Solar Phys
DOI 10.1007/s11207-008-9308-3

Modeling a Shallow Solar Dynamo

Kenneth H. Schatten

Received: 2 September 2008 / Accepted: 19 December 2008

Abstract Photospheric ephemeral regions (EPRs) cover the Sun like a magnetic carpet. From this, we update the Babcock – Leighton solar dynamo. Rather than sunspot fields appearing in the photosphere *de novo* from eruptions originating in the deep interior, we consider that sunspots form directly in the photosphere by a rapid accumulation of like-sign field from EPRs. This would only occur during special circumstances: locations and times when the temperature structure is highly superadiabatic and contains a large subsurface horizontal magnetic field (only present in the Sun's lower latitudes). When these conditions are met, superadiabatic percolation occurs, wherein an inflow and downflow of gas scours the surface of EPRs to form active regions. When these conditions are not met, magnetic elements undergo normal percolation, wherein magnetic elements move about the photosphere in Brownian-type motions. Cellular automata (CA) models are developed that allow these processes to be calculated and thereby both small-scale and large-scale models of magnetic motions can be obtained. The small-scale model is compared with active region development and *Hinode* observations. The large-scale CA model offers a solar dynamo, which suggests that fields from decaying bipolar magnetic regions (BMRs) drift on the photosphere driven by subsurface magnetic forces. These models are related to observations and are shown to support Waldmeier's findings of an inverse relationship between solar cycle length and cycle size. Evidence for significant amounts of deep magnetic activity could disprove the model presented here, but recent helioseismic observations of "butterfly patterns" at depth are likely just a reflection of surface activity. Their existence seems to support the contention made here that the field and flow separate, allowing cool, relatively field-free downdrafts to descend with little field into the nether worlds of the solar interior. There they heat by compression to form a hot solar-type Santa Ana wind deep below active regions.

Keywords Solar dynamo · Solar magnetism · Sunspots · Solar activity · Dynamo · Percolation · Superadiabatic · Hinode

K.H. Schatten (✉)
a.i. solutions, Suite 215, 10001 Derekwood Lane, Lanham, MD 20706, USA
e-mail: ken.schatten@ai-solutions.com

Published online: 28 January 2009

1. Introduction

Sunspots are generally thought to arise from deep within the Sun, through the buoyancy of flux tubes. In Babcock's (1961) original dynamo ideas, however, he advocated a shallow dynamo at the beginning of his abstract: "Shallow submerged lines of force ... produce a spiral wrapping of five turns after ... three years." Leighton (1969) considered both the possibility of a shallow dynamo as well as deep dynamo models. These models were "alpha-Omega" dynamos (see the review by Stix, 1974), meaning that magnetic fields were magnified by differential rotation and magnetic buoyancy; the former is called the Omega effect, and the latter is called the alpha effect. The deeper the model, the greater the role that radial differential rotation would play in magnification, compared with latitudinal differential rotation. More recently, Brandenburg (2005) also advocated a shallow solar dynamo. I too have taken this stand (Schatten, 2007; hereafter called paper I). Here those ideas are taken to provide more realistic detailed models, as opposed to a general theoretical treatment.

Let us begin by describing the essence of these new views. Rather than fields arising from deep within the Sun, we consider that the Sun's surface fields coalesce into sunspots and subsequently these fields develop into the solar dynamo. We refer to the original growth behavior as percolation, to examine whether small-scale magnetic fields (ephemeral regions, EPRs) could gather together to form the larger structured sunspots in the highly superadiabatic environment of the Sun's upper convection zone. Superadiabaticity is a measure of the free energy available in a convecting atmosphere to drive dynamical motions. Percolation was first used to study solar phenomena by Seiden and Wentzel (1996) using a cellular automata (CA) model. They examined how deep fields might develop into active regions using a percolation model for the Sun's convection zone. A model for active regions by Fragos, Rantsiou, and Vlahos (2004) further developed CA models to examine the distribution of energy storage of solar active regions within the corona and solar wind. In this paper, the use of CA models is extended to understand the development of ephemeral regions into sunspots and then, subsequently, the development and motion of larger scale field structures (spots and unipolar magnetic regions, UMRs) to form the solar dynamo. The two CA models used here are differentiated as the small-scale model and the large-scale model. These models are used to study the motion of surface magnetism acted upon by forces, both magnetic and from fluid motions arising below the photosphere. The magnetic forces are straightforward, and are simply associated with the tension of the magnetic field in the Babcock – Leighton dynamo picture, although fictitious magnetic monopoles (called unipoles to distinguish them from real, undiscovered monopoles) will be incorporated as a way of calculating the subsurface magnetic forces for simplicity. Surprisingly, however, although the magnetic forces are presumably an important prime driver for motions of field patterns toward the poles as the Sun's dynamo evolves, it does not appear that they have ever been considered as a driving force for field motion. They will be replacing the standard diffusion ideas of Leighton for how following flux (from active regions) drifts poleward.

The model developed here also uses the effects of fluid motion upon the Sun's surface magnetic field in a manner similar to the way meteorologists study storm system motions in the terrestrial or planetary atmospheres. However, without detailed knowledge of the flows below the photosphere, some reasonable assumptions must be made about how percolation drives field motions. Future models may improve upon the approach used here. Use of CA models for flows in the Sun's outer convection zone, where flows become highly nonlinear, may be thought to play the same role in which similar nonlinear behavior drives instabilities in other areas of science. For example, in ferromagnetism, Ising (1925) developed methods to categorize spin states and undertake these types of calculations. The Ising model became

the standard test-bed, essentially a mathematical-based model, like a biological fruit fly upon which the concepts and future techniques in statistical thermodynamics would advance. Not only are these techniques (CA) used now in lattice statistics, where an occupation variable is best described in a nonlinear fashion (using quantum statistics – which is inherently nonlinear, since the variable jumps from integer to integer without smoothly going through the real number line), but they are now used in forest fires, predator – prey (Lotka – Volterra) equations, and in a broad range of disciplines not amenable to the former techniques of differential equations.

Why use CA, rather than MHD as a methodology for ascertaining solutions to the Sun's sunspot and dynamo behaviors? MHD is an exact science, with differential equations based upon physics, amenable to conditions where parameters are slowly changing, thereby allowing differentials to be obtained. Even though a number of models of the upper surface layers of the Sun have had some success, they have not been totally successful for whole-Sun modeling, particularly with regard to how to include the magnetic field together with the dynamics of fluid flow in a convective and radiative environment. For chaotic conditions, where phenomena are changing so rapidly that differentials may not be possible (as in the nonlinear situations previously cited), nonlinear entities, such as CA, may serve to model interactions more readily. Cellular automata are mathematical entities that obey a set of rules that one ascribes to them. The rules we employ are somewhat *ad hoc*, but our choices are governed so as to mimic the interactions of the magnetic field and fluid dynamics in the Sun's outer convection zone. It is hoped that advances will be made as understanding of the complex dynamics of the interaction between field and plasma in the convective environment of the Sun's outer convection zone increases. Until we can obtain a set of equations that can be applied from the smallest scales to the largest, it may be helpful to use some nonlinear techniques that researchers are currently developing.

Before discussing the use of CA in the Sun's outer convection zone, a review of our knowledge of the smallest scale solar fields is in order. To understand the Sun's dynamo, one must move across many scale sizes, from the smallest scale fields to the largest scales possible on the Sun, a solar diameter. Schrijver and Title (2001) studied some of the smallest scale field structures seen in the photosphere, EPRs, and found that EPRs can be viewed as a magnetic carpet covering the Sun. This carpeting effect forms from the turbulent churning that magnifies weak fields into small-scale field structures (EPRs and pores). In paper I it was shown that like-sign magnetic fields from EPRs could cluster together into larger structured sunspots. To do this, it is necessary for like-sign fields to attract, and it was shown that the Sun's outermost layers may undertake this activity. This was based on Parker (1984) rigorously showing that granules, by sequestering magnetic field into a small area, would allow the solar convective energy flux to be freely transported, uninhibited by the magnetic field, if the small-scale fields were so confined. Zwaan (1978), too, examined the possibility that sunspot formation might result from the convective collapse of magnetic fibrils. These views suggest that inflow and downflow patterns below small like-sign magnetic features draw the photospheric magnetic field together into tiny elements of high field strength, rather than remain in the conventional quiet magnetostatic state (slowly changing, nearly uniform field strength). The magnetostatic environment describes the conventional, commonly acceptable "vacuum" picture (*e.g.*, around Earth and planetary objects). Kitchatinov and Mazure (2000) also examined the stability and convective collapse behavior of spot fields in the stratified atmosphere of the Sun owing to the outflow of solar energy. Thus the convective collapse of magnetism could allow these nonlinear processes to occur over large areas of the Sun, rather than just within granules, and may help explain why the Sun's disk is so mottled with magnetism, rather than having a smooth field as we are accustomed to around planets.

To be consistent with these processes, in paper I it was hypothesized that a dynamical force for solar regions of enhanced superadiabatic gradient would gather like-sign magnetic fields together. The outer layers of the convection zone are the most highly superadiabatic in the Sun, and, for that reason, nonlinear processes may be driven to their utmost there and this may partially account for that region being the birthplace of sunspots. The gases there are the thinnest and most neutral (un-ionized); these properties may allow an inflow and downflow pattern to scour large surface areas of weakly ionized hydrogen to gather magnetic elements into active region structures, like stalks of celery. Why this region may be the primary location in the Sun for magnetic fields to amass is discussed more fully in the concluding discussion.

The name percolation is derived from the Latin term *percolare*, which means to strain or filter. Other scientists have often used it when other nonlinear phenomena occur (*e.g.*, growth patterns in forest fires, crystal structures, *etc.*). Some have used the word clustering in a similar vein as percolation. For the Sun, we distinguish two forms or flavors of percolation: superadiabatic and normal. Superadiabatic percolation is assumed to occur where and when the superadiabatic gradient becomes, for a brief time, exceptional large. This would be in the lower latitude layers of the outer solar convection zone where sunspots form and could grow by same-sign fields gathering together. Normal percolation would occur where and when the superadiabatic gradient was not exceptionally large or where the horizontal magnetic field was not too great, so that no gathering of magnetic elements into a preferred orientation was possible. In these regions, like-sign fields would behave normally; namely, they would disperse. What we call normal percolation may be similar to the dispersal of magnetic field by random granular and supergranular motions acting somewhat diffusively, and this process was utilized to good effect by Leighton (1964) in his dynamo models. In paper I, the physics of superadiabatic percolation is discussed more fully and includes an equation for this force based upon hydrodynamic flows. The superadiabatic percolation force is needed to offset the natural magnetic repulsion of like-sign magnetic elements that are a product of Maxwell's equations. This paper extends paper I, primarily through improved modeling of both small- and large-scale field structures. The motions of large-scale field structures will be developed here into a solar dynamo model. Additionally, observations are related to some of the models, adding support for the current view. Before going into the details of the percolation model, it is important to outline a little more why this highly nonlinear model form was chosen, rather than the more traditional MHD models.

Detailed MHD models have succeeded (*e.g.*, Woodward *et al.*, 2006) in modeling the outer surface of the Sun and stars on scale sizes not seen before. Yet, despite the improvements in these petascale computations, the advances have not, for example, led to definitive results with regard to solar activity predictions that have occurred in fields with similar complexity, such as weather forecasting. For the Sun, on short time scales, "nowcasting" has been used, wherein features on the solar disk lead to improved nowcasts of the space-weather environment near Earth, essentially bypassing computational models. Long-term solar activity models (see Schatten, 2005) have been achieved with the Babcock and Leighton dynamo model, by using the Sun's polar fields to supply a "precursor" measure of the state of the Sun's dynamo to predict the next cycle's activity. This is essentially a direct utilization of the physics-based Babcock – Leighton solar dynamo model, as opposed to the more computationally oriented, but less physics-based models, again bypassing the computational models. Despite the lack of a computationally rich structure, the polar field precursor prediction method has worked quite well over decadal time scales. Rather than comparing this with a weather forecasting model, one may liken it to a climate forecasting model, as the Sun's long-term behavior may be more predictable than that on shorter time scales.

The problem with many of the numerical models may be based, not upon their numerical aspects, but upon the physics they employ. This may be essentially the lack of understanding in how to incorporate magnetic fields in MHD models with convective energy transport. Often MHD models just provide magnetic field with an anisotropic pressure in accord with Maxwell's equations. This denies them the complex capability for nonlinear dynamics that is the hallmark of solar activity and leads to the development of the new methods discussed here. As Shakespeare's Julius Caesar (Act I, Scene II) said, "The fault, dear Brutus, is not in our stars, but in ourselves, that we are underlings." Namely, we have fallen victim to the dazzling power of computation, without sufficiently understanding how convection and magnetic field interact together in MHD on large scales, when enormous energy transport is occurring concomitantly.

The new element not considered in MHD models is the percolation force associated with large-scale convection. The superadiabatic gradient helps control this force. We are unfamiliar with the percolation force because it is only significant in a gravitating conducting atmosphere undergoing convective energy transport, such as the Sun's. It is driven by the convective dynamics and, as such, is not solely electromagnetic; electromagnetism only plays a catalytic role, acting on the nonlinear processes involving the convection of a high-conductivity fluid. The catalytic role transforms the small-scale 3D convective cells, like Bénard cells in the Earth's atmosphere, into 1D, unidirected (inward and outward) flow structures, thereby aiding the Sun to shed its luminosity. We view the detailed motions and energetics to involve the inflow of rarefied, weakly ionized hydrogen gas in the photosphere and downflow below growing field regions, with the magnetic field acting to guide like-field elements together, into celery-stalk geometries, and thereby allow the energy transport to occur with least impediment. In such a case, the percolation force F_{P}, between two field sources 1 and 2, was approximated by modifying the hydrodynamic force F_{H}, as

$$F_{\mathrm{P}} \approx F_{\mathrm{H}} \frac{\mathbf{B}_1 \cdot \mathbf{B}_2}{|\mathbf{B}_1 \cdot \mathbf{B}_2|} S \approx \frac{\rho v_0^2}{2} \frac{\mathbf{B}_1 \cdot \mathbf{B}_2}{|\mathbf{B}_1 \cdot \mathbf{B}_2|} S, \tag{1.1}$$

where the hydrodynamic force F_{H} is the expression for the force on a Pitot tube embedded in a horizontal flow of velocity v_0, S is the superadiabatic temperature gradient discussed later, and where $\mathbf{B}_1$ and $\mathbf{B}_2$ are the radial components of the neighboring flux element. The quantity $(\mathbf{B}_1 \cdot \mathbf{B}_2)/|\mathbf{B}_1 \cdot \mathbf{B}_2|$ provides a highly nonlinear form, changing sign rapidly when either $\mathbf{B}_1$ or $\mathbf{B}_2$ reverses direction, allowing attractive and repulsive forces between differently oriented fields. In this manner, when S is large and negative, F_{P} is negative, namely, attractive for like-sign magnetic fields. When S goes to low or normal values, Equation (1.1) gives $F_{\mathrm{P}} \sim 0$ or if S reverses (when S becomes subadiabatic), the force is repulsive, as like-sign magnetic fields in these cases normally do separate. Clearly Equation (1.1) is highly simplified; it bears no dependence upon the separation of flux elements nor on other possibly significant factors. For example, it does not explain how the flow distinguishes gathering inward and outward magnetic elements. Nevertheless, we make use of Equation (1.1) in the modeling efforts in the remainder of this paper. The percolation cases are divided based on energy transport, by subdividing them into two cases. This enables us to choose the simplest model or to invoke the binary nature of nonlinearities when driven to extremes or to invoke the nonlinear physics associated with superadiabaticity.

