

On Numerical Studies of Solar/Stellar Convection

Kwing L. Chan¹ and Katsuyo Serizawa²

¹Applied Research Corporation Landover, Maryland, USA

²Iowa State University Ames, Iowa, USA

1. Introduction

It has become a cliché to say that convection lies at the foundation of the activities, magnetism, and dynamos of the Sun and cool stars – the title of this colloquium. However, it has not been possible, until recently, to study theoretically these phenomena and their connections with quantitative rigor. The new development owes much to the rapid growth in the power of computers and the spread of sophisticated numerical techniques.

Since the last decade numerical studies of solar and stellar convection problems thrive. The trend towards a greater use of computers in these studies will likely become stronger. Now is a good time to reflect on the numerical approach itself, so that we may understand it better and make better use of it in the future. Here we look at two aspects of this methodology. First (Section 2), we examine its relationships to the analytical approach and observation, for which we make a differentiation between numerical experimentation and simulation. It is obvious that the ultimate test of any theory, numerical or not, is to confront observations, but not generally realized is that the connection to analytical studies is an important requisite for the healthy development of numerical studies. This point is to be elaborated. Second (Section 3), we discuss the prevalence of pitfalls in the numerical approach; two examples of current interest in convection computations are analysed in detail. In Section 4, we go on to suggest that one of these pitfalls might have caused the discrepancy between the results of numerical modelling of the solar differential rotation and helioseismology.

For economy of space and to avoid repetition, a review of the literature is not provided here. The readers are referred to Zahn (1987) for a recent review on the subject of solar/stellar convection. An updated survey of the literature can be found in Chan *et al.* (1990).

2. Connections to the analytical approach and observations

2.1 Numerical experimentation vs. simulation

In many reports of numerical studies, the words ‘numerical experiment’ and ‘simulation’ are used interchangeably, indicating that the differences between these two modes of numerical undertaking are not generally appreciated. When such confusion occurs, it is often accompanied by a failure to acknowledge the physical and numerical limitations of the computation and to qualify the applicability of the results. This is so because numerical experiment and simulation serve very different purposes, invoke different techniques, and generate different kinds of products. Numerical experiments have close link to the development of analytical concepts and theories while simulations are directed towards confronting or interpreting observations. A contrasting comparison of these two modes of study is summarized in Table 1.

Table 1. Numerical experimentation versus simulation

	experimentation	simulation
<i>objective</i>	study principles	reproduce observational results
<i>application</i>	extrapolation to inaccessible ranges	confront observations
<i>model</i>	idealized	realistic
<i>prerequisite</i>	know limitation of computation	also need realistic parameterizations
<i>code design</i>	for 10–20 runs	for 1–2 runs
<i>it failed</i>	prerequisite violated	difficult to identify cause

There is much similarity between numerical and laboratory experiments. They share the procedure of varying the experimental setup or external parameters with the intention to find patterns of scaling relationships for a well-defined physical system or process. The aim is to study the underlying principles. An important application of these principles is to provide estimates for situations which are not yet accessible by the experiments. To quite the contrary, the aim of a simulation is to reproduce or predict the occurrence of an observational event; it has no freedom to transgress the boundaries of the constraints. A good example is weather prediction; based on known principles and conditions, its objective is to model the detailed distributions and evolutions of the weather quantities faithfully. For this reason, the inputs to a simulation must be as realistic as possible and the recipes for the parameterized processes must be reliable.

A computer code designed according to these requirements imposes very heavy demand on the computational resources. Only a very limited number of cases can be computed; sometimes only a fraction of a case can be completed. The design of numerical experiments is quite the opposite. To make easy the extraction of the principles underlying the process of interest, the numerical model should be

as clean from secondary effects as possible. It is not constrained to match the conditions and parameters to those of a realistic situation (quite often that is not feasible). Most importantly, the parameters should be chosen within the range where the numerical code performs efficiently and accurately, so that the model can produce reliable results. It is also important to realize that this range of validity depends on the objective and scope of the study, so extreme care is needed to avoid overlooking the limitations of the model.

