The Future: Where are We Headed

Åke Nordlund

NBI, Univ. of Copenhagen and JILA, Univ. of Colorado

Abstract. I discuss what the future may bring in some of the topics of this meeting. In particular I discuss near-surface dynamics and the solar surface velocity spectrum, sub-surface dynamics and dynamo action, emerging flux, sunspots, and the chromosphere and corona. For each topic, a prediction and/or provocation is made, as a challenge towards bringing about progress in that topic area of solar physics.

1. Introduction

What will the future bring in the topic areas discussed at this meeting? We can be sure of two things: There will be many important advances in these areas, and these advances will be of great importance not only to solar physics but to astrophysics in general.

Advances will come mainly from a combination of new and better observations on the one hand, and new and better numerical simulations on the other hand. In that respect, our research field – solar physics – is no different than so many other branches of astrophysics. We do, however, have a distinct advantage in one important respect; we can observe the Sun in much more detail than most other astrophysical objects, and, moreover, we have the unique advantage of being able to observe the time evolution. Thus, in solar physics, we have access to the “fourth dimension” – time – on both the observational and theoretical side. In most other (much more distant) astrophysical contexts the natural time scales are such that we, in practice, have access only to observational “snapshots”, to be compared with theoretical models and numerical simulations. A common joke is that some theoreticians consider more limited observational details an advantage; let us laugh at that and then come back to order, recognizing that the abundant details and temporal information that the Sun offers is indeed a great advantage. Indeed, using the more abundant and detailed observations available for the Sun, theories and numerical models may be tested and validated on the Sun before being applied in more obscure astrophysical contexts.

Below, I touch briefly upon some key issues among the topics discussed at this meeting. For each topic, I start out from a prediction and/or provocation concerning what we might expect from the future under that topic; future proof that these predictions are wrong will be as useful as the opposite! After briefly discussing the merits of realistic vs. idealized models in Section 2, I go on to discuss near-surface dynamics and the solar surface velocity spectrum in Section 3, sub-surface dynamics and dynamo action in Section 4, emerging flux in Section 5, sunspots in Section 6, the chromosphere in Section 7, and the corona in Section 8.
2. Realistic vs. Idealized Models

Prediction / Provocation: The use of realistic models will increase monotonically, relative to the use of idealized models!

Observations and theory clearly go hand in hand in our efforts to understand physical processes that occur on the Sun. Progress has been rapid on both sides, and progress on one side generally triggers better focussed efforts and progress on the other side. On the theoretical side, numerical simulations have emerged as a powerful tool to test theoretical scenarios and models. The ability to use powerful computers has given theoreticians a tool that is much less limited by practical constraints than analytic theory (the latter is of course still useful and necessary for analyzing basic physical concepts).

Numerical models are also useful for “extending” observations into domains that are not accessible directly. In solar physics, for example, observations of the solar surface are easily obtained, while informations about sub-surface motions are not. In other branches of astrophysics, even the time domain is often not directly accessible. With access to numerical modeling, once a model is doing a good job at reproducing observations of the accessible sub-space, one is free to exploit the access to the full 4-D (space and time) data set to answer questions about the physical mechanisms at work.

Two complementary numerical approaches have been and are being used. One uses simplified physics to explore basic properties and mechanisms (in solar physics e.g. Porter & Woodward 2000; Elliott 2000; Tobias et al. 2001; Cline et al. 2003; Bushby & Houghton 2005), while the other employs as realistic physics as possible, with the goal to be able to compare results directly with observations (e.g. Stein & Nordlund 1998; Rosenthal et al. 1999; Stein & Nordlund 2001; Keller et al. 2004; Carlsson et al. 2004; Vögler et al. 2005). Both approaches are valuable, and to a large extent complementary.

One disadvantage of using realistic models is the relatively complex setup required; one may need to construct tabular equations of state and opacity tables, and add new numerical methods such as those required to treat radiative transfer of energy. These are one-time disadvantages, however; once the methods are there they may be used and reused indefinitely. A more permanent drawback is increased computing time. However, since the computing time anyway scales very steeply with numerical resolution (generally with the inverse fourth power), relatively little resolution has to be given up to compensate for the increased cost per time step. And of course the big advantage is the ability to make forward (direct) comparisons with observations.