Two cases or flavors are developed. Case A is for superadiabatic percolation, where like-sign fields attract with a force given by Equation (1.1), which shows why so much attention must be paid to the superadiabatic gradient, S, as this will play a major role governing where and when like-sign fields gather into larger entities (where the horizontal subsurface field is

of sufficient strength to gather and separate the accumulating EPRs). Case B, for normal percolation, is for situations with more normal adiabaticity and field conditions where like-sign fields repel. Here, percolation reverts to a kind of Brownian motion of field elements, presumably driven by overturning convective patterns (granules and supergranules). As a result, sunspots and other fields disperse into weak unipolar fields, unlike fields attract each other, and cancellation between opposite field often occurs. The CA model used can model either flavor A or flavor B.

Let us now turn our attention to modeling the growth of small-scale ephemeral regions into pores and sunspots and their subsequent motion within relatively small regions of the Sun (active region scales). We also must model how fields may move across the solar surface drawn by subsurface horizontal fields, in a manner similar to how electric charges are propelled by electric fields. Later, the large-scale modeling and its relevance to the Sun's dynamo will be discussed.

2. Sunspot Growth and Development

In the model presented in this section, magnetic field motions are computed using CA. These are small elements with a given set of prescriptions, or properties, that govern their changing form. The oldest cellular automaton model is the game of life, by John Conway, wherein complex behavior was achieved with tiny elements governed by a few simple rules (see Gardner, 1970). It was then recognized that in general complex phenomena or behavior might be achievable from simple sets of primitive rules. This has, of course, led to the burgeoning field of chaos, as well as to other new areas of mathematics and deeper understandings of complexity in many areas of the natural sciences.

The model provided allows simulations of the growth and development of small-scale magnetic fields in the photosphere, by using these CA to calculate magnetic field changes on a simulated percolating surface. In all models presented here, when the surface is said to be percolating, this means that the surface is undergoing random shuffles of cells. As mentioned, percolation comes in two flavors – (A) highly superadiabatic and (B) normal – depending upon the adiabaticity and strength of the horizontal subsurface field. These cases govern magnetic field motions, which will be computed by using CA elements. Each element contains a magnetic field, and these are allowed to undertake a small walk within each model time step. Their chances of cell swap will depend upon the percolation flavor effect upon the energy, and how it governs the cell swap. In the remainder of this section, the details of the small-scale CA modeling are discussed and comparisons are made with an interesting event of sunspot growth and development seen with the *Hinode* satellite. The event has been called the trilobite event, since a growing sunspot develops some interesting finlike structures resembling a trilobite; additionally, it appears to swim across the solar surface in a rather remarkable fashion toward another like-sign sunspot. The apparent swimming motion is only seen in the high-resolution movies and likely represents very active convection and small-scale field motions.

The workings of the small-scale CA model used here is explained next. A square computational area is chosen; within this square, grid points may have one of three values: inward (−), outward (+), or neutral (0). The routine shown was run on a 60×60 square grid, with small randomly oriented bipoles placed in a central area of 15 units on a side. For the models shown, randomly oriented bipoles are placed in the central region of a square grid in the first ΔT time steps. This is, in effect, laying a small region of magnetic carpet, similar to the Schrijver and Title (2001) model of EPRs. The bipoles have the option of being displaced

in longitude or randomly oriented; the latter was chosen. First, one pole is placed, and then an opposite pole is placed at a random angle and small distance, δ, from the initial pole. The bipoles could be added slowly, but for the cases chosen they were added for the first two steps in time and allowed to percolate (described next) during these, as well as subsequent, steps.

The small-scale algorithm used consists of a number of steps and substeps. During each step, all the CA may be considered to percolate and/or move based upon the magnetic forces applied to them. A substep consists in the possible motion/interchange of one cellular automaton with another. During this substep, the automaton is considered for "positional cell swapping," as follows, based upon the parameters n and k, where n is the size of the substep and k is the chance of a random move: *i*) Select a random automaton cell (for the superadiabatic case not surrounded by eight neighbors all in the same state; for the normal case, any cell is allowed). *ii*) Among those locations on the grid that are within n units of distance from the cell, select one at random, other than the cell itself. *iii*) If there is a cell at that location and it either has the same state as the chosen cell or is surrounded by eight neighbors all in the same state as that (second) cell then continue with the next substep. *iv*) If there is no cell at that location, then calculate the "energy gain," E, defined as the increase or decrease associated with the number of neighbors of the cell in the same state as that cell, which would result from moving the cell to that location. This really is the "percolation energy" in our model and would be dependent upon the superadiabatic gradient, S, which can be chosen to be a function of location, field, and/or time. For the present, we have E (and S) constant in this model. If $E \geq 0$ then move the cell to the new location. If $E = -1$ then make the move with a probability of 1 in k. *v*) If there is a cell at that location then calculate the combined energy gain, E, that is, the net increase or decrease in the number of neighbors of the two cells in the same state as those cells that would result from swapping them. If $E \geq 0$ then swap the cells. If $E = -1$ or $E = -2$ then perform the swap with a probability of 1 in k. For all the models, the following parameters were used: δ is 3, s is 60, n is 10, k is 20, and E is 0, -1, or -2 depending upon number of neighbors (as previously stated). In later calculations, the magnetic field B is in dimensionless units, and a value of 1 was simply used. For normal percolation, the same cell swapping is considered; however, the first requirement of not being surrounded by eight same-sign neighbors is removed, and the energy term is set to 0 throughout, effectively making for a free swapping similar to a 2D Brownian motion on a square grid. Thus both the superadiabatic and normal percolation aspects involve swapping of cells via Brownian-type motion; however, in the superadiabatic case, as is discussed next, a like-field attraction is achieved.

Let us outline how these energy changes affect the two types of percolation. The magnetic field motions are computed by using CA elements, wherein they are allowed to undertake a small walk within each model time step. The differences between the two percolation flavors arise from the energies associated with their cell swaps. In case A, when a cell is considered for swapping, the number of same states in the immediate neighborhood is calculated. If a cell has all same-sign neighbors, for case A, the cell is internally situated within a spot, and its position is not changed. Otherwise, a cell swap may occur based upon probabilities for improved placement; this increases the number of like-sign cells, which effectively creates field stickiness. As described earlier, improved placement is obtained by associating a negative energy with the number of same-sign neighbors. The cells in the superadiabatic environment are thus able to generally increase the number of neighbors of same-sign cells, resulting in the emergence of clusters of cells in the same state. This essentially provides for an effective clustering in the case of superadiabatic percolation. For case B, involving normal percolation, all cells are considered for swapping, the number of nearest neighbors

is not considered, there is no enhanced movement or energy associated with increasing like-sign neighbors, and spot cells may also move freely; hence a Brownian-type motion occurs. For both models (cases A and B), the inclusion of a magnetic force associated with an external, Babcock – Leighton, horizontal subsurface field is also added. In the next section the reasons for this force are discussed, but for now, we simply take it as part of our model. This effect is obtained by the addition of a drift added to the percolation motions. Added conditions (*e.g.*, that CA do not occupy the same space at the same time) are also provided.

The results of calculations using this model are discussed next, and an interesting comparison is made with observations. Figure 1 shows six time steps from a series of calculations starting from random EPRs added in the first two time steps. As time passes the random bipoles form larger bipolar-like structures, which gradually coalesce into a single large bipolar magnetic region (BMR) structure. Primarily, one gets BMR structures, but other examples sometimes arise with more complex patterns. The main point is that localized superadiabatic percolating regions do combine numbers of small EPRs into large-scale BMR patterns.

Let us consider why an external field is added to drive motions, rather than the more common use of meridional circulation and diffusion for solar dynamo models. To aid understanding of how fields on the Sun move on a large scale, ask the following question: What allows fields on the Sun to migrate from near equatorial regions to the high latitudes of the poles? The original explanation of Babcock and Leighton (of diffusion driving motions, without meridional transport) seemed questionable, as is discussed when observations of solar fields are shown later. One might also prefer a magnetically driven motion, as opposed to one driven by hydrodynamics, since this will allow opposite-sign fields to drift in opposing directions, which is helpful for dynamo maintenance. Nevertheless, it is recognized that the observations support meridional transport, and it is not ruled out; it is simply not yet employed. Thus I was delighted and a little surprised to see a *Hinode* observation of solar magnetic fields during the birth of a sunspot in the presence of other activity. In addition to supporting the previous contention (paper I) of like-sign fields showing attraction, it also shows behavior that could allow shallow horizontal magnetic forces to play a role in governing field motions. Selected panels from this are shown in Figure 2. One notes that between the two poles of a large region in the left of panel 1, a small bipole starts to form, first appearing in panel 2, but appearing significant in panel 3. A close examination shows that inward oriented field (dark shading) is rather diffuse and it spreads into the surrounding region with no single, identifiable center. The outward field (light shading), however, grows into a spotlike structure. Closer examination of the smaller scale elements may be interpreted as like-sign elements gathering together. The single pole (called a unipole) then moves in the direction of the "like-sign" sunspot in the middle of each panel. In the model presented here, the bipoles grow and the unipole moves in this direction owing to this region being highly superadiabatic. It does seem possible that some purely MHD model (*e.g.*, an emerging field structure, such as an Omega loop) may also be able to explain this field behavior.

To see how the addition of a magnetic force affects the motion and growth of a developing BMR in our model, the external field is chosen to be a horizontal longitudinal field. Any direction for the horizontal field could be employed, as well as any arbitrary spatial numerical field function (*e.g.*, determined from the surrounding field). For illustrative purposes, the simplest nonzero case, a constant longitudinal field is chosen to illustrate the effects of a large-scale field upon the development of EPRs. Random poles were placed during steps 1 and 2, and a constant horizontal subsurface magnetic field is imposed throughout, which applies a magnetic unipole-dependent magnetic force, based on Equation (3.6). The field is considered to be an "external" subsurface field, since this horizontal field is not from CA

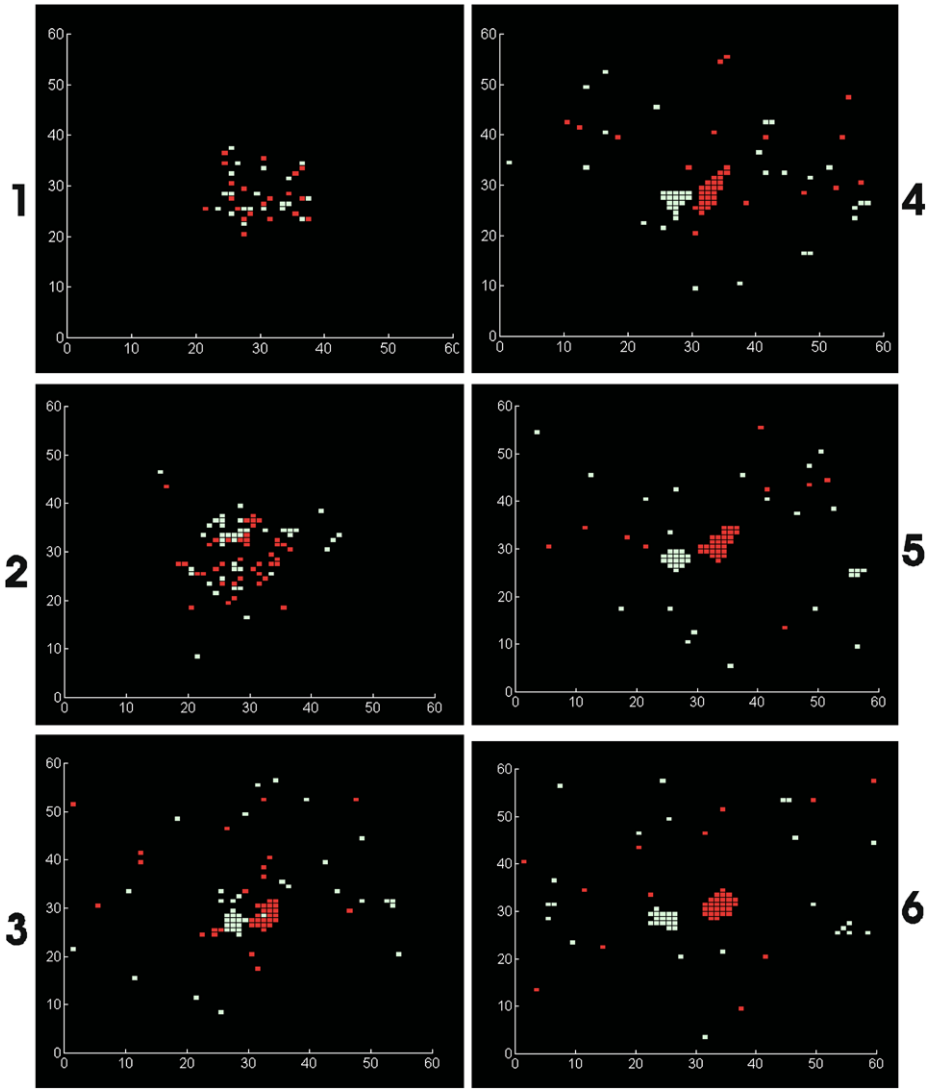

Figure 1 Model of BMR created from a highly superadiabatic percolation of EPRs. It is coincidental that in this figure the main bipolar regions are roughly aligned with the East – West axis. Other cases have been run, with the fields oriented in other orientations. Additionally, there are many cases where complex spot groups result. The tendency of this model is for spot fields to cluster predominantly into bipolar regions. The degree to which this occurs depends upon the parameter n and the size parameter that governs the size within which fields can wander. The quantity δ is 3, s is 60, n is 10, k is 20, and E is as stated in the text (0, -1, or -2 depending upon neighbors). In the numbered panels, time steps are 1, 2, 5, 8, 13, and 18, respectively.

elements. The field from the two external opposite-sign sunspots, outside the growing group in Figure 2, is similar to this external source of subsurface field driving the sunspot motion here.