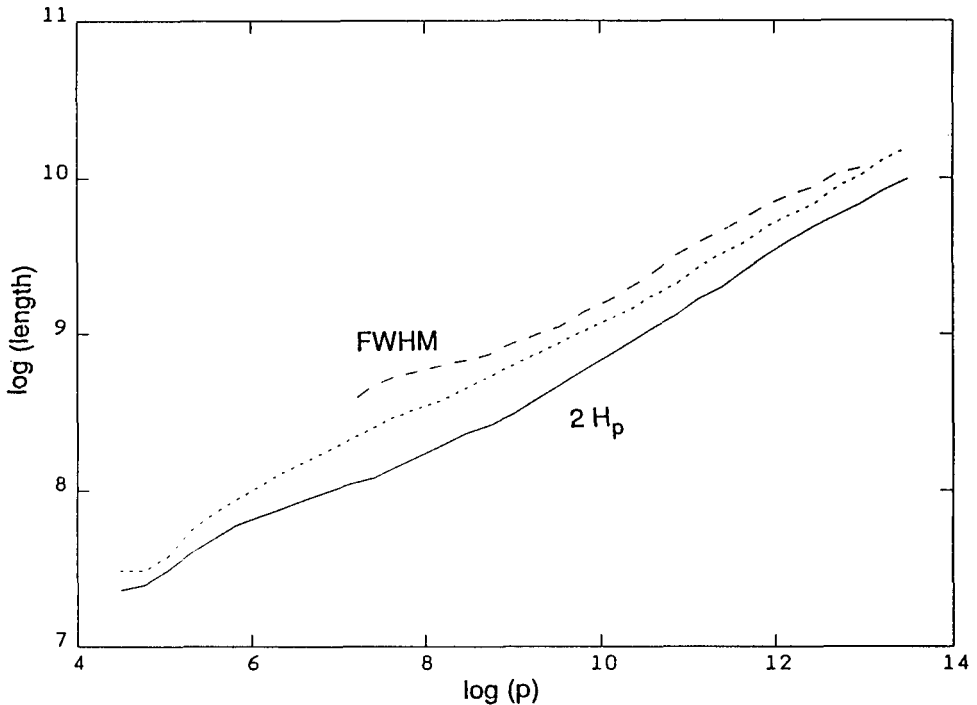


Fig. 1. Comparison of the FWNM of the convective modes with other length scales

When a simulation fails to reproduce observational results correctly, claim is often made that a certain theory is ruled out by the observations. Quite often, such claims are perfunctory. The real culprit may lie in the numerical model, in the input, or in the parameterizations. While the detection of a failure is straightforward in a simulation, its cause or causes are generally difficult to identify. On the other hand, the failure of a numerical experiment is not so clear-cut. As long as the procedure and setup of the experiment is properly documented and followed, the results of the study can be considered to be valid data points. 'Failure' may occur in the interpretation of the results, as inferences are made beyond the numerical and physical limitations of the model.

Therefore, if a numerical study happens to get lost between experimentation and simulation, it is highly susceptible to generating confusing results and misleading claims.

2.2 Connecting numerical experiments to analytical results – an example

Since a theoretical framework cannot be based only on pictures and graphs, it is useful and important to relate numerical works to established concepts and analytical results. This includes several aspects: (i) guide the design of the numerical model; (ii) help to avoid pitfalls; (iii) add insights; and (iv) provide inputs for further analytical study.

Here we discuss a case based on the numerical experiments performed by Chan and Sofia (1986, 1989). Motivated by the controversy over the validity of the mixing length theory (MLT), they studied compressible convection in three-dimensional convective layers with depths around 5 pressure scale heights. Their experiments varied the input flux to the layer, the thermal properties of the gas, and the size of the numerical grids. Quite invariant to these variations, the vertical autocorrelation lengths of the vertical velocity, the temperature fluctuation, and the entropy fluctuation were found to be scaled by the pressure scale height. This provides support to the mixing length picture.