Idealized models have a corresponding disadvantage: direct forward comparisons are not possible. One may jokingly consider that to also be an advantage, along with the more bona fide advantage that lies in a simpler setup, which may allow a quicker attack on new problems, and an easier mapping of a parameter space.

My prediction for the future is that we will see an increased use of the realistic approach in solar physics, as modelers try to keep up with increasingly detailed and sophisticated observations. Also, in order to make decisive tests of various proposed mechanisms, it is often necessary to aim for direct, quantitative comparisons, since several mechanisms may actually be viable explanations for a certain phenomenon, given enough freedom of choice with respect to input parameters. By using actual solar conditions, and by using realistic physics and parameter values, one reduces (ideally to zero) the number of free parameters.

The general strategy in the realistic approach to modeling is to verify (and possibly calibrate) the model by comparing robust diagnostic signatures between observations and models. If possible, the diagnostic signatures should be insensitive to spatial resolution, while still encoding a reasonably complex “fingerprint”, so the chance of having accidental agreement is minimized. One of the most successful examples of this approach is the synthesis of photospheric spectral lines from simulations of near-surface convection (Asplund et al. 2000a,b,c). There, the combination of spectral line width,
Figure 1. The solar velocity spectrum, defined as $v(k) = \sqrt{kP(k)}$, where $P(k)$ is the velocity power per unit wavenumber $k$ and hence $kP(k)$ is the velocity power per unit logarithmic wave number. $v(k)$ is a suitable quantitative measure of the velocity on scales of the order $L = 2\pi/k \approx 4200/\ell$ Mm, where $\ell$ is the spherical harmonic wave number. The curves are, as indicated with labels, based on results obtained by Doppler imaging by Hathaway et al. (2000), by correlation tracking measurements from TRACE and SOHO-MDI (R. Shine & LMSAL group, private communication), and by supercomputer simulation on granular scales (Stein & Nordlund 1998).

strength, shift, and shape provides tight constraints on velocity amplitudes (through the spectral line widths), gradients of velocity along the line of sight (through the spectral line strength), and correlations between velocity and temperature (through the spectral line shift and asymmetry).

3. Near-Surface Dynamics and Large Scale Velocity Spectrum

Prediction / Provocation: We will come to accept that there are no special scales larger than granulation; motions on all scales larger than granulation exhibit an approximately scale-invariant behavior!

For decades “supergranulation” (Simon & Leighton 1964) has been thought of as a relatively well defined pattern of motions on the solar surface, shaping the quiet Sun magnetic field and the related chromospheric emission into a cellular pattern, with a scale of the order of 30 Mm. “Mesogranulation” was introduced as a name for motions on a scale ($\sim 7 – 10$ Mm) intermediate between that of supergranulation and granulation (Toomre et al. 1979). Since these scales are much larger than the photospheric density scale height, the motions are necessarily predominantly horizontal.

However, if one analyzes the solar velocity field, as observed for example by SOHO-MDI and TRACE (cf. Figure 1), one finds that the velocity amplitude is a continuously increasing function of wave number, which joins smoothly with the velocity amplitudes found in numerical simulations on granulation scales (where – as pointed out above – the amplitudes are known with very high precision from photospheric spectral line
Figure 2. Left: Full disk SOHO-MDI dopplergram. An approximately $200 \times 200$ Mm area from which a deprojected flat-fielded and sub-sonic filtered Doppler image has been extracted is shown. Middle: Four panels, showing blow-ups from the $200 \times 200$ Mm image. The horizontal scales differ by a factor of two between panels, but the panels are shuffled into random order. Each panel is filtered to an effective resolution of about $16 \times 16$ resolution elements. Right: Four panels, showing blow-ups from a $24 \times 24$ Mm image of horizontal velocity in a numerical simulation of convection (Benson et al., 2006). The horizontal scales differ by a factor of two between panels, and the panels are again shuffled into random order. Each panel is filtered to an effective resolution of about $8 \times 8$ resolution elements.

There may be a slight excess of velocity on supergranulation scales, but if so it is a small effect – the dominating behavior is a nearly linear trend, over three orders of magnitude in velocity and scale.

