that seemingly disconnected holes are actually connected by observationally unresolved open field corridors. These corridors are likely to be dynamic, with the field continuously opening and closing in response to photospheric motions and flux emergence or submergence. Furthermore, we expect such corridors to be ubiquitous at the boundaries of coronal holes, causing these boundaries to have a fractal-like and inherently dynamic structure (see Fig. 2b). Consequently, the corridors may be an important source of the slow wind.

The prediction is that if a coronal hole develops in an active region (i.e., a nested polarity region), then the polar coronal hole will grow to surround the nested polarity. This may hold even for transient coronal holes associated with CMEs. Note that implicit in the nesting conjecture is the conclusion that nested coronal holes are, indeed, possible. As argued above, however, they are likely to be rare. On the other hand, it would clearly be a useful test of the quasi-steady models to search for evidence of such nested coronal holes and their accompanying current sheets in the heliospheric data.

Related to the nesting conjecture is the corollary: Any coronal hole that opens inside a nested polarity must encompass the spine. Testing this prediction requires unambiguous identification of the null and spine before the field opening, which is generally not possible with available observations. There have been cases of eruptive flares, however, in which there appears to be evidence for strong reconnection at the spine in association with the flare-ribbon reconnection (e.g., Aulanier et al. 2000; Ugarte-Urra et al. 2007), implying that the spine opened along with the filament channel.

Since the quasi-steady models are solarcentric in origin, the discussion above focuses mainly on the coronal consequences of our results, but they also have important implications for the heliosphere. Our central conclusion is that, even for complex photospheric flux distributions, we expect the magnetic field topology to be well behaved, with a topologically smooth separation of open and closed fields. This result is especially true deep in the closed field regions, such as near any photospheric polarity inversion line, for example. Counter to the predictions of the interchange models, we claim that no open flux can exist deep in the closed field regions of the corona, except for transient openings due to CMEs. This conclusion follows from basic force-balance arguments that are likely to hold irrespective of any small-scale photospheric structure and dynamics. In short, the low-lying closed field is definitely closed.

The situation is more interesting, however, in the open field regions. Even in apparently simple, polar coronal holes, we expect that most of the flux is actually closed. Due to the constant emergence of bipoles, as in the magnetic carpet, there must be many small closed field regions with the topology of Figure 2b inside every coronal hole. Some can be seen clearly in Figure 2b, but there are many more that are either too small or too near the poles to be resolved. As a result of this small-scale “mixing” of open and closed flux, many of the concepts of the interchange models may well be valid inside coronal holes. For example, there is bound to be considerable reconnection between the bipolar closed flux and surrounding open field at the separatrices between them—the fan and spines—as the bipoles move and evolve. We argue that most, if not all, interchange reconnection is simply this reconnection at the null point and separatrix of an embedded bipole (Antiochos 1990), and hence, any consideration of the properties of interchange reconnection must be done within the context of the embedded bipole topology. If coronal holes are densely filled with such reconnecting bipoles, then a diffusion model (Fisk et al. 1999; Fisk 2005) for the open field evolution could apply. As suggested by these authors, this diffusion may be the explanation for many of the classic features of the heliosphere, such as the observation of field line random walk (Jokipii & Parker 1968). Furthermore, the embedded-bipole interchange reconnection is physically identical to that described by various authors as being responsible for accelerating the solar wind and launching Alfvén waves into the heliosphere (e.g., Parker 1992; Axford & McKenzie 1992).

We conclude, therefore, that in coronal hole regions the quasi-steady and interchange models may be quite complementary. The interchange model calculates the random motion of the field due to the magnetic carpet “noise,” while the quasi-steady models calculate the global structure. An important aspect of interchange reconnection in coronal holes is that it should release initially closed field plasma into the heliosphere, even deep in a coronal hole where the fast wind originates. This result may appear to be in conflict with the observation that the fast wind has abundances more similar to the photosphere than to the hot, closed corona (e.g., Zurbuchen 2007). The key point, however, is that the bipoles of the magnetic carpet have a short lifetime, too short to acquire a coronal abundance. As discovered by Sheeley (1996), the “salt-and-pepper” (the original term for the magnetic carpet) tends to exhibit photospheric abundances. Therefore, the fast wind abundances do seem to agree with embedded-bipole interchange
reconnection. On the other hand, there may be other composition tests for such reconnection, that one could identify in the fast wind by comparing heliospheric data with solar spectroscopic observations of closed field regions in polar coronal holes.

In terms of heliospheric consequences, perhaps the most important result of this paper is the nature of the interface between open and closed fields. We find that for typically observed photospheric flux distributions, the coronal-hole boundaries must be extremely complex. Furthermore, they must be highly dynamic, with open field corridors continuously appearing and disappearing as embedded bipole move back and forth across these boundaries via magnetic reconnection. Although complex, these dynamics cannot be described by the interchange model, because the open field cannot simply diffuse into the closed field regions. According to our uniqueness conjecture, the open field must always remain topologically connected to the main body of the coronal hole.

It is tempting to speculate that these coronal-hole boundary dynamics are the underlying cause for the slow solar wind, but this hypothesis needs quantitative study, especially with fully time-dependent simulations. Note that such a model would still have the problem of reconciling the coronal evolution with the heliospheric measurements. We would expect that any opening and closing of the coronal magnetic field, even if highly structured, would show up in the heliosphere as bidirectional or dropout electron heat fluxes. Clearly, this problem of relating the heliospheric data to the standard models for the corona needs much more study.

Finally, it should be emphasized that our results on coronal hole topology listed above are still conjectures, even for the quasi-steady models. One may be able to find counterexamples, especially in systems with special symmetries so that structures such as null lines or null surfaces appear in the corona. On the other hand, the coronal magnetic field is generally observed to have smooth structure, without evidence for such topological pathologies. Note also that topologies such as null lines are structurally unstable, in general, so they would exist only as transient structures. Therefore, if our conjectures are valid for the topologies discussed above, it seems likely that they will hold for most observed solar magnetic fields. We further emphasize that, for application to the Sun, all our results depend on the underlying assumption that the large-scale corona can be considered to be in a quasi-steady equilibrium state, as in the source surface and MHD models. If time-dependent effects dominate instead, as in the interchange models, then many of the statements above need not hold. Consequently, observational testing of our conjectures may be the most effective method for determining the correct theory for the solar-heliospheric magnetic field.

This work has been supported by NASA, ONR, and the NSF SHINE Program. The work has benefited greatly from the authors’ participation in the NASA TR&T focused science team on the solar-heliospheric magnetic field. Discussions with the team leader, Thomas Zurhuenle, were especially helpful. The work of Z. Mikić was supported, in part, by the NASA TR&T focused science team on the CME-ICME connection.