More interestingly, the width and shape of the autocorrelation curves of Chan and Sofia (see Figure 1 of their 1986 paper) have striking similarity to those of the convective modes obtained by Narasimha and Antia (see Figure 1 of their 1982 paper). These authors obtained the convective modes from a solar convection zone model calculated by a mixing length theory which approaches the standard MLT in the deeper layer of the convection zone. They showed that the convective flux can be consistently reconstructed from a parilinear superposition of the statistically independent unstable modes, provided the effects of turbulent conductivity and viscosity are taken into account. The sizes of these modes are more or less scaled by the local pressure scale heights as shown in Figure 1 (kindly provided by Dr. Narasimha). In this figure, the solid curve shows the length of 2 pressure scale heights, the dashed curve shows the FWHM (full-width-half-maximum) of the modes, and the dotted curve shows the mixing length used for the background solar model.

The close resemblance of the auto-correlation curves and the convective modes prompted us to interpret the vertical coherence of the fluid motion showing up in the numerical calculations as a trace of the localized linear modes which drive the flow. The turbulence generated by the nonlinear effects may be important in controlling the amplitudes of these modes by providing a dissipation mechanism, but the cause and the general pattern of the flow may be linked back to these modes. Such a scenario is quite consistent with the analytical model of turbulent convection developed by Canuto and his co-workers (see his article in these Proceedings and references therein). In fact, even the generation of the strong downflow filaments (Massaguer and Zahn, 1980; Hurlburt *et al.*, 1984) and their coalescence

in successively deeper layers (Stein and Nordlund, 1989) may be the result of a not-so-random superposition of the modes arranged by nonlinear locking.

Besides providing bases for developing useful insights, analytical results can also help to guide us through the treacheries of computation, a subject to be discussed next.

Table 2. Examples of pitfalls in numerical study of convection

• inappropriate approximations
numerical – how much physics is caused to be lost?
physical – viscosity modelling suitable?
• ill-posed conditions
background structure – consistent with dynamics?
boundaries – open vs. closed top and bottom
• insufficient relaxation
integration time > thermal relaxation time?
• over-confidence and over-interpretation

3. Pitfalls in the numerical approach

More than any other approach, numerical studies are extremely susceptible to pitfalls. The large amount of numbers generated by a calculation makes it very difficult to analyse the meanings and to detect errors. Errors can occur in any of the different stages of a study. A summary of the common sources is listed in Table 2. Some of pitfalls are quite well-known, but some are subtler. Here, we discuss two important, possibly controversial topics in the computation of solar/stellar convection.

3.1 Open versus closed boundaries at top and bottom

Under most circumstances, the domain of a calculation can only be part of the region where interactions take place. This makes the choice of boundary conditions a very difficult problem. The appropriate choice of course depends on the situation. In some studies, closed upper and lower boundaries which keep the fluid to stay inside a layer are used because the system is well-defined, and the conservation of various quantities can more readily be checked. In many other studies, it is argued that open boundaries which allow fluid to go in and out of the domain would represent a situation more realistically. In these cases, large oscillations of the layers generally occur. Oscillations do occur in closed-boundary cases but with much smaller amplitude. This makes us wonder if some of the large amplitude oscillations are caused by the open boundaries.

An open horizontal boundary is usually created by making $\partial \mathbf{v} / \partial z = 0$, and p , $T = \text{constant}$ where \mathbf{v} is the velocity vector, z is the vertical coordinate, p is the pressure, and T is the temperature. The values of the pressure at the top and bottom are chosen such that the initial state of the gas is in hydrostatic equilibrium. Trouble arises as the convection is turned on, then the mean distribution of the gas