Figure 3 shows the model in the presence of a horizontal, x-direction, subsurface magnetic field. Because the region has a strong superadiabaticity, the fields are driven toward

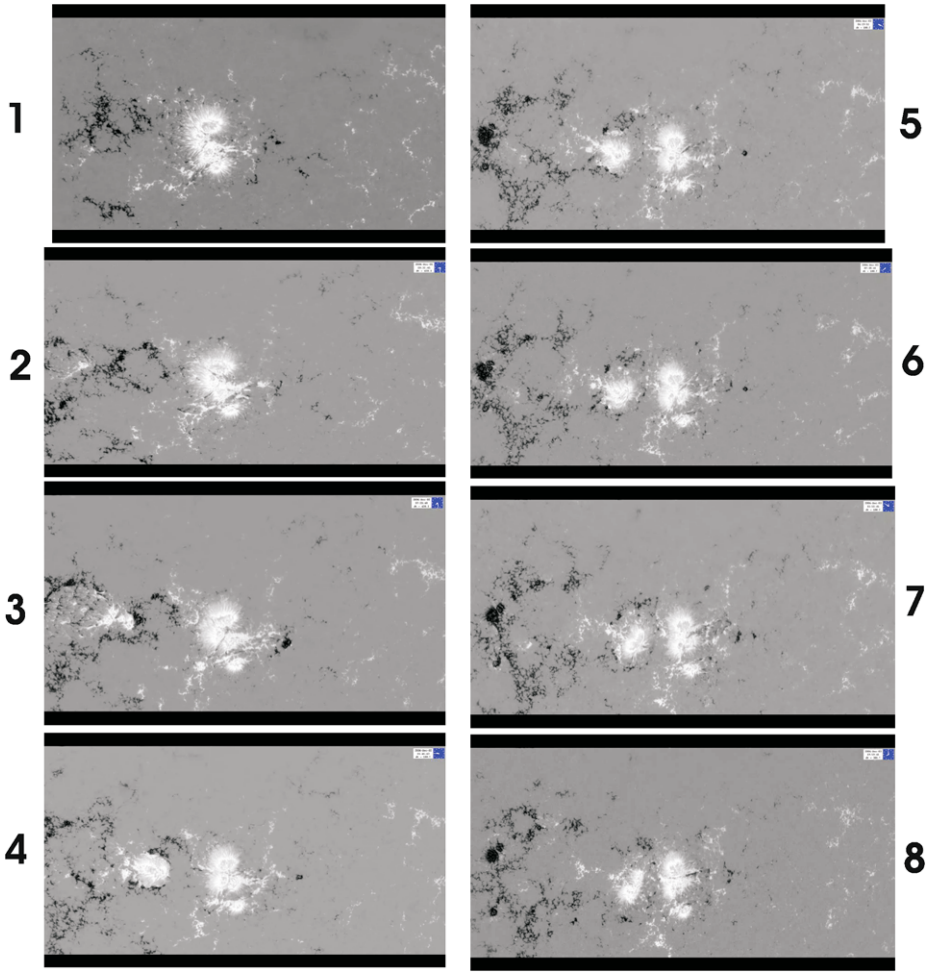

Figure 2 Observations of the Sun's magnetic field taken with the *Hinode* satellite. Original data is at: http://www.nasa.gov/mission_pages/solar-b/trilobite.html. Time frames were selected during the week of 18 September 2007, as the active region crosses the solar disk. These eight panels show the birth and growth of an unusual field structure, known as the trilobite (an ancient organism) for its unusual form. Its earlier development can be seen in panel 2; however, the trilobite really starts to grow appreciably in panels 3 – 8. As it grows, it also moves toward the like-sign spot region located near the center of the figure. The attraction of like-sign flux is one of the hallmarks of superadiabatic percolation discussed in paper I. The time steps are increased in time for these eight panels, as the earlier panels show more rapid development, and at the end, few changes occur. Additionally, there is increasing foreshortening as the regions near the Sun's west limb in panel 8.

same-sign fields, with a behavior similar to that shown in Figure 2. This modeling also shows that CA field structures can result in "field drift" and form BMR structures aligned with subsurface fields. This allows Babcock's toroidal field geometry, originally designed as a shallow field structure, to amalgamate with our CA model to form localized BMRs in accordance with Hale's polarity laws and, further, also to align with the tilts seen in Joy's law.

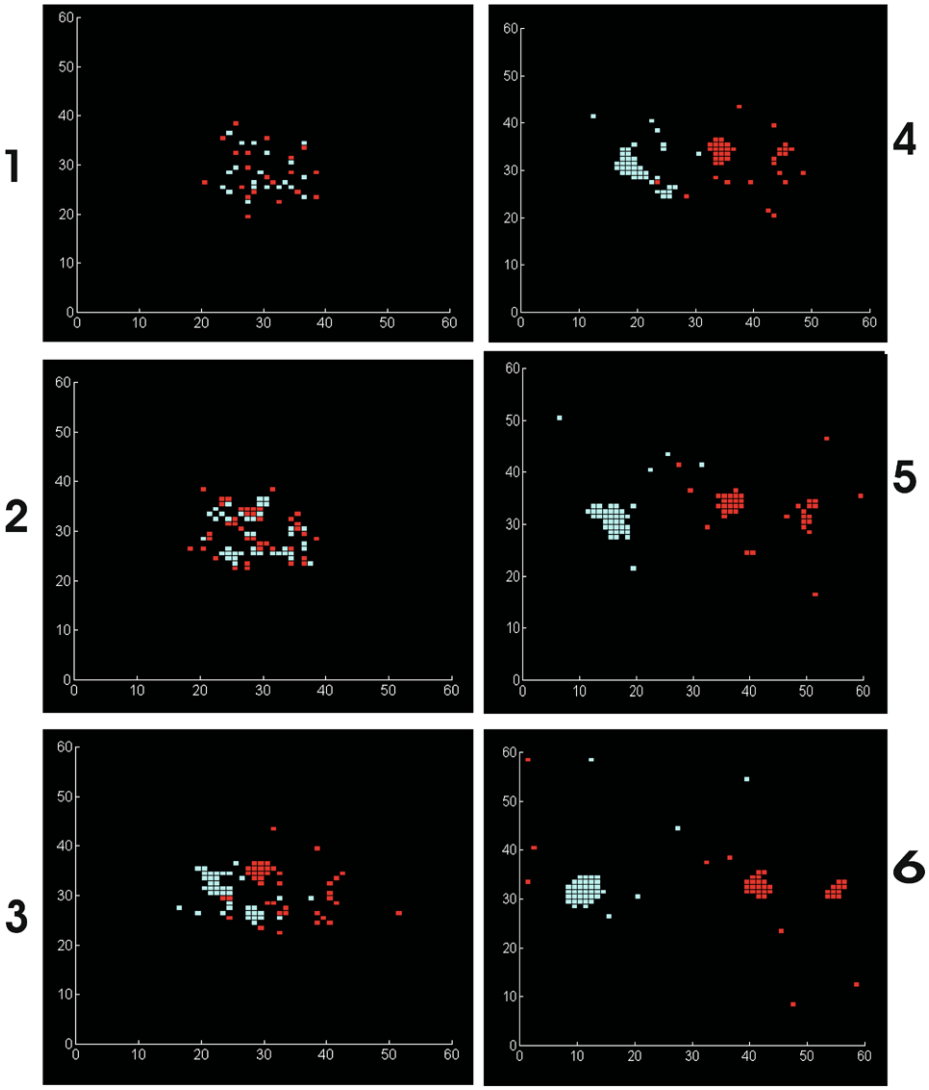

Figure 3 A series of panels showing the temporal evolution of the field. The redistribution of the field starts from a localized region of random dipoles to clustering of the field in the highly superadiabatic percolation model. This process proceeds along with the separation of the field of the evolving BMR as a result of the driving force of a subsurface X-direction external field. The horizontal longitudinal magnetic field guides opposite fields in opposite directions. The fields expand to the left and right as a result. Since the fields may separate in this model, we refer to the two halves of the BMR as unipoles. The calculations were carried out on a toroidal grid, so a few flux elements disappear on the right in panel 5 and appear on the left of panel 6 (an artifact associated with the toroidal grid). The quantity δ is 3, s is 60, n is 10, k is 20, and E is as given in the text (0, -1, or -2 depending upon neighbors). Time steps are 1, 2, 5, 8, 13, and 18, respectively.

Running the small-scale model illustrates how EPRs can lead to sunspots that subsequently disperse into the large-scale magnetic field. This is done by considering the combined effects of both normal and superadiabatic percolation in the presence of an external field. Six views of the temporal development of initial random small dipoles, which then un-

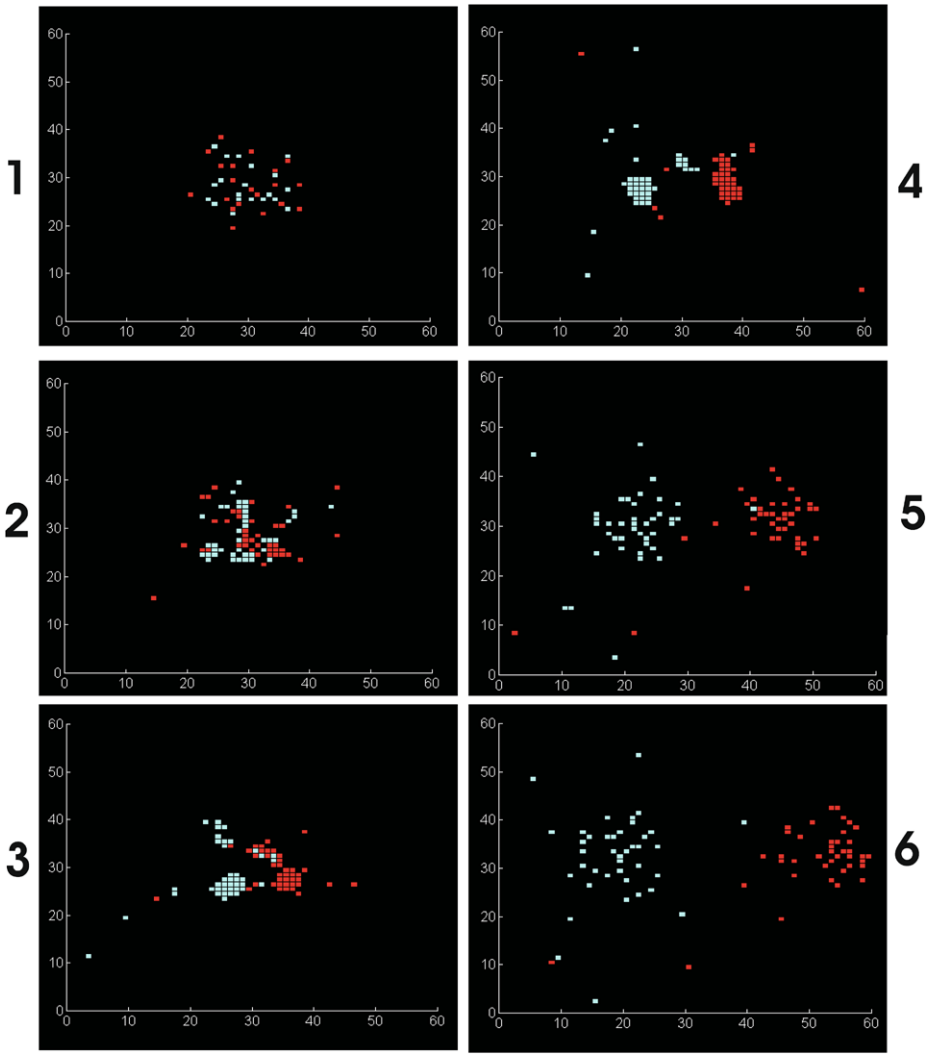

Figure 4 Model of large superadiabatic percolation of EPRs, including field drift into BMRs, followed by dispersal associated with normal percolation. Shown are six views of the temporal development of initial random small dipoles, which then undergo two motions associated first with superadiabatic percolation in an external field. As in Figure 3, a horizontal subsurface longitudinal magnetic field guides opposite fields in opposite directions. The bipolar regions expand to the left and right as a result. The small-scale EPRs/pores gather or clump into spots and then separate until step 4. After this time, the superadiabatic gradient is deemed quenched and returns to normal, owing to the inflow of neutral hydrogen from the surface. This allows the spot fields to disperse into two large-scale field patterns, essentially UMRs. The quantity δ is 3, s is 60, n is 10, k is 20, and E is as in the text (0, -1, or -2). Time steps are 1, 2, 5, 8, 13, and 18, respectively.

dergo two motions associated with percolation and an external force (field drift), are shown in Figure 4. As in Figure 3, a horizontal subsurface longitudinal magnetic field guides opposite fields in opposite directions. The bipolar fields undergoing field drift expand to the left and right as a result. After step 4, the superadiabatic gradient is deemed to diminish owing to the inflow of neutral hydrogen from the surface. In this model, normal percolation is al-

lowed. Note that, although the bipolar fields separate because of the external magnetic force, the individual spots disperse owing to the normal percolation. Namely, the fields of individual poles break up into a series of small magnetic features, which are small unipoles. On the Sun, such features would likely first appear as small bright faculae emanating from the sunspots. The change from dark features (spots) to bright features (faculae) occurs as a consequence of the release of energy (from upflows that disperse field) previously blocked by the inflows of cool hydrogen gas from the near-surface layers surrounding the spot, when the active region was in its more active phase (see Schatten and Mayr, 1985). After sunspots finish emitting faculae, the upflows and facular brightness decrease and the magnetic field then simply resides in small magnetic elements within the calcium plage network. This allows the large-scale field to disperse into two large-scale field patterns, essentially UMRs. More time is allowed to pass between numbered panels so earlier changes are more apparent. The slow field drift is allowed to continue in this model. It is remarkable that starting mainly from random fields, the model is able to generate a large-scale surface field. This likely occurs because the relatively weak subsurface field "breaks the symmetry," which thereby allows large-scale magnetic field to separate and draw energy from the superadiabatic gradient.

This remarkable phenomenon, namely, sunspot creation of strong regions, is compared with remarkably large magnetic field and flux, from relatively weaker regions (EPRs) as a type of transistor. A transistor has three wires going into it, one carrying the signal and two carrying the power. Because of the nonlinearity of the system, the small signal current magnifies the flow of a much larger current. In the same manner, owing to the large amount of power transported through the nonlinear layers of the outer convection zone, we see in Figure 4, a similar amplification process (of magnetic field). This analogy is a useful concept to understand how magnetism can be amplified through flow energy and superadiabaticity. In the next section, this result will be used along with the inclusion of normal percolation following spot development.

Having gained some knowledge of how spots may move in the presence of a subsurface magnetic field, let us now see how this idea relates to the overall solar dynamo. This aspect is for us a key litmus test to ascertain whether or not all these ideas are useful. If not, then the model presented here is just a novel way to consider active region field drift, but its role in the Sun's dynamo would be nonexistent.

3. Large-Scale Solar Dynamo Model

The basis for this model is the modeling work of Schatten *et al.* (1972), Sheeley, DeVore, and Boris (1985), and Schrijver, DeRosa, and Title (2002), wherein observed photospheric features correlated with future observed large-scale field. These models were based upon one of Leighton's (1964) equations (Equation (12)), where he uses the density n of magnetic lines of force with advection and diffusion to calculate field change:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \cdot (\mathbf{B}\mathbf{v}_s) + \kappa \nabla_s^2 \mathbf{B} + \mathbf{S}_s, \tag{3.1}$$

where $\mathbf{B}$ is now the surface magnetic field strength, rather than Leighton's number of field lines, n; the quantity $\mathbf{v}_s$ is the surface velocity field (which can include the differential rotation and possibly meridional flow, which, however, is not used here), the term $\mathbf{B}\mathbf{v}_s$ is tensor notation, κ is the diffusion constant, ∇_s^2 is the spherical Laplacian operator, and $\mathbf{S}$ is a source function describing new solar features (in particular, sunspots) that appear on the solar disk. This term was not in Leighton's Equation (12), as he was not using his model to attempt to directly mimic solar features, as is done here. In the earlier work, these were all observable

parameters. Here they are used as modeled features to ascertain whether surface features can, by advection, propagate by Equation (3.1) forward in time and whether such propagation can lead to dynamo behavior. Schrijver (2001) has also provided simulations of the Sun's dynamo via surface features, with ephemeral regions, differential rotation, and meridional flow patterns. These are more detailed than the dynamo model presented here since illustrating the importance of combining shallow fields (via percolation) is important and thus many elements that Schrijver and others have used to describe the real solar dynamo have been left out. Thus the model here is somewhat simplified and contains less phenomenology. This section outlines the development of this dynamo model from its physical basis to its detailed workings and then on to model results.

3.1. Energy Transport Associated with Magnetic Field

First the physical basis for how surface magnetic fields interact with atmospheric flows below the photosphere will be discussed. Despite the general knowledge of MHD and plasma physics, for increased understanding of field – fluid interactions in the thin, outer solar atmosphere, the solar corona, there has been a lack of incorporation of MHD or meteorological ideas into the dynamics of the deeper solar atmosphere, the convection zone. In particular, a discussion of the physics of energy transport in the Sun's upper convection zone and its relation to the Sun's dynamo is in order. We start with the first obvious signs of the solar dynamo, namely, sunspots and their energetics. As with all known atmospheres of real objects, the actual temperature and its gradients vary in space and time. Thus we consider the temperature of the outer solar layers to vary in space and time, rather than just possessing a mathematically uniform average value as many astronomical stellar models provide. The low density of the photosphere would allow significant variations fairly easily. The density of the photosphere is akin to the density of the Earth's ionosphere (a region notorious for its variations and ephemeral behavior). When the low gas density of the photosphere is combined with consideration of the large energy flux transported through the photosphere, it would be surprising if temperature variations did not ensue. The powering and energetics of sunspots using flow dynamics were first discussed by Parker (1978) in terms of superadiabaticity and later by Schatten and Mayr (1985) in terms of hydrogen ionization. Both views are equivalent, the former being more astronomical and the latter being more meteorological. This sets the stage for discussing the role the radial temperature gradient (superadiabaticity) plays in sunspot formation, growth, and decay by first discussing what superadiabaticity is and then the energetics of sunspot dynamics related to flow and temperature gradients.