So, there are motions on all scales, with an amplitude that is approximately inversely proportional to size, and thus with turn over times that scale approximately as the square of the size. Moreover, as may be demonstrated with a “blind test” such as the one in Figure 2, the patterns of motions are similar on all scales larger than granulation (the one panel that stands out in Figure 2 does so because its scale – $3 \times 3$ Mm – is small enough to fall outside the regime of near scale-invariance). How can this be compatible with the well-established existence of supergranulation and mesogranulation, and why is there a nearly scale-invariant spectrum?

First, convective motions in a highly stratified medium tend to like a certain aspect ratio of horizontal scales to vertical scale height, of the order $2 - 4$. This is dictated by mass conservation on the one hand, which makes a very large aspect ratio sub-optimal, because the vertical motions tend to be suppressed. A very small aspect ratio (narrow vertical columns of gas) is, on the other hand, also not preferred, since such columns tend to expand due to the inertia of horizontal motions. The scale heights in the solar convection zone increase systematically with depth, so for any one depth there is an associated scale height, and hence an associated preferred horizontal scale of motion. There is therefore a hierarchy (a “stack” if you like) of scales available, with larger scales associated with larger depths. Due to the increase in heat capacity per unit volume with depth, the velocity amplitude of the convective motions decrease with depth.

Secondly, for reasons that have to do with local pressure balance, horizontal motions belonging to each depth extend with approximately constant amplitude up into more shallow layers. Vertical motions on the associated horizontal scale, on the other hand, are forced by the requirement of mass conservation to decline in proportion to the decreasing vertical scale height. Instead, the smaller scales of motions closer to the surface are able to take care of the vertical heat transport that can no longer be accommodated by the decreasing vertical velocities of the larger scales.

The bottom line is that larger scale motions, rooted at larger depths, make themselves visible in the horizontal velocity field at the surface, which thus contains a superposition
of contributions from all depths, with amplitudes decreasing with increasing scale (cf. Stein & Nordlund 1989). Since the behavior of the amplitudes as a function of scale is determined mostly by the run of vertical scale height with depth, one may expect the imprint of the first and second helium ionization – which has long been expected to be the “cause” of supergranulation – to be quite modest, consistent with the behavior in Figure 1.

Quite distinct impressions of patterns on supergranulation and mesogranulation scales are not at all incompatible with a nearly scale-free solar velocity spectrum. Given the shape of the spectrum, any method of observation that is fully sensitive only down to a certain scale is going to see that scale as dominating, with smaller scales attenuated by decreasing visibility. In an extended sense, the magnetic network behaves in a similar manner. There it is the lifetime of the “corks”; i.e. the magnetic network elements, that ultimately determine which scales one can see. It is well known from “cork experiments” that if corks are given a long lifetime they tend to pick up mesogranulation and granulation scales rather than supergranulation scales.

### 4. Sub-surface Dynamics, Dynamo Action

**Prediction / Provocation:** The solar dynamo works in the bulk of the solar convection zone – Babcock-Leighton style!

The gist of this provocation is that, despite the many years that have passed since the original Babcock-Leighton concept of the solar dynamo emerged (Babcock 1961, 1963; Leighton 1964, 1969), that general picture may still turn out to be closer to the truth than many of the theories that have been put forward in the meantime.

Specifically, for a number of years it has been popular to place the site of the solar dynamo in the overshoot layer below the convection zone, and to associate the dynamo action mainly with the strong radial shear that exists there (I will spare the many authors that have worked on this concept the embarrassment of being cited here). The main reasons in favor of the *interface dynamo* concept appear to have been the perceived need to avoid the magnetic buoyancy of the bulk convection zone, combined with the news value associated with the discovery of the rapid change of rotation speed with depth in the overshoot layer, with the subsequent adoption of the *tachocline* designation (Spiegel & Zahn 1992).