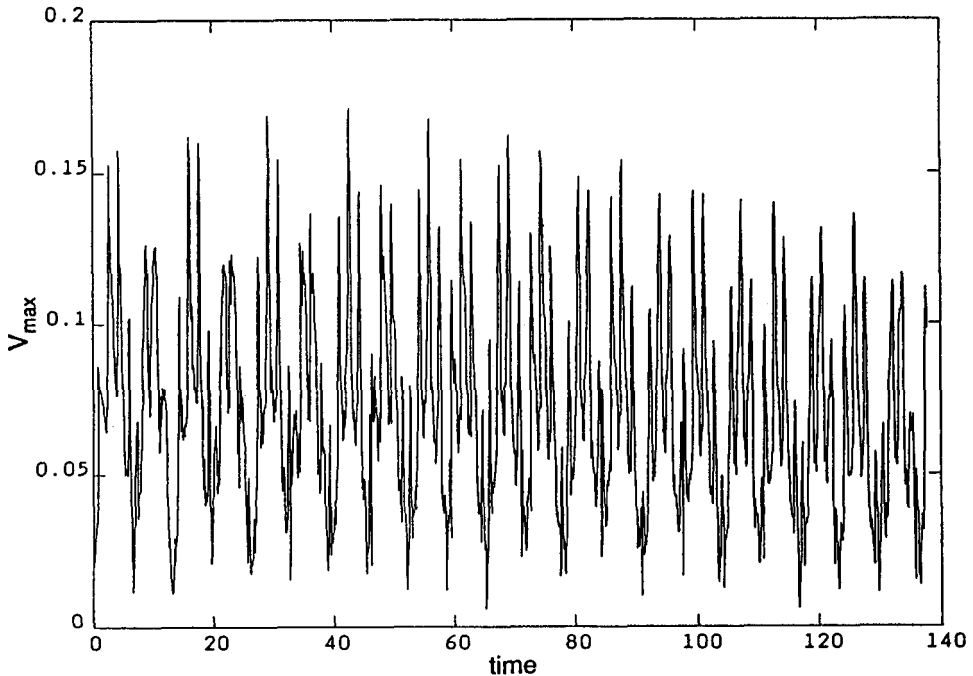


Fig. 2. Absolute value of the maximum vertical velocity (unit velocity = sound speed at the top) vs. time = depth of the layer/unit velocity)

will be changed by the development of convection. The structure of the whole layer will evolve towards a new configuration to be determined self-consistently with the convection (see, for example, Figure 1b of Sofia and Chan, 1984). The original values of the pressure used for the upper and lower boundaries most likely will not be consistent with the new configuration. Will this generate spurious oscillations? To explore this possibility, we performed some simple numerical experiments.

Oscillations can be excited by the convective turbulence; to avoid contamination, we study a convectively stable layer with open boundaries which are created by the conditions as described above ($\gamma = 5/3$, $T/T = 2$, polytropic index = 2, Prandtl number $Pr = 1$; the subscripts 'bot' and 'top' denote values at bottom and top respectively). When everything is in balance, no significant motion is generated, as should be. However, what if the boundary pressure is perturbed, as may be caused by an adjustment of the overall mean structure? This is imitated by a 2% shift of the boundary pressure at the bottom of the layer. The first thing that happens is the appearance of a lot of vertical oscillations (see Figure 2), and they persist for a very long time.

When the viscosity of the gas is increased by a factor of 100, the damping of the motion can be traced (solid curve in Figure 3), but a constant vertical flow through the layer remains, carrying with it an artificial energy flow. We can see the appeal and advantages of allowing flows to go through the boundaries of a domain, but such extraneous behaviour is quite undesirable. Is there some way to improve the situation? One thing to try is to seal off an end so that no constant net flow can go through. We therefore close the top boundary in the above model (set $v_z = 0$) and re-do the experiment. The outcome is illustrated by the dashed curve in Figure 3. The amplitude of the boundary induced oscillations becomes much smaller, and indeed no net flow is sustained. This way, the total amount of material inside the layer can automatically adjust for a new mean structure with less side-effects.