The energy brought into a volume of gas of number density N_{T} by transport associated with a steady vertical flow of hydrogen of mass M_{H}, with velocity $\bar{\mathbf{v}}$ and fractional ionization α, in a gravitating atmosphere, $\mathbf{g}$, satisfying $\nabla \cdot N_{\mathrm{T}}\bar{\mathbf{v}} = 0$, under hydrostatic pressure balance, $\nabla \cdot N_{\mathrm{T}}\bar{\mathbf{v}} = N_{\mathrm{T}}M_{\mathrm{H}}\mathbf{g}$, can be written (see Cox and Giuli, 1968, and Mihalas, 1970) as

$$\frac{\partial E}{\partial t} = -N_{\mathrm{T}}\bar{\mathbf{v}} \cdot \left[(1+\alpha)\frac{5}{2}k\nabla T + \left(I + \frac{5}{2}kT \right)\nabla\alpha + M_{\mathrm{H}}\mathbf{g} \right], \tag{3.2}$$

where one neglects the smaller turbulent, $\frac{1}{2}\mathbf{v}^2$, and electron excitation energies, χ_{m}. Although Equation (3.2) shows that energy transport has its leading term proportional to the temperature gradient, because of the work done by gravity in gas compression associated with the last term, this energy flux equation is reduced significantly. It turns out that this reduction leads to the bracketed expression being proportional to only the superadiabatic gradient, S. This makes the energy flux much more sensitive to the temperature gradient (relative to the adiabat), than it would otherwise be. This is because Equation (3.2) provides

that, in a superadiabatic environment ($S < 0$), an upflow ($\bar{\mathbf{v}} > 0$) results in heating, and a downflow results in cooling. In a subadiabatic environment, the reverse is true.

The superadiabatic gradient is the difference in radial temperature gradient from that of a slowly convecting gravitating atmosphere (an adiabatic gradient). For sunspot energetics, the energy transport equation (Cox and Giuli, 1968; Schatten, 1988) in its simplified astronomical form can be written as

$$F_{\text{conv}} = -\rho C_{\text{p}} K S, \tag{3.3}$$

where ρ is the density, K is the eddy thermal diffusivity of order $K \sim \frac{1}{2} v_{\text{t}} \lambda$, where v_{t} is the turbulent viscosity, λ is the mixing length, and S is the superadiabatic gradient, with negative values indicating energy flux outward. The quantity

$$S = \left(\frac{\mathrm{d}T}{\mathrm{d}r}\right) - \left(\frac{\mathrm{d}T}{\mathrm{d}r}\right)_{\text{AD}},$$

where the subscript AD refers to the adiabatic value, is the difference between the temperature gradient and the adiabatic value. Thus when S is negative, energy is transported outward; when S is small or near zero, little energy is transported outward. The quantity S is often expressed in dimensionless log pressure units:

$$\left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right) - \left(\frac{\mathrm{d}\log T}{\mathrm{d}\log P}\right)_{\text{AD}}.$$

In these cases S is usually written as $\nabla - \nabla_{\text{AD}}$, with each term referring to the previously bracketed logarithmic quantities. Expressed in this pressure coordinate system, the adiabatic lapse rate can then be obtained from thermodynamics; that is, $\nabla - \nabla_{\text{AD}} = (\gamma - 1)/\gamma$, where γ is the ratio of specific heats. For example, γ is 5/3 for a monatomic gas without ionization or any other latent energy sources, and $\nabla - \nabla_{\text{AD}} = 0.4$.

The superadiabatic gradient governs the direction of convective energy transport. For planetary atmospheres, meteorologists have needed to advance this theory well beyond astronomical theory, as they do not assume constancy of chemical composition but allow for variations in particular species [*e.g.*, humidity changes associated with varying water vapor (giving rise to wet and dry adiabats)]. They also consider subadiabatic as well as superadiabatic atmospheres. For the Earth, the atmosphere is predominantly subadiabatic, but under rare conditions, superadiabatic, as during the northern hemisphere autumn when cool air overlies warm, low-latitude oceanic basins, allow hurricanes to form. For the Sun the convective flux, F_{conv}, is predominantly outward (positive), so the temperature gradient is, in the mean convection zone, steeper (more negative) than the adiabatic gradient. In the upper layers of the convection zone near the photosphere, $(\frac{\mathrm{d}T}{\mathrm{d}r})_{\text{AD}}$ has a mean value of order $-5\ \mathrm{K\,km^{-1}}$ (see Schatten and Mayr, 1985). At deeper layers, below about 2000 km from the photosphere, $(\frac{\mathrm{d}T}{\mathrm{d}r})_{\text{AD}}$ has a mean value whose magnitude is smaller than $0.1\ \mathrm{K\,km^{-1}}$. In addition to S varying with depth, we postulate that it may vary over the Sun's surface, much as weather systems in the Earth's atmosphere have spatial and temporal temperature variations. Let us see how this relates to sunspots.

Below sunspots the downward advective energy transport can offset the outward convective energy transport (Parker, 1979; Schatten and Mayr, 1985). This picture augments the older Biermann view that the magnetic field simply inhibits convective energy transport into sunspots. In the Schatten and Mayr view, inflow and downflow provide a cooling to sunspots by feeding neutral hydrogen into the sunspot area via a large-scale downflow pattern (see Figure 5, panel B). The neutral gases are funneled and heated by compressive

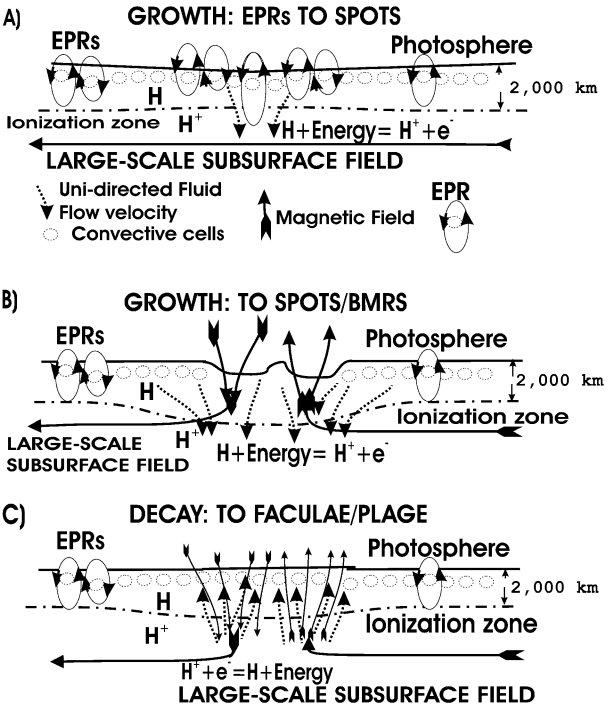

Figure 5 (A) Growth from EPRs to sunspots, (B) to spots and BMRs, and (C) to larger scale faculae and plage. In A, the initial growth of pores to spots begins in a region where the superadiabatic gradient S is larger than average, shown by the thinning region between the photosphere and ionization layers. As a consequence, convective cells allow downflows (dashed lines) to reach the ionization layer. In A, the large-scale subsurface field connects from right to left. The legend shown below A shows velocity, fields, *etc.* In B, the growth toward spots and BMRs is enhanced as unidirected downdrafts (dashed arrows) reach below the ionization layer. The cool downdrafts below active regions can extend deep into the Sun, and these correlate with the magnetism above, without the field extending deeply. The cooling ionization reaction helps develop a cool "finger" of gas, which drives the dynamics. The cool finger becomes sufficient to pull more gas into the vortex below sunspots. In C, the cool finger of gas quenches the temperature gradient, and as a result the flow reverses. This results in an exothermic reaction and the extra energy takes the form of faculae and plage.

forces as gravity pulls the cool, thin photospheric gases deeper; they then become ionized in an endothermic chemical reaction:

$$\mathrm{H} + \text{energy} \rightarrow \mathrm{H}^{+} + e^{-}. \tag{3.4}$$

This can be viewed as a thermal buffer reaction, to offset the compressive heating and thereby quench the outward convective energy flux, as well as to keep the descending gases cool, as they make their way into the nether regions deep within the Sun. Although the gases descending below sunspots heat by compression, they remain cooler and therefore denser than their new surroundings; the associated pressure gradient then drives the downflow dynamics. The chemical reaction of Equation (3.4) is highly reminiscent of the latent energy reaction of water vapor to liquid water that helps drive terrestrial hurricanes, a similarity that led to my calling sunspots "inverse ion hurricanes." Because of the low density of the gases near the photosphere, magnetic flux from a broad area may be swept into any particu-

lar downflow, thereby allowing a rapid accumulation of EPRs from a vast surface area into activity centers.

Before reaching the pattern shown in Figure 5B, where sunspots have reached full maturity, a discussion of how sunspots first form from small EPRs is needed. The sunspot pattern seen in Figure 5B first develops from the top panel (A) as ephemeral regions are drawn together merging their downflow patterns as the flows reach the ionization zone, shown by the dashed-dotted line. Reaching this zone means the downflow gases can cool themselves through ionization and thereby offset gravitational compressive heating, allowing the inflow and downflow dynamics to strengthen. Without this cooling, compressive heating would quench the downflow patterns, resulting in small-scale overturning bubbles, as in standard mixing-length theory. The directed pattern of an inflow converting into a cool downflow pattern would tend to start in regions of the photosphere where there is a preferentially thinning of the isothermal structures in the outer layers of the Sun, allowing enhanced S values. These are the solar equivalent of low-pressure regions in the Earth's atmosphere, allowing surrounding gases to be drawn into a vortex of flow. A horizontal magnetic field may play a catalytic role by allowing some stratification of the flow pattern, away from the normal granulation pattern.

One of the main differences between the model presented here and conventional dynamo models is that the magnetic fields here are considered to originate in the shallow layers below the photosphere, whereas most other models consider their birthplace to be deep within the Sun's convection zone. Although some reasons are provided in the discussion section of this paper for why the Sun's magnetic field collects in the thin layers below the photosphere, nevertheless, perhaps superior to any theoretical justification is observational support, which often trumps great theories. One of America's most famous solar observers, George Ellery Hale (see Hale *et al.*, 1918) describes the sunspot development process with these words: "[P]hotographs (show) masses of hydrogen in the act of being drawn from a great distance toward the center of sun-spots, as though sucked into a vortex. These photographs suggested the hypothesis that a sun-spot is a vortex, in which electrified particles, produced by ionization in the solar atmosphere, are whirled at high velocity." Horizontal subsurface magnetic fields, of the type envisioned by Babcock and Leighton, may guide vast amounts of low-density, weakly ionized gases to gather EPR fields toward centers of downflow patterns. The magnetic fields are simply sucked into the vortices, and remain as spot fields, until the downdraft is sated toward the end of an active region's lifetime.

When the small-scale EPR and spot fields connect to the buried subsurface field, as shown in Figure 5B, subsurface tensions can then draw the BMRs apart, allowing the fields of the individual spot flux tubes to separate and freely form faculae and plage, shown in Figure 5C. In the absence of this field separation, no major restructuring of the large-scale solar field could occur. Faculae and plage are the remaining residual field structures that memorialize the now fallow surfaces, where once stood proud dark sunspots. With terrestrial hurricanes drawing in volumes of moist air, for equilibrium to return, the moisture and gas must descend back to the surface of the Earth. The moisture does so in the form of rain, which may be widely scattered from the oceanic source of the hurricane's moisture. In a similar fashion, for the Sun, the gas densities, temperatures, and ionization fraction, on average, would tend to return to their average values. Thus neutral hydrogen would, on average, be required to return to the surface of the Sun. It does so by return upflows of ionized hydrogen, which recombine, thereby carrying the sunspots' blocked energy flux (beyond the normal surface irradiance, F_0).

Let us consider what happens to the field and energy flux as a sunspot region ages. As the dynamical transport of cool neutral hydrogen flows into the central area of the sunspot

region, it begins to replace the earlier hotter gases that initiated the spot formation. The superadiabatic gradient is thus eventually quenched, and the dynamical motions cease and are replaced with more normal flow patterns of smaller scale turbulence. Overall then, superadiabaticity leads to increasing growth, which in a closed system is not indefinitely sustainable. Consequently, a collapse occurs, resulting in a breakup of the originally enhanced temperature structure. This leads to a return to the normal temperature structure. As a result, the energy supplying the sunspot's enhanced electrical currents supporting its magnetic field is reduced. The field lines as well as the heated gases below the sunspot region, however, are still intimately connected to the surface, with the magnetic flux conserved. Consequently, both energy and field lines emanate into numerous diffuse regions surrounding the original group, as it ages. These field and energy concentrations mark what are commonly called faculae or plage, when seen in either white light or spectral lines, respectively. Our computer model allows for this type of behavior, as we will see, by undergoing "normal percolation," wherein photospheric fields move on the surface in a Brownian-type motion, conserving flux but randomizing concentrated patterns. This acts, as previously discussed, in a dispersive manner, so that older field structures no longer have the concentrated energy from strong downflows, and the concentrated fields become larger scale weakened field structures of faculae and plage. The association of gaseous upflows with extra luminous energy is in accordance with Equation (3.2).

Overall, the large-scale pattern of flow from an inward directed surface plunging of cold gas, drawing in cool neutral flows, and the later return of heat as the ionized hot gases rise to the surface serves to transport more heat outward than would otherwise be transported by a steady-state, small-scale granulation. This aspect is given testimony by the net enhancement of the Sun's irradiance associated with active regions (see Lean *et al.*, 1998). The increase in outward energy transport associated with solar activity may be understood in the current model, as a process that essentially "stirs the pot" of hot solar gases in the Sun's upper convection zone, thereby tending to aid energy transport. For the magnetic field, the entire process is a gradual transformation of small-scale random bipolar fields of EPRs to larger scale organized magnetic structures. The connection of this gradual transformation of small magnetic features to the organized large-scale solar dynamo will now be discussed.