A strong argument against placing the main site of dynamo action in the Sun at the interface between the convection zone and the stable layer below is that magnetic activity appears to be continuous across the line in the HR-diagram where stars become fully convective (Linsky 1985). A similarly strong argument against the interface dynamo is that magnetic activity occurs also at those mid-latitudes where the radial shear vanishes or is small. A more indirect argument is that there is actually no need to place the magnetic field in the overshoot layer to prevent it from rising due to magnetic buoyancy. As shown by Tobias et al. (2001) and Dorch & Nordlund (2001) a magnetic field immersed in the convection zone is held down by “convective pumping”, and is perfectly happy to stay in the low convection zone and make itself available to dynamo processes there. Further arguments for and against a solar convection zone interface dynamo are given by Brandenburg (2005,2006).

The main argument in favor of a bulk-CZ dynamo is that it is compatible with the conceptual framework of the original Babcock-Leighton theory, which relies to a considerable extent on observed behavior of active region magnetic fields, and on the observed transport and diffusion of magnetic fields at the solar surface (see Sheeley 2005; Schrijver et al. 2005, and references therein). In the Babcock-Leighton scenario the dynamo action is mainly due to latitudinal shear rather than radial shear, which is more consistent with the latitudinal distribution of flux, as well as with the time scale of the solar activity cycle.
Arguments against a bulk-CZ location of the main source of dynamo action, and favoring instead a location in the upper or surface layers of the solar convection zone are given by Brandenburg (2005, 2006). The main argument has to do with matching the apparent speed of rotation of magnetic structures with the rotation speed of the source region.

Arguments against an upper CZ dynamo include that the convection zone has global dynamics – so there is little chance to constrain dynamo action to a localized dynamo layer near the surface. Also, the motion time scales in the upper CZ are inappropriate for the cycle, for the tilt, and for the persistence of active regions. The upper CZ dynamo shares these difficulties with mean field dynamos. See Brandenburg (2006) for additional discussion.

5. Emerging Flux

*Prediction / Provocation:* The “flux tube mafia” needs to realize that “there is an end to flux tubes” – both in a direct and indirect sense!

This provocation contains dual messages. The first is that, despite the usefulness of the flux tube concept, it is now time to move on; i.e., to put an end to the era of flux tubes. The second message is that one of the main reasons to move on is that, in a sense, flux tubes actually have “ends”, and that this has important consequences that need to be explored.

With respect to the first message: A number of suggestive and useful results regarding the dynamics and emergence of active region magnetic fields have been obtained over the years, based on the thin flux tube concept. However, to advance beyond what has already been learnt, we need to now continue exploring fully three-dimensional models. This is particularly urgent with respect to the modeling of the final rise phase of emerging active region flux, from the point where the thin flux tube approximation breaks down and until new flux emerges through the solar surface. However, even the rise of flux structures from near the bottom of the convection zone to near the surface needs to be modeled in three dimension, in realistic environments that are actively convecting and turbulent.

So, what is wrong with traditional flux tube models of active region flux emergence, which utilize the thin flux tube approximation, and start out from coherent longitudinal flux tubes circling the entire Sun? Clearly, there are a number of weaknesses in these models, both with regard to assumptions and initial conditions, and with respect to the results. One of the main weaknesses is the assumption that the initial condition consists of a flux tube that is coherent all around the Sun, closing back onto itself. In nature, we may safely assume that magnetic field lines never close back on themselves, and that magnetic field structures do not remain coherent along their individual magnetic field lines.

This leads over to the second message in the provocation: Intermittent flux tube like structures generated in turbulent and convective fluids have “ends”, in the sense that flux surfaces traced from a cross section of an intermittent flux concentration only remain coherent for a limited distance, after which the field lines lose their coherence and connect to regions with much weaker magnetic field strength. Such a “flux tube” may be thought of as having frayed ends, with loose threads that may connect to very diverse regions of space. Such a topology has important implications for both the dynamics and energy balance of flux tubes. Thus, whereas thin flux tubes that are assumed to remain coherent mainly interact with the surrounding, field free plasma through drag forces, real flux tubes are likely to couple much more efficiently to the surrounding plasma, through their frayed ends, and through some amount of “leakage” of flux, even from relatively coherent parts of the structures. The “ends” of flux structures also allow a much more efficient energy exchange with the surrounding plasma, which can flow in
and out of coherent parts of a flux structure, along field lines connecting to weaker, surrounding plasma.