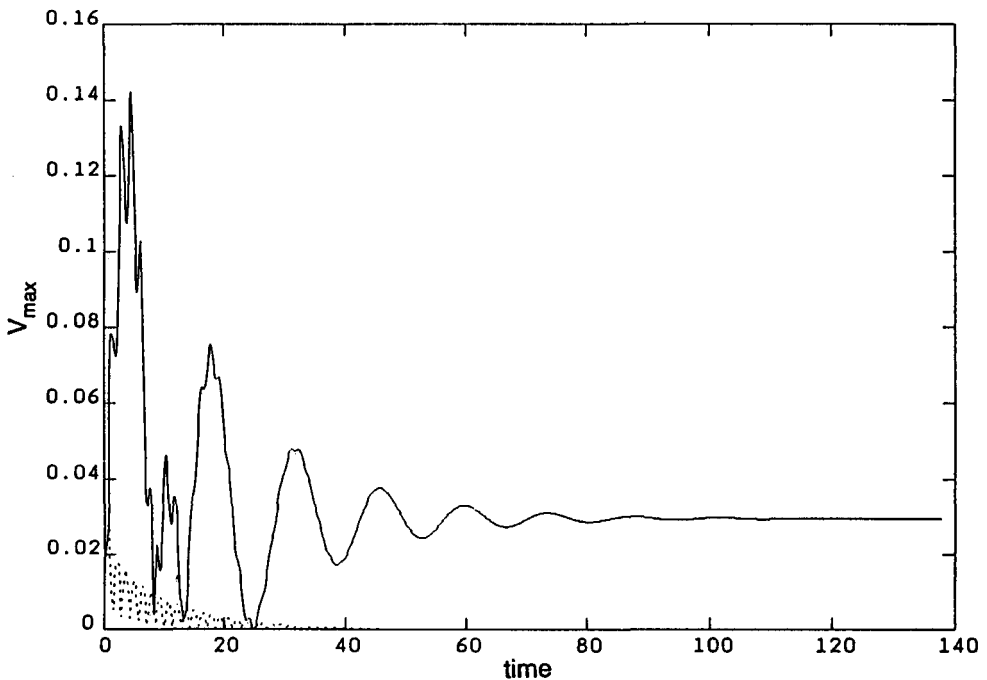


Fig. 3. Absolute values of maximum vertical velocities vs. time for the fully open (solid curve) and half open (dashed curve) boundary cases.

3.2 Significance of thermal relaxation

Now we turn to a computational requirement that is generally acknowledged but not always observed. The convective velocity, temperature fluctuation etc. depend on the structure of the convection zone (e.g. the entropy gradient), and vice versa. When the structure is evolving, the convection is evolving, and there is no guaranteed way to choose an initial structure that is sufficiently close to the developed state. To obtain consistency between the convective process and the structure of the convection zone, it is necessary to integrate the model beyond a thermal relaxation (Kelvin-Helmholtz) time scale. This time scale is given by

$$\tau_{\text{th}} = \int e dz / F, \quad (1)$$

where e is the internal energy density and F is the energy flux.

Here, we illustrate the significance of this time scale by looking at the relaxation of a two-dimensional calculation (adequate for our purpose). The model starts with a slightly superadiabatic, polytropic distribution ($\gamma = 5/3$, $T_{\text{bot}}/T_{\text{top}} = 8$) whose energy content can be estimated as

$$\int e dz \sim p_{\text{bot}} d / (2\gamma - 1), \quad (2)$$

where d is the total depth of the layer. (Note that the estimate provided by Eq. 34 of Chan and Sofia, 1986, contains an extraneous factor Z at the denominator). A constant flux is applied at the lower boundary, and the entropy is fixed at the upper boundary. Some small velocities are used to generate the initial perturbation.

The evolution of the height distribution of the horizontally averaged entropy is shown in Figure 4. The long dot-dashed, short dot-dashed, dashed, solid, and dotted curves show the distributions at the instances $1/4$, $1/2$, $3/4$, 1 , and $5/4$ respectively; a trace of the initial entropy distribution can be detected near the bottom of the figure. All of the distributions are quite flat in the deeper region of the convection zone, but the steady adiabat is not obtained until the integration goes beyond $3/4$.

The evolution of the energy flux flowing out the top of the region is shown by the solid curve in Figure 5; the dashed line shows the fixed input flux. It is clear that the two fluxes do not match until about. Therefore the layer is not in statistical energy balance during the transient phase which lasts about one thermal relaxation time.

The effect of thermal imbalance on the reliability of numerical results again depends on the problem of interest. For properties that are sensitive to the structure of the convection zone, the results of calculations that do not satisfy the relaxation requirement should be viewed with caution.