Figure 6A shows the field structure in the early phase of a cycle in the Babcock (1961) and Leighton (1969) dynamo models. Poloidal field is stretched into giant toroidal flux ropes, which erupt from the interior into bipolar magnetic regions, where the p and f letters indicate the preceding and following fluxes. These fields are postulated to move equatorward and poleward, respectively, through diffusive processes. The model presented here, shown in Figure 6B, also uses Babcock's subsurface toroidal field geometry, for the first half of an odd-numbered cycle, during which time the magnetic fields are outwardly oriented from the North Pole, as shown, with BMR field geometry oriented as shown. Rather than erupting, however, EPR fields gather together into BMRs, along Babcock flux ropes. The large-scale motion of these fields is subsequently governed by tensions on these flux elements. The 11-year period of the solar dynamo in this model may be estimated, as we will do shortly, by considering the motion of magnetic elements in this model. When the fluxes from the two halves of a BMR exit the Sun through the photosphere, as shown in the northern group, the subsurface field tensions can guide the P and F group toward opposite poles, along field lines, but neither goes specifically to the equator. Rather, fields are drawn toward both poles dependent upon the subsurface magnetic tensions. When conditions are otherwise (*e.g.*, one group containing an excess of flux, as shown in the southern hemispheric group), the entire group can be pulled toward one pole or the other, dependent upon the sign of the majority flux. The reason for this difference is that, for an unbalanced group, subsurface field lines

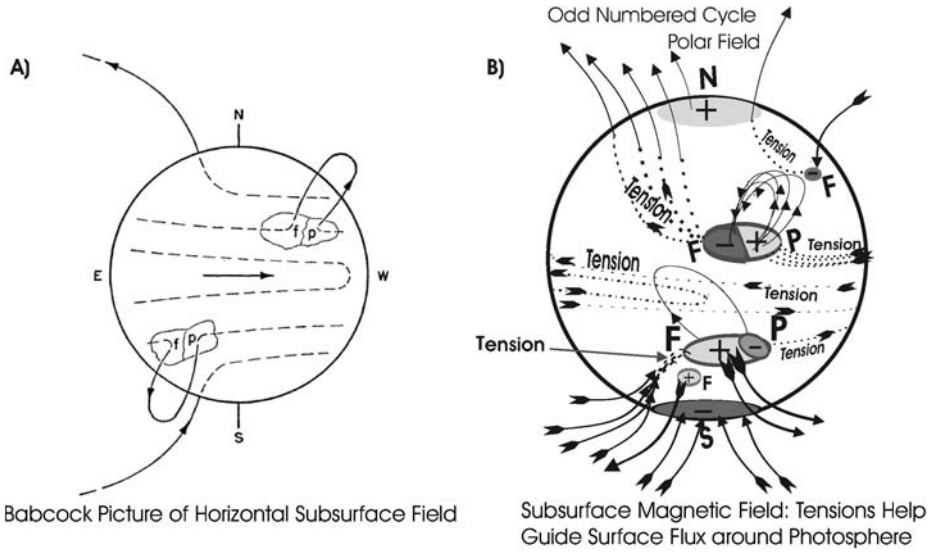

Figure 6 (A) Babcock's view of subsurface magnetic fields. The dashed line shows the subsurface field, with p and f distinguishing the preceding and following BMR polarities. (B) The current view for the early to middle phase of an odd-numbered solar cycle (during which times the magnetic fields are outwardly oriented from the North Pole during the first half of such a cycle). Light shading and a + sign indicates outward pointing magnetic field; dark shading and a − sign indicates inward pointing magnetic field. In this model magnetic tensions guide the surface flux around the photosphere. In the upper right region a small F, −, region is guided toward the N pole. The large northern hemispheric group contains equal + and − fluxes. When most of the field, as from this group, exits the photosphere, it becomes herniated. Consequently, the tensions below the spot group can no longer hold it together, and the two halves separate. The F flux moves toward the same hemisphere pole, and the P flux to the opposite hemisphere pole. This is shown in greater detail in Figure 7.

connect the two halves and may serve to glue it together. Thus both sign fluxes can, at times, be pulled toward the same pole, but the majority of flux motions are such that a given sign flux moves toward the opposite sign polar flux. (Note that these motions are associated with field attraction in the presence of normal temperature gradients.)

3.2. Global Cellular Automata 2D Dynamo Model

The foundation for the large-scale dynamo field 2D model described here comes from the works of Schatten *et al.* (1972), Sheeley, DeVore, and Boris (1985), and Schrijver, DeRosa, and Title (2002) combined with our current understanding of percolation. These earlier works were not dynamos; they simply used observed photospheric features and evolved them into future large-scale structures. The most recent work of Schrijver, DeRosa, and Title (2002) was more sophisticated than the earlier models and matched many solar cycle features. The model included flux emergence, random walk dispersal, meridional advection, differential rotation, and flux cancellation. The model presented here makes use of similar processes (field birth, death, and motions). Field motions include normal percolation, wherein spots disperse and are thereby converted into the weaker large-scale field as faculae/plage, and differential rotation (and other possible flows), which carries all field elements with it. The Schrijver *et al.* model simulates many aspects of the observed long-term solar cycle with only a few inherent solar parameters as input. Although their model is developed

using deep dynamo visualizations, nevertheless, there is nothing inherent in their methodology that is explicitly deep, and their input observations are totally either photospheric or use proxy data (*e.g.*, cosmic rays). The point is that despite visualizations incorporating the deep dynamo, there seems to be no evidence of solar magnetism significantly below the hydrogen ionization zone. There is evidence for flows going into the deep interior, with signatures of the surface activity leading to thermal increases below active regions, in accord with the butterfly diagram; however, these may well be associated with descending columns of gas below active regions, becoming heated at depth by compression (see González Hernández *et al.*, 2008, and the related work of coauthors Howe and Kommm and other references included) as a solar equivalent to a Santa Ana wind.

Before the details of the model are presented, note that the main result in using the dynamo model presented here is the following: As soon as a BMR finishes developing, the cool downflow quenches the original thermal gradient; hence the superadiabatic gradient is relieved, and the two halves of the BMR separate, emitting unipoles from the activity center into the surrounding area. Perhaps an added justification for this is the following: Although many bipoles on the actual Sun may not separate, the primary concern here is those that do since the bipoles that do not separate will not contribute to large-scale field changes for the simple reason that these fields will, by definition, remain linked. The fields that separate become the unipole field structures under consideration, and these are free to move about the Sun's surface like ships under power in a vast sea. We shall consider that they are driven by magnetic forces and fluid motions (*e.g.*, the differential rotation, meridional flows, *etc.*). Our dynamo model employs a larger scale coarse grid (than that of the small-scale model) where sunspots and unipolar regions are the primary field entities; however, CA are employed in a manner consistent with the earlier model.

Let us now consider how the magnetic forces are derived. We provide mathematical manipulation with one of Maxwell's equations, as it allows us to employ a method to calculate forces on photospheric fields. Although we shall be talking about photospheric fields and subsurface fields, it is best to consider both fields as if applied to the same spherical shell, just below the photosphere. Since we are dealing with a shallow dynamo, choosing a single surface is not problematic but just a convenience to distinguish the radial field, which is observable above the photosphere from the horizontal field, which we call the subsurface field, since it is not generally directly observable, as far as we know, because it is buried below the photosphere. Since we really do not know the geometry, or likely even the topology, of subsurface fields, our approach is just one of many that may be attempted. Let us consider the Maxwell divergence-free equation $\nabla \cdot \mathbf{B} = 0$, with the field split into two components: $\mathbf{B} = \mathbf{B}_{\mathrm{R}} + \mathbf{B}_{\mathrm{SS}}$, using a radial component, $\mathbf{B}_{\mathrm{R}}$, and a horizontal (subsurface, SS) component, $\mathbf{B}_{\mathrm{SS}}$. As a mathematical convenience (namely, by not modifying or negating the divergence-free Maxwell equation, just breaking it up into two forms, so that the summation of the separate components has zero divergence, but individually they do not), we consider that the divergence of each of these field components emanates from fictitious monopoles similar to the electric field divergence (see Jackson, 1962), as follows:

$$\nabla \cdot \mathbf{B}_{\mathrm{SS}} = m_{\mathrm{SS}} = -m_{\mathrm{R}} = -\nabla \cdot \mathbf{B}_{\mathrm{R}}, \tag{3.5}$$

where we now refer to the monopoles as "unipoles," to remove any question of their being real. So far this is nothing more than a mathematical convenience. The subsurface unipole, m_{SS}, is balanced by the opposite sign radial unipole, m_{R}, as shown by Equation (3.5). We recognize that the photospheric field does not emanate radially nor is the subsurface field purely tangential; however, this division allows us to name these two separate components in

a convenient fashion, and of course the actual value of this mathematical trick will depend upon its usefulness, as we model these fields. We shall be considering the forces on the subsurface unipoles, m_{SS}. The force on a magnetic charge depends upon the subsurface field; this is calculated by using the force equation, but this force arises solely from the magnetic tension of the field lines (where we neglect the field tension for any field line emanating through the photosphere, since the field strength weakens dramatically as the external pressure drops and the field expands):

$$\mathbf{F}_{BL} = m_{SS}\mathbf{B}_{SS}. \tag{3.6}$$

As this force is dependent upon the Babcock – Leighton subsurface field geometry, we refer to this magnetic force, $\mathbf{F}_{BL}$, as the Babcock – Leighton force. Using this equation one can obtain a period consistent with the 11-year solar cycle. If we equate Equation (3.6) with the viscous force on a length L of magnetic rope, the unipoles move with a velocity v_{SS}; consequently, the time scale for a solar cycle may be estimated from

$$\tau = L/v_{SS} = \frac{\kappa\rho L^2}{m_{SS}\mathbf{B}_{SS}}, \tag{3.7}$$

where κ is the turbulent viscosity, ρ is the gas density, m_{SS} is the unipole flux, and $\mathbf{B}_{SS}$ is the subsurface field strength of the magnetic flux ropes. Parnell (2002) estimates the average emergence rate for the magnetic carpet to be between 6×10^{-2} and 10^{-5} Mx cm^{-2} s and EPR fields to be of strength $10^{16} - 10^{21}$ Mx. In cgs units, by using a density of 10^{-6} g cm^{-3}, a viscosity of 10^{13} g cm^{-1} s, a distance scale of 10^{11} cm, a flux rope strength of 100 G, and a flux unipole strength of $10^{18.5}$ Mx, the solar cycle period, τ, is ~ 10 years. The viscosity is obtained from granulation parameters (a size of 1000 km and a turbulent velocity of 1 km s^{-1}) and a photospheric density. A wide range of values for these parameters is possible, and a 10- or 11-year period is, at best, consistent. Additionally, the variation in parameters is so wide that, from this viewpoint, it is not clear why the solar cycle period is so stable. This cycle length stability may result from the large number of flux elements on the Sun. We also see in Equation (3.7) the inverse relationship between solar cycle length and field magnitude that has long been known to exist (Waldmeier, 1961).

The effect of individual active regions on the dynamo may be likened to the effect of a child pumping or kicking his or her legs on a swing to energize the motion. In this analogy, the child's kicks provide little bits of extra energy that add to the large-scale normal mode of oscillation. Similarly, the growth of individual sunspot groups, in the present scenario, provides extra flux that drives the dynamo oscillation. Their growth, as envisioned, augments on average the growth of a new cycle since the field is added so as to encourage this large-scale oscillation. Figures 7A – 7E show in greater detail the various aspects and stages of odd-numbered (*e.g.*, 19, 21, 23, *etc.*) solar cycles that illustrate the various field components and relation to these unipole field elements. Figure 7A shows an early stage where the field is poloidal. By using our description, one sees that the surface is also peppered with small ephemeral regions; all panels should convey this, but for convenience, the EPRs are not drawn in other panels. Figures 7B – 7D display the photospheric and subsurface unipoles more clearly. The radial unipoles allow flux to emanate from the photosphere into the corona, and the subsurface unipoles draw the flux from subsurface flux ropes, transferring it into the photosphere.

For odd-numbered cycles (*e.g.*, 19, 21, 23, *etc.*), as shown in Figure 7D, the following, F, spots have + unipoles in the subsurface region. The fields there are attached to those from the North Pole of the Sun; hence they are drawn along Babcock's field lines in the direction

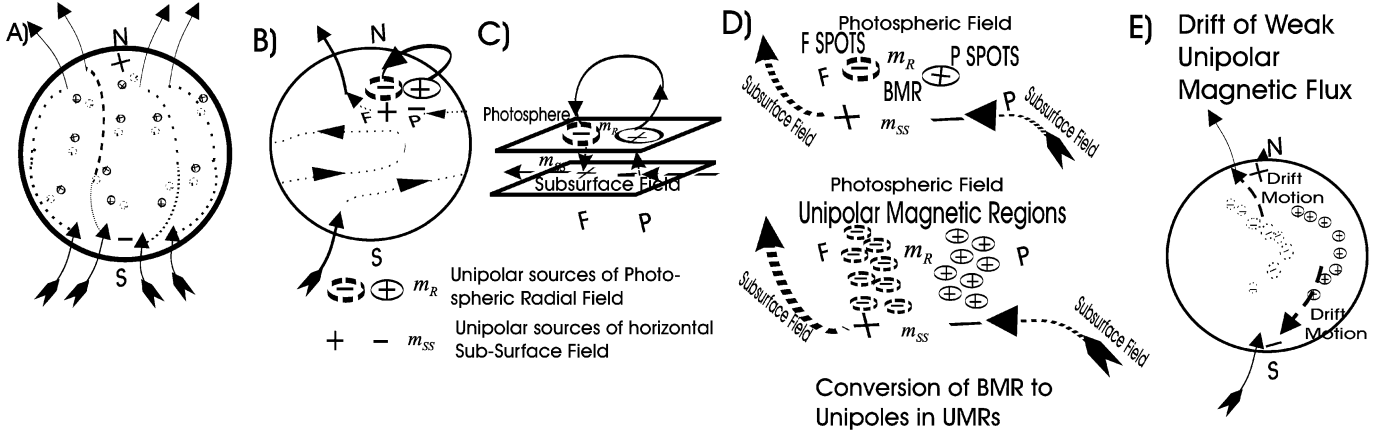

Figure 7 The development of fields during the early phase of odd-numbered solar cycles. Throughout the cycle the photosphere is filled with small EPRs, shown only in A. As in Babcock's model, the poloidal subsurface field (dashed lines) goes from south to north in the outer layers of the Sun. In B, the subsurface field is twisted into a toroidal form. In our model small EPRs coalesce and reconnect with elements of the subsurface field. We draw sign fields (values of + and −) with surrounding ovals to indicate that the unipoles (field sources) are at the photospheric level. Subsurface sources of flux are displayed without the surrounding ovals. Following, F, field regions in the northern hemisphere show inward directed field. (B)–(D) The F regions in the northern hemisphere, having + sign poles in the subsurface layer, are pulled directly along Babcock's subsurface field lines, westward and northward. The northern hemispheric P poles, having − sign poles in the subsurface layer, indicate that the force is opposite to the field direction, pulling structures eastward and southward. (D) The spot fields break up into small unipoles, individual field elements that become part of UMRs once the fields spread out beyond the activity center. Panel E shows how the + and − sign fields subsequently drift toward opposite poles as a consequence of magnetic forces upon the unipoles, serving to reverse the dynamo.

toward the Sun's North Pole. The preceding, P, northern hemisphere field structures are drawn toward the South Pole. Southern hemispheric fields behave in a similar manner. These rules show motions consistent with the Babcock – Leighton mechanism, but being dependent upon field forces rather than diffusion, the motions are more direct. In the BL model, motions start with Joy's law and depend upon weak diffusive motions to transport magnetic flux. Note that subsurface fields are shown herniated below the active region. Any ballooning of field through the photosphere into the larger volume of the corona reduces field tension and hence cohesiveness between the two halves of active regions, thereby freeing the two halves to disconnect and move separately. Without this separation, BMRs would not be able to transport significant flux poleward. When the flux is not totally disconnected below the photosphere, the flux motions may be more complex than in our simplified model. As a consequence, much room for improvements in our simplified model exists; for example, it does not include meridional circulation, an important flow component. Additionally, only the field geometry is shown and only the forces associated with field during the first half of an odd-numbered solar cycle are discussed. A brief discussion dealing with the second half of solar cycles is given in the concluding section.

The bottom half of Figure 7D shows that BMRs break apart into larger structure unipolar regions. This allows the subsurface magnetic forces to induce movement of magnetic elements toward the poles. This, combined with the differential rotation, stretches field structures into "backward C structures," called UMRs by Bumba and Howard (1965), as shown in Figure 7E. Sheeley, Nash, and Wang (1987) also studied these field patterns from a theoretical viewpoint.

In our model, these large-scale UMR fields preferentially migrate toward fields of the opposite sign (the sign to which they are normally attracted). Since their numbers usually exceed the polar magnetic flux, they can reverse the Sun's polar fields. At the end of an ~ 11-year time period, the structure shown in Figure 7A repeats itself with reversed field and the next cycle can begin anew with reversed polarity. None of the processes shown in Figure 7 is as clear-cut as drawn; there are mixed fields of intertwined polarity throughout the cycle, with active regions in various stages of individual evolution, and most fields coalesce with opposite-sign field and cancel before reaching the opposite pole. Our model incorporates some general field processes that seem to be key elements in our view of the solar cycle. Our large-scale modeling effort will now use the ideas discussed here as the basis for its development.