Thus, even though intermittent coherent pieces of flux tend to occur spontaneously also in fully three-dimensional simulations of MHD-turbulence and magneto-convection, we may expect to learn that part of what has been deduced from experiments with thin flux tubes will have to be revised. It will be extremely interesting to see, if, when and how such models succeed in explaining the well known properties of active region magnetic fields. For one thing, we may hope to find the explanation of “active longitudes” along this line of investigation, as well as explanations for why active region emerging flux structures appear to have a shorter extent in longitude than would be surmised from the instability properties of thin flux tubes, which generally tend to generate emerging loops with too large extents in the longitudinal direction.

6. Sunspots

_Prediction / Provocation:_ Flux pumping is _not_ the cause of penumbral filaments!

The Nature paper by Thomas et al. (2002) has as a title “Downward pumping of magnetic flux as the cause of filamentary structures in penumbrae”. It contains a nice cartoon picture of the assumed structure of sunspot penumbra, including an “interlocking-comb” magnetic field, where some field lines have a large inclination while others run nearly horizontally, and eventually dip down below the surface, apparently being held down by the same process of convective downward pumping that was mentioned above. The general structure of the penumbral field is supported by observations, which require large variations in the inclination angle over very small azimuthal scales in the penumbra. What is much less clear is if downward pumping of field lines in the convection zone has a direct bearing on the arrangements of the penumbral magnetic field.

Lites et al. (2004) pointed out very interesting details in the small scale structure in active regions, suggesting that umbral light bridges and penumbral filaments are closely related. Both light bridges and penumbral filaments have, as revealed by the remarkable, high resolution observations from the Swedish one-meter Solar Telescope on La Palma (Rouppe van der Voort et al. 2004; Langhans et al. 2005), very thin dark lanes running along their entire lengths. I have made simple numerical experiments that illustrate how such thin dark lanes may come about (Nordlund & Stein 2006, in prep). I find that the dark center lane is an optical effect, which does not correspond to either up- or down-flow, but instead corresponds to the top of a nearly field-free cusp, where the gas pressure and density are high, so the opacity is large there, and causes a dark “shadow” along the top of the cusp.

Spruit & Scharmer (2006) demonstrate that even a potential field, straddling almost field-free “gaps”, can have the required large variations of inclination angle. They argue that the nearly field-free gaps are supported by convective energy transport, which is also needed to support the relatively large penumbral energy flux. Enhanced density and opacity in the cusp above the nearly field-free gaps causes the central dark lane. Spruit & Scharmer thus argue convincingly that the penumbral structure must be tightly coupled to the mechanism of penumbral energy transport, and that the structure is likely to be locally controlled (as opposed to being influenced by convective pumping in the convection zone).
7. Chromosphere

*Prediction / Provocation:* The magnetic chromosphere is heated the same way as the corona!

The chromosphere is, as Judge (2006) has pointed out, *not a mess!* The upper chromosphere “looks like” the corona, while the lower chromosphere is decidedly non-force-free, but has an apparent complexity that comes largely from huge inhomogeneities in energy and density.

Some of the main questions here are: What drives the flows in the chromosphere; particularly the cool upflows? How is the chromosphere heated, and what is the explanation for the solar and stellar so called flux-flux relations? The flux-flux relations are well known relations between coronal and chromospheric diagnostics; particularly between chromospheric line emission and coronal X-ray emission (Schrijver 1987; Hempelmann et al. 1996). Particularly striking is the fact that local solar flux-flux relations extend to global relations among the corresponding stellar diagnostics.

The prediction/provocation made above is attractive precisely because of these observed correlations; they would indeed be natural consequences of a situation where the chromosphere and corona are heated by essentially the same mechanism. Moreover, based on experience from numerical modeling of coronal heating (Gudiksen & Nordlund 2002, 2005a,b) it is clear that it is even hard to avoid such a conclusion; models of coronal heating tend to dump much more energy in the chromosphere as a side effect.

Space does not allow a more extended discussion of the potentials and obstacles involved in modeling the chromosphere, but it is clear that in order to make progress one needs to find economical and yet reasonably realistic methods for treating the radiative energy losses from the chromosphere.

8. Corona

*Prediction / Provocation:* The corona is heated by magnetic dissipation – primarily due to flux braiding and flux emergence!