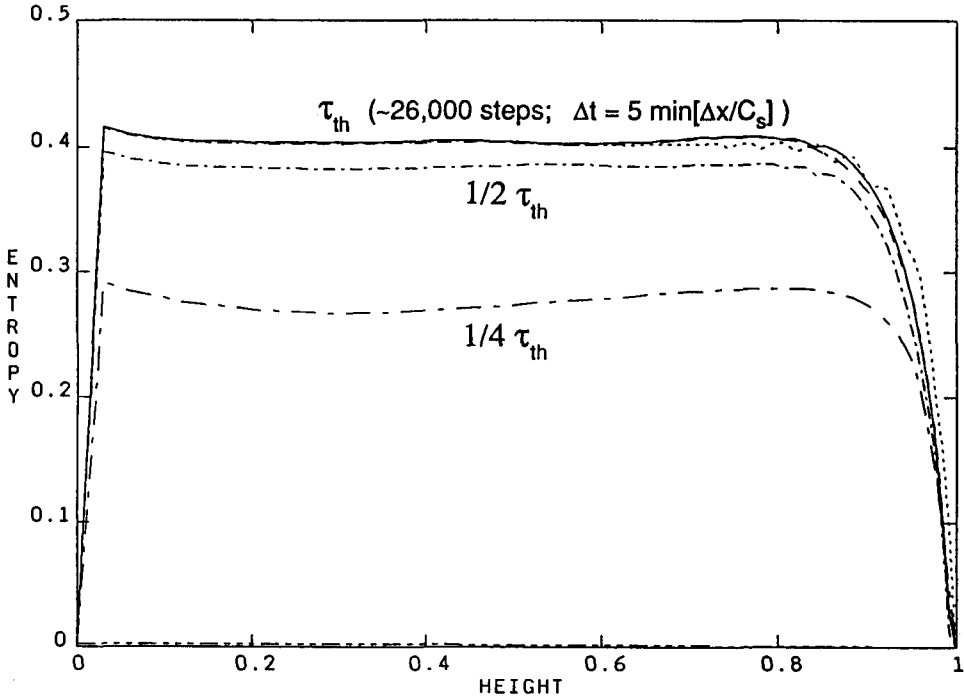


Fig. 4. Thermal relaxation of the entropy distribution

4. Taylor columns in numerical models of solar differential rotation

The lack of thermal relaxation may be the problem that plagues existing numerical models of the solar differential rotation (Glatzmaier, 1985; Gilman and Miller, 1986). These models predicted angular velocity contours mostly parallel to the rotation axis (Taylor columns), but recent results from helioseismology indicate that the contours are more like radial in the solar convection zone (Brown *et al.*, 1989; Libbrecht, 1989; also see the article by Goode in these Proceedings). This discrepancy is a puzzle. For this, let us examine the conditions under which Taylor columns are generated. In quasi-stationary situation the momentum balance in a rotating shell is governed by:

$$(1/\rho)[\nabla p + \rho \mathbf{g} + \nabla \cdot (\rho \mathbf{v} \mathbf{v} - \boldsymbol{\sigma})] + 2\boldsymbol{\Omega} \times \mathbf{v} = 0, \quad (3)$$

where \mathbf{v} is the velocity, \mathbf{g} is the gravitational acceleration, $\boldsymbol{\sigma}$ is the viscous stress tensor, and $\boldsymbol{\Omega}$ is the angular velocity of the shell. If the Coriolis term *really* dominates,

$$\begin{aligned}\nabla \times (\boldsymbol{\Omega} \times \mathbf{v}) &= \boldsymbol{\Omega}(\nabla \cdot \mathbf{v}) - \boldsymbol{\Omega} \cdot (\nabla \mathbf{v}) + \mathbf{v} \cdot (\nabla \boldsymbol{\Omega}) - \mathbf{v}(\nabla \cdot \boldsymbol{\Omega}) \\ &= -\boldsymbol{\Omega} \partial_z \mathbf{v} = 0,\end{aligned}\tag{4}$$

and the Taylor columns follow. This happens if: (i), i.e., where L is the characteristic length, is the kinematic viscosity, and Ta is the Taylor number, so that the viscous terms are unimportant; (ii), i.e. heliostrophic, where V is the characteristic velocity associated with the scale L , so that the nonlinear advection terms also are unimportant; (iii), i.e. close to isentropic, since density and pressure fluctuations where S is the entropy. In the global scale, the first two conditions are generally believed to be satisfied, but the last condition is highly questionable (Durney, 1987). As indicated by the non-existence of Taylor columns in the solar convection zone, this may just be the condition that has been violated in the numerical models.