The global modeling effort will simulate the following physical processes affecting the photospheric field: birth, death, and field motions. Birth in our small-scale model occurs in areas of large superadiabaticity from the accumulation of EPRs in the lower latitudes into sunspots; we simulate this process here by generating randomly oriented sunspots in a probabilistic manner. On the time scales of our model, birth is quickly followed by a percolation motion, wherein spots disperse into the weaker large-scale field as faculae/plage. Death, or field cancellation, also occurs, in a probabilistic fashion, based upon the flux density of $+$ and $-$ fluxes in an area. Field motions are obtained from our small-scale CA model. We use percolation, differential rotation, and surface field motions guided by the local horizontal subsurface field.

This leads us to a more detailed description of the CA dynamo model presented here. Pairs of sunspots are placed in a broad rectangular grid. The grid is chosen to resemble the whole Sun, from $-90°$ to $+90°$ latitude, and from $0°$ to $360°$ longitude. Sunspots are placed probabilistically, but their placement depends upon the poloidal field, within "active latitudes" in the range from $\pm 5°$ to $\pm 40°$ latitude zones for $\sim \frac{1}{2}$ of each period of a solar cycle. They are allowed to undergo the following processes, occurring with probabilistic dependencies: birth, movement, conversion to weaker large-scale field, and death.

Sunspot birth, as mentioned, is field strength dependent and stochastic. To avoid any long-term secular growth or oscillations from occurring, a low-frequency bandpass filter, similar to Leighton's (1969) term, $e^{-t/\tau}$, is included. To avoid too much growth or decay, a "floor" and a "ceiling" are also included. These are values to restrict the amplitudes the model can attain. This is done for two practical reasons: *i*) Nonlinear physical phenomena always have some limit to their strength, and *ii*) observationally, a floor in the interplanetary field has been reported by Svalgaard and Cliver (2007). If a floor were not used, the duration of our broad activity minima periods would lengthen.

For simplicity, new sunspots have been placed with a uniform probability in the sunspot birth region (without the declining latitude with time – the butterfly diagram), and without any tilt of active regions (Joy's law). Even though these features are well-known properties of the solar dynamo, the model presented here has been simplified to ascertain the minimal attributes needed to obtain some large scale dynamo process. This helps us distinguish the minimal attributes from the nonessential features that are simply observed to occur. Joy's law, which is the average tilt of sunspot regions away from parallels of longitude, is considered to be nonessential in the model presented here, as the magnetic force used here far surpasses this effect. Probably, Joy's law arises from the reorientation of EPRs into spots as they gather (percolate) along the Babcock – Leighton subsurface field, which is tilted in the same direction as active regions. There is a coupling between the fact that both active region tilts and their latitudes are strong functions of time. This means that latitude is not an independent variable, as it is usually taken to be, when tilt is expressed (in Joy's law) as a function of latitude.

Our model places sunspots in the birth zones of sunspot fields, and then moves the fields in accord with the model previously discussed (with the spot fields percolating into larger scale, weaker fluxes). Initial sunspot placements are proportional to poloidal field strength, in accord with the Babcock – Leighton dynamo, and the view that the subsurface toroidal field (from the poloidal field) plays a role in gathering and orienting sunspot groups (by their poleward motion). General migration of all fields of one sign, in one particular direction, is a simplification, ignoring meridional flow, *etc.* It is hoped that percolation, in the absence of other associated solar effects upon the Sun's large-scale fields, can sufficiently buttress the computed field changes.

Figure 8 illustrates the primary workings of this model. For illustrative purposes, extra large sunspots are placed on a grid representing the solar surface. Here, an initial uniform poloidal subsurface magnetic field, based upon a positive field emanating from the North Pole, is placed as in an odd-numbered solar cycle (*e.g.*, 19, 21, 23, *etc.*). This polar field is shown in panel 1, and although present throughout this illustration, it is only shown in panel 1. The sunspots then undergo normal percolation and differential rotation and experience the forces associated with the poloidal field. As can be seen, the reversed spot fields migrate toward opposite poles. Additionally, there is a death process of magnetic field, based on probability with a functional dependence on the density of the two signed fields (in each region). This is incorporated so that not all fluxes reach the opposite pole. Model parameters are chosen so that, on average, births are canceled by deaths. As mentioned earlier, Leighton's techniques are also used to accomplish this. This condition (births and deaths canceling), however, does not exist for each and every cycle, since, as we will see, secular changes do occur.

One interesting feature, not seen in the Babcock – Leighton dynamo, is that weak fields cross the equator unimpeded and do not necessarily cancel there, nor are they pulled toward it. There seems to be nothing special about the Sun's equator that drives field toward it nor impedes the field from further motion toward the opposite pole, once they reach this mathematical line. Our utilization of magnetic forces thus seems more physical than the traditional

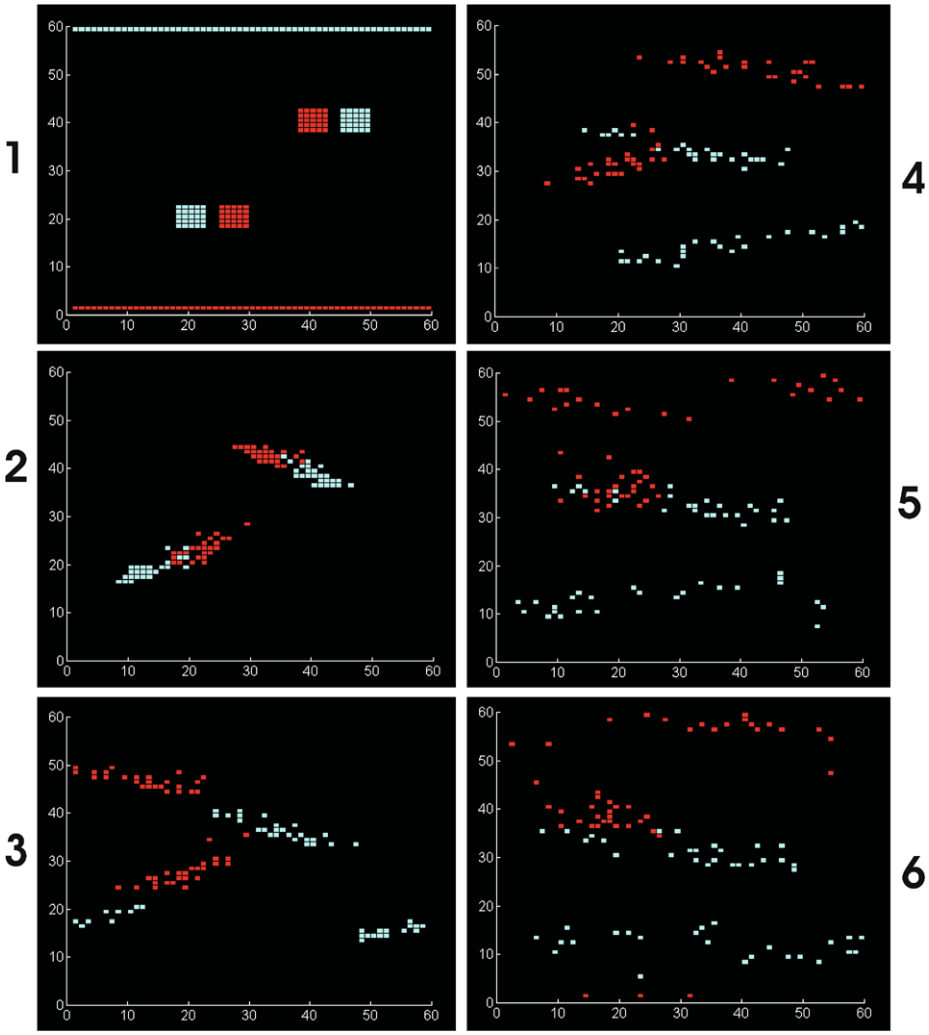

Figure 8 The evolution of large sunspot-type field patterns. These patterns were chosen to illustrate the associated motions of field from smaller scale sunspots in accordance with the current large-scale field model. The model contains the following processes: normal percolation, drift from subsurface dipole fields, and associated differential rotation affecting all the structures. In panel 1, we have placed two pairs of extra large sunspot field patterns, so their evolution may be seen, as well as viewing the polar fields, which exist throughout the high-latitude zones. For small-scale motions, the quantity δ is 3, s is 60, n is 10, k is 20, and E is as given in text (0, -1, or -2 depending upon neighbors). Time steps are 1, 2, 5, 8, 13, and 18, respectively.

description of preceding field motions "toward the equator." Further, this aspect seems to be supported by observations of the Sun's large-scale field, which is examined next. Subsurface magnetic forces have been made an essential feature of the model here, as opposed to drifts from flows, which here are limited to differential rotation and the random motions of percolation. In the model presented here, meridional motions occur primarily through the magnetic force and a small amount from percolation, but none come from meridional fluid

flows directly. Now let us turn our attention to the behavior of the Sun's large-scale field and look at the model results.

3.3. Solar Cycle Observations and Large-Scale Model Results

Shown in Figure 9 (supplied by Leif Svalgaard) is a solar cycle's treasure of magnetic field observations. It uses a novel coordinate system described in Ulrich and Boyden (2006), called a supersynoptic map. These maps allow one not to have extraneous vertical cuts, which normal Carrington synoptic maps traditionally have. The vertical cuts are removed by having time, within each map, run backward. For orientation purposes, the supersynoptic map shows the entire cycle 22 at the bottom. Examining the enlarged period in 1994 shown in Figure 9C, we see a key feature is that large-scale magnetic field does not appear to diffuse, but rather it moves rapidly and linearly with time, as if driven by advection. Diffusion is a slower process, with features spreading outward as $\sqrt{t}$, whereas we see + and − features moving linearly with time toward both poles. The patterns are rather sharp, rather than the accustomed gradual changes associated with diffusion. Figure 9C also shows that a number of field patterns move across the equator, attracted to the opposite pole, as discussed in our model, quite different from the standard Babcock – Leighton dynamo picture. Additionally, we also see that there are surges of both field signs to any given pole, indicating that the attraction of only one field sign to a particular pole at a given time is more simplified than the Sun's behavior warrants. It is as if some flux from both the preceding and following portions of a BMR may sometimes both be propelled toward a pole. This can occur from meridional flows (which have not been included in the model here) or from the two halves of some BMRs not separating. This can occur, as shown in Figure 6, in the southern hemispheric BMR, when there is an imbalance of active region flux, so a small amount of preceding flux may be carried along with the following flux. The sunspots, which do not last very long, are not carried poleward themselves, just the faculae/plage field remnants. The field structures are more easily displayed in the groupings drawn.

Large-scale magnetic fields on the Sun may not likely be driven solely by a single given force. Various mechanisms seem to be at play so that the fields seem to run a gauntlet, where flux can cancel, before most flux elements can reach either geometric pole and then serve as seeds for the next cycle's activity. Observations support that only a small fraction of solar fields reach the poles during a cycle. This is generally known and is supported by the following estimates. From early 1972 to mid 2002, NOAA has labeled 10 000 active regions. Hence there are roughly 3000 active regions in a cycle, with an average of about 10^{21} Mx of + and − flux, making a total flux of one sign of $\sim 3 \times 10^{24}$ Mx. With each polar region containing $\sim 2 \times 10^{22}$ Mx, this means that only $\sim 1\% - 2\%$ of the bipolar group flux can be used to erase and form new polar flux. Thus, without significant cancellation of active region flux before reaching a pole, too much field, by a factor of ~ 100, would converge toward the poles. Thus our model contains significant chance of flux cancellation before reaching the poles.

The motions of large-scale fields on the Sun are not likely governed solely by the magnetic force we focus on, but also on hydrodynamic forces, such as those associated with meridional, turbulent, and other flows. Thus, although field motions toward the poles seem relentless, like the tides, turbulent and other stochastic effects may play significant roles in their motion. All the effects of fluid motion are swept into our percolation term, which gives rise to stochastic-type effects. The cases we see in Figure 9, where both polarities move toward a particular pole, provide examples that are not directly incorporated within our model. In other words, our model oversimplifies field motions by only allowing these types of motions as statistical events. The addition of meridional flow would directly allow such field

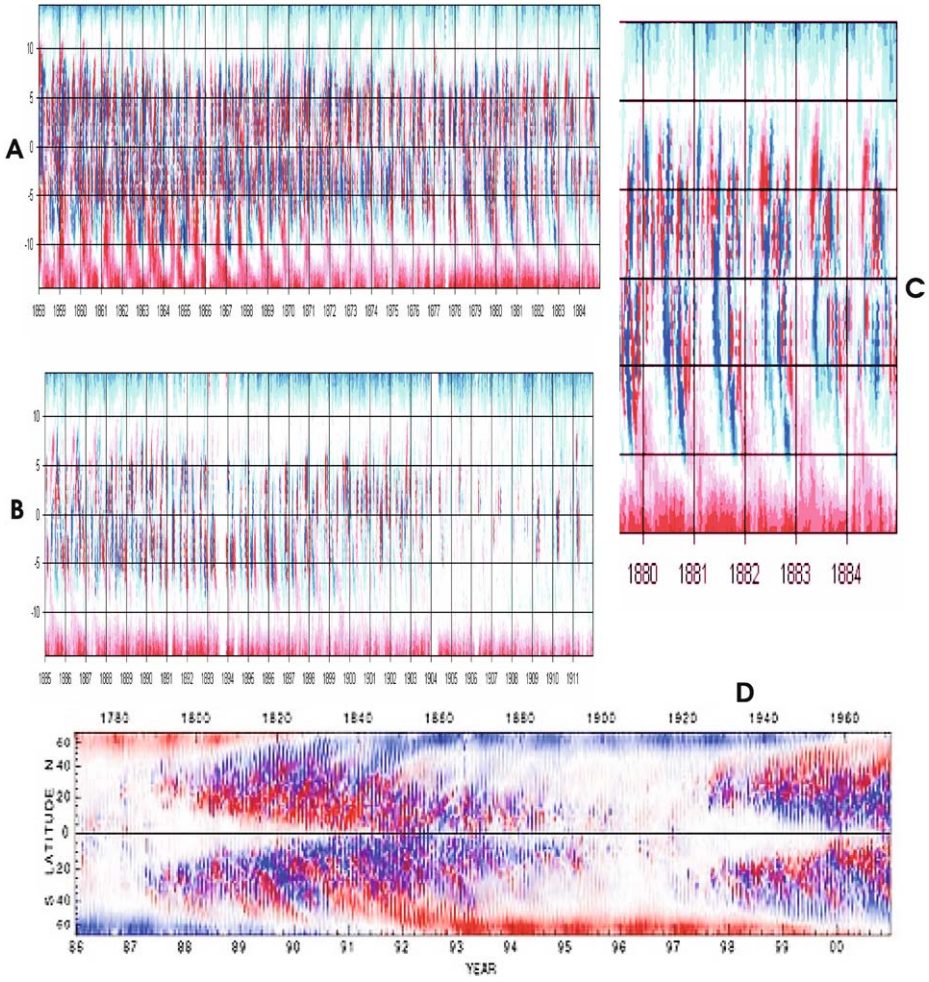

Figure 9 (A) – (C) Supersynoptic maps, basically a higher time resolution of the synoptic maps, for periods in cycle 22, where the detailed times can be ascertained by the Carrington rotation numbers (courtesy of Leif Svalgaard and the Wilcox Solar Observatory). Panel C shows a period of time in 1994. Each Carrington rotation is reversed, so that time is running uniformly from left to right. This allows a continuous time map, rather than the typical synoptic maps that have discontinuities each rotation. The latitudes are in terms of sine latitude. (D) Mount Wilson's supersynoptic map format for latitude versus time for solar cycle 22. As can be seen, the fields really move in a fairly linear fashion, which looks far from diffusive.

motions to occur, as meridional flows would transport both polarities toward a pole. Thus our model might be broadened to incorporate such flows. The goal here, however, is not to try to incorporate every motion, but rather to try to incorporate only the minimal amount needed to obtain a simplified shallow dynamo. Magnetic forces, as previously described in relation to Figure 6, can pull both polarity fields toward a given pole, but this effect has not been incorporated here. There are a number of possibilities to explain how both polarities may be dragged toward the same pole. I favor the magnetic force explanation with opposite signed fields simply remaining connected together, as discussed, but our model, at present, does not use it.