A remarkable and wonderful point, made already by Parker a long time ago (Parker 1988), is that in a situation with boundary driven magnetic dissipation, if the dissipation depends on the resistivity at all, it must *increase* with decreasing resistivity! This has been verified in numerical experiments by at least three different groups (Galsgaard & Nordlund 1996; Hendrix et al. 1996; Dmitruk & Gómez 1999). This result has far reaching implications, in that it allows modeling of boundary driven magnetically dominated system (such as the solar corona!) at relatively low – and hence computationally affordable – numerical resolutions. One is then guaranteed that the resulting magnetic dissipation is a *lower bound* on the actual dissipation. Galsgaard & Nordlund (1996) furthermore showed that, because the level of dissipation is controlled by a robust condition on the winding of the magnet field from boundary to boundary, the dissipation is actually close to the asymptotic value already at affordable numerical resolutions. At higher resolutions large scale current structures generate subsidiary small scale structures, which do it again (on shorter time scales), and again, until the spatial scales are small enough to support the dissipation as either normal Joule dissipation or through other small scale and dissipative plasma instabilities.

These results paved the way for the first MHD-simulations of a small active region (Gudiksen & Nordlund 2002, 2005a,b), which used realistic initial- and boundary conditions and a photospheric driving calibrated to be consistent with the solar photospheric velocity spectrum. The experiments gave lower bounds on coronal heating that were already sufficient to produce MK loops at realistic particle densities, leaving little room for additional heating mechanisms.
Spectral line diagnostics using transition region lines is an important method to check models of coronal heating (Peter et al. 2004, 2005). One may use spectral lines emitting in narrow temperature ranges to probe the structure of the transition region and corona at different temperatures. For example, as shown by Figure 3, Silicon IV picks up cooling condensations that are supported by magnetic forces. By using a number of such lines, one can also construct diagnostic fingerprints that encode the complicated temperature and velocity structure of the transition region and lower corona into simple $x$-$y$ plots. Figure 4, for example, shows a plot of emission measure versus temperature, which compares very nicely against corresponding data for the quiet Sun.

In the future we may expect larger models of this kind, and also models where the driving is obtained directly from self-consistent models of the photosphere and upper convection zone, with embedded model active regions.
9. Summary and Concluding Remarks

The topics that I have touched upon briefly here are all interrelated. The solar surface velocity spectrum is an important driver of chromospheric and coronal activity, and the convection zone dynamics that it reflects provides the environment in which dynamo action and the resulting emerging flux dynamics takes place. Future work will be forced to face the couplings among these topic areas, and is likely to benefit from having to do so.

Most of the urgent questions that have been raised here can in principle be attacked and solved by employing larger and more realistic numerical simulations, where the sub-surface convection zone dynamics is coupled directly to chromospheric and coronal dynamics. It is, for example, possible already with the computer power available today to model the subsurface dynamics of the convection zone on scales up to and beyond that of supergranulation, while still resolving the granular scales that control the energy transport in the surface layers. Likewise, and over a similar range of scales, it should be possible to model, from first principles, the sub-surface dynamics of active regions. Such models may then be used as boundary conditions and drivers for studies of chromospheric and coronal dynamics. A project along these lines is being currently being pursued on the Columbia computing system at NASA/Ames by Bob Stein, my self, and others.

In addition to the classical topics discussed here, which can all more or less be addressed within the magnetohydrodynamic (MHD) approximation, there are closely related topics where new numerical modeling techniques, such as particle-in-cell modeling of charged particle dynamics, are needed to make progress. The corona is really a nearly collisionless environment, and to fully understand the limitations of MHD in the coronal environment one must step outside it, into a model environment where particles and electromagnetic fields are treated together, self-consistently.

Acknowledgments. It is a pleasure to acknowledge decades of fruitful collaboration with the guest of honor at the NSO 2005 meeting: Bob Stein!

References

Linsky, J. L. 1985, Solar Phys., 100, 333
Peter, H., Gudiksen, B. V., & Nordlund, Å. 2005, ArXiv Astrophysics e-prints
Toomre, J., November, L. J., Gebbie, K. B., & Simon, G. W. 1979, BAAS, 11, 641