In the light of recent results from helioseismology the entropy gradient in the solar convection zone therefore appears to be a crucial factor of the mechanism that generates the global differential rotation. It is important to determine the distribution of the gradient correctly, consistently with the convection. However, this is extremely difficult because according to our previous discussion, this takes a Kelvin-Helmholtz time scale. While the hydrostatic adjustment requires only a few years (Spiegel, 1987), the thermal relaxation takes about 10 years. Existing numerical models can only compute through a hydrostatic adjustment time scale. As a result, the entropy distribution obtained by a model represents only an initial state of a very long transient.

There are two ways that allow global scale motions to develop in the numerical models despite the relative meagerness of the buoyance term compared to the Coriolis terms. First, the buoyance driving is performed in smaller scales where the hindering effects of the Coriolis terms are smaller, and then energy is transferred to the global scale through nonlinear interactions (Gilman and Miller, 1986). Second, kinetic energy is generated at the expense of the entropy stratification which becomes even less superadiabatic (Glatzmaier, 1987). The kinetic energy residing in the global scale can then be organized by the Coriolis force into Taylor columns.

At this point, it is also important to point out that while a mixing length type theory may provide a good estimate for the structure of the upper convective region where the Coriolis effects are unimportant, it is not applicable to regimes where the Coriolis effects are large. Most likely, when the Coriolis terms become significant, the superadiabatic gradient needs to be larger than that given by MLT, so that the hindering effects of rotation can be countered.

5. Summary

In this article, we reflect upon the numerical methodology of studying solar/stellar convection and stress the importance and difficulty of avoiding pitfalls. While these seemingly technical "fine points" are usually ignored in preference to the final numerical products, their appreciation is crucial to the proper comprehension of the numerical results. In developing our discussion, we touch upon a variety of subjects; the major points are recapitulated here so that the readers can more easily see their interconnections:

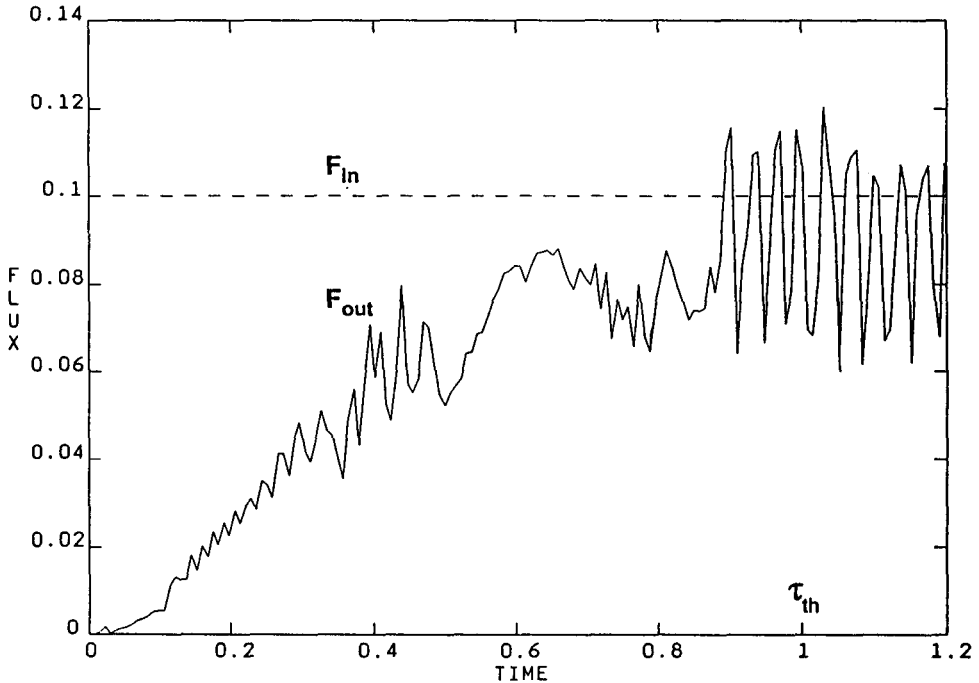


Fig. 5. Thermal relaxation of the energy flux

(i) Numerical studies can be divided into two classes: experimentation and simulation, each serving very different purposes. Conscientious effort to specify and tailor a numerical study to one of these classes can greatly clarify the contribution and limitation of the work.