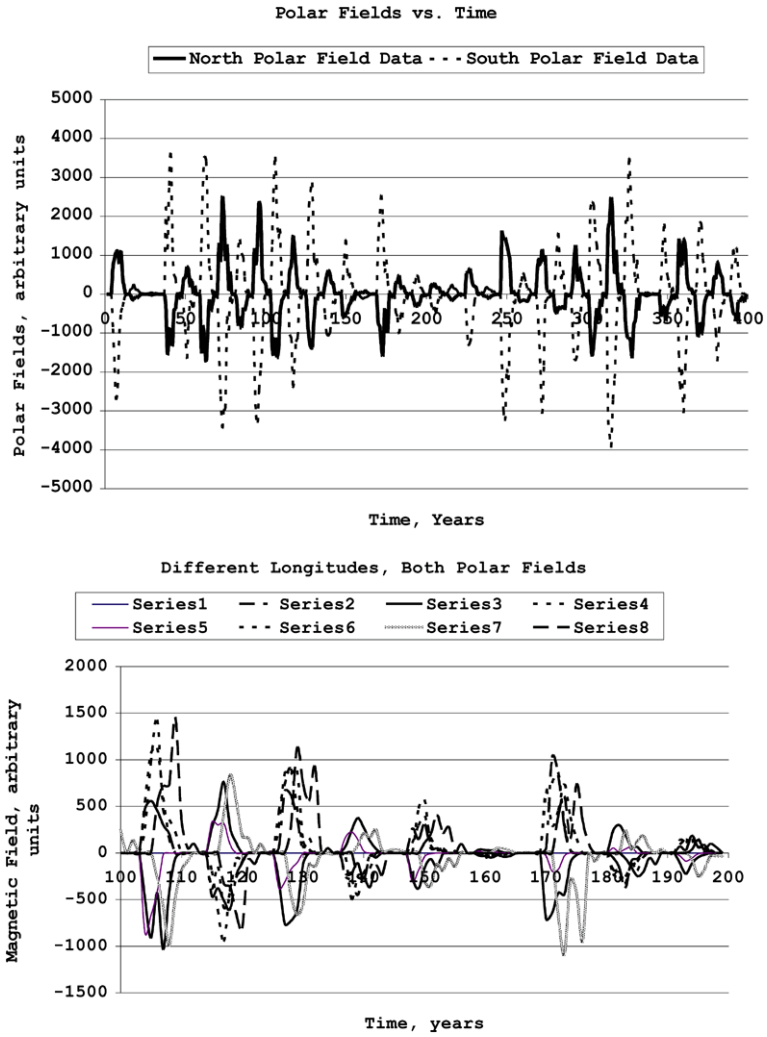

Figure 10 High-latitude/polar fields in our modeling effort are those poleward of about 60°. This graph allows modeled solar cycles to be viewed. (Top) Both poles for 400 years, with fields summed for all longitudes. (Bottom) A shorter period of 100 years, showing various longitudes separated for each pole. This allows variations and the manner in which fields are drawn into the poles to be better seen. Each pole is divided into four quadrants, and the fields are summed. The light lines are from one hemisphere and the heavy lines from the opposite hemisphere. The period of time, being only 100 years, allows solar cycle details to be viewed.

Sample results from this 2D modeling effort are shown in Figure 10. To clarify the large-scale field structures' temporal behavior, we have chosen to look at fields poleward of 60°. This provides a convenient way to separate the wheat from the chaff. It is sufficient to gather polar field data and shows enough variations to see temporal effects of cycles in a form that is readily discernable.

The top panel shows a 400-year record, wherein secular changes are easily seen in this display, namely, overall growth and decay of solar cycles, as the fluxes wax and wane. The

model shows both 11-year time-scale variations, as well as periods when a dearth of activity persists, such as Maunder minimum behavior. Such high and low levels last many cycles, usually, before the behavior changes. The cycle seems to have approximately the right level of persistence and also chaotic behavior.

The lower panel displays a 100-year record, providing a reasonable facsimile to yearly variations in the Sun's activity. Now, however, in addition to greater temporal variations exposed, we also see longitudinal variability. There are eight curves, but now from four different longitudes in the northern and southern hemispheres, again at high latitudes, to minimize noise. One sees here the yearly variations, as well as structural variations. The variations are reminiscent of how UMRs (also known as sectors, when their fields project into interplanetary space) might appear as they cross the solar disk.

Although we have seen some overall workings of our shallow dynamo model, it needs to be tested in greater detail than the average fluxes displayed in Figure 10. Thus a consideration of some observational tests of the dynamo may distinguish this shallow dynamo viewpoint from the conventional deep dynamo paradigm.

4. Observational Tests of the Solar Dynamo

Many of the workings of a deep dynamo are, by definition, hidden from view. The shallow dynamo proposed here has only its toroidal field buried, and it is even possible that the nonradial tilts of sunspots, which are observed, are direct evidence of a shallow toroidal field. Here observations that may distinguish the shallow from the deep dynamo are discussed. Comparisons with observations are necessary for any theory to be more than a tautology. For ideas to succeed one must allow them the possibility of failure.

The observational differences between the shallow and deep dynamo models are listed in Table 1. For the deep field origin of activity, consider the picture of an Omega loop rising to the surface to initiate sunspots. In this view, we do not expect the loop to rise with any preference for particular surface features (*e.g.*, large-scale field geometry, *etc.*). So for the deep dynamo view, we expect the following: There should be near-zero correlation of sunspot births with surface features. We would expect sunspots (from Omega loops) to arise in equal and opposite pairs, as a flux tube crossing through a surface adds a value of $\iint_{\mathrm{A}} \mathbf{B} \cdot \mathrm{d}\mathbf{A} = 0$ from Gauss's theorem. Thus there would be little tendency toward an imbalance (in terms of $+$ and $-$ signs) of new sunspot fluxes, the most notable examples of which are the small percentage of unipolar sunspots. Thus, although one expects the flux rising from both sides of an Omega loop, rising from the deep interior of the Sun to contain equal and opposite flux, actual observed field passing through the photosphere could well differ from this, since observations of the real photospheric field through the real photospheric surface are not sampled in "ideal conditions," owing to photospheric dynamics, radiative transport, solar turbulence, terrestrial atmospheric seeing, 5-minute fluctuations, *etc.*

For shallow dynamos, our shallow dynamo model in particular, the view is taken of EPR flux being gathered together by vast inflows and downflows of tenuous neutral gases into centers of activity. The amount of magnetic flux (of both signs) gathered into an active region would depend upon the amounts of these fluxes in the surrounding surface regions and the strength of the superadiabatic gradient powering the inflow gases. In general, for any surface (or shallow) dynamo, we would expect changes in the surface (or shallow) layers to be reflected by other changes in those layers. Thus changes in the magnetic flux in the photosphere should relate to other alterations of this surface. A nice example of this is that changes in the polar fields, in our model, should correlate with the aging and transport of

Table 1 Observations and dynamo models. Several observed effects could distinguish the shallow dynamo model from the conventional deep-seated dynamo view. No detailed references seem to be available for testing deep-seated dynamos against shallow dynamo theories, so the deep-seated dynamo signatures are only suggestive.

Observations	Deep Seated Dynamo	Shallow Dynamo
Sunspot attributes	Eruptions occur with little or no relation to surface features	Coalescence of spot fluxes relate to flux in surrounding surface neighborhood
Sunspot fluxes	Bipolar pairs contain equal and opposite fluxes	More sunspot flux of the same sign as in the surrounding UMR (unipolar region)
Sunspot birth/size	Little or no relation to surface features	BMR numbers and sizes (larger + more numerous) relate to Hale (UMR) boundaries
Large-scale field (including growth and decay of polar fields)	Diffusive and meridional flow processes	Relation to other surface field effects: field-driven motions, plus percolation, and advection (including meridional flow) of flux [*e.g.*, UMRs and polar fields grow (or shrink) as active regions supply same (or opposite) sign flux]
EPRs/bright points	No relation of EPRs and bright points to active regions	EPRs and bright point number densities may correlate with (decrease near active centers) the vicinity of large growing activity centers

active region flux toward and away from the poles. For deep dynamos, contrariwise, we generally expect surface magnetic flux changes to appear *de novo*, from deep-seated events. Generally then, for shallow dynamos, we expect correlations between spot appearances and large-scale field patterns to emerge.

To test our model, consider such a possible correlation with two types of patterns: *i*) that sunspot polarities would have flux imbalances based upon the sign of a UMR and *ii*) that the boundaries between unipolar regions, also known as solar sector boundaries, might aid the preceding and following sunspot growth in accord with how that polarity region fell into accord or discord with the UMR boundary patterns.

To help illustrate some of the behavior of sunspots in relation to surface fields, two modeling studies are undertaken of EPRs undergoing superadiabatic percolation in the presence of surface features. Figure 11 shows the results of these studies. The top series of four panels starts with a large-scale UMR, of randomly distributed + fields (shown as light blue). To this are added random EPRs, again near the center of the diagram. The fields then undergo superadiabatic percolation throughout the entire grid. Not surprisingly, as suggested in paper I, a bipolar pair forms with a preferential + field. In our present model, when undergoing superadiabatic percolation, no field is ever "lost"; hence the − flux just collects, and the smaller spot simply develops. The size of the + anomaly is governed in this model by the amount of excess + field and the size of the regions allowing pair swapping and other parameters. The point of this model is to illustrate that spot regions with unbalanced fluxes may form, and they do so predominantly in regions of UMRs of that sign. This then illustrates a neat test of our model.

In the second study, shown on the bottom of Figure 11, we see an initial UMR boundary, also called a sector boundary. Near the center again, EPRs have been added and a superadiabatic percolation has been allowed to occur. Dipoles are augmented by the fields from the sector boundary. Such an effect was observed by Svalgaard and Wilcox (1976). When the subsurface field direction is associated with the sector boundary (called a Hale boundary), more sunspots are observed to form compared to the case when the subsurface field opposes

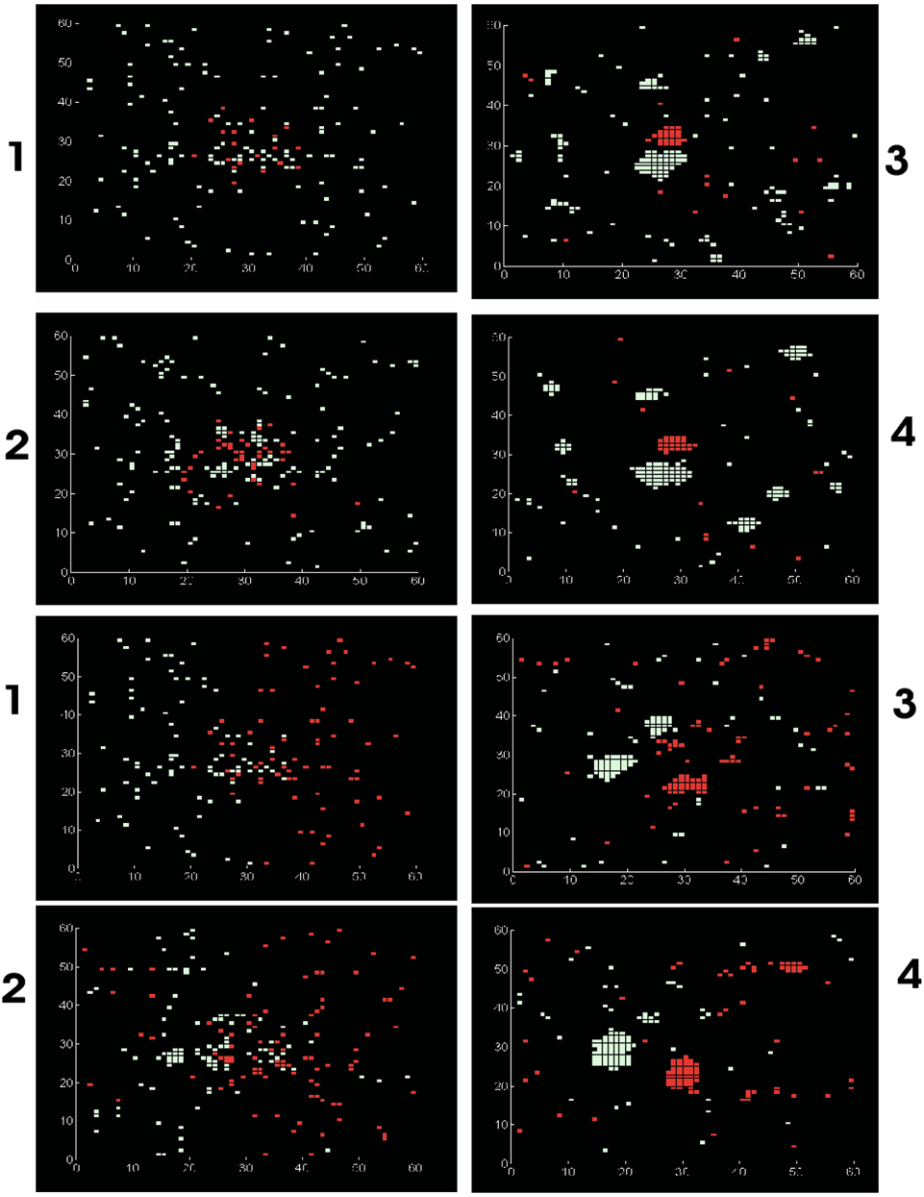

Figure 11 Two time series of percolation models. (Top) The series starts with a large-scale UMR, of randomly distributed + fields (shown as light blue). Near the center are added random EPRs. The fields then undergo superadiabatic percolation throughout the entire grid. One sees that an unequal BMR is generated (oriented randomly, with more of the same sign contained in the UMR). If the percolation grid were larger, a greater inequality would emerge. (Bottom) Series in which the magnetic boundary +/− between two UMRs is aligned vertically. Again random EPRs are initiated and the entire grid is allowed to undergo superadiabatic percolation. One sees that a large-scale BMR aligns somewhat along the boundary of the UMR. This supports the Svalgaard – Wilcox Hale boundary effect. The quantity δ is 3, s is 60, n is 10, k is 20, and E is as given in text (0, −1, or −2 depending upon neighbors). Time steps in both models are 1, 2, 10, and 20, respectively.

the sector boundary (called a non-Hale boundary). Thus there is not only a correlation with a particular boundary orientation but an anticorrelation associated with the reverse boundary orientation. It would be a valuable test of surface versus deep field dynamo models to undertake the Svalgaard – Wilcox (1976) study with modern data and ascertain the persistence of their correlation. This may shed further light on the important question of whether the source of the Sun's dynamo is deep or shallow.

The three correlations discussed previously are outlined in the first three rows of Table 1. The fourth row shows that large-scale fields could be correlated with differing types of motions and processes. Additionally, the fifth row outlines another testable way to distinguish the particular shallow dynamo model developed here against deep-seated theories. In our model, EPRs (and associated bright points) should be swept into activity centers when the active region is growing. If they are not repopulated quickly, it seems possible that a dearth of them may exist around large activity centers during their growth phase. Of course our theory suggests that active regions attain their magnetic flux by the accumulation of neighboring magnetic flux, and thus they also contain pores (EPRs) and are themselves bright in soft X rays. Hence to calculate whether a dearth of neighboring EPR flux exists, one might need to "blot out" the active region flux itself and only engage in a number count just outside the region, being careful not to allow coronal holes to affect number count.