(ii) The numerical approach is prone to pitfalls. The cardinal one is probably overconfidence; it leads to overlooking the numerical traps and to overinterpreting the results.

(iii) Numerical studies, at least so far, have to base on the concepts and understandings developed by analytical or semi-analytical studies. It is important to acknowledge and utilize this wealth of resources for consolidating the numerical results.

(iv) As an example, we point out that the existence of strong autocorrelations in the turbulent quantities within a pressure scale height is compatible with the dominance of localized convective modes obtained by a semi-analytical theory. The numerical and analytical results reinforce each other.

(v) Among the many possible trouble spots in the numerical study of solar/stellar convection, we discuss two in detail: first, the hazard of open upper and lower boundaries; second, the peril of not achieving thermal relaxation. We

test for these problems in idealised situations, so that the effects can be cleanly demonstrated. It is not clear to what extent such problems might have (or have not) affected existing numerical studies, as the situations may be very different. Our discussion is intended only to be a caution.

(vi) The disagreement between the results of numerical modelling of the solar differential rotation and helioseismology is presently an important puzzle. We speculate that this may be caused by the lack of thermal relaxation in the numerical models, since we believe a correct entropy gradient can break the dominance of the Coriolis effects.

(vii) Due to the difficulty of obtaining thermal relaxation in a numerical model of the solar convection zone, it is probably more useful, at least at the moment, to explore further analytical and semi-analytical models for the solar differential rotation.

Acknowledgements. KLC thanks Dr. Ilkka Tuominen for the enjoyable meeting and NSF for support (AST-8815457). KS thanks NASA for a summer student research fellowship at Goddard Space Flight Center. Furthermore, we thank Drs. D. Narasimha, Hans Mayer, and Detlef Gigas for helpful discussions. Part of the computing was performed at the Pittsburgh Supercomputing Center.

References

- Brown, T.M., Christensen-Dalsgaard, J., Dziembowski, W.A., Goode, P., Gough, D.O., Morrow, C.A.: 1989, *Astrophys. J.* **343**, 526
- Chan, K.L., Sofia, S.: 1986, *Astrophys. J.* **307**, 222
- Chan, K.L., Sofia, S.: 1989, *Astrophys. J.* **336**, 1022
- Chan, K.L., Nordlund, A., Steffen, M., Stein, R.F.: 1990, in *Solar interior and atmosphere*, eds. A.N. Cox and W.C. Livingston, U. Arizona Press, Tucson, in press
- Durney, B.R.: 1987, in *The internal solar angular velocity*, eds. B.R. Durney and S. Sofia, Reidel, Dordrecht, p. 235
- Gilman, P.A., Miller, J.: 1986, *Astrophys. J. Suppl.* **61**, 585
- Glatzmeier, G. A.: 1985, *Astrophys. J.* **291**, 300
- Glatzmeier, G. A.: 1987, in *The internal solar angular velocity*, eds. B.R. Durney and S. Sofia, Reidel, Dordrecht, p. 263
- Hurlburt, N.E., Toomre, J., Massaguer, J.M.: 1984, *Astrophys. J.* **282**, 557
- Libbrecht, K.G.: 1989, *Astrophys. J.* **336**, 1092
- Massaguer, J.M., Zahn, J.-P.: 1980, *Astron. Astrophys.* **87**, 315
- Narasimha, D., Antia, H.M.: 1982, *Astrophys. J.* **262**, 358
- Sofia, S., Chan, K.L.: 1984, *Astrophys. J.* **282**, 550
- Spiegel, E.A.: 1987, in *The internal solar angular velocity*, eds. B.R. Durney and S. Sofia, Reidel, Dordrecht, p. 321
- Stein, R.F., Nordlund, A.: 1989, *Astrophys. J. (Letters)* **342**, L95
- Zahn, J.-P.: 1987, in *Solar and stellar physics*, eds. E.H. Schröter and M.S. Schüssler, Springer-Verlag, New York, p. 55