Observations, of course, could settle this question. Are there any observations that show significant solar magnetism at depths appreciably ($>\sim$ factor of three) below the bottom of the hydrogen ionization zone, at a few thousand kilometers? As discussed previously, evidence for flows below active regions descending into the deep interior provide a carbon-copy signature of the surface activity but so far no evidence that the magnetic field itself extends deep into the interior. To settle this important question, one needs evidence of deeply seated field that is more than a shadow of surface activity.

5. Discussion and Conclusions

This paper offers models of numerical percolation for small- and large-scale fields in the photosphere. These models assume that superadiabatic percolation occurs for brief periods of time in special locations, where the temperature gradient in a localized region just below the photosphere is exceptionally high and where a large horizontal subsurface field exists, allowing the EPR fields to be collected. When these conditions are met, like-sign fields from EPRs gather quickly into growing pores and sunspots. Thus Babcock's toroidal field geometry may amalgamate with our small-scale CA model and form localized BMRs in accordance with Hale's polarity laws. Their fluxes disperse following sunspot decay, after which they alter the large-scale field of the Sun.

One assumption inherent in the shallow dynamo model is that magnetic field resides predominantly in the shallow layers below the photosphere. Factors that may contribute to this are *i*) field buoyancy, which could directly inhibit the field from descending more deeply than the outer convection zone; *ii*) the need for flow and field to separate, as elements of cold dense flow are drawn more deeply toward the inward radial direction, whereas elements of field would tend to be driven perpendicular to the flow field (as a leaf falling under gravity falls bowed side down, increasing its area normal to the flow) toward the horizontal direction, as determined by the differential flow forces on its elements; and *iii*) fibril field parameters working against field from descending too deeply into the Sun's convection zone. This last point is understood simply as follows. Consider how a fibril flux tube's parameters vary with depth: The magnetic pressure varies as $P_{\mathrm{B}} = B^2/8\pi$, the field tension as

$T_B = AB^2/8\pi \approx \phi B$ (where the flux $\phi = AB$ is constant in a flux tube), and the buoyancy force as $F_B \sim g\Delta\rho V$. By starting from a small value below the photosphere, the pressure, the tension, and the buoyancy increase exponentially with depth. Near the photosphere ρ and $\Delta\rho$ are low, leading to vanishing buoyancy there. The magnetic field thus has increased buoyancy with depth, which might make the field close to unsinkable, like trying to sink a giant raft to the bottom of the ocean. In our view then, if we follow the flow and field toward greater depths, the field's increasing tautness and buoyancy tend to inhibit it from following the flow to great depths. Alternatively, others (see Nordlund *et al.*, 1992) suggest that, in convective dynamos, magnetic flux tubes may be wrapped around swirling downdraft motions resembling vortices. This is sometimes referred to as magnetic pumping, and the difficulties of sinking magnetic flux might be overcome.

Thus, the magnetic field may not simply be directed along flow lines, as Alfvén's frozen-field approximation stipulates. Instead, as in Babcock's original picture, the magnetic field may reside close to the Sun's surface, allowing the flow to descend unrestrained by the field buoyancy, and thus the flow may transport the solar luminosity, unhampered by the magnetic field. If the magnetic field stays sequestered near the Sun's surface and above (the corona), it moves toward low-β regions, with β being the ratio of kinetic to magnetic pressure; it thus behaves like plasmas we have seen on Earth, for example, in tokomak-type plasma machines. In these machines, high-β plasmas and magnetic fields abhor (separate from) each other (*e.g.*, Müller *et al.*, 1999). If the field of the Sun does descend deep into the solar interior, it is behaving in a manner unlike the manner it displays in terrestrial plasma machines, where the plasma and field are immiscible.

The questions of how and why flow and field might not be directly aligned, as one usually expects from Alfvén's frozen-field approximation, may be due to the unusual environment of magnetic field present in the Sun's outer convection zone (with high turbulence, lowered ionization levels, velocities near the sound speed, and its remarkably exotic environment with distinguished disequilibrium, magnetic field pressure exceeding the gas pressure, magnetic stresses and buoyant forces allowing the magnetic elements to behave like buoyant, stiff wires, *etc.*). Sinking magnetic field into the Sun's interior may be a difficult task, as mentioned earlier. So, both sinking and floating the flux tubes are problematic, the former because of flux buoyancy, the latter because of Alfvén's flux condition. Theory may push our views in one direction or another, but only observations can answer this question.

This paper shows CA models that allow field growth and motion to be calculated. The models are strongly dependent upon the superadiabatic temperature gradient, as discussed. When an active region first develops, the superadiabatic gradient is large, allowing superadiabatic percolation to coalesce like-sign field. This effect is compared with the workings of a transistor, taking power from an energy source and magnifying a signal, owing to the nonlinear internal workings. For active regions, after the superadiabatic gradient is lessened, normal percolation occurs and the sunspots previously formed disperse into large-scale UMRs. Our model is not complete and does not start from the fluid dynamics or MHD directly; rather, it uses CA models to obtain working models of magnetic field – fluid dynamical interactions associated with the energy transport. Without a complete tying together of MHD including solar convective zone dynamics and radiative transfer into our CA models there is room to tie down the model's details within a theoretical framework.

The large-scale dynamo model also has areas for improvement. In particular, the motions of large-scale fields on the Sun have been simplified by just including percolation, differential rotation, and magnetic forces. The first category force acts in a stochastic fashion. The only pure hydrodynamic force included is the differential rotation. Magnetic forces are calculated based on the connection of subsurface fields with each other. With the model being

simplified, our description was similarly described. The astute reader may have noticed in Section 3.2 that the motions and forces on bipolar regions were described only during the first half of a solar cycle. This was not an oversight, since if one examines the argument during the second half of a cycle, and uses the new cycle polar field, the unipolar fields go toward the "wrong pole." This is not a problem, but it does require a few machinations to explain. Consider that when the polar field reverses, rather than the dipole just going to zero, and the entire field reversing, the highest latitude fields are reversed, but the lower latitude remainder of the old cycle polar field does not reverse until later. One can think of this as the poloidal field simply moving equatorward at special longitudes, with opposite-sign field moving poleward at different longitudes. It is likely in this manner that the solar dynamo progresses, rather than through a uniform axisymmetric behavior. As a consequence, the lower latitude sunspots are still affected locally by forces associated with the "old polar field sign" until the poloidal field reverses at the lowest latitudes, rather than at the poles. This situation (of a simplified description) is also present in Babcock – Leighton-based dynamos. In these models, the polar portion of their flux ropes is removed by the polar field reversal. The surface sunspot zones are still bathed in the old cycle field and thus still are driven toward the appropriate old cycle pole and reach there in due course. This dilemma may require further examination, but for now the situation seems reasonably explained.

For any general solar dynamo, the energy for all relative motion and field on the Sun must ultimately come from the solar luminosity. In the model presented here this starts with the magnetic carpet view of EPRs put forth by Schrijver, Title, deRosa, Berger, and many others in the Lockheed group and described by the convective isolation of magnetic flux, in accord with Zwaan and Parker's models. We start, as did Babcock and Leighton, by considering the dynamo in its poloidal field phase. Toroidal field is then formed by differential rotation of a shallow subsurface field. EPRs are magnified into sunspots by superadiabatic percolation using the Sun's convective energy, in the same way transistors magnify a signal. For active regions, the process is through the enhanced scale (rather than small granule and supergranule cells) of convective energy transport by allowing unidirected flows below active regions (with subsequent return upflows), guided by growing field regions. Growing sunspots thus drain away the cool nearly neutral hydrogen surrounding individual active regions and thereby gather like-sign flux into and below the sunspots. As a result of the vast amount of cooling associated with the cool downdrafts, the enhanced superadiabatic temperature structure is quenched and returns to normal. Normal percolation returns, acting to disperse the concentrated sunspot magnetic fields. In this process, the energy stored below the sunspot's ionization zone is transported outward along individual fibril fields in upward flows. These return flows help transport the solar luminosity outward and are associated with the bright faculae released from the active region. Faculae, being associated with magnetic field, do have a small well in the photospheric field location, but the associated outward energy transport raises the temperature and scale height of the surrounding photosphere, raising these features into a hillock geometry (see Figure 5 of Schatten *et al.*, 1986). Without this hillock geometry, faculae would not be seen near the solar limbs, as they appear to be. As faculae lose their brightness, normal percolation continues to disperse the field and they help form the large-scale UMRs, which Bumba and Howard suggested are the main source of the Sun's large-scale field patterns. The entire process creates a synergy, aiding energy transport from the Sun's interior, using the magnetic field as a catalyst, to create more efficient unidirected flow patterns than the small-scale overturning pattern of mixing-length bubbles that standard astronomical theory suggests. Overall, active regions thus create large-scale flows that effectively "stir the pot" and thereby help transport the solar luminosity outward. The added energy transport associated with solar activity is observed in the solar irradiance (see Lean *et*

al., 1998). The active region large-scale flow patterns, without the magnetic field's presence, would not be possible, since convective patterns break up into smaller scale eddies, and the magnetic field serves to guide the large-scale flows. Thus, as Bob Leighton said, "If the Sun did not have a magnetic field, it would be as uninteresting a star as most astronomers believe it to be."

One intriguing question about the shallow dynamo model should be raised. We have seen solar variability on very long secular time scales. It might be disturbing to try and have the model answer the following question: How can a shallow solar dynamo result in long-time-scale variability, without access to the large energy reservoirs contained in the deeper layers of the Sun? One possible answer is that the downflows below active regions transport significant energy to great depths. Nevertheless, this problem is left to the next generation in the hope of a more definitive answer. Observers may settle the deep versus shallow dynamo question by finding evidence, either directly in the photosphere or in the deeper layers from helioseismology, of significant magnetism (pro or con) existing at depths appreciably ($>\sim$ factor of three) below the bottom of the hydrogen ionization zone, at a few thousand kilometers. As discussed earlier, these downdrafts below active regions may not be evidence that the magnetic field itself extends deep into the interior. Rather, the butterfly diagrams of heated gases at depth below active regions (see González Hernández *et al.*, 2008) are viewed as a solar equivalent to a Santa Ana type wind. To settle this important question, one needs evidence of deeply seated field that is more than a shadow of surface activity.

Acknowledgements We appreciate Leif Svalgaard for fruitful discussions and supplying us with his supersynoptic charts, originally developed by Roger Ulrich. We also thank Todd Hoeksema, Phil Scherrer, and the Wilcox Solar Observatory staff as well as the Mount Wilson Observatory staff for use of their solar magnetic data. Further, we appreciate the help of Marc deRosa and Alan Title of the Lockheed group and the SOT group of the *Hinode* satellite. *Hinode* is a Japanese mission developed and launched by ISAS/JAXA, with NAOJ as domestic partner and NASA and STFC (UK) as international partners. It is operated by these agencies in cooperation with ESA and NSC (Norway). We are further appreciative to Gene Parker for his insights on solar magnetism. We also have benefited in the past from discussions on solar magnetism with Robert Leighton, Robert Howard, and John Wilcox. Additionally, we thank Norman Ness, who served as a mentor for many years. Last, but not least, we have benefited from our many discussions with Hans Mayr, who sparked our interest to view sunspots as a meteorological phenomenon, rather than from the traditional magnetostatic viewpoint. Any mistakes in this paper are the responsibility of the author and not due to past collaborators.

References

Babcock, H.W.: 1961, *Astrophys. J.* **133**, 572.
Brandenburg, A.: 2005, *Astrophys. J.* **625**, 539.
Bumba, V., Howard, R.F.: 1965, *Astrophys. J.* **141**, 1502.
Cox, J.P., Giuli, R.T.: 1968, *Principles of Stellar Structure*, Gordon & Breach, New York.
Fragos, T., Rantsiou, E., Vlahos, L.: 2004, *Astron. Astrophys.* **420**, 719.
Gardner, M.: 1970, *Sci. Am.* **223**, 120.
González Hernández, I., Kholikov, S., Hill, F., Howe, R., Komm, R.: 2008, *Solar Phys.* **252**, 221.
Hale, G.E., Ellerman, F., Nicholson, S.B., Joy, A.H.: 1918, *Astrophys. J.* **49**, 153.
Ising, E.: 1925, *Z. Phys.* **31**, 253.
Jackson, J.D.: 1962, *Classical Electrodynamics*, 2nd edn., Wiley, New York, 12.
Kitchatinov, L.L., Mazure, M.V.: 2000, *Solar Phys.* **191**(2), 325.
Lean, J.L., Cook, J., Marquette, W., Johannesson, A.: 1998, *Astrophys. J.* **492**, 390.
Leighton, R.B.: 1964, *Astrophys. J.* **140**, 1547L.
Leighton, R.B.: 1969, *Astrophys. J.* **156**, 1L.
Mihalas, D.: 1970, *Stellar Atmospheres*, Freeman, San Francisco.

Müller, H.W., Büchl, K., Kaufmann, M., Lang, P.T., Lang, R.S., Lorenz, A., *et al.*: 1999, *Phys. Rev. Lett.* **83**(11), 2199.
Nordlund, Å., Brandenburg, A., Jennings, R.L., Rieutord, M., Ruokolainen, J., Stein, R.F., Tuominen, I.: 1992, *Astrophys. J.* **392**, 647 – 652.
Parker, E.N.: 1978, *Astrophys. J.* **221**, 368.
Parker, E.N.: 1979, *Astrophys. J.* **232**, 291.
Parker, E.N.: 1984, *Astrophys. J.* **283**, 343.
Parnell, C.: 2002, *Astron. Geophys.* **43**(4), 4.16.
Schatten, K.H.: 1988, *Astrophys. J.* **329**, 1028.
Schatten, K.: 2005, *Geophys. Res. Lett.* **32**, L21106.
Schatten, K.H.: 2007, *Astrophys. J. Suppl.* **169**, 137.
Schatten, K.H., Mayr, H.G.: 1985, *Astrophys. J.* **299**, 1051.
Schatten, K.H., Leighton, R.B., Howard, R., Wilcox, J.M.: 1972, *Solar Phys.* **26**, 283.
Schatten, K.H., Mayr, H.G., Omidvar, K., Maier, E.: 1986, *Astrophys. J.* **311**, 460.
Schrijver, C.J.: 2001, *Astrophys. J.* **547**, 475.
Schrijver, C.J., Title, A.M.: 2001, *Astrophys. J.* **551**, 1099.
Schrijver, C.J., DeRosa, M.L., Title, A.M.: 2002, *Astrophys. J.* **577**, 1006.
Seiden, P.E., Wentzel, D.G.: 1996, *Astrophys. J.* **460**, 522.
Sheeley, N.J., DeVore, C.R., Boris, J.P.: 1985, *Solar Phys.* **98**, 219.
Sheeley, N.R. Jr., Nash, A.G., Wang, Y.-M.: 1987, *Astrophys. J.* **319**, 481.
Stix, M.: 1974, *Astron. Astrophys.* **37**, 121.
Svalgaard, L., Cliver, E.W.: 2007, *Astrophys. J.* **661**, L203.
Svalgaard, L., Wilcox, J.M.: 1976, *Solar Phys.* **49**, 177.
Ulrich, R., Boyden, J.: 2006, *Solar Phys.* **235**(1 – 2), 17.
Waldmeier, M.: 1961, *The Sunspot Activity in the Years 1610 – 1960*, Zurich Schultess and Company, Zurich.
Woodward, P., Porter, D., Anderson, S., Fuchs, T., Herwig, F.: 2006, *J. Phys.* **46**, 370.
Zwaan, C.: 1978, *Solar Phys.* **60**, 213.